OwMatch: Conditional Self-Labeling with Consistency for Open-world Semi-Supervised Learning

Shengjie Niu[1], Lifan Lin[2], Huang Jian[1], Wang Chao[2]

[1]The Hong Kong Polytechnic University and [2]Southern University of Science and Technology

November 13, 2024

Outlines

Outlines

Open-world Semi-Supervised Learning (OwSSL)

- Expensive and time-consuming labeling process limits real-world deep-learning applications.
- Semi-supervised learning (SSL) reduces the dependency on labeled data by exploring the inherent structure of unlabeled data.
- Existing SSL methods typically assume a closed-world where all classes possess labeled instances.
- A common case is the presence of novel classes in the unlabeled data.
- **Objective of OwSSL**: classify seen-class samples, or discover novel-class samples and clustering them.

Challenges

- Discover novel classes and assign instances to them. (clustering task)
- Confirmation Bias: The model is biased towards seen classes because it has been exposed only to instances from seen classes.
- Synchronize the varying learning pace that result from the diverse learning style between seen and novel classes. (novel class tend to be slower than seen class)
    - The learning of seen classes base on the supervision of ground-truth labels.
    - For unseen classes, the model can only learn from the clustering objective.

Existing Works

- Pairwise-similarity-based methodologies.
  - Construct a pairwise objective on representation space.
  - Pairwise similarity is calculated and proximity in the prediction of paired instances is encouraged by Binary Cross-Entropy (BCE) loss.
  - Examples: ORCA [1], NACH [1], OpenLDN [2].
- Contrastive-based methodlogies.
  - Construct unsupervised contrastive objective (source from SimCLR [3]) for all data and supervised contrastive objective (source from SupCon [4]) for labeled data.
  - Examples: GCD [5], SimGCD [6].
- Other clustering techniques: Self-labeling-based, TRSSL [7].

Outlines

Flowchart of OwMatch

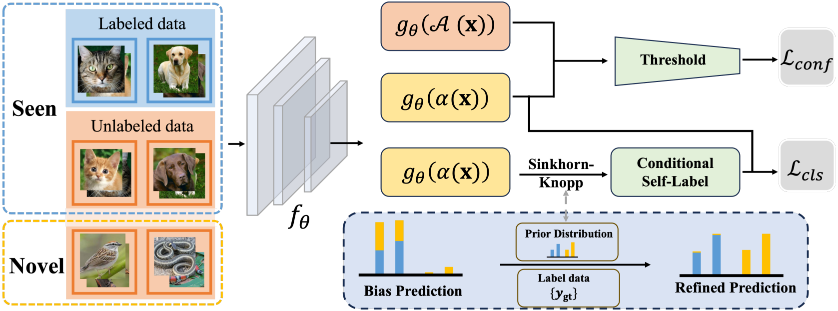

Figure 1: Overview of the OwMatch framework, which is fundamentally composed of three objectives: a) **standard supervised objective**; b) **clustering objective**, which discovers novel-class samples; c) **confidence objective**, which balances the different learning pace between seen and novel classes.

Clustering Objective

Consider the input data $\mathbf{x}^{(i)}$, denote $\mathbf{p}^{(i)}, \mathbf{q}^{(i)} \in \mathbb{R}^K$ as the **model prediction** and **soft self-label** for $\mathbf{x}^{(i)}$, the clustering loss for $\mathbf{x}^{(i)}$ is defined as $\mathcal{L}_{cls}(\mathbf{x}^{(i)}) := H(\mathbf{q}^{(i)}, \mathbf{p}^{(i)})$, where H refers to the Cross-Entropy.

Self-labeling to optimize $\mathbf{q}$: Denote $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{K \times N}$ as the prediction and self-label for $\{\mathbf{x}^{(i)}\}_{i=1}^N$. $\mathbf{Q}$ is enforced to follow a desired partition by constraining it to belong to the transportation polytope: $\mathcal{Q}_1 := \{\mathbf{Q} \in \mathbb{R}_+^{K \times N} | \mathbf{Q}\mathbf{1}_N = N\mathcal{P}, \mathbf{Q}^T\mathbf{1}_K = \mathbf{1}_N\}$, where $\mathbf{1}_v$ is v-dimensional vector of all ones, $\mathcal{P}$ denotes the desired class distribution.

The self-label assignment generation can be understood as an **optimal transportation** problem as $\min_{\mathbf{Q} \in \mathcal{Q}_1} \mathrm{Tr}(\mathbf{Q} \log(\mathbf{P}^T))$.

Conditional Self-labeling

Core idea of **conditional** self-labeling method to refine the self-label assignment under partial supervision. Specifically, we exploit the ground-truth in the labeled dataset and introduce another constraint:

$$\mathcal{Q}_2 := \{\mathbf{Q} \in \mathbb{R}_+^{K\times N} | \mathbf{q}^{(i)} = \mathbf{y}_{\mathrm{gt}}^{(i)}, i = 1, \ldots, N^l\}, \tag{1}$$

Combining these two constraints, we generate the conditional self-label assignment by optimizing,

$$\min_{\mathbf{Q}\in\mathcal{Q}_1\cap\mathcal{Q}_2} \mathrm{Tr}(\mathbf{Q}\log(\mathbf{P}^T)) + \epsilon E(\mathbf{Q}), \tag{2}$$

where $E(\cdot)$ is the entropy function, ϵ is a hyperparameter controlling the smoothness of $\mathbf{Q}$. We denote the optimal solution of (2) as $\tilde{\mathbf{Q}}$.

Theoretical Analysis

Definition 1

Expectation of chi-square statistics (ECS) for $\hat{\boldsymbol{\mu}}$ are defined as the population deviation between the estimator of unlabeled class distribution $\hat{\boldsymbol{\mu}}$ and its true distribution $\mathcal{P}^u$:

$$\mathrm{ECS}(\hat{\boldsymbol{\mu}}) := \mathbb{E}[\chi^2(\mathbf{A})] = \mathbb{E}\left[\sum_{i=1}^{K} \frac{(A_i - \mathbb{E}_{\mathcal{P}}[N_i^u])^2}{\mathbb{E}_{\mathcal{P}}[N_i^u]}\right], \quad (3)$$

where $\mathbf{A}$ are estimators based on $N_1^l, N_2^l, \cdots, N_K^l$, thus are still random variables.

Theorem 1

Consider two estimators for class distribution on unlabeled data, $\hat{\boldsymbol{\mu}}_{\mathrm{uncon}}$ and $\hat{\boldsymbol{\mu}}_{\mathrm{con}}$, we have $\hat{\boldsymbol{\mu}}_{\mathrm{uncon}}$ is a biased estimator and $\hat{\boldsymbol{\mu}}_{\mathrm{con}}$ is an unbiased estimator.

Theoretical Analysis

Theorem 2

Suppose $r_i := \frac{N^l \cdot p_i^l}{N}$ denote the ratio of label samples of the i-th class to the whole samples, $r := \sum_i r_i$ denotes the ratio of labeled samples to the whole samples. For unlabeled sample size N^u, if $\sqrt{N^u} > \frac{1}{\max(|r_i - r \cdot p_i^u|, r \cdot p_j)}$ for $\forall i \in \mathcal{C}_l, \forall j \in \mathcal{C}_u$, then $\mathrm{ECS}(\hat{\boldsymbol{\mu}}_{\mathrm{con}}) \leq \mathrm{ECS}(\hat{\boldsymbol{\mu}}_{\mathrm{uncon}})$.

Conclusion

Following rigorous statistical analysis, the generated label assignments from conditional self-labeling method are closer to the true class distribution in the following scenarios:

- Estimation based on large unlabeled sample size (N^u);
- The difference between prior distribution $\mathcal{P}$ and class distribution of unlabeled data $\mathcal{P}^u$ is not negligible.

Confidence Objective

Unlabeled data are typically used to enhance model performance through consistency regularization [8]:

$$\sum_{i=1}^{N} \mathbb{I}(\max(\mathbf{p}^{(i)}) \geq \tau) H(\hat{\mathbf{p}}^{(i)}, \mathbf{p}^{(i)}),$$

where τ is a scalar hyperparameter denoting the threshold above which we retain a one-hot pseudo-label $\hat{\mathbf{p}}^{(i)}$.

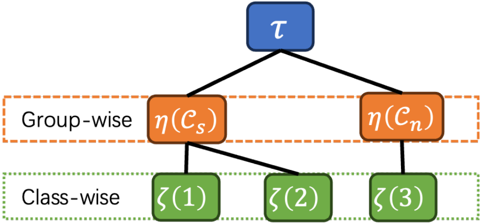

Figure 2: An illustration of the hierarchical thresholding scheme, which involves first estimating the overall learning conditions of two groups and then hierarchically modulating the thresholds in a class-specific manner.

Hierarchical thresholding scheme

Group-wise learning condition for a set of classes $\mathcal{C}_i = \mathcal{C}_s$ or $\mathcal{C}_n$ as

$$\eta(\mathcal{C}_i) = \frac{1}{N_{\mathcal{C}_i}} \sum_{i=1}^{N} \max(\mathbf{p}^{(i)}) \mathbb{I}(\hat{p}^{(i)} \in \mathcal{C}_i), \ \mathcal{C}_i = \mathcal{C}_s, \mathcal{C}_n, \tag{4}$$

where $N_{\mathcal{C}_i} = \sum \mathbb{I}(\hat{p}^{(i)} \in \mathcal{C}_i)$ denotes the number of samples whose predictive labels $\hat{p}^{(i)}$ belong to the group $\mathcal{C}_i$. Similarly, the **class-wise** learning conditions can be defined as

$$\zeta_c = \frac{1}{N_c} \sum_{i=1}^{N} \max(\mathbf{p}^{(i)}) \mathbb{I}(\hat{p}^{(i)} = c), \ c = 1, \ldots, K, \tag{5}$$

where $N_c = \sum \mathbb{I}(\hat{p}^{(i)} = c)$ denotes the number of samples whose predicted labels belong to the c-th class.

Objective

We merge these two learning conditions and obtain the open-world hierarchical threshold as $\tau(c) = \frac{\zeta_c}{\max_{c \in \mathcal{C}_i} \zeta_c} \cdot \eta(\mathcal{C}_i)$. And the confidence objective has the form of

$$\mathcal{L}_{conf} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(\max(\mathbf{p}^{(i)}) > \tau(\hat{p}^{(i)})) \cdot H(\hat{\mathbf{p}}^{(i)}, g_\theta(\mathcal{A}(\mathbf{x}^{(i)}))). \tag{6}$$

Together with the supervised objective $\mathcal{L}_{sup} = \frac{1}{N^l} \sum_{i=1}^{N^l} H(\mathbf{y}_{gt}^{(i)}, \mathbf{p}^{(i)})$ and clustering objective $\mathcal{L}_{cls} = \frac{1}{N} \sum_{i=1}^{N} H(\tilde{\mathbf{q}}^{(i)}, \mathbf{p}^{(i)})$, the overall objective for OwMatch is defined as:

$$\mathcal{L} = \mathcal{L}_{sup} + \mathcal{L}_{cls} + \mathcal{L}_{conf}. \tag{7}$$

Outlines

Main Result

Table 1: Average accuracy on the CIFAR-10/100 and ImageNet100 with 50% novel classes and 50% labeled data within seen classes.

Method	**CIFAR-10**			**CIFAR-100**			**ImageNet100**		
	Seen	Novel	All	Seen	Novel	All	Seen	Novel	All
FixMatch [8]	71.5	50.4	49.5	39.6	23.5	20.3	65.8	36.7	34.9
DS^{3}L [9]	77.6	45.3	40.2	55.1	23.7	24.0	71.2	32.5	30.8
CGDL [10]	72.3	44.6	39.7	49.3	22.5	23.5	67.3	33.8	31.9
DTC [11]	53.9	39.5	38.3	31.3	22.9	18.3	25.6	20.8	21.3
RankStats [12]	86.6	81.0	82.9	36.4	28.4	23.1	47.3	28.7	40.3
SimCLR [13]	58.3	63.4	51.7	28.6	21.1	22.3	39.5	35.7	36.9
UNO [14]	91.6	69.3	80.5	68.3	36.5	51.5	-	-	-
ORCA [15]	88.2	90.4	89.7	66.9	43.0	48.1	89.1	72.1	77.8
NACH [1]	89.5	92.2	91.3	68.7	47.0	52.1	91.0	75.5	79.6
OpenLDN [2]	95.7	95.1	95.4	73.5	46.8	60.1	89.6	68.6	79.1
TRSSL [7]	**96.8**	92.8	94.8	80.0	49.3	64.7	-	-	-
OpenCon [16]	89.3	91.1	90.4	69.1	47.8	52.7	90.6	80.8	83.8
OwMatch	93.0	95.9	94.4	74.5	55.9	65.1	**91.7**	72.0	81.8
OwMatch+	96.5	**97.1**	**96.8**	**80.1**	**63.9**	**71.9**	91.5	**79.6**	**85.5**

Ablation on Components

Table 2: Ablation study on datasets with both novel class ratio and label ratio of 50%. Here, **ConSL** refers to conditional self-labeling, **PLCR** refers to consistency regularization, and **OwAT** refers to an open-world hierarchical thresholding scheme.

Components			**CIFAR-10**			**CIFAR-100**			**Tiny-ImageNet**		
ConSL	PLCR	OwAT	Seen	Novel	All	Seen	Novel	All	Seen	Novel	All
×	×	×	96.5	90.2	93.3	78.8	56.7	67.7	66.5	38.1	52.0
✓	×	×	95.4	96.4	95.9	79.2	58.5	68.7	66.0	39.4	52.4
✓	✓	×	96.3	97.3	96.8	80.1	59.4	69.6	68.6	42.0	54.2
×	✓	✓	97.1	90.4	93.8	80.7	59.7	69.9	69.7	41.4	54.6
✓	✓	✓	96.5	97.1	96.8	80.1	63.9	71.9	68.8	42.4	55.0

Practical and Robust Settings

Table 3: Performance on benchmarks with different imbalance factors (IF) with/without prior class distribution.

Dataset	Prior	Uniform (IF=1)			IF=10			IF=20		
		Seen	Novel	All	Seen	Novel	All	Seen	Novel	All
CIFAR10	w/	96.5	97.1	96.8	93.7	72.1	82.5	92.9	70.1	80.9
	w/o	96.9	90.9	93.9	95.8	66.5	80.3	95.3	64.2	78.8
CIFAR100	w/	80.1	63.9	71.9	76.8	42.0	57.3	76.1	35.2	51.9
	w/o	82.5	57.9	69.2	74.6	39.7	54.1	73.9	33.9	49.2
Tiny-ImageNet	w/	68.8	42.4	55.0	61.7	25.1	41.6	62.4	21.7	38.3
	w/o	69.6	40.6	54.8	61.0	24.9	40.1	61.3	20.3	36.9

Outlines

[1] Lan-Zhe Guo, Yi-Ge Zhang, Zhi-Fan Wu, Jie-Jing Shao, and Yu-Feng Li.
Robust semi-supervised learning when not all classes have labels.
Advances in Neural Information Processing Systems, 35:3305–3317, 2022.

[2] Mamshad Nayeem Rizve, Navid Kardan, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah.
OpenLDN: learning to discover novel classes for open-world semi-supervised learning.
In *European Conference on Computer Vision (ECCV)*, pages 382–401. Springer, 2022.

[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton.
A simple framework for contrastive learning of visual representations.
In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[4] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan.
Supervised contrastive learning.
Advances in neural information processing systems, 33:18661–18673, 2020.

[5] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman.
Generalized category discovery.

In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7492–7501, 2022.

[6] Xin Wen, Bingchen Zhao, and Xiaojuan Qi.
Parametric classification for generalized category discovery: A baseline study.
In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16590–16600, 2023.

[7] Mamshad Nayeem Rizve, Navid Kardan, and Mubarak Shah.
Towards realistic semi-supervised learning.
In *European Conference on Computer Vision (ECCV)*, pages 437–455. Springer, 2022.

[8] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li.
Fixmatch: Simplifying semi-supervised learning with consistency and confidence.
Advances in neural information processing systems, 33:596–608, 2020.

[9] Lan-Zhe Guo, Zhen-Yu Zhang, Yuan Jiang, Yu-Feng Li, and Zhi-Hua Zhou.
Safe Deep Semi-Supervised Learning for Unseen-Class Unlabeled Data.
In *Proceedings of the 37th International Conference on Machine Learning*, pages 3897–3906. PMLR, November 2020.

[10] Xin Sun, Zhenning Yang, Chi Zhang, Keck-Voon Ling, and Guohao Peng.
Conditional gaussian distribution learning for open set recognition.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13480–13489, 2020.

[11] Kai Han, Andrea Vedaldi, and Andrew Zisserman.
Learning to discover novel visual categories via deep transfer clustering.
In *Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR)*, pages 8401–8409, 2019.

[12] Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman.
Automatically discovering and learning new visual categories with ranking statistics.
In *International Conference on Learning Representations (ICLR)*, 2020.

[13] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton.
A simple framework for contrastive learning of visual representations.
In *International Conference on Machine Learning (ICML*, pages 1597–1607. PMLR, 2020.

[14] Enrico Fini, Enver Sangineto, Stéphane Lathuiliere, Zhun Zhong, Moin Nabi, and Elisa Ricci.
A unified objective for novel class discovery.

In *Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR*, pages 9284–9292, 2021.

[15] Kaidi Cao, Maria Brbic, and Jure Leskovec.
Open-world semi-supervised learning.
arXiv preprint arXiv:2102.03526, 2021.

[16] Yiyou Sun and Yixuan Li.
OpenCon: open-world contrastive learning.
Transactions on Machine Learning Research, 2023.

[17] Lilian Weng.
Learning with not enough data part 1: Semi-supervised learning.
lilianweng.github.io, Dec 2021.

[18] Yuki M. Asano, Christian Rupprecht, and Andrea Vedaldi.
Self-labelling via simultaneous clustering and representation learning.
In *International Conference on Learning Representations (ICLR)*, 2020.

[19] Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki.

Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling.
Advances in Neural Information Processing Systems (NeurIPS), 34:18408–18419, 2021.

[20] Yi Xu, Lei Shang, Jinxing Ye, Qi Qian, Yu-Feng Li, Baigui Sun, Hao Li, and Rong Jin.
Dash: Semi-supervised learning with dynamic thresholding.
In *International Conference on Machine Learning*, pages 11525–11536. PMLR, 2021.

[21] Harold W Kuhn.
The hungarian method for the assignment problem.
Naval research logistics quarterly, 2(1-2):83–97, 1955.

[22] Kuniaki Saito, Donghyun Kim, and Kate Saenko.
Openmatch: Open-set consistency regularization for semi-supervised learning with outliers.
arXiv preprint arXiv:2105.14148, 2021.

[23] Lan-Zhe Guo, Zhen-Yu Zhang, Yuan Jiang, Yu-Feng Li, and Zhi-Hua Zhou.
Safe deep semi-supervised learning for unseen-class unlabeled data.
In *International Conference on Machine Learning*, pages 3897–3906. PMLR, 2020.

[24] Yue Fan, Dengxin Dai, Anna Kukleva, and Bernt Schiele.

Cossl: Co-learning of representation and classifier for imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14574–14584, 2022.

[25] Zhengfeng Lai, Chao Wang, Sen-ching Cheung, and Chen-Nee Chuah. Sar: Self-adaptive refinement on pseudo labels for multiclass-imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4091–4100, June 2022.

[26] Hyuck Lee, Seungjae Shin, and Heeyoung Kim. Abc: Auxiliary balanced classifier for class-imbalanced semi-supervised learning. *Advances in Neural Information Processing Systems*, 34:7082–7094, 2021.