IOMatch: Simplifying Open-Set Semi-Supervised Learning with Joint Inliers and Outliers Utilization
Supplementary Material

Zekun Li[1] Lei Qi[2] Yinghuan Shi[1,*] Yang Gao[1]
[1] Nanjing University [2] Southeast University
https://github.com/nukezil/IOMatch

A. Open-Set Semi-Supervised Learning Setting

A.1. Class Space Mismatch

Open-Set Semi-Supervised Learning (OSSL) assumes that labeled and unlabeled data have different class spaces, which can be referred by the term, *Class Space Mismatch*. Let $\mathcal{C}_l$ and $\mathcal{C}_u$ be the class sets of labeled and unlabeled data. Several pioneer works [3, 16] assume that $\mathcal{C}_l \not\subseteq \mathcal{C}_u$ and $\mathcal{C}_u \not\subseteq \mathcal{C}_l$, while more recent OSSL works [6, 9, 11, 22] focus on the case where $\mathcal{C}_l \subset \mathcal{C}_u$. For this point, we share a similar opinion with [7]: As it is usually much easier to collect unlabeled data than labeled data, it is more likely for unlabeled data to have more categories than labeled data. Thus, we assume $\mathcal{C}_l \subset \mathcal{C}_u$ in this work.

Remark. A broader concept is *Class Distribution Mismatch* [4, 23]. If we denote the marginal class distributions of labeled and unlabeled data as $\boldsymbol{p}_l(y)$ and $\boldsymbol{p}_u(y)$, then the class distribution mismatch in SSL indicates that $p_l(y) \neq p_u(y)$. The class space mismatch can be also viewed as such a case, where $p_l(y \in \mathcal{C}_u/\mathcal{C}_l) = 0 \neq p_u(y \in \mathcal{C}_u/\mathcal{C}_l)$. In this work, we just focus on the class space mismatch, which is the most common and problematic case of class distribution mismatch [7].

A.2. Connections to Out-of-Distribution Detection

Out-of-distribution (OOD) detection [10] aims to detect OOD samples existing in test data by assigning higher OOD scores to OOD samples than ID samples. Representative works design the OOD scores using the predicted logits and probabilities [14, 18], or using the information in feature space [12], or combining both of them [19]. More comprehensive reviews can be found in [20].

Although unseen-class outliers can be also regarded as a kind of OOD samples, OOD detection is largely different from open-set SSL in the following aspects. Firstly, OOD detection tasks usually assume that sufficient labeled ID samples are provided for training (and no OOD sample exists), which cannot be satisfied in OSSL. It is a key reason why OOD detection methods cannot be directly applied in OSSL for detecting outliers. Secondly, the main objective of OOD detection is to separate OOD samples from ID samples, which can viewed as a binary classification task. However, the motivation of OSSL is to fully exploit open-set unlabeled samples for improving the model's performance on multi-class classification tasks. Therefore, a model good at OOD detection could not perform well on ID (seen-class) classification. This is the reason why we adopt Balanced Accuracy (BA) rather than AUROC, which is widely used in OOD detection, for open-set evaluation.

B. Distribution Alignment Strategy

For the distribution alignment (DA) strategy, we simply follow the implementation from ReMixMatch [1]. Specifically, we maintain a running average of the model's predictions on unlabeled data, denoted by $\boldsymbol{p}_{avg}$. The marginal class distribution $\boldsymbol{p}_{mrgl}$ is estimated based on the labeled samples in training (which is the uniform distribution in our setting). Given the model's prediction $\boldsymbol{p}_i^w = \phi(f(\mathcal{T}_w(\boldsymbol{u}_i)))$ on an weakly augmented unlabeled sample $\mathcal{T}_w(\boldsymbol{u}_i)$, we scale $\boldsymbol{p}_i^w$ by the ratio $\boldsymbol{p}_{mrgl}/\boldsymbol{p}_{avg}$ and normalize the result as a valid probability distribution:

$$\widetilde{\boldsymbol{p}}_i = \text{Normalize}(\boldsymbol{p}_i^w \cdot \frac{\boldsymbol{p}_{mrgl}}{\boldsymbol{p}_{avg}}), \tag{1}$$

where $\text{Normalize}(\boldsymbol{p})_i = p_i / \sum_j p_j$. $\boldsymbol{p}_i^w$ is then used as the seen-class prediction for producing the unified open-set target and training the closed-set classifier via pseudo-labeling. $\boldsymbol{p}_{avg}$ is computed with the predictions over the last 128 batches.

In practice, we find the DA strategy is effective when the number of classes is relatively large (*e.g.*, for CIFAR-100 and ImageNet-30). However, for CIFAR-10 with fewer classes, the DA strategy may lead to performance degradation instead. The reason could be that the presence of unseen-class outliers interferes with the estimation of $\boldsymbol{p}_{avg}$. Thus, we do not apply the DA strategy in the tasks on CIFAR-10.

C. Extensions with Self-Supervision

IOMatch is such a simple framework that we can easily incorporate other powerful techniques with it to further improve the performance. Recently, self-supervised learning objectives including pretext tasks [5] and contrastive learning [2, 8] have shown strong performance in SSL [1, 13, 24]. We find experimentally that the self-supervised modules can also bring performance gains to IOMatch (see Table 5 in the paper). Here we introduce the details of the extensions of IOMatch.

It is quite easy to incorporate the rotation recognition pretext task with IOMatch. For each unlabeled image $\boldsymbol{u}_i$, we rotate $\boldsymbol{u}_i$ by an angle of $\angle_i$ degrees and obtain Rotate($\boldsymbol{u}_i, \angle_i$), where $\angle_i$ is sampled uniformly from $\angle_i \sim \{0, 90, 180, 270\}$. We add an auxiliary classifier $\theta(\cdot)$ (implemented as a fully connected layer) connected to the backbone encoder, which predicts the rotation degree among the four options, *i.e.*, $\boldsymbol{a} = \theta(f(\text{Rotate}(\boldsymbol{u}_i, \angle_i))) \in \mathbb{R}^4$. The rotation prediction loss is defined as:

$$\mathcal{L}_{rot} = \frac{1}{\mu B} \sum_{i=1}^{\mu B} \text{H}(\text{OneHot}(\angle_i), \boldsymbol{a}). \tag{2}$$

We implement the contrastive learning objective following SimMatch [24]. Given the projected features of all labeled samples $\{\boldsymbol{z}_l : l \in (1, \ldots, N_l)\}$ (maintained in a memory bank), the instance similarities between each unlabeled sample $\boldsymbol{u}_i$ and all labeled samples are defined as $\boldsymbol{r}_i$:

$$\boldsymbol{r}_{i,l}^{w/s} = \frac{\exp(\text{sim}(\boldsymbol{z}_i^{w/s}, \boldsymbol{z}_l))}{\sum_{j=1}^{N_l} \exp(\text{sim}(\boldsymbol{z}_i^{w/s}, \boldsymbol{z}_j))}, \tag{3}$$

where $\text{sim}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^\intercal \boldsymbol{v} / \|\boldsymbol{u}\| \, \|\boldsymbol{v}\|$, and $t = 0.1$ is the temperature parameter. The similarity target $\widetilde{\boldsymbol{r}}$ is then generated by scaling $\boldsymbol{r}_i^w$ with $\widetilde{\boldsymbol{p}}_i$. The contrastive loss is defined as:

$$\mathcal{L}_{con} = \frac{1}{\mu B} \sum_{i=1}^{\mu B} \text{H}(\widetilde{\boldsymbol{r}}_i, \boldsymbol{r}_i^s). \tag{4}$$

As the above two self-supervised objectives are both standard cross-entropy losses, we can simply add them to the total loss with the weights $\mathcal{L}_{rot} = \mathcal{L}_{con} = 1$. In spite of the promising results, the extensions of IOMatch introduce extra network modules (*e.g.*, the rotation classifier and the memory bank) and thus extra training costs. It is noteworthy that, as a simple yet effective OSSL framework, IOMatch can outperform the complicated baselines on most tasks even without these extra learning objectives.

D. Inference

We use the standard closed-set classifier for the inference in the closed-set classification task, in order to ensure fair comparisons with other baselines. In fact, the open-set classifier can also be used for closed-set classification by ignoring the last item of $\boldsymbol{q}_t$. We find experimentally that in this case, the predictions made by $\phi(\cdot)$ and $\psi(\cdot)$ are mostly the same. The difference in closed-set accuracy is usually less than 0.5%. In the paper, we evaluate the closed-set performance using the closed-set classifier to keep consistent with other methods. However, we can just employ a single open-set classifier $\psi(\cdot)$ for both the close-set and open-set classification tasks for the sake of simplicity.

E. Open-Set Evaluation with Foreign Outliers

We have performed open-set evaluation with the test sets of CIFAR-10/100 (see Table 2 of the paper), which consist of all seen and unseen classes observed during training. In such case, unseen-class outliers in testing are similar to those in training. As the seen and unseen classes come from the same dataset, we denote them as the **intra-dataset** test data. Here we also consider the **inter-class** case where additional foreign outliers come from different datasets than CIFAR10/100. In particular, we add samples from SVHN [15], LSUN [21], and synthetic Gaussian and uniform noise images [22] as part of the testing data.

The results are shown in Table S1. Since the added foreign outliers are more dissimilar to the inliers, they are easier to identify. Therefore, the open-set accuracy on the inter-dataset test data is a little higher than that on the intra-dataset test data, while the difference is not significant.

Table S1. Open-set classification balanced accuracy (%) on the **inter-dataset** open-set test data, which contain samples from different datasets than CIFAR10/100.

	Dataset		CIFAR-10		CIFAR-100					
	Class split (Seen / Unseen)		6 / 4		20 / 80		50 / 50		80 / 20	
	Number of labels per class		4	25	4	25	4	25	4	25
Open-Set SSL	UASD [3]	AAAI'20	18.32 ± 0.61	35.78 ± 0.22	11.03 ± 0.43	27.35 ± 0.33	7.03 ± 0.45	31.94 ± 0.74	5.92 ± 0.35	27.83 ± 0.85
Open-Set SSL	DS3L [6]	ICML'20	31.38 ± 0.52	40.92 ± 0.68	13.05 ± 1.03	35.03 ± 0.47	11.84 ± 0.79	34.88 ± 0.57	11.38 ± 0.89	29.32 ± 0.38
Open-Set SSL	MTCF [22]	ECCV'20	28.35 ± 4.84	46.06 ± 0.69	8.16 ± 2.12	26.77 ± 3.70	4.14 ± 0.38	38.04 ± 0.15	1.46 ± 0.17	30.51 ± 0.27
Open-Set SSL	T2T [11]	ICCV'21	<u>51.35 ± 1.76</u>	<u>61.78 ± 0.89</u>	<u>17.82 ± 1.57</u>	37.78 ± 0.73	12.33 ± 1.87	43.86 ± 0.71	<u>34.45 ± 0.67</u>	<u>51.77 ± 1.03</u>
Open-Set SSL	OpenMatch [17]	NeurIPS'21	14.37 ± 0.05	20.31 ± 3.49	8.77 ± 2.83	<u>39.96 ± 1.17</u>	9.97 ± 0.37	<u>49.56 ± 1.15</u>	6.31 ± 0.88	44.77 ± 0.58
Open-Set SSL	SAFE-STUDENT [9]	CVPR'22	46.37 ± 0.61	54.23 ± 0.42	16.31 ± 0.88	29.44 ± 0.56	<u>23.31 ± 0.93</u>	46.91 ± 1.42	29.52 ± 0.55	50.83 ± 0.41
	IOMatch	**Ours**	**77.82 ± 2.48**	**82.44 ± 0.54**	**46.97 ± 2.05**	**60.30 ± 0.99**	**46.09 ± 1.98**	**60.64 ± 0.79**	**40.08 ± 0.75**	**54.57 ± 0.30**

References

[1] David Berthelot, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. In *ICLR*, 2020. 1, 2

[2] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607, 2020. 2

[3] Yanbei Chen, Xiatian Zhu, Wei Li, and Shaogang Gong. Semi-supervised learning under class distribution mismatch. In *AAAI*, pages 3569–3576, 2020. 1, 3

[4] Yue Duan, Lei Qi, Lei Wang, Luping Zhou, and Yinghuan Shi. Rda: Reciprocal distribution alignment for robust semi-supervised learning. In *ECCV*, pages 533–549, 2022. 1

[5] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. In *ICLR*, 2018. 2

[6] 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 *ICML*, pages 3897–3906, 2020. 1, 3

[7] Lan-Zhe Guo, Zhi Zhou, and Yu-Feng Li. Robust deep semi-supervised learning: A brief introduction. *arXiv preprint arXiv:2202.05975*, 2022. 1

[8] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, pages 9729–9738, 2020. 2

[9] Rundong He, Zhongyi Han, Xiankai Lu, and Yilong Yin. Safe-student for safe deep semi-supervised learning with unseen-class unlabeled data. In *CVPR*, pages 14585–14594, 2022. 1, 3

[10] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *ICLR*, 2017. 1

[11] Junkai Huang, Chaowei Fang, Weikai Chen, Zhenhua Chai, Xiaolin Wei, Pengxu Wei, Liang Lin, and Guanbin Li. Trash to treasure: harvesting ood data with cross-modal matching for open-set semi-supervised learning. In *ICCV*, pages 8310–8319, 2021. 1, 3

[12] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *NeurIPS*, 2018. 1

[13] Junnan Li, Caiming Xiong, and Steven CH Hoi. Comatch: Semi-supervised learning with contrastive graph regularization. In *ICCV*, pages 9475–9484, 2021. 2

[14] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. In *NeurIPS*, pages 21464–21475, 2020. 1

[15] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. 2011. 2

[16] Avital Oliver, Augustus Odena, Colin A Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. Realistic evaluation of deep semi-supervised learning algorithms. In *NeurIPS*, 2018. 1

[17] Kuniaki Saito, Donghyun Kim, and Kate Saenko. Openmatch: Open-set consistency regularization for semi-supervised learning with outliers. In *NeurIPS*, 2021. 3

[18] Yiyou Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. In *NeurIPS*, pages 144–157, 2021. 1

[19] Haoqi Wang, Zhizhong Li, Litong Feng, and Wayne Zhang. Vim: Out-of-distribution with virtual-logit matching. In *CVPR*, pages 4921–4930, 2022. 1

[20] Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334*, 2021. 1

[21] Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015. 2

[22] Qing Yu, Daiki Ikami, Go Irie, and Kiyoharu Aizawa. Multi-task curriculum framework for open-set semi-supervised learning. In *ECCV*, pages 438–454, 2020. 1, 2, 3

[23] Zhen Zhao, Luping Zhou, Yue Duan, Lei Wang, Lei Qi, and Yinghuan Shi. Dc-ssl: Addressing mismatched class distribution in semi-supervised learning. In *CVPR*, pages 9757–9765, 2022. 1

[24] Mingkai Zheng, Shan You, Lang Huang, Fei Wang, Chen Qian, and Chang Xu. Simmatch: Semi-supervised learning with similarity matching. In *CVPR*, pages 14471–14481, 2022. 2