Binary Decomposition: A Problem Transformation Perspective for Open-Set Semi-Supervised Learning

Jun-Yi Hang [1 2] **Min-Ling Zhang** [1 2]

Abstract

Semi-supervised learning (SSL) is a classical machine learning paradigm dealing with labeled and unlabeled data. However, it often suffers performance degradation in real-world open-set scenarios, where unlabeled data contains outliers from novel categories that do not appear in labeled data. Existing studies commonly tackle this challenging open-set SSL problem with detect-and-filter strategy, which attempts to purify unlabeled data by detecting and filtering outliers. In this paper, we propose a novel binary decomposition strategy, which refrains from error-prone procedure of outlier detection by directly transforming the original open-set SSL problem into a number of standard binary SSL problems. Accordingly, a concise yet effective approach named BDMatch is presented. BDMatch confronts two attendant issues brought by binary decomposition, i.e. class-imbalance and representation-compromise, with adaptive logit adjustment and label-specific feature learning respectively. Comprehensive experiments on diversified benchmarks clearly validate the superiority of BDMatch as well as the effectiveness of our binary decomposition strategy.

1. Introduction

Semi-supervised learning (SSL) allows to improve model performance by leveraging unlabeled data when available labeled data is insufficient (Engelen & Hoos, 2020; Yang et al., 2023). As a classical machine learning paradigm, SSL has been widely studied in various tasks, such as image annotation (Sohn et al., 2020; Shi et al., 2023) and text categorization (Xie et al., 2020).

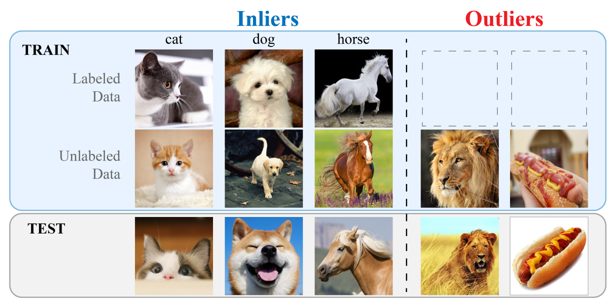

Figure 1. An exemplary open-set semi-supervised learning scenario. Unlabeled set contains outliers from novel categories that do not appear in labeled set.

Most existing SSL approaches are designed under the closed-world assumption, i.e. labeled and unlabeled data share exactly the same label space. However, such a closed-world assumption is rather strict or even impractical in real-world scenarios. Without the intervention of human annotators, unlabeled set inevitably contains outliers from novel categories that do not appear in labeled set, as shown in Figure 1. Thus, a realistic setting called open-set semi-supervised learning (Oliver et al., 2018) arises, which aims to induce a prediction model that can both accurately classify inliers from seen categories and correctly reject outliers from novel categories.

A straightforward strategy for open-set SSL is to purify unlabeled set by detecting and filtering outliers. Following this strategy, existing works either rely on heuristic criterions (Chen et al., 2020; He et al., 2022a; Du et al., 2023), or learn a parameterized detector (Guo et al., 2020; Yu et al., 2020; Saito et al., 2021; Wang et al., 2023c) to filter out potential outliers from unlabeled set so that classifier can be induced on purified unlabeled data with standard SSL approaches. However, this detect-and-filter strategy may lead to suboptimal performance, since it is not an easy job to obtain a reliable detection function with no labeled outlier but only limited inliers.

In this paper, a novel strategy to tackle open-set SSL problem is proposed, which refrains from error-prone procedure of outlier detection in existing detect-and-filter strategy. From the perspective of problem transformation, our binary

[1]School of Computer Science and Engineering, Southeast University, Nanjing 210096, China [2]Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, China. Correspondence to: Min-Ling Zhang <zhangml@seu.edu.cn>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

decomposition strategy decomposes the original open-set SSL problem into a number of standard binary SSL problems, each of which aims to derive a one-vs-all classifier for a seen category from a binary labeled set and an unlabeled set consisting of potential positive and negative samples. This strategy provides a concise way for open-set SSL, but two attendant issues should be carefully considered before it works well.

Firstly, the binary SSL problems may suffer from a high degree of *class-imbalance*, even though the original class distributions of labeled and unlabeled data are balanced. For example, imbalance ratio of a binary labeled set derived from a balanced multi-class labeled set is $\frac{1}{K-1}$ with K being the size of label space in labeled set. What makes the problem more challenging is that class distribution of the binary unlabeled set is different from that of binary labeled set and concealed from the learning algorithm. Secondly, solving theses binary problems together could lead to a *compromised representation*, as these problems may possess distinct discriminative preferences. For example, in image annotation task, *shape*-based features would be more essential in recognizing the *plane* category, while *color*-based features might be preferred in discriminating the *sky* category.

Taking the above two issues into considering, we provide the first instantiation, BDMatch, for **B**inary **D**ecomposition strategy. Specifically, BDMatch learns class-balanced binary classifiers via an adaptive logit adjustment mechanism, which accumulates model predictions during training to estimate a proxy of the potential class distributions of each binary training set and adjusts classifiers' outputs to be appropriately biased toward the minority classes. Instead of sharing a compromised representation among all binary classifiers, BDMatch learns label-specific features, i.e. the most pertinent and discriminative features for each class label, to fully consider distinct discriminative preferences of each binary problem. Comprehensive experiments on various benchmark data sets show that BDMatch performs better than existing open-set SSL algorithms.

The rest of this paper is organized as follows. Section 2 briefly reviews related works. Section 3 presents details of the BDMatch approach. Section 4 reports experimental results over a wide range of benchmark data sets. Section 5 concludes this paper.

2. Related Works

Semi-Supervised Learning. With the ability to exploit both labeled and unlabeled data, semi-supervised learning is one of the research hotspots in the last decade (Engelen & Hoos, 2020; Yang et al., 2023). As the most popular SSL technique, consistency regularization is widely adopted and investigated in recent literatures (Xie et al., 2020; Xu et al., 2021; Wang et al., 2023b). The basic idea is to enforce the model to output similar predictions among different perturbations of a sample, which endows model with local smoothness so that decision boundary would go through low-density data regions (Miyato et al., 2019). As a representative SSL approaches, FixMatch (Sohn et al., 2020) combines consistency regularization with a confidence-based filtering mechanism. Follow-up works further enhance it by customized strategies, such as debiasing the model predictions with distribution alignment (Berthelot et al., 2020) or neighbor voting (Li et al., 2021), selecting reliable samples with adaptive thresholding (Zhang et al., 2021; Chen et al., 2023), incorporating extra learning objectives from self-supervised pretext tasks (Zheng et al., 2022; Nassar et al., 2023). However, all these SSL approaches are developed with the assumption that labeled and unlabeled data share the same label space, which prevents them from deploying in real-world SSL scenarios.

Open-Set Semi-Supervised Learning. The problem of open-set SSL is firstly studied in (Oliver et al., 2018), which shows standard SSL approaches would suffer performance degradation when unlabeled set contains outliers. As an intuitive solution, existing works commonly attempt to alleviate the effect of outliers with detect-and-filter strategy. Generally speaking, these approaches can be roughly grouped into two categories, which differ in the way to detect outliers, namely criterion-based approaches and detector-based approaches. Criterion-based approaches rely on heuristic criterions, such as prediction confidence (Chen et al., 2020; Huang et al., 2023c), sample similarity (Du et al., 2023), or energy discrepancy (He et al., 2022a), to detect outliers. While detector-based approaches turn to a parameterized detector to filter outliers. Various ways have been developed to train the detector. For example, MTCF (Yu et al., 2020) trains it under a noisy label learning framework, where all the unlabeled samples are treated as noisy outliers. Some works induce detector with self-training techniques, e.g. entropy minimization regularization (Huang et al., 2021; Wang et al., 2023c) and pseudo-label assignment (Saito et al., 2021; Fan et al., 2023). While other works (Guo et al., 2020; He et al., 2022b) formalize the training procedure as a bi-level optimization problem, where detector is learned by optimizing derived classifier on labeled data.

Along this line, some recent studies point out outliers can be useful and reconsider filtered outliers for model training by self-supervised representation learning (Huang et al., 2021; Mo et al., 2023; Wang et al., 2023c), adversarial feature adaptation (Huang et al., 2023c), or style-based data augmentation (Huang et al., 2023b). However, since there is no labeled outlier but only limited inliers, it is an inherently hard task to obtain a reliable detection function. As a consequence, errors of outlier detection will propagate to derived classifier, leading to suboptimal performance in inlier classification.

To circumvent the error-prone procedure of outlier detection, some studies attempt to transfer the open-set SSL problem to closed-set one (Li et al., 2023; Ma et al., 2023). They regard all outliers as samples from a new virtual category and train a $(K+1)$-way classifier with pseudo-labels constructed by heuristics. While Fix-A-Step (Huang et al., 2023a) introduces the mixup trick to blend samples from both inliers and outliers. Instead, our binary decomposition strategy directly decomposes the open-set SSL problem into a number of standard binary SSL problems, which is more straightforward and has no need to design complicated heuristics for constructing pseudo-labels.

Class-Imbalanced Learning. Class-imbalanced learning has been widely studied by machine learning community, since real-world data sets are typically long-tailed (Zhang et al., 2023). In supervised learning, imbalanced class distributions can be easily tackled by re-balancing methods, including data re-sampling (Kang et al., 2020) and loss reweighting (Cui et al., 2019). However, they become rather intractable when the data set contains unlabeled data, e.g. in SSL scenarios. In this challenging direction, early attempts (Wei et al., 2021; Lee et al., 2021) assume the labeled and unlabeled data share the same class distribution, so that classical re-balancing methods can be adapted for class-imbalanced SSL. Subsequent studies (Guo & Li, 2022; Lai et al., 2022; Wang et al., 2022a; 2023a; Wei & Gan, 2023) move one step further. They resort to distribution estimation techniques (Kim et al., 2020; Zhao et al., 2022), with which fine-grained balance can be achieved when the real distributions are mismatched between the labeled and unlabeled data. Both as realistic settings of SSL, class-imbalanced SSL and open-set SSL are investigated separately. Our work makes a first attempt to find the point of intersection between these two lines of studies. With binary decomposition strategy, open-set SSL problem is transformed into a number of class-imbalanced SSL problems, so that advanced techniques in class-imbalanced SSL can be exploited for open-set SSL.

3. The BDMatch Approach

3.1. Preliminaries

In open-set SSL, the training set consists of a labeled set $\mathcal{D}^l = \{(\mathbf{x}_i^{(l)}, y_i)\}_{i=1}^N \subset \mathcal{X} \times \mathcal{Y}$ and an unlabeled set $\mathcal{D}^u = \{\mathbf{x}_i^{(u)}\}_{i=1}^M \subset \mathcal{X}$, where $\mathcal{X}$ denotes the input space and $\mathcal{Y} = \{1, 2, \ldots, K\}$ denotes the label space with K categories. The labeled set contains and only contains samples belonging to K categories in label space $\mathcal{Y}$. While the unlabeled set contains both inliers from K categories in label space $\mathcal{Y}$ and outliers from novel categories that do not appear in labeled set. For notation briefness, we denote the set of all novel categories by $\mathcal{U}$ and thus the potential label space of unlabeled set is $\mathcal{Y} \cup \mathcal{U}$ with $\mathcal{Y} \cap \mathcal{U} = \emptyset$. Formally, open-set SSL aims to derive a prediction model which can accurately classify inliers from seen categories and correctly reject outliers from novel categories.

3.2. Limitations of Detect-and-Filter Strategy

We firstly review the general framework of existing detect-and-filter strategy and analyze its limitations, which motivates us to investigate a more straightforward strategy for tackling open-set SSL problem.

Given a detection function $d : \mathcal{X} \to \{0, 1\}$[1], the detect-and-filter strategy derives classifier $f : \mathcal{X} \to \mathbb{R}^K$ by minimizing a supervised classification loss on labeled set $\mathcal{D}^l$ and a masked unsupervised regularization term on unlabeled set $\mathcal{D}^u$. Formally, the learning objective can be formalized as follows

$$\min_\theta \sum_{i=1}^N \ell(f(\mathbf{x}_i^{(l)}; \theta), y_i) + \sum_{i=1}^M d(\mathbf{x}_i^{(u)}) \cdot \Omega(\mathbf{x}_i^{(u)}; \theta), \quad (1)$$

where θ parametrizes the K-way classifier f, and ℓ denotes the cross-entropy loss. $d(\mathbf{x}_i^{(u)})$ describes the detection and filtering procedure, where $d(\mathbf{x}_i^{(u)}) = 1$ indicates $\mathbf{x}_i^{(u)}$ is detected as an inlier and $d(\mathbf{x}_i^{(u)}) = 0$ otherwise. For inliers, a per-sample regularization term $\Omega(\mathbf{x}_i^{(u)}; \theta)$ incorporates some expected properties (e.g. smoothness) into derived decision boundary, which is generally implemented as the consistency regularization in literatures

$$\Omega(\mathbf{x}_i^{(u)}; \theta) = \ell(f(\mathcal{A}(\mathbf{x}_i^{(u)}); \theta), q_i), \quad (2)$$

where $q_i = \arg\max_{y \in \mathcal{Y}} f(\mathbf{x}_i^{(u)}; \theta)_y$ denotes the pseudo-label constructed from predictions of classifier f, and $\mathcal{A}(\cdot)$ is the data augmentation scheme (Cubuk et al., 2020) imposed to obtain a perturbed version of sample $\mathbf{x}_i^{(u)}$.

The above general framework discloses limitations of detect-and-filter strategy in dealing with open-set SSL problem. (1) *Errors of outlier detection will propagate to derived classifier*: If an outlier is detected as an inlier, then the pseudo-label assigned to it dooms to be a wrong one, since its real label is not in the label space $\mathcal{Y}$. Such a mislabeled example will become a permanent noise for classifier induction. (2) *Outliers are wasted*: Even with an oracle detection function which can perfectly distinguish outliers from inliers, classifier is only trained on inliers, which means outliers contribute no useful information for classifier induction.

[1] Existing approaches have different implementations of detection function, e.g. heuristic criterions or parameterized models. We omit technical details here and suppose it is ready to use for brevity.

Instead, our binary decomposition strategy does not have these limitations. We will describe it in the next subsection.

3.3. The Binary Decomposition Strategy

From the perspective of problem transformation, our binary decomposition strategy directly decomposes open-set SSL problem into a number of standard binary SSL problems. Specifically, given an open-set SSL data set consisting of a labeled set $\mathcal{D}^l = \{(\mathbf{x}_i^{(l)}, y_i)\}_{i=1}^N$ and an unlabeled set $\mathcal{D}^u = \{\mathbf{x}_i^{(u)}\}_{i=1}^M$, we decompose it into K binary SSL data sets $\{(\mathcal{D}_c^l, \mathcal{D}_c^u)\}_{c=1}^K$ as follows

$$\begin{aligned} & \mathcal{D}_c^l = \mathcal{P}_c^l \cup \mathcal{N}_c^l, \quad \mathcal{D}_c^u = \mathcal{D}^u \\ with \quad & \mathcal{P}_c^l = \{(\mathbf{x}_i^{(l)}, 1) | (\mathbf{x}_i^{(l)}, y_i) \in \mathcal{D}^l, y_i = c\} \\ & \mathcal{N}_c^l = \{(\mathbf{x}_i^{(l)}, 0) | (\mathbf{x}_i^{(l)}, y_i) \in \mathcal{D}^l, y_i \neq c\}, \end{aligned} \tag{3}$$

where $\mathcal{D}_c^l, \mathcal{D}_c^u$ denote binary labeled and unlabeled sets for the c^{th} seen category in original label space $\mathcal{Y}$ respectively. Labeled set $\mathcal{D}_c^l$ consists of a positive sample set $\mathcal{P}_c^l$ and a negative sample set $\mathcal{N}_c^l$, which correspond to the samples in the original labeled set $\mathcal{D}^l$ with the c^{th} seen category and without the c^{th} seen category respectively.

It is worth noting that the unlabeled set $\mathcal{D}_c^u$ contains only potential positive and negative samples for each seen category. Namely, labeled set $\mathcal{D}_c^l$ and unlabeled set $\mathcal{D}_c^u$ now share the same binary label space, which means that these binary SSL problems follow standard SSL setting.

We solve these binary SSL problems with the popular consistency regularization framework

$$\min_{\Theta} \sum_{c=1}^K [\sum_{i=1}^N \ell_{BCE}(f_c(\mathbf{x}_i^{(l)}; \theta_c), y_i^c) + \sum_{i=1}^M \Omega_c(\mathbf{x}_i^{(u)}; \theta_c)], \tag{4}$$

where a one-vs-all classifier $f_c : \mathcal{X} \to \mathbb{R}, (1 \le c \le K)$ is learned for each binary problem and $\Theta = \{\theta_1, \theta_2, \dots, \theta_K\}$ is the set to parametrize these classifiers. ℓ_{BCE} denotes the binary cross-entropy loss and y_i^c is the binary label of sample $\mathbf{x}_i^{(l)}$, which can be determined according to whether $\mathbf{x}_i^{(l)}$ is a positive sample or not for the c^{th} binary problem

$$y_i^c = \begin{cases} 1, & if\, \mathbf{x}_i^{(l)} \in \mathcal{P}_c^l \\ 0, & otherwise. \end{cases} \tag{5}$$

Following FixMatch (Sohn et al., 2020), the per-sample regularization term is implemented as follows

$$\Omega_c(\mathbf{x}_i^{(u)}; \theta_c) = \ell_{BCE}(f_c(\mathcal{A}(\mathbf{x}_i^{(u)}); \theta_c), q_i^c), \tag{6}$$

where we denote by $p_i^c = \sigma(f_c(\mathbf{x}_i^{(u)}; \theta_c))$ the prediction probability that sample $\mathbf{x}_i^{(u)}$ belongs to the c^{th} seen category, with $\sigma(\cdot)$ being the sigmoid function. Then, the pseudo-label q_i^c can be constructed via $q_i^c = \mathbb{I}[p_i^c > 0.5]$. Here, $\mathbb{I}[P] = 1$ if predicate P holds and $\mathbb{I}[P] = 0$ otherwise. Following (Sohn et al., 2020), we mask out samples that the current classifier is not confident about to alleviate confirmation bias (Arazo et al., 2020) in self-training. The mask m_i^c is generated via a simple thresholding process

$$m_i^c = \mathbb{I}[(p_i^c > \rho) \vee (p_i^c < 1 - \rho)], \tag{7}$$

where $p_i^c > \rho$ (resp. $p_i^c < 1 - \rho$) means that $\mathbf{x}_i^{(u)}$ is a confident positive (resp. negative) sample. The threshold ρ is set to 0.99 in this paper.

With binary decomposition strategy, inherent limitations of detect-and-filter strategy are completely overcome. By decomposing the original multi-class classification problem into a number of binary classification problems, outliers become potential negative samples of each binary classification problem. Therefore, classifiers can assign right pseudo-labels for outliers and include them as a normal part of training data to improve the performance of prediction model. However, as a cost for its conciseness, binary decomposition strategy brings new issues, i.e. class-imbalance and representation-compromise. We will describe our considerations towards these two issues in detail.

3.3.1. Adaptive Logit Adjustment for Class-Imbalanced Issue

The derived binary SSL data sets in Eq.(3) may suffer from a high degree of class-imbalance, which will lead to biased classifiers with low accuracy on minority classes. We will firstly take a deep look at the class-imbalance issue arising with the binary decomposition procedure, and then present our solution tailored for this specific binary SSL problem.

Let $N_c, M_c\, (1 \le c \le K)$ denote the number of samples for the c^{th} seen category in the original labeled set $\mathcal{D}^l$ and unlabeled set $\mathcal{D}^u$ respectively. And we denote the number of outliers in $\mathcal{D}^u$ by M_o. Then, imbalance ratios of the binary SSL data set $(\mathcal{D}_c^l, \mathcal{D}_c^u)$ can be derived as follows

$$\begin{aligned} \gamma_c^l &= \frac{|\mathcal{P}_c^l|}{|\mathcal{N}_c^l|} = \frac{N_c}{\sum_{i=1, i\neq c}^K N_i} \\ \gamma_c^u &= \frac{|\mathcal{P}_c^u|}{|\mathcal{N}_c^u|} = \frac{M_c}{\sum_{i=1, i\neq c}^K M_i + M_o}, \end{aligned} \tag{8}$$

where $|\cdot|$ denotes the cardinality of a set. It indicates that derived binary SSL data sets are generally imbalanced and class distributions are different between the labeled and unlabeled data. Particularly, suppose the original labeled and unlabeled sets are balanced (i.e. $N_1 = N_2 = \cdots = N_K$ and $M_1 = M_2 = \cdots = M_K$), then we have $\gamma_c^l = \frac{1}{K-1}$ and $\gamma_c^u < \frac{1}{K-1}$, which is highly imbalanced when the number of the seen categories is large.

To improve classification performance in such a challenging scenario, training techniques against class-imbalance are necessary. For balancing the supervised classification loss on labeled set, we replace the naive binary cross-entropy loss with a re-margining version (Ren et al., 2020; Menon et al., 2021) as follows

$$\mathcal{L}_{sup} = \sum_{c=1}^{K}\sum_{i=1}^{N} -y_i^c \log\sigma[f_c(\mathbf{x}_i^{(l)};\theta_c) + \tau\cdot\log\gamma_c^l] \\ -(1-y_i^c)\log\sigma[-f_c(\mathbf{x}_i^{(l)};\theta_c) - \tau\cdot\log\gamma_c^l], \quad (9)$$

where the imbalance ratio γ_c^l can be obtained from labeled set, and $\tau > 0$ is a scaling parameter controlling the strength of balance. The above re-margining loss encourages training process to pay more attention to the tail class (i.e. the positive class when $\gamma_c^l < 1$), so that bias of model can be alleviated.

However, it is not an easy work to balance the unsupervised regularization term on unlabeled data, since imbalance ratio of unlabeled set γ_c^u is unknown and different from γ_c^l. Instead of trying to estimate the concealed imbalance ratio γ_c^u, we attempt to capture the current learning state of each class. Intuitively, if learning state of the positive class lags behind the negative class, then subsequent learning process should be biased toward the positive class to confront the imbalance. Following (Lai et al., 2022; Wang et al., 2022a), we use the empirical prediction probability of classifier on unlabeled set, i.e. $\tilde{p}^c = \frac{1}{M}\sum_{i=1}^{M}\sigma(f_c(\mathbf{x}_i^{(u)};\theta_c))$, as a proxy of learning state of the positive class. And for the negative class, the proxy can be derived as $1-\tilde{p}^c$. To improve efficiency, we adopt a running averaging mechanism to approximate the proxy

$$\tilde{p}^c \leftarrow \mu\cdot\tilde{p}^c + (1-\mu)\cdot\frac{1}{|\mathcal{B}^{(u)}|}\sum_{\mathbf{x}^{(u)}\sim\mathcal{B}^{(u)}}\sigma(f_c(\mathbf{x}^{(u)};\theta_c)), \quad (10)$$

where $\mu \in [0,1]$ is a momentum factor and $\mathcal{B}^{(u)}$ denotes a batch of unlabeled data. With the proxy, we can derive a counterpart of the imbalance ratio of unlabeled set as follows

$$\tilde{\gamma}_c^u = \frac{\tilde{p}^c}{1-\tilde{p}^c}. \quad (11)$$

Recent works (Kim et al., 2020; Wei & Gan, 2023) find that predictions of the classifier induced on imbalanced data are biased toward the majority classes. Such bias may affect the quality of generated pseudo-labels and thus accumulates during the self-training process. To prevent accumulation of bias, we adapt the logit adjustment technique (Menon et al., 2021) with the proxy in Eq.(11) to adaptively adjust classifiers' outputs to be appropriately biased toward the minority classes. Specifically, we fix the prediction probability on unlabeled sample $\mathbf{x}_i^{(u)}$ by

$$\hat{p}_i^c = \sigma(f_c(\mathbf{x}_i^{(u)};\theta_c) - \tau\cdot\log\tilde{\gamma}_c^u). \quad (12)$$

The above adjusted prediction probability is used to replace the original prediction probability p_i^c to obtain the pseudo-label and mask used in Eq.(6).

3.3.2. Label-Specific Features for Representation-Compromise Issue

Existing SSL approaches generally obey the K-way softmax classification paradigm, where a linear softmax classifier transfers the representation $\mathbf{z}$ extracted by a backbone into prediction probabilities for each class label. Analogously, in binary decomposition scenario, one can achieve classification by attaching a group of linear one-vs-all classifiers to the identical representation $\mathbf{z}$ extracted by a backbone.

Though such an implementation is feasible, it might be suboptimal as it fails to consider that each binary problem may possess its own discriminative preferences. For example, in image annotation task, *color*-based features would be preferred in recognizing the sky and non-sky images, while they can be nuisance factors in discriminating the plane and non-plane images. Intuitively, these potentially distinct discriminative preferences will result in a compromised representation, if we tackle these binary problems with a shared feature extractor. Therefore, we hypothesize that if label-specific features, i.e. the most pertinent and discriminative features for each class label, could be used in the learning process, a more effective approach to solve these different binary problems could be achieved. The methodology can be formalized as

$$l^c = h_c(g_c(\mathbf{x})), \quad 1 \le c \le K \quad (13)$$

where each linear one-vs-all classifier h_c is fed with the tailored representation $\mathbf{z}^c$ (i.e. label-specific features) extracted by a feature extractor g_c specific to each binary problem, instead of the identical representation $\mathbf{z}$ shared among all binary problems.

Here, we present the implementation to learn label-specific features for image data, since existing studies on open-set SSL all evaluate their approaches on image annotation task. Implementation recommendations for text and tabular data classifications can be found in the appendix.

For image data $\mathbf{x}$, a feature map $\mathbf{Z}_{map} \in \mathbb{R}^{e\times h\times w}$ can be obtained by removing the top pooling layer of widely-used backbones (e.g. Wide ResNet-28-2 (Zagoruyko & Komodakis, 2016) and ResNet-18 (He et al., 2016)). After flattening the spatial dimension of the feature map, we adopt a dot-product attention module to extract the most pertinent and discriminative features for each binary problem

$$\mathbf{z}^c = \mathbf{Z}_{map}^f \cdot softmax(\frac{\mathbf{K}^{\mathrm{T}}\mathbf{r}^c}{\sqrt{e}}) \\ with \quad \mathbf{Z}_{map}^f = Flatten(\mathbf{Z}_{map}) \in \mathbb{R}^{e\times(h\cdot w)}, \quad (14)$$

where $\mathbf{K} = \mathbf{W}_{key}\mathbf{Z}_{map}^{f} \in \mathbb{R}^{e\times(h\cdot w)}$ is the key matrix and $\mathbf{W}_{key}$ denotes the parameter of a linear layer. $\mathbf{R} = [\mathbf{r}^1, \mathbf{r}^2, \ldots, \mathbf{r}^K] \in \mathbb{R}^{e\times K}$ is a learnable query matrix modelling the discriminative preferences of each binary problem. The above module can be regarded as a weighted-averaging pooling layer which uses different weights to aggregate the feature map $\mathbf{Z}_{map}$ for each class label, while the compromised representation $\mathbf{z}$ is simply obtained by aggregating the feature map $\mathbf{Z}_{map}$ with uniform weights.

3.3.3. Inference

For an unseen sample $\mathbf{u} \in \mathcal{X}$, the inference procedure is rather simple, which can be formalized as follows

$$y = \begin{cases} K+1, & if\, p_m < 0.5 \\ \arg\max_{1\leq c\leq K} \sigma(f_c(\mathbf{u};\theta_c)), & otherwise \end{cases}$$
$$with \quad p_m = \max_{1\leq c\leq K} \sigma(f_c(\mathbf{u};\theta_c)), \tag{15}$$

where $y = K+1$ means the prediction model assigns unseen sample $\mathbf{u}$ to a novel category, since all the one-vs-all classifiers predict it as a negative sample.

4. Experiments

4.1. Experimental Setup

4.1.1. Data Sets

Following (Saito et al., 2021; Fan et al., 2023; Li et al., 2023), experiments are conducted on benchmark data sets CIFAR-10, CIFAR-100 (Krizhevsky et al., 2009), and ImageNet (Deng et al., 2009) with different numbers of labeled data and diversified open-set settings.

CIFAR-10: The 6 animal classes are used as seen categories and other 4 classes as unseen categories. For each seen category, we randomly select 4 or 25 samples from training set as the labeled set. The rest of training set is used as the unlabeled set. For evaluation, we report algorithm performance on test set of CIFAR-10.

CIFAR-100: Seen/unseen categories are split on the hierarchy of super classes. Three settings are considered by using the first 4, 10, or 16 super-classes as seen categories, resulting in seen/unseen category split of 20/80, 50/50, and 80/20 respectively. For each seen category, we randomly select 4 or 25 samples from training set as the labeled set. The rest of training set is used as the unlabeled set. For evaluation, we report algorithm performance on test set of CIFAR-100.

ImageNet: Following (Saito et al., 2021; Fan et al., 2023; Li et al., 2023), we use ImageNet-30, which is a subset of ImageNet containing 30 categories. The first 20 classes are used as seen categories and other 10 classes as unseen categories. For each seen category, we randomly select 1% or 5% samples from training set as the labeled set (13 or 65 samples per category respectively). The rest of training set is used as the unlabeled set. For evaluation, we report algorithm performance on test set of ImageNet-30.

4.1.2. Evaluation Metrics

For comprehensive performance evaluation, we follow (Li et al., 2023) to evaluate an algorithm with the following two evaluation metrics.

Inlier accuracy: It is the test accuracy on seen categories, which reflects the ability of an algorithm to utilize open-set unlabeled set for improving classification performance on inliers.

Overall accuracy: It measures classification performance on the whole open-set test set including both seen and unseen categories, which reflects the real performance of an algorithm in open-set environment. In evaluation, all unseen categories are regarded as a single category, i.e. the $(K+1)^{th}$ category. Since open-set test set may have class-imbalance problem, the balanced accuracy (BA) (Brodersen et al., 2010) is calculated, which is defined as follows

$$BA = \frac{1}{K+1}\sum_{c=1}^{K+1} Recall^c, \tag{16}$$

where $Recall^c$ denotes the recall score of the c^{th} category.

4.1.3. Implementation Details

Following (Li et al., 2023), we use Wide ResNet-28-2 (Zagoruyko & Komodakis, 2016) as the backbone for experiments on CIFAR-10/100, and ResNet-18 (He et al., 2016) for ImageNet. The scaling parameter τ controlling the strength of balance is set as 0.5 and the momentum factor μ in Eq.(10) is set as 0.999. Models are all trained by the SGD optimizer and the learning rate is set as 0.03 with a cosine decay. For CIFAR experiments, models are trained for $256 * 1024$ iterations and each iteration contains a batch of 64 labeled samples and $64 * 7$ unlabeled samples. For ImageNet experiments, models are trained for $100 * 1024$ iterations and each iteration contains a batch of 32 labeled samples and 32 unlabeled samples.

To tackle class-imbalance, we find it to be helpful to adopt the dual-branch architecture (Kang et al., 2020; Wei & Gan, 2023) during training, where two groups of one-vs-all classifiers are learned simultaneously on shared backbone. One group of classifiers (balanced classifiers) is trained with the adaptive logit adjustment mechanism, responsible for learning unbiased classifiers. While another group of classifiers (standard classifiers) is trained with the standard consistency regularization framework in Eq.(4) to elicit good representation from the original data distribution. After training, we only use balanced classifiers for inference.

Table 1. Inlier accuracy of each comparing approach (mean±std. deviation) on data sets with varying seen/unseen class splits and labeled set sizes. Best results are highlighted in **boldface**.

	Data set		CIFAR-10		CIFAR-100						ImageNet-30	
	Class split (Seen / Unseen)		6 / 4		20 / 80		50 / 50		80 / 20		20 / 10	
	Number of labels per class		4	25	4	25	4	25	4	25	1%	5%
Standard SSL	MixMatch	NeurIPS'19	43.08±1.79	63.13±0.64	28.13±5.06	51.28±1.45	26.97±0.46	56.93±0.84	28.35±0.83	53.77±0.97	-	-
	ReMixMatch	ICLR'20	72.82±1.81	87.08±1.12	36.02±3.56	61.83±0.81	37.57±1.54	65.80±1.33	40.64±2.97	62.90±1.07	-	-
	FixMatch	NeurIPS'20	81.58±6.63	92.94±0.80	46.27±0.64	66.45±0.74	48.93±5.05	68.77±0.89	43.06±1.21	64.44±0.51	52.52±3.82	78.55±1.46
	CoMatch	ICCV'21	86.08±1.08	92.57±0.47	43.53±3.01	66.82±1.37	43.17±0.55	67.85±1.17	37.89±1.22	62.04±0.08	62.92±0.90	79.17±0.42
	FlexMatch	NeurIPS'21	73.34±4.42	86.44±3.72	37.93±4.49	62.68±2.02	44.10±1.88	68.98±0.94	43.44±2.40	64.34±0.64	-	-
	SimMatch	CVPR'22	79.84±4.76	90.07±2.44	36.93±5.72	67.23±1.13	51.53±2.02	69.71±1.44	50.32±2.57	65.68±1.43	64.15±0.94	80.23±0.53
	FreeMatch	ICLR'23	79.26±4.11	92.27±0.15	45.18±8.36	64.62±0.79	50.26±1.92	68.57±0.27	47.34±0.57	64.41±0.55	-	-
Open-Set SSL	UASD	AAAI'20	35.25±1.07	56.42±1.34	29.78±4.28	53.78±0.67	29.08±1.44	54.24±1.10	26.41±2.16	50.33±0.62	-	-
	DS3L	ICML'20	39.09±1.24	51.83±1.06	19.70±1.98	41.78±1.45	21.62±0.54	47.41±0.61	20.10±0.48	40.51±1.02	-	-
	MTCF	ECCV'20	49.15±6.12	74.42±2.95	32.58±3.36	55.93±1.66	35.35±2.39	57.72±0.20	25.40±1.20	54.59±0.49	-	-
	T2T	ICCV'21	73.89±1.55	85.69±1.90	44.23±2.27	65.60±0.71	39.31±1.16	68.59±0.92	38.16±0.59	63.86±0.32	63.70±0.83	78.87±0.49
	OpenMatch	NeurIPS'21	43.63±3.26	66.27±1.86	37.45±2.67	62.70±1.76	33.74±0.38	66.53±0.54	28.54±1.15	61.23±0.81	56.35±3.35	73.90±1.05
	SAFE-STUDENT	CVPR'22	59.28±1.18	77.87±0.14	34.53±0.67	58.07±1.40	35.84±0.86	62.75±0.38	34.17±0.69	57.99±0.34	58.38±2.34	75.85±0.99
	SSB	ICCV'23	70.93±1.73	93.52±0.23	45.25±3.06	66.30±0.83	49.87±2.64	68.86±0.82	45.89±1.58	44.85±1.05	42.07±2.45	78.08±3.28
	OSP	CVPR'23	78.14±0.54	85.41±3.56	43.65±0.97	66.13±1.37	43.14±1.35	68.69±1.11	37.27±2.24	62.55±0.23	72.12±1.50	78.45±0.48
	IOMatch	ICCV'23	89.68±2.04	**93.87±0.16**	53.73±2.12	67.28±1.10	56.31±2.29	69.77±0.58	50.83±0.99	64.75±0.52	69.18±1.68	81.43±0.78
	BDMatch	Ours	**91.77±1.62**	93.86±0.57	**57.40±2.44**	**69.58±0.75**	**59.37±1.57**	**71.79±0.16**	**52.30±1.43**	**66.53±0.47**	**75.18±0.98**	**82.88±1.35**

Table 2. Overall accuracy of each comparing approach (mean±std. deviation) on data sets with varying seen/unseen class splits and labeled set sizes. Best results are highlighted in **boldface**.

	Data set		CIFAR-10		CIFAR-100						ImageNet-30	
	Class split (Seen / Unseen)		6 / 4		20 / 80		50 / 50		80 / 20		20 / 10	
	Number of labels per class		4	25	4	25	4	25	4	25	1%	5%
Open-Set SSL	UASD	AAAI'20	17.10±0.32	36.01±0.22	10.50±0.83	26.96±0.53	6.92±0.55	32.23±0.54	5.77±0.21	27.61±1.15	-	-
	DS3L	ICML'20	30.89±0.33	40.45±0.77	12.56±1.21	34.35±0.41	12.14±0.39	35.17±0.48	11.10±1.27	29.09±0.31	-	-
	MTCF	ECCV'20	33.35±7.21	46.13±0.54	8.12±2.10	26.60±3.66	4.13±0.37	38.36±0.29	1.46±0.17	30.75±0.52	-	-
	T2T	ICCV'21	50.57±0.38	61.10±0.39	17.17±1.37	37.18±0.60	12.74±2.66	44.24±0.42	34.23±0.57	51.41±0.96	48.81±0.88	58.51±0.41
	OpenMatch	NeurIPS'21	14.37±0.05	20.35±3.50	8.77±2.84	39.89±1.16	7.00±0.02	49.75±1.08	6.30±0.87	44.83±0.62	21.80±1.90	57.25±0.76
	SAFE-STUDENT	CVPR'22	45.27±0.36	52.78±0.64	15.94±1.07	28.83±0.46	23.98±0.88	46.71±1.74	29.43±0.66	50.48±0.61	44.08±2.09	55.25±1.46
	SSB	ICCV'23	62.87±1.80	62.98±3.32	25.76±2.78	33.28±1.68	26.61±1.56	36.25±3.83	19.61±1.10	24.68±0.15	25.81±1.96	62.71±0.44
	OSP	CVPR'23	66.32±3.67	75.23±3.48	35.61±0.86	63.52±1.49	41.01±2.00	66.38±0.82	32.23±3.56	56.27±3.68	52.62±1.51	62.91±1.64
	IOMatch	ICCV'23	75.08±1.92	78.96±0.08	45.94±1.70	58.52±0.48	46.36±1.93	60.78±0.71	39.96±0.95	54.39±0.38	57.71±2.69	73.94±0.99
	BDMatch	Ours	**78.31±0.89**	**79.81±0.57**	**50.46±2.14**	**63.71±0.87**	**55.17±1.52**	**68.20±0.58**	**48.09±1.92**	**63.14±0.19**	**65.48±1.03**	**74.17±1.38**

4.2. Comparative Studies

We compare BDMatch[2] with existing SSL approaches, including seven standard SSL approaches and nine open-set SSL approaches. For fair comparison, experiments are conducted on a unified codebase based on USB (Wang et al., 2022b). Common hyperparameters sharing among different approaches are set as the same values, while other hyperparameters specific to each approach are set by parameter configurations suggested in respective literatures. Experimental results are averaged across three independent runs with different random seeds.

Table 1 and Table 2 report detailed experimental results in terms of inlier accuracy and overall accuracy respectively. Based on these results, it is impressive to observe that

- For inlier classification, BDMatch achieves superior performance to existing standard SSL approaches as well as open-set SSL approaches. For example, when labels are extremely scarce, BDMatch outperforms the second best approach with up to 2.09%, 3.67%, 3.06%, 1.47%, and 3.06% absolute increase in test accuracy on CIFAR-10-6-4[3], CIFAR-100-20-4, CIFAR-100-50-4, CIFAR-100-80-4, ImageNet-30-20-p1 respectively. These consistently better results demonstrate the effectiveness of our binary decomposition strategy for open-set SSL.
- Meantime, across various open-set SSL settings, BDMatch significantly outperforms existing SSL approaches in terms of overall accuracy, which is a more comprehensive evaluation metric to measure the ability of an algorithm to achieve both good inlier classification and accurate outlier detection. For example, on CIFAR-100-80-25 and ImageNet-30-20-p1, BDMatch improves the existing best results by 12.21% and 13.46% respectively. And performance of BDMatch keeps relatively stable as the number of labeled samples decreases, while most of existing open-set SSL approaches suffer severe performance degradation when labels are extremely scarce.

We also consider a cross-dataset open-set evaluation setting, where the test set contains samples from categories that do

[2]Code package of BDMatch is publicly available at: `http://palm.seu.edu.cn/zhangml/files/BDMatch.rar`.

[3]For briefness, the open-set setting on CIFAR-10 with 6 seen categories and 4 labeled samples per category is denoted by CIFAR-10-6-4. Other open-set settings are abbreviated similarly.

Table 3. Predictive performance of BDMatch and its variants (mean±std. deviation) in terms of inlier accuracy with a fixed random seed. LA is the abbreviation of logit adjustment.

Algorithm	CIFAR-100-20-4	CIFAR-100-80-4
BDMatch	**58.05**	**53.83**
(a) w.o. Outliers (Purified data)	54.75	52.36
(b) w.o. Label-specific features	54.45	51.24
(c) w.o. Dual-branch architecture	55.85	52.19
(d) w.o. Proxy $\tilde{\gamma}_c^u$ for adaptive LA	54.50	49.28
(e) w.o. Non-adaptive LA with γ_c^l	52.55	43.86
(f) w.o. Re-margining in Eq.(9)	55.70	50.14

not appear even in the unlabeled training set. Following (Li et al., 2023), we expand the original test set with foreign outliers from data sets, including SVHN (Yuval, 2011), LSUN (Yu et al., 2015), Synthetic Gaussian (Yu et al., 2020), and Synthetic Uniform (Yu et al., 2020). Table A.1 reports detailed experimental results in this evaluation setting, which shows that BDMatch can still achieve consistently superior open-set classification performance to existing open-set SSL approaches.

4.3. Further Analyses

4.3.1. Ablation Studies

Comprehensive ablation studies are conducted to have a deeper look at important designing elements in BDMatch.

Consideration on representation-compromise issue. BDMatch accounts for each binary problem's own discriminative preferences via learning label-specific features. To validate the effectiveness of this designing element, we implement a variant of BDMatch, which removes the dot-product attention module in Eq.(14) and directly performs classification on the identical representation $\mathbf{z}$ extracted by the backbone. As shown in Table 3(b), without consideration on representation-compromise issue, performance of this variant lags behind BDMatch significantly.

Consideration on class-imbalance issue. We tackle class-imbalance issue with an adaptive logit adjustment mechanism and the dual-branch architecture. To examine the effectiveness of these designing elements, we remove them successively and derive a series of variants. We firstly remove the dual-branch architecture. As shown in Table 3(c), dual-branch architecture is essential to confront class-imbalance. Then, we replace the adaptive logit adjustment mechanism with a non-adaptive one by directly using the imbalance ratio γ_c^l of labeled set as a substitute of the adaptive proxy $\tilde{\gamma}_c^u$ of unlabeled set (i.e. variant (d) in Table 3). While variant (e) further removes the non-adaptive logit adjustment mechanism, which is actually a baseline without any consideration on class-imbalance. We also implement an variant (f) which removes the re-margining loss for balancing supervised classification on labeled set. Results in Table 3 show

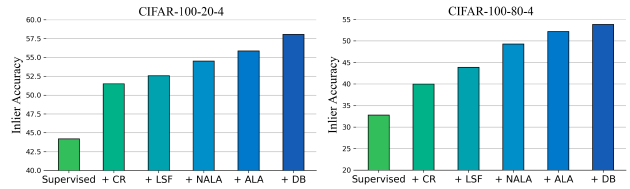

Figure 2. Ablation studies of BDMatch. Based on a supervised baseline, we successively add consistency regularization (CR), label-specific features (LSF), non-adaptive logit adjustment (NALA), adaptive logit adjustment (ALA), and dual-branch architecture (DB).

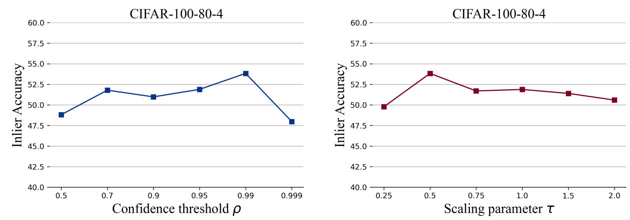

Figure 3. Predictive performance of BDMatch with varying confidence threshold ρ and scaling parameter τ.

that the class-imbalance issue is worth careful treatment and the adaptive proxy is quite effective.

Are outliers utilized effectively by the binary decomposition strategy? We implement a variant of BDMatch, which is trained on purified data set containing only inliers. Such a purified data set is obtained by an oracle detection function. As shown in Table 3(a), BDMatch achieves better result than this variant model, which demonstrates that our binary decomposition strategy does have the ability to utilize outliers to improve classification performance.

Figure 2 gives an illustrative example on how the performance of a supervised baseline, which learned from only the labeled data, gradually improves when we successively add designing elements in BDMatch.

4.3.2. Parameter Sensitivity

Figure 3 gives illustrative examples on how the performance of BDMatch changes when the values of confidence threshold ρ (in Eq.(7)) and scaling parameter τ (in Eq.(12)) change. When $\rho = 0.5$, no sample is masked out during training. So confirmation bias accumulates and thus leads to degraded performance. And scaling parameter τ should be treated carefully, since both under-balancing and over-balancing may be harmful. For practicers, we recommend a high enough confidence threshold ρ (e.g. 0.99) and a proper scaling parameter τ around 0.5, which can obtain reasonable results according to our studies.

5. Conclusion

In this paper, we propose to tackle open-set semi-supervised learning with a novel binary decomposition strategy. Our strategy directly decomposes the original open-set SSL problem into a number of standard binary SSL problems, and thus refrains from error-prone procedure of outlier detection in existing detect-and-filter strategy. Following this binary decomposition strategy, we present an open-set SSL approach BDMatch, which confronts two attendant issues, i.e. class-imbalance and representation-compromise, with an adaptive logit adjustment mechanism and a label-specific feature learning mechanism respectively. Comprehensive experiments against current competing algorithms show the superiority of our approach and indicate that the binary decomposition strategy is a promising direction for future studies.

Acknowledgements

The authors wish to thank the anonymous reviewers for their helpful comments and suggestions. This work was supported by the National Science Foundation of China (62225602), the Fundamental Research Funds for the Central Universities (2242024K30035), and the Big Data Computing Center of Southeast University.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Arazo, E., Ortego, D., Albert, P., O'Connor, N. E., and McGuinness, K. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In *Proceedings of the International Joint Conference on Neural Networks*, pp. 1–8, Glasgow, UK, 2020.

Berthelot, D., Carlini, N., Cubuk, E. D., Kurakin, A., Sohn, K., Zhang, H., and Raffel, C. ReMixMatch: Semi-supervised learning with distribution matching and augmentation anchoring. In *Proceedings of the 8th International Conference on Learning Representations*, Addis Ababa, Ethiopia, 2020.

Brodersen, K. H., Ong, C. S., Stephan, K. E., and Buhmann, J. M. The balanced accuracy and its posterior distribution. In *Proceedings of the 20th International Conference on Pattern Recognition*, pp. 3121–3124, Istanbul, Turkey, 2010.

Chen, H., Tao, R., Fan, Y., Wang, Y., Wang, J., Schiele, B., Xie, X., Raj, B., and Savvides, M. SoftMatch: Addressing the quantity-quality tradeoff in semi-supervised learning. In *Proceedings of the 11th International Conference on Learning Representations*, Kigali, Rwanda, 2023.

Chen, Y., Zhu, X., Li, W., and Gong, S. Semi-supervised learning under class distribution mismatch. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, pp. 3569–3576, New York, NY, 2020.

Cubuk, E. D., Zoph, B., Shlens, J., and Le, Q. V. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3008–3017, Seattle, WA, 2020.

Cui, Y., Jia, M., Lin, T.-Y., Song, Y., and Belongie, S. J. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9268–9277, Long Beach, CA, 2019.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 248–255, Miami, Florida, 2009.

Du, P., Zhao, S., Sheng, Z., Li, C., and Chen, H. Semi-supervised learning via weight-aware distillation under class distribution mismatch. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16364–16374, Paris, France, 2023.

Engelen, J. E. V. and Hoos, H. H. A survey on semi-supervised learning. *Machine Learning*, 109(2):373–440, 2020.

Fan, Y., Kukleva, A., Dai, D., and Schiele, B. SSB: Simple but strong baseline for boosting performance of open-set semi-supervised learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16022–16032, Paris, France, 2023.

Guo, L.-Z. and Li, Y.-F. Class-imbalanced semi-supervised learning with adaptive thresholding. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 8082–8094, Baltimore, MD, 2022.

Guo, L.-Z., Zhang, Z., Jiang, Y., Li, Y.-F., and Zhou, Z.-H. Safe deep semi-supervised learning for unseen-class unlabeled data. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 3897–3906, Virtual Event, 2020.

Guo, Y., Chung, F., Li, G., Wang, J., and Gee, J. C. Leveraging label-specific discriminant mapping features for multi-label learning. *ACM Transactions on Knowledge Discovery from Data*, 13(2):24:1–24:23, 2019.

Hang, J.-Y. and Zhang, M.-L. Collaborative learning of label semantics and deep label-specific features for multi-label classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9860–9871, 2022a.

Hang, J.-Y. and Zhang, M.-L. Dual perspective of label-specific feature learning for multi-label classification. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 8375–8386, Baltimore, MD, 2022b.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, 2016.

He, R., Han, Z., Lu, X., and Yin, Y. Safe-Student for safe deep semi-supervised learning with unseen-class unlabeled data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14565–14574, New Orleans, LA, 2022a.

He, R., Han, Z., Yang, Y., and Yin, Y. Not all parameters should be treated equally: Deep safe semi-supervised learning under class distribution mismatch. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, pp. 6874–6883, Virtual Event, 2022b.

Huang, J., Li, G., Huang, Q., and Wu, X. Learning label-specific features and class-dependent labels for multi-label classification. *IEEE Transactions on Knowledge and Data Engineering*, 28(12):3309–3323, 2016.

Huang, J., Fang, C., Chen, W., Chai, Z., Wei, X., Wei, P., Lin, L., and Li, G. Trash to treasure: Harvesting OOD data with cross-modal matching for open-set semi-supervised learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8290–8299, Montréal, Canada, 2021.

Huang, Z., Sidhom, M.-J., Wessler, B., and Hughes, M. C. Fix-A-Step: Semi-supervised learning from uncurated unlabeled data. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 8373–8394, Valencia, Spain, 2023a.

Huang, Z., Xia, X., Shen, L., Han, B., Gong, M., Gong, C., and Liu, T. Harnessing out-of-distribution examples via augmenting content and style. In *Proceedings of the 11th International Conference on Learning Representations*, Kigali, Rwanda, 2023b.

Huang, Z., Yang, J., and Gong, C. They are not completely useless: Towards recycling transferable unlabeled data for class-mismatched semi-supervised learning. *IEEE Transactions on Multimedia*, 25:1844–1857, 2023c.

Jia, B.-B., Liu, J.-Y., Hang, J.-Y., and Zhang, M.-L. Learning label-specific features for decomposition-based multi-class classification. *Frontiers of Computer Science*, 17 (6):176348, 2023.

Kang, B., Xie, S., Rohrbach, M., Yan, Z., Gordo, A., Feng, J., and Kalantidis, Y. Decoupling representation and classifier for long-tailed recognition. In *Proceedings of the 8th International Conference on Learning Representations*, Addis Ababa, Ethiopia, 2020.

Kharbanda, S., Banerjee, A., Schultheis, E., and Babbar, R. CascadeXML: Rethinking transformers for end-to-end multi-resolution training in extreme multi-label classification. In *Advances in Neural Information Processing Systems 35*, pp. 2074–2087, New Orleans, LA, 2022.

Kim, J., Hur, Y., Park, S., Yang, E., Hwang, S. J., and Shin, J. Distribution aligning refinery of pseudo-label for imbalanced semi-supervised learning. In *Advances in Neural Information Processing Systems 33*, pp. 14567–14579, virtual, 2020.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.

Lai, Z., Wang, C., Gunawan, H., Cheung, S.-C. S., and Chuah, C.-N. Smoothed adaptive weighting for imbalanced semi-supervised learning: Improve reliability against unknown distribution data. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 11828–11843, Baltimore, MD, 2022.

Lee, H., Shin, S., and Kim, H. ABC: Auxiliary balanced classifier for class-imbalanced semi-supervised learning. In *Advances in Neural Information Processing Systems 34*, pp. 7082–7094, virtual, 2021.

Li, J., Xiong, C., and Hoi, S. C. H. CoMatch: Semi-supervised learning with contrastive graph regularization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9455–9464, Montréal, Canada, 2021.

Li, Z., Qi, L., Shi, Y., and Gao, Y. IOMatch: Simplifying open-set semi-supervised learning with joint inliers and outliers utilization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 15824–15833, Paris, France, 2023.

Liu, W., Wang, H., Shen, X., and Tsang, I. W. The emerging trends of multi-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7955–7974, 2022.

Ma, Q., Gao, J., Zhan, B., Guo, Y., Zhou, J., and Wang, Y. Rethinking safe semi-supervised learning: Transferring the open-set problem to a close-set one. In *Proceedings* *of the IEEE/CVF International Conference on Computer Vision*, pp. 16324–16333, Paris, France, 2023.

Menon, A. K., Jayasumana, S., Rawat, A. S., Jain, H., Veit, A., and Kumar, S. Long-tail learning via logit adjustment. In *Proceedings of the 9th International Conference on Learning Representations*, Virtual Event, 2021.

Miyato, T., ichi Maeda, S., Koyama, M., and Ishii, S. Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41 (8):1979–1993, 2019.

Mo, S., Su, J.-C., Ma, C.-Y., Assran, M., Misra, I., Yu, L., and Bell, S. RoPAWS: Robust semi-supervised representation learning from uncurated data. In *Proceedings of the 11th International Conference on Learning Representations*, Kigali, Rwanda, 2023.

Nassar, I., Hayat, M., Abbasnejad, E., Rezatofighi, H., and Haffari, G. Protocon: Pseudo-label refinement via online clustering and prototypical consistency for efficient semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11641–11650, Vancouver, Canada, 2023.

Oliver, A., Odena, A., Raffel, C., Cubuk, E. D., and Goodfellow, I. J. Realistic evaluation of deep semi-supervised learning algorithms. In *Advances in Neural Information Processing Systems 31*, pp. 3239–3250, Montréal, Canada, 2018.

Ren, J., Yu, C., Sheng, S., Ma, X., Zhao, H., Yi, S., and Li, H. Balanced meta-softmax for long-tailed visual recognition. In *Advances in Neural Information Processing Systems 33*, pp. 4175–4186, virtual, 2020.

Saito, K., Kim, D., and Saenko, K. OpenMatch: Open-set semi-supervised learning with open-set consistency regularization. In *Advances in Neural Information Processing Systems 34*, pp. 25956–25967, virtual, 2021.

Shi, J., Wu, T., Yu, H., Qin, A. K., Jeon, G., and Lei, Y. Multi-layer composite autoencoders for semi-supervised change detection in heterogeneous remote sensing images. *Science China Information Sciences*, 66(4), 2023.

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C., Cubuk, E. D., Kurakin, A., and Li, C.-L. FixMatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems 33*, pp. 596–608, virtual, 2020.

Wang, R., Jia, X., Wang, Q., Wu, Y., and Meng, D. Imbalanced semi-supervised learning with bias adaptive classifier. In *Proceedings of the 11th International Conference on Learning Representations*, Kigali, Rwanda, 2023a.

Wang, X., Wu, Z., Lian, L., and Yu, S. X. Debiased learning from naturally imbalanced pseudo-labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14627–14637, New Orleans, LA, 2022a.

Wang, Y., Chen, H., Fan, Y., Sun, W., Tao, R., Hou, W., Wang, R., Yang, L., Zhou, Z., Guo, L.-Z., Qi, H., Wu, Z., Li, Y.-F., Nakamura, S., Ye, W., Savvides, M., Raj, B., Shinozaki, T., Schiele, B., Wang, J., Xie, X., and Zhang, Y. USB: A unified semi-supervised learning benchmark for classification. In *Advances in Neural Information Processing Systems 35*, pp. 3938–3961, New Orleans, LA, 2022b.

Wang, Y., Chen, H., Heng, Q., Hou, W., Fan, Y., Wu, Z., Wang, J., Savvides, M., Shinozaki, T., Raj, B., Schiele, B., and Xie, X. FreeMatch: Self-adaptive thresholding for semi-supervised learning. In *Proceedings of the 11th International Conference on Learning Representations*, Kigali, Rwanda, 2023b.

Wang, Y., Qiao, P., Liu, C., Song, G., Zheng, X., and Chen, J. Out-of-distributed semantic pruning for robust semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23849–23858, Vancouver, Canada, 2023c.

Wei, C., Sohn, K., Mellina, C., Yuille, A. L., and Yang, F. CReST: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10857–10866, virtual, 2021.

Wei, T. and Gan, K. Towards realistic long-tailed semi-supervised learning: Consistency is all you need. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3469–3478, Vancouver, Canada, 2023.

Wu, X., Jiang, B., Yu, K., Chen, H., and Miao, C. Multi-label causal feature selection. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pp. 6430–6437, New York, NY, 2020.

Xie, Q., Dai, Z., Hovy, E. H., Luong, T., and Le, Q. Unsupervised data augmentation for consistency training. In *Advances in Neural Information Processing Systems 33*, pp. 6256–6268, virtual, 2020.

Xu, Y., Shang, L., Ye, J., Qian, Q., Li, Y., Sun, B., Li, H., and Jin, R. Dash: Semi-supervised learning with dynamic thresholding. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 11525–11536, Virtual Event, 2021.

Yang, X., Song, Z., King, I., and Xu, Z. A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(9):8934–8954, 2023.

You, R., Zhang, Z., Wang, Z., Dai, S., Mamitsuka, H., and Zhu, S. AttentionXML: Label tree-based attention-aware deep model for high-performance extreme multi-label text classification. In *Advances in Neural Information Processing Systems 32*, pp. 5812–5822, Vancouver, Canada, 2019.

Yu, F., Zhang, Y., Song, S., Seff, A., and Xiao, J. LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint*, arXiv:1506.03365, 2015.

Yu, Q., Ikami, D., Irie, G., and Aizawa, K. Multi-task curriculum framework for open-set semi-supervised learning. In *Proceedings of the 16th European Conference of Computer Vision*, pp. 438–454, Glasgow, UK, 2020.

Yuval, N. Reading digits in natural images with unsupervised feature learning. In *Proceedings of the NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.

Zagoruyko, S. and Komodakis, N. Wide residual networks. In *Proceedings of the British Machine Vision Conference*, York, UK, 2016.

Zhang, B., Wang, Y., Hou, W., Wu, H., Wang, J., Okumura, M., and Shinozaki, T. FlexMatch: Boosting semi-supervised learning with curriculum pseudo labeling. In *Advances in Neural Information Processing Systems 34*, pp. 18408–18419, virtual, 2021.

Zhang, M.-L. and Wu, L. LIFT: Multi-label learning with label-specific features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1):107–120, 2015.

Zhang, M.-L. and Zhou, Z.-H. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837, 2014.

Zhang, Y., Kang, B., Hooi, B., Yan, S., and Feng, J. Deep long-tailed learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10795–10816, 2023.

Zhao, Z., Zhou, L., Duan, Y., Wang, L., Qi, L., and Shi, Y. DC-SSL: Addressing mismatched class distribution in semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9747–9755, New Orleans, LA, 2022.

Zheng, M., You, S., Huang, L., Wang, F., Qian, C., and Xu, C. SimMatch: Semi-supervised learning with similarity matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14451–14461, New Orleans, LA, 2022.

A. Cross-Dataset Open-Set Evaluation

In this section, we consider a cross-dataset open-set evaluation setting, where the test set contains samples from categories that do not appear even in the unlabeled training set. Such an evaluation setting aims to measure the generalization ability of an algorithm to outliers from brand-new categories. Following (Li et al., 2023), we expand the original test set with foreign outliers from data sets, including SVHN (Yuval, 2011), LSUN (Yu et al., 2015), Synthetic Gaussian (Yu et al., 2020), and Synthetic Uniform (Yu et al., 2020). Table A.1 reports detailed experimental results in this evaluation setting, which shows that BDMatch can still achieve consistently superior open-set classification performance to existing open-set SSL approaches.

Table A.1. Overall accuracy of each comparing approach (mean±std. deviation) on cross-dataset open-set test sets with varying seen/unseen class splits and labeled set sizes. Best results are highlighted in **boldface**.

	Data set		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	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
	DS3L	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
	MTCF	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
	T2T	ICCV'21	51.35±1.76	61.78±0.89	17.82±1.57	37.78±0.73	12.33±1.87	43.86±0.71	34.45±0.67	51.77±1.03
	OpenMatch	NeurIPS'21	14.37±0.05	20.31±3.49	8.77±2.83	39.96±1.17	9.97±0.37	49.56±1.15	6.31±0.88	44.77±0.58
	SAFE-STUDENT	CVPR'22	46.37±0.61	54.23±0.42	16.31±0.88	29.44±0.56	23.31±0.93	46.91±1.42	29.52±0.55	50.83±0.41
	SSB	ICCV'23	64.91±3.47	55.18±3.53	25.84±3.26	33.56±1.13	26.88±1.30	36.29±4.29	19.42±1.09	24.56±0.18
	OSP	CVPR'23	64.75±1.54	76.81±2.72	35.02±0.66	63.45±1.50	41.13±1.82	66.29±0.86	32.18±3.37	56.31±3.50
	IOMatch	ICCV'23	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
	BDMatch	Ours	**82.90±0.56**	**83.48±2.99**	**51.47±1.96**	**64.05±0.85**	**56.09±1.59**	**69.05±0.73**	**48.55±2.13**	**63.71±0.08**

B. Further Discussion and Comparison with Related Works

Although OpenMatch (Saito et al., 2021), IOMatch (Li et al., 2023) and BDMatch all have one-vs-all classifiers in their prediction models, the methodologies for tackling open-set SSL problem are totally different.

OpenMatch obeys detect-and-filter strategy. Accordingly, its one-vs-all classifiers are used as parameterized detector to filter out outliers in unlabeled set so that another K-way softmax classifier can be induced on purified unlabeled set. IOMatch transfers K-way open-set SSL problem to $(K+1)$-way closed-set one, where one-vs-all classifiers are used to generate pseudo-labels with heuristics for training a closed-set $(K+1)$-way softmax classifier. While our approach follows a concise binary decomposition strategy, which both avoids error-prone outlier detection process in OpenMatch and has no need to design complicated heuristics for constructing pseudo-labels in IOMatch.

Table B.1 reports further experimental results on the superiority of binary decomposition strategy. To show how much the performance of detect-and-filter approaches is affected by error-prone outlier detection, we conduct an experiment which provides oracle detection function for these approaches. We also analyze how much the performance of a $(K+1)$-way classifier can be improved if ground-truth $(K+1)$-way labels of outliers are provided. Significant performance improvement is witnessed, which demonstrates both outlier detection and pseudo-label construction are obstacles for good performance.

Table B.1. Further experimental results on the superiority of binary decomposition strategy. Experiments are conducted on CIFAR-100-80-4.

Algorithm	Inlier accuracy	Overall accuracy
OpenMatch (detect-and-filter strategy)	28.75	6.42
OpenMatch w. oracle detection function	49.23	49.85
OSP (detect-and-filter strategy)	37.46	28.76
OSP w. oracle detection function	47.40	48.05
IOMatch ($(K+1)$-way classification)	51.09	40.49
IOMatch w. oracle labels of outliers	56.73	57.26
BDMatch	53.83	50.14

C. Label-Specific Feature Learning

Label-specific feature learning works as an effective technique to conciliate the learning conflicts among multiple classification problems in multi-label learning (Zhang & Zhou, 2014; Liu et al., 2022). The basic idea is to facilitate the discrimination of each class label by tailoring its own features, instead of sharing an identical feature among all class labels. A theoretical interpretation of this technique has been provided in recent study (Hang & Zhang, 2022b). For tabular data classification, label-specific features can be constructed either by feature transformation (Zhang & Wu, 2015; Guo et al., 2019; Jia et al., 2023), or by feature selection (Huang et al., 2016; Wu et al., 2020; Hang & Zhang, 2022a). While for text classification, we would recommend the attention mechanism (You et al., 2019; Kharbanda et al., 2022) to mining the most pertinent features for each classification problem.

In this paper, we make a first attempt to introduce this technique and validate its effectiveness for open-set SSL problem. With the binary decomposition strategy, the original open-set problem is transformed into multiple different binary problems. These binary problems may possess distinct discriminative preferences, and thus have natural requirements for label-specific features to confront the attendant representation-compromise issue.