Unknown-Aware Graph Regularization for Robust Semi-supervised Learning from Uncurated Data

Heejo Kong1, Suneung Kim2, Ho-Joong Kim2, Seong-Whan Lee2

1Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea
2Department of Artificial Intelligence, Korea University, Seoul, South Korea
hj_kong@korea.ac.kr, se_kim@korea.ac.kr, hojoong_kim@korea.ac.kr, sw.lee@korea.ac.kr

Abstract
Recent advances in semi-supervised learning (SSL) have relied on the optimistic assumption that labeled and unlabeled data share the same class distribution. However, this assumption is often violated in real-world scenarios, where unlabeled data may contain out-of-class samples. SSL with such uncurated unlabeled data leads training models to be corrupted. In this paper, we propose a robust SSL method for learning from uncurated real-world data within the context of open-set semi-supervised learning (OSSL). Unlike previous works that rely on feature similarity distance, our method exploits uncertainty in logits. By leveraging task-dependent predictions of logits, our method is capable of robust learning even in the presence of highly correlated outliers. Our key contribution is to present an unknown-aware graph regularization (UAG), a novel technique that enhances the performance of uncertainty-based OSSL frameworks. The technique addresses not only the conflict between training objectives for inliers and outliers but also the limitation of applying the same training rule for all outlier classes, which are considered on previous uncertainty-based approaches. Extensive experiments demonstrate that UAG surpasses state-of-the-art OSSL methods by a large margin across various protocols. Codes are available at https://github.com/heejokong/UAGreg.

Introduction
Recent advances in deep supervised learning have been driven by the availability of large-scale annotated datasets. However, constructing such training data is labor-intensive and time-consuming due to the labeling process. As a remedy, significant efforts have been dedicated to the field of semi-supervised learning (SSL) (Lee et al. 2013; Xie et al. 2020; Sohn et al. 2020; Li, Xiong, and Hoi 2021; Zheng et al. 2022). They have provided effective solutions to leverage abundant unlabeled data with only a fraction of manual annotations, and shown their promising performances.

All the positive results observed in SSL are typically based on the optimistic assumption that both labeled and unlabeled data are drawn from the identical class distribution. However, in practical scenarios, the unlabeled dataset often includes out-of-class data, i.e., outlier, which easily violates this assumption. This uncurated unlabeled data severely degrade the performance of SSL (Oliver et al. 2018). Hence, it is desirable that the training models not only classify samples from known categories, i.e., inliers, but also identify samples from novel classes as outliers. This task is known as open-set semi-supervised learning (OSSL) (Yu et al. 2020). While OSSL is more realistic and practical for various applications, it has been rarely considered in previous literature.

In OSSL problem, the challenge lies in the absence of supervision for distinguishing unknown from known samples. Recent notable approaches (Yu et al. 2020; Saito, Kim, and Saenko 2021) have tackled this problem by utilizing similarity distances in feature space. Based on intra- or inter-class distances of known categories, they aim to identify unknown samples, which deviate significantly from in-distribution (ID) data, and consider the samples as outliers. Although the techniques have substantially improved the performance of OSSL, they are still quite limited for general use. As discussed in (Saito, Kim, and Saenko 2021), similarity-based measures can fail to detect highly correlated outliers that exhibit similar visual characteristics to ID data. It is evident that this limitation is more pronounced when unknown classes share the same superclass with labeled classes, whereas in a correlated setting, they share the same superclass space.

Figure 1: An illustration of the example for uncorrelated and correlated outliers. The classes marked with red and blue box share the same superclass, animal and transportation, respectively. In an uncorrelated setting, out-of-class data do not share a same superclass with labeled classes, whereas in a correlated setting, they share the same superclass space.
We conduct extensive experiments on CIFAR-10/100 (Krizhevsky, Hinton et al. 2009) and ImageNet-30 (Hendrycks et al. 2019) datasets, following the previous benchmarks. The results demonstrate that our method successfully addresses the limitations of uncertainty-based approaches and significantly surpasses the performance of previous state-of-the-art methods. Notably, our approach exhibits superior performance, even in scenarios where the outliers are highly correlated with the ID data.

**Related Works**

**Semi-supervised learning.** A standard SSL assumes that all training and testing data share identical class distribution, regardless of whether they are labeled or not, and aims to classify unlabeled examples into known classes. The mainstream of SSL can be broadly categorized into entropy minimization (Lee et al. 2013; Grandvalet and Bengio 2004; Cascante-Bonilla et al. 2021; Maeng et al. 2013), consistency regularization (Tarvainen and Valpola 2017; Sajjadi, Javanmardi, and Tasdizen 2016; Laine and Aila 2016; Miyato et al. 2018; Xie et al. 2020; Nam et al. 2020), and holistic methods (Berthelot et al. 2019a,b; Sohn et al. 2020; Kong et al. 2023). Recently, several studies (Li, Xiong, and Hoi 2021; Zheng et al. 2022) have attempted to combine consistency regularization with contrastive learning (Chen et al. 2020a; Khosla et al. 2020) by exploiting the instance-level similarity relationships as a target. Note that most outstanding works (Sohn et al. 2020; Li, Xiong, and Hoi 2021; Zheng et al. 2022) for SSL are based on self-training. Hence, when not all classes have labels in the training data, these methods will train outlier examples from unlabeled data as the known categories, and this learning causes the return of the corrupted SSL model (Oliver et al. 2018).

**Open-set semi-supervised learning.** A OSSL problem considers a more practical scenario using uncurated unlabeled data, where the training data often contains out-of-class instances. Early works focused solely on preventing the degradation of closed-set accuracy by adaptively assigning the weights of examples expected to be out-of-class data (Guo et al. 2020) or completely filtering them (Chen et al. 2020b). However, these methods do not have any objective for separating outliers from in-distribution data. Representative OSSL works (Yu et al. 2020; Huang et al. 2021; Saito, Kim, and Saenko 2021) adopted the similarity distance in feature space for addressing the limitation. In the concept of intra-class (Yu et al. 2020) or inter-class (Saito, Kim, and Saenko 2021) distance for known categories, they attempted to distinguish the outliers from known ones. As another alternative, some works exploited uncertainty scores of predictions in logit space using confidence (Sun and Wang 2022) and energy function (He et al. 2022; Liu et al. 2020). However, as discussed in this paper, the similarity-based methods are difficult to cope with highly correlated outliers, and uncertainty-based methods do not show competitive performance compared to (Yu et al. 2020; Saito, Kim, and Saenko 2021). In contrast, the proposed UAG exhibits relatively consistent performance improvement across all configurations, irrespective of the correlation between inliner and outlier classes.
Our Approach

Problem definition. Given a batch of labeled set $\mathcal{X} = \{(x_b, y_b)\}_{b=1}^B$, where $y_b$ is the corresponding label, and a batch of unlabeled set $\mathcal{U} = \{(u_b)\}_{b=1}^B$, where $\mu$ determines the relative size of $\mathcal{X}$ and $\mathcal{U}$, the objective of SSL is to learn a classification model by effectively leveraging both labeled and unlabeled data. Unlike standard SSL, we believe that the unlabeled data is more likely to be uncurated, where the unlabeled set may contain out-of-class data unseen in the labeled set. Hence, this work contributes to solving an OSSL problem. We aim to train a model not only to classify known $K$-way categories, i.e., inlier, but also to identify novel classes, i.e., outliers, from the known ones.

Approach overview. Our proposed model consists of a shared encoder $f(\cdot)$ with three heads: the closed-set classifier $h(\cdot)$, the open-set classifier $\tilde{h}(\cdot)$, and the projection head $p(\cdot)$. At test time, the closed-set classifier first predicts the $K$-way label for closed-set classification. Then, the open-set classifier measures the uncertainty of each prediction for open-set recognition. The projection head is deployed only for the training phase. A key technical novelty comes from the choice of a multi-head structure for enhancing the uncertainty-based approach as well as training them with contrastive graph regularization.

Uncertainty-based Outlier Detection

We exploit the uncertainty of logits to construct an outlier detector. Due to the mapping process into class-dependent space, the logits focus on activating only high-level semantic attributions for the target task (Wang et al. 2022). It effectively suppresses the interference of task-irrelevant correlation. Our work adopts a maximum softmax probability (MSP) (Hendrycks and Gimpel 2016) as an uncertainty score, and considers the samples with the low scores as outliers. Specifically, the score is derived from the composite function of encoder $f(\cdot)$ and the open-set classifier $\tilde{h}(\cdot)$ as follows:

$$s(x; T) = \max_i \frac{\exp \left( \frac{[\tilde{h} \circ f(x)]_i}{T} \right)}{\sum_{j=1}^K \exp \left( \frac{[\tilde{h} \circ f(x)]_j}{T} \right)},$$

(1)

where $T$ is a temperature parameter and set to larger than 1, for further enlarging the score gap between in- and out-of-distribution data, as discussed in (Liang, Li, and Srikant 2017). $[\tilde{h} \circ f(x)]_i$ indicates the logits of the $i$-th class.

To detect inliers and outliers with the estimated score $s(x; T)$, we utilize the thresholding, i.e., the sample $x$ is determined as inlier if $s(x; T) > \tau_{in}$ or outlier if $s(x; T) < \tau_{out}$. The thresholds are adaptively decided by employing a two-component Gaussian Mixture Model (GMM). For each training epoch $t$, GMM is fit on the predicted scores for the unlabeled set $\mathcal{U}$ with the Expectation-Maximization algorithm. The expectations of two gaussian distributions are assigned $\tau_{in}$ and $\tau_{out}$, respectively. Instead of using the current thresholds directly, we employ EMA thresholds which are updated by averaging the consecutive ones with a momentum parameter $m$. Their initial values, $\tau_{0,in}$ and $\tau_{0,out}$, are set to 1. The EMA update is applied as a warm-up process to stabilize the thresholds at the early stage of the training as:

$$\hat{\tau}_{t}^{in} \leftarrow m_{t}^{in} \hat{\tau}_{t-1}^{in} + (1 - m_{t}) \tau_{t}^{in},$$

(2)

$$\hat{\tau}_{t}^{out} \leftarrow m_{t}^{out} \hat{\tau}_{t-1}^{out} + (1 - m_{t}) \tau_{t}^{out}.$$  

(3)

Unknown-Aware Graph Regularization

As discussed in introduction, this work aims to enhance the uncertainty-based OSSL framework by addressing two aspects: i) Conflicts between objectives for learning inliers and outliers, ii) The convergence of all outliers into a single representative space, even though the outliers may consist of various novel classes. Inspired by the above drawbacks, we propose a couple of learning approaches as follows, and an overview of the proposed framework is presented in Fig. 3.

Figure 3: Framework of the proposed UAG. Weakly augmented views are used to generate model predictions for closed-set ($p$) and open-set ($\tilde{p}$) contexts. Based on $\tilde{p}$, pseudo-inliers and -outliers are assigned, $p$ are used as targets for training of pseudo-inliers (Eq. 5), while $\tilde{p}$ is used to train pseudo-outliers (Eq. 6). Pseudo-label ($W^{pbg}$) and pseudo-outlier ($W^{pog}$) graphs are constructed to measure similarity between samples for known and unknown contexts, respectively. Both graphs are integrated into an unknown-aware pseudo graph ($W^{uag}$) to train an embedding graph (Eq. 10).
Exclusive multi-head training. To address the first aspect, we adopt a multi-head structure that exclusively learns inliers and outliers. Note that existing methods (He et al. 2022; Sun and Wang 2022) only depend on a single classifier which trains to maximize the entropy of pseudo-outliers and minimize the entropy of pseudo-inliers. However, this strategy allows the model to learn a significant amount of erroneous predictions, even though they are confidently predicted as outliers and inliers. Note that the outlier type, whether uncorrelated (Uncorr.) or correlated (Corr.), is shown in each column.

Whereas, in our method, two classifier heads learn the objectives for inliers and outliers independently. Specifically, they share the same feature extractor \( f(\cdot) \), and train the labeled data as the cross-entropy loss \( H(\cdot, \cdot) \) between ground-truth labels \( y_b \) and predictions:

\[
L_s = \frac{1}{B} \sum_{b=1}^{B} \{ H(y_b, p(A(x_b))) + H(y_b, \hat{p}(A(x_b))) \},
\]

where \( A(\cdot) \) refers to weak augmentation, \( p \) and \( \hat{p} \) represent the softmax probabilities predicted by closed-set classifier \( h \circ f \) and open-set classifier \( \hat{h} \circ f \), respectively.

For training the closed-set classifier \( h \), the samples predicted as inliers, i.e., pseudo-inliers, are leveraged only. We adopt FixMatch (Sohn et al. 2020) as our learning objective due to its simplicity yet effectiveness. It is defined as the cross-entropy between pseudo-labels \( \hat{y}_b \) and predictions:

\[
L_{in}^u = \frac{1}{\mu_B} \sum_{b=1}^{\mu_B} \mathbb{I} \left( s_b \geq \hat{z}_i^{in} \right) \cdot H \left( \hat{y}_b, \hat{p}(\hat{A}(u_b)) \right),
\]

where \( \hat{A}(\cdot) \) refers to strong augmentation, \( s_b = s(u_b; T) \) represents MSP score (Eq. 1) for \( b \)-th unlabeled example.

The open-set classifier \( \hat{h} \) learns only the examples assigned as outliers from unlabeled data. Following the previous work (Sun and Wang 2022), we directly maximize the mean entropy for the outliers, in order to separate them from the inliers in the task-dependent space. Specifically, we optimize the following objective:

\[
L_{out}^u = -\frac{1}{\mu_B} \sum_{b=1}^{\mu_B} \mathbb{I} \left( s_b < \hat{z}_i^{out} \right) \cdot H \left( \hat{p}(A(u_b)) \right).
\]

By contrast with Eq. 5, since outliers are not assigned any labels, only weak augmentation is utilized for stability.

Contrastive graph regularization. Existing works have adopted only the entropy maximization for entire outliers, despite the fact that they comprise multiple novel classes with diverse semantic information. When considering updates only for the weights \( \omega \in \mathbb{R}^{k \times d} \) of the last classifier in this learning process, it can be expressed as follows. For simplicity, let’s momentarily exclude the influence of bias.

\[
\arg \min_{\omega \in \Omega} \mathbb{E}_{(x,y) \sim D} [H(y, f(x, \theta); \omega)]. \tag{7}
\]

By minimizing the cross-entropy loss for inliers, \( H(y, f(x, \theta); \omega) = -y \log \sigma (f(x, \theta) \cdot \omega^T) \), the weights converge to points in the embedding space that maximize the similarity with samples corresponding to the \( k \)-th label. In theoretical terms, the weights function as representative points for the embedding features \( z = f(x, \theta) \) of each class.

\[
\arg \max_{\omega \in \Omega} \mathbb{E}_{(x, y) \sim D} [H(f(x, \theta); \omega)]. \tag{8}
\]

From this perspective, maximizing the entropy of outliers aims to minimize the similarity between their embedding features \( z \) and the \( k \)-way representative points \( \omega \). Hence, the intermediate features of outlier examples converge to a single cluster orthogonal to the weights \( \omega \). Such training inhibits the model’s ability to encode outlier data precisely, degrading the performance of outlier detection.

Interestingly, as shown in Fig. 4, we found that conventional SSL have low discriminative ability between outliers and inliers, while high discriminative power among outlier classes in embedding space. This suggests that semantic information about the known context can benefit learning about out-of-classes. Based on this empirical rationale, we derive an unknown-aware graph regularization that allows the outliers to form multiple clusters in their embeddings.

Specifically, following a previous CoMatch (Li, Xiong, and Hoi 2021), we build a pseudo-label graph \( W^{plg} \in \mathbb{R}^{N_B \times \mu_B} \) by leveraging the batch-wise predictions of a closed-set classifier. The graph \( W^{plg} \) is a similarity matrix of the closed-set predictions, and assigns only examples with similar predictions for a known context to the same cluster:

\[
W_b^{plg} = \begin{cases} 
1 & \text{if } b = j \\
 p_b \cdot p_j & \text{if } b \neq j \text{ and } p_b \cdot p_j \geq \tau_g , \\
0 & \text{otherwise} 
\end{cases}
\]

Table 1: Results of the models trained solely on labeled data for CIFAR-10 (100 labeled data per class). FPR/FNR (%) indicate the proportions of samples with erroneous predictions, even though they are confidently predicted as outliers and inliers. Note that the outlier type, whether uncorrelated (Uncorr.) or correlated (Corr.), is shown in each column.
where $\tau_g$ denotes a pre-defined threshold. $p_h$ and $p_j$ represent the softmax probabilities predicted by $h \circ f$ for $u_h$ and $u_j$, respectively. Each sample is connected to itself with the strongest edge of value 1 as a self-loop, while the samples with similarity less than $\tau_g$ are not connected.

Unlike CoMatch, we utilize the pseudo-outlier graph $W^{\text{pog}} \in \mathbb{R}^{nB \times nB}$, which is information about the unknown context, in the process of forming the target graph. The graph $W^{\text{pog}}$ is built by employing the batch-wise predictions of an open-set classifier. Note that $\eta_b = I(s_b < s^\text{out}_t)$ is an outlier indicator, and returns 1 for values predicted as outliers and 0 for the others. The graph $W^{\text{pog}}$ assigns only samples with the same prediction for the unknown context to the same cluster by connecting the edges with the same predictions of the outlier indicator to the value of 1 as:

$$W^{\text{pog}}_{bj} = \begin{cases} 1 & \text{if } \eta_b = \eta_j \\ 0 & \text{otherwise} \end{cases}. \quad (10)$$

An unknown-aware pseudo graph is obtained by matrix multiplication of two pseudo graphs: $W^{\text{aug}} = W^{\text{pog}} \cdot W^{\text{pog}}$. That is, only samples with the same predictions in unknown context as well as the similar predictions for known context are assigned to the same cluster in the graph. The graph $W^{\text{aug}}$ serves as a target to train embedding graph.

To construct the embedding graph, we first obtain a pair of images with different augmentations $A(\cdot)$ and $A(\cdot)$. The embeddings are extracted from a composite function of encoder and projection head $g$: $z_b = g \circ f(A(u_b))$. The embedding graph $W^z$ is derived as the batch-wise similarity of the two projected embeddings, where $t_e$ is a scalar temperature parameter:

$$W^z_{bj} = \begin{cases} \exp \left( \frac{z_b \cdot z_j}{t_e} \right) & \text{if } b = j \\ \exp \left( \frac{z_b \cdot z_j}{t_e} \right) & \text{otherwise} \end{cases}. \quad (11)$$

Subsequently, the graphs are normalized with $\hat{W}_{bj} = W_{bj} / \sum W_{bj}$, so that each row of the similarity matrix sums to 1. The graph contrastive loss is derived to minimize the cross-entropy between the two normalized graphs as:

$$L_g = \frac{1}{\mu B} \sum_{b=1}^{nB} \mu_B H(\hat{W}^{\text{aug}}_{bj}, \hat{W}^z_{bj}). \quad (12)$$

By minimizing the loss $L_g$, the embedding graph is enforced to have the same structure as $W^{\text{aug}}$. This learning encourages the model to have similar embeddings for samples with similar predictions in both closed-set and open-set recognition. In other words, outlier samples are positioned within a single space that minimizes their similarity with the known context, while samples with similar semantic information belong to the same cluster.

**Overall objectives.** The overall objective of our proposed framework is the weighted sum of the supervised loss $L_s$, the unsupervised loss $L_u$, and the contrastive loss $L_g$. We summarize the unsupervised loss as the sum of $L_u^{\text{in}}$ and $L_u^{\text{out}}$. Hence, the overall loss can be written as follows:

$$\mathcal{L} = L_s + \lambda_u L_u + \lambda_g L_g, \quad (13)$$

where $\lambda_u$ and $\lambda_g$ are the hyperparameters to control the weight of each objective.

**Experiments**

**Experimental Settings.**

**Setup.** We follow the default setting of the previous work (Saito, Kim, and Saenko 2021): CIFAR-10/100 (Krizhevsky, Hinton et al. 2009) setup on WRN-28 (Zagoruyko and Komodakis 2016) for small-scale experiments; ImageNet (Deng et al. 2009) and Semi-iNature-2021 (Su and Maji 2021) setup on ResNet-18 (He et al. 2016) for large-scale experiments. Note that we use an identical set of hyperparameters for whole experiments except for projection head $g$, where the output dimension is scaled on each dataset. For an outlier detector, $T$ and $m$ are set to 1.5 and 0.9, respectively. The settings for graph contrastive loss are fixed across all experiments for simplicity, e.g., a graph threshold $\tau_g$, a scalar temperature $t_e$. More information regarding implementation details can be found in Appendix.

**Compared methods.** We compare the proposed UAG with representative SSL, similarity-based OSSL and uncertainty-based OSSL methods. For the SSL baselines, we employ FixMatch (Sohn et al. 2020) and CoMatch (Li, Xiong, and Hof 2021). Since UAG is incrementally applied to these works (Sohn et al. 2020; Li, Xiong, and Hof 2021), a comparison with them effectively reveals the strength of our proposed approach. As the similarity-based OSSL baselines, we employ MTC (Yu et al. 2020) and OpenMatch (Saito, Kim, and Saenko 2021) using the author’s implementations. Referring to the papers for the uncertainty-based methods (He et al. 2022; Sun and Wang 2022), we implement two objectives, which are entropy maximization (Ent-Max) (Sun and Wang 2022) and uniform distribution matching (UDM) (He et al. 2022), and adopt the objectives in ablation studies to demonstrate the effectiveness of our proposed UAG. In addition, the most recent OSSL work, OSP (Wang et al. 2023), is also compared.

**Evaluation metric.** We evaluate our approach with two common metrics: Error rates for closed-set accuracy, and AUROC for open-set recognition. We reported the results averaged over three runs and their standard deviations.

**Experimental Results.**

**CIFAR-10/100.** To assess the CIFAR datasets, we consider two setups: (a) uncorrelated, where unknown classes have no superclass relation with known ones, and (b) correlated, where known and unknown classes share superclass space. For CIFAR-10 (CIFAR-100), we divide classes into 6 known (60 known) and 4 unknown (40 unknown) classes. All samples, except for the labeled data, are designated as unlabeled data. Tables 2 and 3 show error rates and AUROC results, respectively. Our method consistently outperforms OSSL and SSL baselines in error rates. Notably, UAG shows significant AUROC improvement, contrasting with minimal gains from similarity-based OSSL baselines in the correlated setting, indicating UAG’s robustness across outlier correlations.

**ImageNet.** We evaluate the performance of our proposed approach on ImageNet, a more challenging and complex dataset. Due to computational constraints, we utilize a subset called ImageNet-30 (Tack et al. 2020), which consists of 30 classes, for training instead of the complete ImageNet.
Entropy maximization alone (EMX) improves AUROC but serves as our baseline, with its results reported at the top. Ablation Studies

Table 4: Ablation studies on the individual modules.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>CIFAR-10</th>
<th>CIFAR-100</th>
<th>ImageNet-30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeled Only</td>
<td>34.3±1.2</td>
<td>29.4±0.8</td>
<td>30.9±1.3</td>
</tr>
<tr>
<td>FixMatch</td>
<td>16.8±1.1</td>
<td>10.7±0.9</td>
<td>17.5±0.9</td>
</tr>
<tr>
<td>CoMatch</td>
<td>12.7±0.7</td>
<td>9.5±0.5</td>
<td>14.8±0.8</td>
</tr>
<tr>
<td>MTC</td>
<td>20.4±0.9</td>
<td>13.5±0.8</td>
<td>21.8±1.2</td>
</tr>
<tr>
<td>OpenMatch</td>
<td>10.2±0.9</td>
<td>7.1±0.5</td>
<td>11.7±0.8</td>
</tr>
<tr>
<td>OSP</td>
<td>12.1±0.8</td>
<td>9.2±0.6</td>
<td>11.1±0.8</td>
</tr>
<tr>
<td>Ours</td>
<td>9.6±0.7</td>
<td>5.8±0.4</td>
<td>8.1±0.9</td>
</tr>
</tbody>
</table>

Table 2: Error rates with standard deviation for CIFAR-10/100 and ImageNet on 3 different folds.

Table 3: AUROC performances of Table 2.

Ablation Studies

To better understand the benefits of UAG, we conducted ablation studies on quantitative and qualitative evaluations. All experiments are performed on the test set of CIFAR datasets using models trained with 100 labeled data per each class.

Table 4 compares ablated models on the CIFAR-100 dataset. FixMatch serves as our baseline, with its results reported at the top. Entropy maximization alone (EMX) improves AUROC but not error rates. Multi-head training (MHT) significantly enhances both error rates and AUROC. Incorporating the pseudo-label graph (PLG) reduces error rates but worsens AUROC. Further addition of the pseudo-outlier graph (POG) achieves substantial progress in both metrics. UAG surpasses the baseline, showing notable improvements of 5.8% (5.0%) in error rates and 25.1% (18.8%) in AUROC over the baseline in the uncorrelated (correlated) setting.

Multi-Head Training Effectiveness. Fig. 5 shows the effect of exclusive multi-head training by comparing logit histograms between single- and multi-head structures. All the models are trained on uncorrelated settings. Multi-head training shows a clearer distinction in logit distribution between known and unknown classes compared to the single-head method, indicating the effectiveness of our approach in
Figure 6: t-SNE visualization of embeddings obtained from the ablated models. Pink points denote outliers, while the other colored points represent distinct known classes. (a) A model trained solely with FixMatch. (b) A model trained with entropy maximization further applied on (a). (c) A model trained using the proposed method UAG.

Effectiveness of the contrastive graph regularization. Fig. 6 demonstrates t-SNE visualization of feature distribution extracted from CIFAR-10 models trained in the correlated setting. Results reveal that a FixMatch-only model (a) struggles with inlier-outlier differentiation, whereas our proposed UAG (c) excels. Notably, our model identifies multiple centroids where outlier examples cluster, showcasing its robust discrimination compared to entropy maximization (b), which aims to collapse all outliers into a single cluster.

Further Analysis

Results on various mismatch ratio. We evaluate the robustness of our method against uncurated data corruption by analyzing performances across various mismatch ratios with the CIFAR-100 dataset (100 label per class) in the correlated setting. The results are depicted in Fig. 7. Across all mismatched scenarios, our UAG model consistently outperforms alternative approaches in terms of error rates and AUROC. These findings strongly affirm the effectiveness and superiority of our proposed method.

Results on the different number of known classes. Experiments were conducted with different numbers of known classes, and results are presented in Table 5. All the results are derived from CIFAR-100 (100 label per class) in the correlated setting. The table illustrates the consistently superior performance of our proposed method across all cases, regardless of the known and unknown class ratio.

Table 5: Results with the different number of known classes.

<table>
<thead>
<tr>
<th>Method</th>
<th>Known / Unknown</th>
<th>40 / 60</th>
<th>80 / 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeled Only</td>
<td>30.8</td>
<td>70.4</td>
<td>38.5</td>
</tr>
<tr>
<td>FixMatch</td>
<td>21.9</td>
<td>64.7</td>
<td>31.4</td>
</tr>
<tr>
<td>OpenMatch</td>
<td>19.7</td>
<td>69.1</td>
<td>28.9</td>
</tr>
<tr>
<td>Ours</td>
<td><strong>18.1</strong></td>
<td><strong>80.7</strong></td>
<td><strong>27.6</strong></td>
</tr>
</tbody>
</table>

Table 6: Top-1 and Top-5 accuracy for Semi-iNature 2021.

<table>
<thead>
<tr>
<th>Method</th>
<th>Top-1 Acc.</th>
<th>Top-5 Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeled Only</td>
<td>15.43</td>
<td>30.86</td>
</tr>
<tr>
<td>FixMatch</td>
<td>17.00</td>
<td>32.90</td>
</tr>
<tr>
<td>CoMatch</td>
<td>19.17</td>
<td>36.97</td>
</tr>
<tr>
<td>OpenMatch</td>
<td>18.65</td>
<td>35.68</td>
</tr>
<tr>
<td>Ent-Max</td>
<td>13.35</td>
<td>27.33</td>
</tr>
<tr>
<td>UAG (Ours)</td>
<td><strong>24.16</strong></td>
<td><strong>44.32</strong></td>
</tr>
</tbody>
</table>

Results on real-world dataset. We conducted additional experiments on the Semi-iNature-2021 dataset to verify the robustness of our framework. Employing ResNet-18 as our encoder, trained from scratch for 100 epochs, yielded results presented in Table 6. Our method demonstrates superior performance, surpassing existing baseline models even when applied to large-scale real-world datasets.

Conclusion

In this paper, we introduce unknown-aware graph regularization (UAG), a novel approach for open-set semi-supervised learning (OSSL). We concentrate on highly correlated scenarios, where inlier and outlier classes share the same superclass space, providing a challenging and realistic OSSL benchmark. Extensive experiments show UAG’s outstanding performance across various OSSL scenarios. We consider our work a valuable baseline guiding future research in both semi-supervised learning (SSL) and OSSL.
Acknowledgements

This research was supported by the Challengeable Future Defense Technology Research and Development Program (912911601) of Agency for Defense Development in 2020 and was partly supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant, funded by the Korea government (MSIT) (No. 2019-0-00079, Artificial Intelligence Graduate School Program (Korea University)).

References


