

# ANEDL: Adaptive Negative Evidential Deep Learning for Open-Set Semi-supervised Learning

Yang Yu<sup>1,\*</sup>, Danruo Deng<sup>1,\*</sup>, Furui Liu<sup>2</sup>, Qi Dou<sup>1</sup>,  
Yueming Jin<sup>3,†</sup>, Guangyong Chen<sup>2,†</sup>, Pheng Ann Heng<sup>1</sup>

<sup>1</sup>The Chinese University of Hong Kong, Hong Kong, China

<sup>2</sup>Zhejiang Lab, Hangzhou, China

<sup>3</sup>National University of Singapore, Singapore

## Abstract

Semi-supervised learning (SSL) methods assume that labeled data, unlabeled data and test data are from the same distribution. Open-set semi-supervised learning (Open-set SSL) considers a more practical scenario, where unlabeled data and test data contain new categories (outliers) not observed in labeled data (inliers). Most previous works focused on outlier detection via binary classifiers, which suffer from insufficient scalability and inability to distinguish different types of uncertainty. In this paper, we propose a novel framework, Adaptive Negative Evidential Deep Learning (ANEDL) to tackle these limitations. Concretely, we first introduce evidential deep learning (EDL) as an outlier detector to quantify different types of uncertainty, and design different uncertainty metrics for self-training and inference. Furthermore, we propose a novel adaptive negative optimization strategy, making EDL more tailored to the unlabeled dataset containing both inliers and outliers. As demonstrated empirically, our proposed method outperforms existing state-of-the-art methods across four datasets.

## Introduction

Semi-supervised learning (SSL) has recently witnessed significant progress by propagating the label information from labeled data to unlabeled data (Berthelot et al. 2019; Xu et al. 2021; Wang et al. 2022b; Zheng et al. 2022). Despite the success, SSL methods are deeply rooted in the closed-set assumption that labeled data, unlabeled data and test data share the same predefined label set. In reality (Yu et al. 2020), such an assumption may not always hold as we can only accurately control the label set of labeled data, while unlabeled and test data may include outliers that belong to the novel classes that are not seen in labeled data. During self-training, imprecisely propagating label information to these outliers shall interfere with model learning and lead to performance degradation. To this end, open-set semi-supervised learning (Open-set SSL) (Yu et al. 2020) has been emerging to tackle the problem. It first detects out-of-distribution (OOD) samples as outliers and then recognizes in-distribution (ID) samples as belonging to the classes from the predefined label set.

\*Equal contribution: {yangyu, drdeng}@cse.cuhk.edu.hk

†Co-corresponding: ymjn@nus.edu.sg, gychen@zhejianglab.com  
Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Existing Open-set SSL methods (Guo et al. 2020; Chen et al. 2020) generally consist of an outlier detector and a classifier. To avoid propagating wrong label information to outliers during SSL training of the classifier, they first detect outliers and then optimize SSL loss only for unlabeled instances taken as inliers. To detect outliers, they aim to estimate confidence score of a sample being an inlier (reversely, uncertainty score for outliers). For example, T2T (Huang et al. 2021) proposes to detect outliers by cross-modal matching over binary detector. OpenMatch (Saito, Kim, and Saenko 2021) deploys a one-vs-all (OVA) classifier (Saito and Saenko 2021) as outlier detector to output the confidence score of considering a sample as an inlier and reject outliers with confidence score below a threshold for self-training in SSL. With outliers rejected, they can improve recognition accuracy of SSL methods.

We identify two main limitations in current state-of-the-art literature to tackle this task. First, most previous works estimate confidence scores based on binary likelihoods generated from binary detectors activated by Softmax function. Such softmax-based networks only provide a point estimation of predicted class probabilities and cannot quantify whether the uncertainty stems from the lack of knowledge of outliers (epistemic uncertainty) or from the complexity of the samples within the distribution (aleatoric uncertainty) (Malinin and Gales 2018). Moreover, when we tackle a K-way classification problem with a large K, the binary detectors are less robust to identify outliers from such a complex dataset that contains multi-class information (Carbonneau et al. 2018). One advanced method, evidential deep learning (EDL) (Sensoy, Kaplan, and Kandemir 2018) can explicitly quantify the classification uncertainty corresponding to the unknown class, by treating the network’s output as evidence for parameterizing the Dirichlet distribution according to subjective logic (Jøsang 2016). Compared with Softmax which only provides probabilistic comparison values between classes, EDL models second-order probability and uncertainty for each class. In this work, we take the *first* step to introduce EDL into the outlier detector. As shown in Figure 1, EDL enables a multinomial detector to directly output evidence for each class, which can alleviate the insufficient scalability of existing binary detectors. Moreover, different types of uncertainties can be distinguished by the output vector, so we propose to design different uncertainty met-

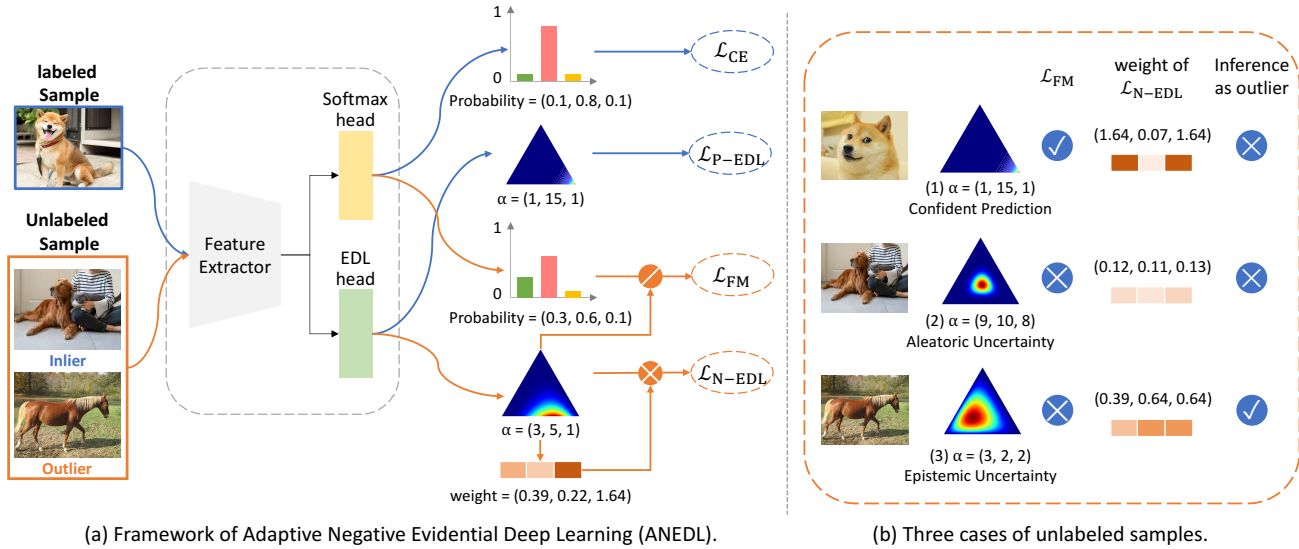


Figure 1: Overview of our proposed Adaptive Negative Evidential Deep Learning (ANEDL). (a) The framework of ANEDL consists of a shared feature extractor, a Softmax head and an EDL head. EDL head is used to detect outliers, while Softmax head is used to classify. To effectively leverage the information of inlier and outlier from unlabeled samples, we introduce Negative Optimization to explicitly regularize our EDL detector to output low evidence values for uncertain classes, while proposing adaptive loss weights to encourage the detector to pay more attention to these uncertain classes. (b) Three cases of unlabeled samples. Our model can quantify different types of uncertainty, including epistemic uncertainty due to lack of knowledge, and aleatoric uncertainty due to the complexity of samples from the distribution. Among them, only samples with confident predictions are used jointly with labeled samples to train the classification head.

rics for self-training and inference. Note that though EDL shows an impressive performance in uncertainty quantification, we empirically find its deficiency in learning representation preferred by classification tasks and consequent low performance in classification accuracy. Therefore, we propose to take advantage of both Softmax and EDL, by keeping Softmax for representation learning.

The other limitation of prior methods is that they do not effectively leverage the information contained in unlabeled data. The detector should output low confidence scores for outliers, but this information is underutilized for model regularization, which may result in suboptimal performance (Malinin and Gales 2018). In this regard, we propose *adaptive negative optimization* tailored for unlabeled data, where the negative optimization is to explicitly regulate our EDL detector to output  $K$  low evidence values for a given outlier sample. However, negative optimization can not directly apply to inliers in unlabeled data. Therefore, we derive an adaptive loss weight for encouraging the model to separately treat inliers and outliers in the unlabeled data in the learning process. It can largely avoid the interference between the ID and OOD features as well as the leading problem of uncertainty availability reduction. Specifically, we achieve this by using the Fisher information matrix (FIM) introduced by  $\mathcal{I}$ -EDL (Deng et al. 2023) to identify the amount of information for each class of unlabeled samples, while imposing smaller constraints on the confident classes with high evidence. Therefore, such optimization can sufficiently com-

press evidence values of unlabeled outliers without interfering with the learning of unlabeled inliers.

To our knowledge, we are the first to exploit evidential deep learning in Open-set SSL and propose a novel Adaptive Negative Evidential Deep Learning (ANEDL) framework. Our main contributions can be summarized as follows:

- We introduce EDL as an outlier detector for Open-set SSL, which enables quantification of different types of uncertainty and design of different uncertainty metrics for self-training and inference, respectively.
- We propose a novel adaptive negative optimization strategy, which tailors EDL on the unlabeled dataset containing both inliers and outliers.
- Extensive experiments on four public datasets, including CIFAR-10, CIFAR-100, ImageNet-30, and Mini-ImageNet, show that our method outperforms state-of-the-art methods in different settings.

## Related Works

### Semi-supervised Learning

Semi-supervised learning (SSL) aims to ease deep networks' reliance on massive and expensive labeled data by utilizing limited labeled data together with a large amount of unlabeled data. Existing works can be generally categorized into two paradigms: consistency regularization and pseudo labeling (Lee et al. 2013). Consistency regularization assumes that the network should be invariant to different per-

turbations of the same instance. Pseudo-labeling takes confidently pseudo-labeled unlabeled data as labeled data to train the network. And many pseudo labeling methods like Fix-Match (Sohn et al. 2020), FlexMatch (Zhang et al. 2021) and AdaMatch (Berthelot et al. 2021) focus on how to select confidently pseudo-labeled data. Although pseudo-labeling methods have achieved tremendous success in SSL, they are deeply rooted in the close-set assumption and can not be directly applied to open-set SSL (Yu et al. 2020; Saito, Kim, and Saenko 2021). Because noisy pseudo-labeled outliers may degrade the performance of self-supervised training. On the other hand, OpenMatch (Saito, Kim, and Saenko 2021) proposes that soft consistency regularization by minimizing the distance between its predictions on two augmentations of the same image can help SSL methods get rid of the close-set assumption can be kept for open-set SSL. Following OpenMatch, our method also applies soft consistency regularization to the outlier detector.

### Open-set Semi-supervised Learning

Open-set SSL (Guo et al. 2020; Chen et al. 2020) aims to detect outliers while categorizing inliers into correct close-set classes during training and inference. Existing open-set SSL methods recognize the weakness of SSL methods in terms of detecting outliers and they develop an extra outlier detector to exclude outliers from self-supervised training. Their outlier detectors are based on binary classifier activated by softmax function. For example, MTC (Yu et al. 2020) deploys a binary classifier to predict the probability of the instance belonging to outliers and updates the network parameters and the outlier score alternately. OpenMatch (Saito, Kim, and Saenko 2021) takes one-vs-all (OVA) classifier as its outlier detector and proposes to apply soft consistency regularization to the outlier detector. An OVA classifier is composed of K binary classifiers and each classifier predicts the probability of the sample belonging to a specific class. However, softmax-based networks are notorious for inflating the probabilities of predicted classes (Szegedy et al. 2016; Guo et al. 2017; Wilson and Izmailov 2020), and binary detectors perform poorly on tasks with a large number of classes (Carbonneau et al. 2018). To address these limitations, we adopt evidential deep learning (EDL) (Sensoy, Kaplan, and Kandemir 2018) as the outlier detector for Open-set SSL.

### Evidential Deep Learning

Evidential neural networks (Sensoy, Kaplan, and Kandemir 2018) model the outputs as evidence to quantify belief masses and uncertainty based on the Dempster-Shafer Theory of Evidence (DST) (Sentz and Ferson 2002) and subjective logic (Jøsang 2016). Similar to this work, (Malinin and Gales 2018) propose Prior Networks that explicitly consider the distributional uncertainty to quantify the mismatch of ID and OOD distribution. Compared with the standard neural network classifiers direct output the probability distribution of each sample, EDL obtains the density of classification probability assignments by parameterizing the Dirichlet distribution. Therefore, EDL can use the properties of Dirichlet distribution to distinguish different types of uncertainties, which shows an impressive performance in uncertainty

quantification and is widely used in various applications. For example, (Zhao et al. 2020) proposes a multi-source uncertainty framework combined with DST for semi-supervised node classification with GNNs. (Soleimany et al. 2021) introduces evidential priors over the original Gaussian likelihood function to model the uncertainty of regression networks. (Bao, Yu, and Kong 2021, 2022) propose a general framework based on EDL for Open Set Recognition (OSR) and Open Set Temporal Action Localization (OSTAL), respectively. Compared with previous efforts, our work is the first to exploit EDL to Open-set SSL. For Open-set SSL, in addition to needing to distinguish unknown and known classes at test time, more importantly is how to better utilize the unlabeled data during training to learn the features of known classes.

## Method

### Problem Setting

Open-set semi-supervised learning (Open-set SSL) aims to learn a classifier by using a set of labeled and unlabeled datasets. Unlike traditional semi-supervised learning (SSL), the unlabeled dataset of Open-set SSL contains both ID and OOD samples, i.e., inliers and outliers. More specifically, for a K-way classification problem, let  $\mathcal{D}_l = \{(\mathbf{x}_j^l, y_j^l)\}_{j=1}^{N_l}$  be the labeled dataset, containing  $N_l$  labeled samples randomly generated from a latent distribution  $\mathcal{P}$ . The labeled example  $\mathbf{x}_j^l$  can be classified into one of K classes, meaning that its corresponding label  $y_j^l \in \{c_1, \dots, c_K\}$ . Let  $\mathcal{D}_u = \{\mathbf{x}_j^i\}_{j=1}^{N_i} \cup \{\mathbf{x}_j^o\}_{j=1}^{N_o}$  be the unlabeled dataset consisting of  $N_i + N_o$  examples with labels either from the K known classes (inliers) or never been seen (outliers). The goal of Open-set SSL is to detect outliers while classifying inliers into the correct class.

### Adaptive Negative Evidential Deep Learning

**Approach Overview.** As shown in Figure 1, our proposed framework contains three components: (1) a shared feature extractor  $F(\cdot)$ ; (2) a Softmax head  $C(\cdot)$  that classifies samples by outputting K-way probabilities; (3) an EDL head  $D(\cdot)$  that quantifies uncertainty by outputting K evidence values to detect outliers in unlabeled data. To adopt our framework to Open-set SSL, we propose an Adaptive Negative Optimization strategy to train EDL head by using all labeled and unlabeled samples. We introduce negative optimization to explicitly regularize our EDL detector to output low evidence values for uncertain classes, while deriving an adaptive loss weight to encourage the detector to pay more attention to these uncertain classes. For unlabeled data with confident prediction, we will use them jointly to train the Softmax classifier with labeled data.

**Evidential Outlier Detector.** Traditional outlier detectors for Open-set SSL use Softmax activation function to predict confidence scores and achieve impressive performance. However, such softmax-based detectors only provide a point estimation of outlier probabilities, unable to quantify different types of uncertainty. To tackle the limitations of Softmax, we introduce evidential deep learning (EDL) (Sensoy, Kaplan, and Kandemir 2018) to our framework as an

outlier detection module. EDL proposes to integrate deep classification networks into the evidence framework of subjective logic (Jøsang 2016), to jointly predict classification probabilities and quantify uncertainty. Specifically, for K-way classification, EDL uses a network  $g(\theta)$  to calculate its evidence vector  $e = g(x|\theta)$ . Note that an activation function, e.g. Relu or Softplus, is on top of  $g(\theta)$  to guarantee non-negative evidence. Then EDL correlates its derived evidence vector  $e \in \mathbb{R}_+^K$  to the concentration parameter  $\alpha \in \mathbb{R}_+^K$  of Dirichlet distribution  $D(p|\alpha)$  with the equality  $\alpha_i = e_i + 1$ . Dirichlet distribution parameterized by  $\alpha$  models the density of classification probability assignments and uncertainty. The expected probability of the  $k^{\text{th}}$  class and uncertainty can be derived by

$$\hat{p}_k = \frac{\alpha_k}{\alpha_0} \quad \text{and} \quad u = \frac{K}{\alpha_0},$$

where  $\alpha_0 = \sum_{k=1}^K \alpha_k$ . Higher values of  $\alpha_0$  lead to more confident distributions. Given a sample,  $\alpha_k$  is incremented to update the Dirichlet distribution of the sample when evidence of  $k^{\text{th}}$  class is observed.

**Joint Optimization with Softmax and EDL.** Although EDL-related works (Sensoy, Kaplan, and Kandemir 2018; Deng et al. 2023) are capable of jointly classifying and quantifying uncertainty, we empirically find its deficiency in optimizing a network to learn representation for classification. That is, the classification performance of the evidential network is lower than that of a network activated by Softmax and optimized with only cross-entropy loss. In this way, we propose to take the strength of both Softmax and EDL and jointly optimize their losses. In our proposed framework, Softmax head takes charge of learning representation and predicting classification probabilities, while EDL is responsible for quantifying uncertainty. During training, with EDL excluding outliers, we adopt an SSL method (i.e., FixMatch following OpenMatch (Saito, Kim, and Saenko 2021)) to our Softmax head to enhance representation quality and classification accuracy. We also utilize counterfactual reasoning and adaptive margins proposed in DeBiasPL (Wang et al. 2022a) to remove the bias of pseudo label in FixMatch.

**Adaptive Negative Optimization.** The core challenge of Open-set SSL is that unlabeled data contains novel categories that are not seen in labeled training data. To enhance the separation between inliers and outliers, detectors should output low confidence scores for outliers, but traditional binary detectors do not sufficiently leverage this information for model regularization, which may result in suboptimal performance (Malinin and Gales 2018). Inspired by negative learning (Ishida et al. 2017), we propose Adaptive Negative Optimization (ANO) to avoid providing misinformation to the EDL detector by focusing on uncertain classes. In particular, we introduce the negative optimization to explicitly regulate our detector to output low evidence values for uncertain classes. Meanwhile, adaptive loss weights are proposed to encourage EDL detectors to pay more attention to uncertain classes during the learning process.

More specifically, we adopt the Fisher information matrix (FIM) introduced by  $\mathcal{I}$ -EDL (Deng et al. 2023) to identify the amount of information contained in each class of each

sample, while imposing explicit constraints on the informative uncertainty classes.  $\mathcal{I}$ -EDL proves that the evidence of a class exhibits a negative correlation with its Fisher information, i.e., the class with more evidence corresponds to less Fisher information, and shows an impressive performance in uncertainty quantification (Deng et al. 2023). Therefore, we propose to use FIM as an indicator to distinguish inliers from outliers to adaptively regulate our model to pay more attention to uncertain classes. For those uncertain classes with less evidence in unlabeled samples, we impose explicit constraints to output K low evidence values through the Kullback-Leibler (KL) divergence term. The objective function for unlabeled data is defined as

$$\min_{\theta} \mathbb{E}_{x \sim \mathcal{D}_u} [D_{KL}(\mathcal{U} || \mathcal{N}(\mathbb{E}_{Dir(\alpha)}[p], \sigma^2 \mathcal{I}(\alpha)^{-1})] \quad (1)$$

where  $\mathcal{U} = \mathcal{N}(1/K, \lambda^2 \mathbf{I})$ ,  $\alpha = g(x|\theta) + \mathbf{1}$ , and  $\mathcal{I}(\alpha)$  denotes the FIM of  $Dir(\alpha)$  (the closed-form expression is provided in Appendix A.1). Eq.(1) can be understood through the lens of probabilistic graphical models, where the observed labels  $\hat{y}$  are generated from the Dirichlet distribution  $p$  with its parameter  $\alpha$  calculated by passing the input sample  $x$  through the network. Since our unlabeled data contains both novel and known classes, we assume  $\hat{y} \sim \mathcal{N}(p, \sigma^2 \mathcal{I}(\alpha)^{-1})$  to make the model pay more attention to uncertain classes. For the target variable  $y$ , we expect its Dirichlet distribution  $p$  to obey a uniform distribution, i.e.  $y \sim \mathcal{N}(1/K, \lambda^2 \mathbf{I})$ . Finally, we use KL divergence to constrain the observed labels  $\hat{y}$  and the target variable  $y$  to obey the same distribution. Due to the analyzable nature of the Dirichlet distribution, Eq.(1) can be simplified as

$$\mathcal{L}_j^{\text{N-EDL}} = \sum_{k=1}^K \left( \frac{1}{K} - \frac{\alpha_{jk}}{\alpha_{j0}} \right)^2 \psi_1(\alpha_{jk}) - \lambda_1 \log |\mathcal{I}(\alpha_j)|, \quad (2)$$

where  $\alpha_{j0} = \sum_{k=1}^K \alpha_{jk}$ , and  $\psi_1(\cdot)$  represents the *trigamma* function, defined as  $\psi_1(x) = d\psi(x)/dx = d^2 \ln \Gamma(x)/dx^2$  (See Appendix A.2 for the detailed derivation process). We name the above weighted loss function as adaptive negative EDL loss. Since  $\psi_1(x)$  is a monotonically decreasing function when  $x > 0$ , class labels with less evidence would be subject to greater penalties to achieve a flatter output. Conversely, once a certain class of evidence is learned, the weight will be reduced (see Fig.1), so that the inlier features in the unlabeled dataset do not interfere too much with the learning of outlier features. We also use labeled dataset to enhance the learning of inlier features. We follow  $\mathcal{I}$ -EDL (Deng et al. 2023) to impose constraints on the uncertain class of labeled samples, outputting a sharp distribution that fits the ground truth. The loss function for an labeled sample  $(x_j^l, y_j^l) \in \mathcal{D}_l$  can be expressed as:

$$\mathcal{L}_j^{\text{P-EDL}} = \sum_{k=1}^K \left[ \left( y_{jk} - \frac{\alpha_{jk}}{\alpha_{j0}} \right)^2 + \frac{\alpha_{jk}(\alpha_{j0} - \alpha_{jk})}{\alpha_{j0}^2(\alpha_{j0} + 1)} \right] \psi_1(\alpha_{jk}) - \lambda_2 \log |\mathcal{I}(\alpha_j)|, \quad (3)$$

where  $y_j^l$  denotes one-hot encoded ground-truth of  $x_j^l$ .

**Strengthened KL Loss.** KL divergence loss  $\mathcal{L}_j^{\text{KL-ORI}}$  as part

of classical EDL loss, aims to penalize evidence for misleading classes that a sample does not belong to:

$$\mathcal{L}_j^{\text{KL-ORI}} = D_{KL}(\text{Dir}(\mathbf{p}_j|\hat{\alpha}_j)\|\text{Dir}(\mathbf{p}_j|\mathbf{1})). \quad (4)$$

where  $\hat{\alpha}_j = \alpha_j \odot (1 - \mathbf{y}_j^l) + \mathbf{y}_j^l$ . However,  $\mathcal{L}_j^{\text{KL-ORI}}$  only shrinks misleading evidence but leaves non-misleading evidence ignored. To help enhance non-misleading evidence, we propose to utilize a strengthened KL-divergence to force our predicted Dirichlet distribution  $\text{Dir}(\mathbf{p}_j|\alpha_j)$  to approach a new target Dirichlet distribution  $\text{Dir}(\mathbf{p}_j|\beta)$ . In particular, we introduce a new KL divergence term as follows,

$$\mathcal{L}_j^{\text{KL}} = D_{KL}(\text{Dir}(\mathbf{p}_j|\alpha_j)\|\text{Dir}(\mathbf{p}_j|\beta)). \quad (5)$$

For labeled data, we set  $\beta = [1, \dots, P, \dots, 1] \in \mathbb{R}_+^K$ , where  $P$  is a large target evidence value for ground-truth, e.g. 100. For unlabeled data, we set  $\beta = \mathbf{1}$  as the target distribution. The detailed closed-form expression of  $\mathcal{L}_j^{\text{KL}}$  can be found in Appendix A.3.

**Overall Loss Function.** For the EDL outlier detector, except Eq. (2), (3), (5), we add a consistency loss  $\mathcal{L}_{\text{CON}}$  to enhance the smoothness of detector, which is defined as

$$\mathcal{L}_{\text{CON}}(\mathcal{D}_u) = \frac{1}{N_i + N_o} \sum_{j=1}^{N_i+N_o} \|\alpha_j^s - \alpha_j^w\|_2^2, \quad (6)$$

where  $\alpha_j^s, \alpha_j^w$  represent the predicted evidence vector of strong and weak augmentations of the same unlabeled sample. For the Softmax classifier, we compute the standard cross-entropy loss  $\mathcal{L}_{\text{CE}}$  for labeled data and the FixMatch loss  $\mathcal{L}_{\text{FM}}$  for the high-certainty inliers  $\mathcal{I}_u$  identified by the EDL detector. Therefore, the overall loss function of our proposed ANEDL can be expressed as

$$\begin{aligned} \mathcal{L}_{\text{ANEDL}} = & \mathcal{L}_{\text{CE}}(\mathcal{D}_l) + \mathcal{L}_{\text{FM}}(\mathcal{I}_u) \\ & + \mathcal{L}_{\text{ANO}}(\mathcal{D}_l, \mathcal{D}_u) + \lambda_{\text{CON}} \mathcal{L}_{\text{CON}}(\mathcal{D}_u), \end{aligned} \quad (7)$$

where

$$\begin{aligned} \mathcal{L}_{\text{ANO}}(\mathcal{D}_l, \mathcal{D}_u) = & \lambda_{\text{P-EDL}} \frac{1}{N_l} \sum_{j=1}^{N_l} (\mathcal{L}_j^{\text{P-EDL}} + \mathcal{L}_j^{\text{KL}}) \\ & + \lambda_{\text{N-EDL}} \frac{1}{N_i + N_o} \sum_{j=1}^{N_i+N_o} (\mathcal{L}_j^{\text{N-EDL}} + \mathcal{L}_j^{\text{KL}}), \end{aligned} \quad (8)$$

$\lambda_{\text{P-EDL}}, \lambda_{\text{N-EDL}}$  and  $\lambda_{\text{CON}}$  are the hyperparameters used to control the trade-off for each objective.

To minimize our overall loss, we train our model for two stages. In the first stage, we pre-train our model with  $\mathcal{L} = \mathcal{L}_{\text{CE}}(\mathcal{D}_l) + \mathcal{L}_{\text{ANO}}(\mathcal{D}_l, \mathcal{D}_u) + \lambda_{\text{CON}} \mathcal{L}_{\text{CON}}(\mathcal{D}_u)$  which drops  $\mathcal{L}_{\text{FM}}$  for  $E_{\text{FM}}$  epochs. In second stage, we start self-training of classifier and use  $\mathcal{L}_{\text{ANEDL}}$  as loss function. Before every self-training epoch, we calibrate softmax head and EDL head to calculate uncertainty metric to select the inliers for softmax head to conduct self-training.

### Uncertainty Metric for Open-set SSL

**Inlier Selection in Self-training.** To help better enhance recognition performance with unlabeled data, we aim to select inliers with accurate pseudo labels from unlabeled data

for self-training. In FixMatch, pseudo label is generated by Softmax head and will also be used to softmax head as a supervision signal. Therefore, we propose to leverage EDL head to quantify the uncertainty of the class predicted by softmax head and select inliers with

confident prediction which are more likely to be accurately pseudo-labeled.

Specifically, given an unlabeled sample  $x_u$ , the Softmax head produce a one-hot pseudo label  $\hat{\mathbf{y}} \in \{0, 1\}^K$ , the EDL head output evidence  $\alpha \in \mathbb{R}_+^K$ . Then we use

$$\mathbf{M}_{\text{Self-training}} = \hat{\mathbf{y}} \cdot \alpha$$

as the confidence value for  $x_u$ . Larger values mean higher confidence, and we select top- $O$  most certain unlabeled samples as inliers. Here,  $O$  is a hyperparameter.

**Outlier Detection for Inference.** Detecting outliers in test data is also one of the tasks of Open-set SSL. Unlike selecting inliers from unlabeled data, detected inliers in test data do not have to consider calibrating with Softmax. Therefore, the uncertainty metric for inference is to accurately quantify epistemic uncertainty for better separation between inliers and outliers. Based on the evidence vector  $\alpha$  predicted by EDL head, we can derive various uncertainty metrics, i.e. mutual information, differential entropy, and total evidence.

Previous EDL methods demonstrate total evidence  $\alpha_0 = \sum_{k=1}^K \alpha_k$  achieves the best outlier detection performance (Sensoy, Kaplan, and Kandemir 2018; Deng et al. 2023). However, we empirically find that classes with small evidence values may form a long-tail effect because the EDL detector assigns a non-negative evidence value to each class. Although these evidence values are small, their total amount is not negligible, especially when dealing with tasks with a large number of classes. To avoid these long-tailed evidence values disturbing the separation between inliers and outliers, we propose to sort the evidence vectors in descending order and use only the sum of the top- $M$  evidence values as the confidence value for inference, which is defined as

$$\mathbf{M}_{\text{Inference}} = \sum_{i=1}^M \bar{\alpha}_i, \quad (9)$$

where  $\bar{\alpha}$  is the descending order of  $\alpha$ .

## Experiments

### Datasets

We evaluate our method on four datasets, including CIFAR-10, CIFAR-100 (Krizhevsky, Hinton et al. 2009), ImageNet-30 (Hendrycks and Gimpel 2016) and Mini-ImageNet (Vinyals et al. 2016). Appendix C.1 provides details about the dataset. We follow OpenMatch (2021) to tailor the dataset for Open-set SSL. Given a dataset, we first split its class set into inlier and outlier class sets. Next, we randomly select inliers to form labeled, unlabeled, validation, and test dataset. Then all outliers will be randomly sampled into unlabeled and test datasets. For the four datasets, we set several experimental settings with different numbers of labeled samples and different amounts of inliers/outliers classes.

Dataset	CIFAR-10			CIFAR-100		CIFAR-100		ImageNet-30
Inlier/Outlier Classes	6/4			55/45		80/20		20/10
No. of labeled samples	50	100	400	50	100	50	100	10%
FixMatch	56.1±0.6	60.4±0.4	71.8±0.4	72.0±1.3	79.9±0.9	64.3±1.0	66.1±0.5	88.6±0.5
MTC	96.6±0.6	98.2±0.9	98.9±0.1	81.2±3.4	80.7±4.6	79.4±1.0	73.2±3.5	93.8±0.8
T2T	43.1±8.8	43.1±14.1	56.2±1.4	60.4±8.9	62.2±4.4	74.2±6.3	65.4±13.4	55.7±10.8
OpenMatch	<b>99.3±0.9</b>	<b>99.7±0.2</b>	99.3±0.2	87.0±1.1	86.5±2.1	86.2±0.6	86.8±1.4	96.4±0.7
Ours	98.5±0.2	99.0±0.2	<b>99.4±0.1</b>	<b>90.3±0.1</b>	<b>91.1±0.4</b>	<b>89.2±0.5</b>	<b>91.1±1.0</b>	<b>96.6±0.3</b>

Table 1: Mean and standard deviation of AUROC (%) on CIFAR-10, CIFAR-100 and ImageNet-30. Higher is better. For CIFAR-10 and CIFAR-100, the three runs correspond to three different folds. For ImageNet-30, the three runs are conducted on the same folder but with different random seeds.

Dataset	CIFAR-10			CIFAR-100		CIFAR-100		ImageNet-30
Inlier/Outlier Classes	6/4			55/45		80/20		20/10
No. of labeled samples	50	100	400	50	100	50	100	10%
FixMatch	43.2±1.2	29.8±0.6	16.3±0.5	35.4±0.7	27.3±0.8	41.2±0.7	34.1±0.4	12.9±0.4
MTC	20.3±0.9	13.7±0.9	9.0±0.5	33.5±1.2	27.9±0.5	40.1±0.8	33.6±0.3	13.6±0.7
T2T	10.4±1.2	11.3±1.3	9.5±1.4	29.1±0.9	27.0±0.4	35.7±1.4	31.6±1.0	12.2±0.9
OpenMatch	10.4±0.9	7.1±0.5	5.9±0.5	27.7±0.4	24.1±0.6	33.4±0.2	29.5±0.3	<b>10.4±1.0</b>
Ours	<b>9.6±0.4</b>	<b>7.0±0.4</b>	<b>5.5±0.2</b>	<b>26.7±0.2</b>	<b>23.2±0.6</b>	<b>32.3±0.3</b>	<b>28.5±0.2</b>	11.3±0.4

Table 2: Error rate (%) corresponding to Table 1. Lower is better.

## Implementation Details

We employ a randomly initialized Wide ResNet-28-2 (Zagoruyko and Komodakis 2016) to conduct experiments on CIFAR-10 and CIFAR-100, and a randomly initialized ResNet-18 (He et al. 2016) for experiments on ImageNet-30 and Mini-ImageNet. The Softmax head is a linear layer and the EDL head is non-linear, which is composed of 4 MLP layers. For the pre-training stage, we set its length  $E_{FM}$  as 10 for all experiments. We set the length of an epoch as 1024 steps, which means we select a new set of inliers every 1024 steps during self-training. For hyper-parameters of FixMatch, we set them the same as in OpenMatch (Saito, Kim, and Saenko 2021). We adopt SGD optimizer with 0.0001 weight decay and 0.9 momentum to train our model by setting initial learning rate as 0.03 with the cosine decrease policy. We set the batch size of labeled and unlabeled samples as 64 and 128 respectively for all experiments. Experiments on ImageNet-30 are conducted with a single 24-GB GPU and other experiments are conducted with a single 12-GB GPU. All experiments in this work are conducted over three runs with different random seeds, and we report the mean and standard derivation of three runs.

## Comparison with Other Methods

To illustrate the effectiveness of our method on Open-set SSL, we compare it against several baseline methods, including FixMatch (Sohn et al. 2020), MTC (Yu et al. 2020), T2T (Huang et al. 2021) and OpenMatch (Saito, Kim, and Saenko 2021). We use classification error rate to evaluate the performance of classifier in inliers and AUROC to eval-

uate the performance of outlier detection. AUROC is the general evaluation metric for novelty detection (Hendrycks and Gimpel 2016). Table 1 and Table 2 show the AUROC and error rate on the CIFAR-10, CIFAR-100, and ImageNet-30, respectively, where the results of other methods are quoted from OpenMatch (Saito, Kim, and Saenko 2021). Figure 2 (a) and (b) reports the experimental results on Mini-ImageNet, where the results of OpenMatch are obtained based on the author’s own implementation.

As shown in Table 1, we can see that our method outperforms the SOTA method, i.e., OpenMatch, in most experimental settings with respect to AUROC. In Table 2, our proposed method also achieves SOTA performance in terms of error rate in 7 out of 8 settings. These results illustrate that our method is more advanced in epistemic uncertainty quantification and has superior performance in outlier detection as well as inlier classification. Notably, in the more challenging settings with a large number of classes, i.e., settings on CIFAR-100, our method significantly outperforms OpenMatch in both error rate and AUROC, with improvements of  $\sim 1\%$  and  $\sim 3\%$ . It shows our multinomial detector is more robust than the binary detector in identifying outliers from complex modalities when dealing with K-way classification tasks with large K. To further demonstrate our superiority in settings with a large number of inlier classes, we compare our method with OpenMatch on Mini-ImageNet, which is even more difficult than CIFAR-100 with larger image size and more complex patterns. As shown in Figure 2 (a) and (b), our method consistently outperforms OpenMatch with respect to both AUROC and error rate.

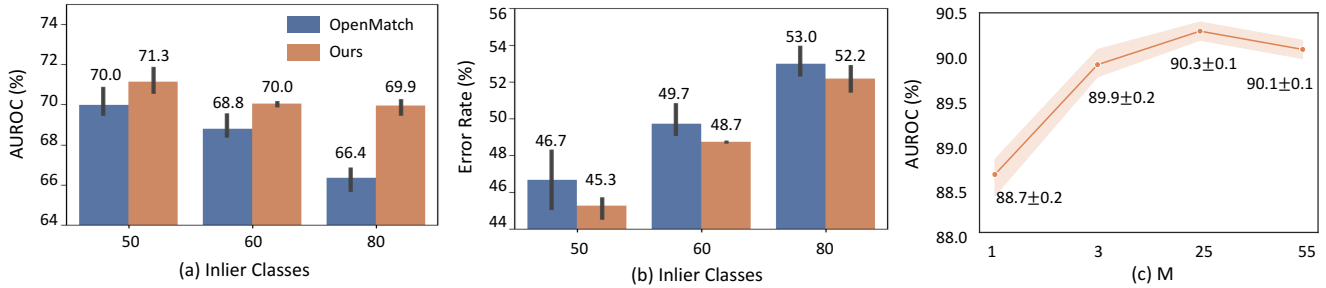


Figure 2: (a) AUROC (%) and (b) Error rate (%) compared to OpenMatch on Mini-ImageNet. (c) The effect of  $M$  using in  $M_{\text{Inference}}$  during inference. We use the same fold to conduct three runs and report the mean of three runs at the top of each bar.

NO	Adaptive	AUROC	Error rate
		88.4±0.4	27.0±0.3
✓		89.0±0.3	26.9±0.3
✓	✓	<b>90.3±0.1</b>	<b>26.7±0.2</b>

Table 3: Component analysis of Adaptive Negative Optimization (ANO).

Self-training	Inference	AUROC	Error rate
$M_{\text{Inference}}$	$M_{\text{Inference}}$	89.8±0.2	27.8±0.2
$M_{\text{Self-training}}$	$M_{\text{Self-training}}$	88.4±0.3	26.7±0.2
$M_{\text{Self-training}}$	$M_{\text{Inference}}$	<b>90.3±0.1</b>	<b>26.7±0.2</b>

Table 4: Ablation study of uncertainty metric in self-training and inference.

## Ablation Studies

All experiments of our ablation studies are conducted under the setting of 55 inlier classes and 50 labeled samples per class on CIFAR-100. We report the mean and standard deviation of three runs on three different folds.

**Component Analysis of ANO.** Our ANO loss consists of two components, i.e. negative optimization and adaptive loss weight. For “without NO and Adaptive”, we directly replace  $\mathcal{L}_{\text{ANO}}$  to classical EDL loss, which means we apply no explicit constraints to outliers in unlabeled data. Since the information contained in unlabeled data keeps underutilized, both classification and outlier detection performance are suboptimal as shown in Table 3. For “without Adaptive”, we remove our adaptive loss weight from  $\mathcal{L}_{\text{ANO}}$ , which means not using the FIM to identify the amount of information for each class of unlabeled samples. As a result, all unlabeled samples including inliers are equally regulated to output K-low evidence values. Although negative optimization compresses evidence values of unlabeled outliers, it simultaneously depresses the evidence of unlabeled inliers, making inlier features and outlier features interfere with each other. Therefore, its improvement compared to “without NO and Adaptive” is still limited. Our proposed  $\mathcal{L}_{\text{ANO}}$ , i.e., “with NO and Adaptive”, peaks the best performance by sufficiently compressing evidence values of unlabeled outliers without interfering with the learning of unlabeled inliers.

**Evaluation of Uncertainty Metric.** To illustrate the effectiveness of designing different uncertainty metrics for self-training and inference, we conduct an ablation study and summarize it in Table 4. Results show that using  $M_{\text{Self-training}}$  and  $M_{\text{Inference}}$  for self-training and inference, respectively, achieves the best performance in both AUROC and error rate. Since  $M_{\text{Self-training}}$  cannot distinguish outliers and

inliers with high aleatoric uncertainty, the AUROC using  $M_{\text{Self-training}}$  is significantly lower than that of  $M_{\text{Inference}}$ . However, the effect of using  $M_{\text{Inference}}$  during self-training outperforms using  $M_{\text{Self-training}}$ , probably due to the fact that inliers with high aleatoric uncertainty interfere with self-training. Note that using different metrics during inference only affects the effectiveness of outlier detection, not inlier classification. Furthermore, since  $M_{\text{Inference}}$  uses the sum of top- $M$  evidence values as the metric, we also conduct an ablation study to evaluate the effect of  $M$ . As shown in Figure 2 (c), the sum of top  $M = 25$  evidence values is a better metric to detect outliers, compared with the sum of all  $K = 55$  evidence values. Note that the improvements steadily appear in each trained model. Comparison with other values of  $M$  also shows that too small  $M$  will miss useful evidence value and lead to insufficient measurement of uncertainty.

## Conclusion

In this work, we propose a novel framework, adaptive negative evidential deep learning (ANEDL), for open-set semi-supervised learning. ANEDL adopts evidential deep learning (EDL) to Open-set SSL for the first time and designs novel adaptive negative optimization method. In particular, EDL, as an advanced uncertainty quantification method, is deployed to estimate different types of uncertainty for inliers selection in self-training and outlier detection in inference. Furthermore, to enhance the separation between inliers and outliers, we propose adaptive negative optimization to explicitly compress the evidence value of outliers in unlabeled data and avoid interfering with the learning of inliers in unlabeled data with adaptive loss weight. Our extensive experiments on four datasets demonstrate that our method is superior to other SOTAs.

## Acknowledgments

This work was supported by the National Key R&D Program of China (2022YFE0200700), the Research Grants Council of the Hong Kong SAR, China (Project No. T45-401/22-N and Project No.14201620), the National Natural Science Foundation of China (Project No. 6237073934, 3234101132 62006219) and the Natural Science Foundation of Guangdong Province (2022A1515011579).

## References

- Bao, W.; Yu, Q.; and Kong, Y. 2021. Evidential deep learning for open set action recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, 13349–13358.
- Bao, W.; Yu, Q.; and Kong, Y. 2022. OpenTAL: Towards Open Set Temporal Action Localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2979–2989.
- Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; and Raffel, C. A. 2019. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32.
- Berthelot, D.; Roelofs, R.; Sohn, K.; Carlini, N.; and Kurakin, A. 2021. Adamatch: A unified approach to semi-supervised learning and domain adaptation. *arXiv preprint arXiv:2106.04732*.
- Carboneau, M.-A.; Cheplygina, V.; Granger, E.; and Gagnon, G. 2018. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*, 77: 329–353.
- Chen, Y.; Zhu, X.; Li, W.; and Gong, S. 2020. Semi-supervised learning under class distribution mismatch. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 3569–3576.
- Deng, D.; Chen, G.; Yu, Y.; Liu, F.; and Heng, P.-A. 2023. uncertainty estimation by fisher information-based evidential deep learning. *arXiv preprint arXiv:0712.0248*.
- Guo, C.; Pleiss, G.; Sun, Y.; and Weinberger, K. Q. 2017. On calibration of modern neural networks. In *International conference on machine learning*, 1321–1330. PMLR.
- Guo, L.-Z.; Zhang, Z.-Y.; Jiang, Y.; Li, Y.-F.; and Zhou, Z.-H. 2020. Safe deep semi-supervised learning for unseen-class unlabeled data. In *International conference on machine learning*, 3897–3906. PMLR.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Hendrycks, D.; and Gimpel, K. 2016. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*.
- Huang, J.; Fang, C.; Chen, W.; Chai, Z.; Wei, X.; Wei, P.; Lin, L.; and Li, G. 2021. 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*, 8310–8319.
- Ishida, T.; Niu, G.; Hu, W.; and Sugiyama, M. 2017. Learning from complementary labels. *Advances in neural information processing systems*, 30.
- Jøssang, A. 2016. *Subjective logic*, volume 3. Springer.
- Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images.
- Lee, D.-H.; et al. 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, 896.
- Malinin, A.; and Gales, M. 2018. Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, 31.
- Saito, K.; Kim, D.; and Saenko, K. 2021. Openmatch: Open-set semi-supervised learning with open-set consistency regularization. *Advances in neural information processing systems*, 34: 25956–25967.
- Saito, K.; and Saenko, K. 2021. Ovanet: One-vs-all network for universal domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9000–9009.
- Sensoy, M.; Kaplan, L.; and Kandemir, M. 2018. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31.
- Sentz, K.; and Ferson, S. 2002. Combination of evidence in Dempster-Shafer theory.
- Sohn, K.; Berthelot, D.; Carlini, N.; Zhang, Z.; Zhang, H.; Raffel, C. A.; Cubuk, E. D.; Kurakin, A.; and Li, C.-L. 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33: 596–608.
- Soleimany, A. P.; Amini, A.; Goldman, S.; Rus, D.; Bhatia, S. N.; and Coley, C. W. 2021. Evidential deep learning for guided molecular property prediction and discovery. *ACS central science*, 7(8): 1356–1367.
- Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; and Wojna, Z. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826.
- Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D.; et al. 2016. Matching networks for one shot learning. *Advances in neural information processing systems*, 29.
- Wang, X.; Wu, Z.; Lian, L.; and Yu, S. X. 2022a. Debiased learning from naturally imbalanced pseudo-labels. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 14647–14657.
- Wang, Y.; Chen, H.; Heng, Q.; Hou, W.; Savvides, M.; Shinzaki, T.; Raj, B.; Wu, Z.; and Wang, J. 2022b. Freematch: Self-adaptive thresholding for semi-supervised learning. *arXiv preprint arXiv:2205.07246*.
- Wilson, A. G.; and Izmailov, P. 2020. Bayesian deep learning and a probabilistic perspective of generalization. *Advances in neural information processing systems*, 33: 4697–4708.

- Xu, Y.; Shang, L.; Ye, J.; Qian, Q.; Li, Y.-F.; Sun, B.; Li, H.; and Jin, R. 2021. Dash: Semi-supervised learning with dynamic thresholding. In *International conference on machine learning*, 11525–11536. PMLR.
- Yu, Q.; Ikami, D.; Irie, G.; and Aizawa, K. 2020. Multi-task curriculum framework for open-set semi-supervised learning. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*, 438–454. Springer.
- Zagoruyko, S.; and Komodakis, N. 2016. Wide residual networks. *arXiv preprint arXiv:1605.07146*.
- Zhang, B.; Wang, Y.; Hou, W.; Wu, H.; Wang, J.; Okumura, M.; and Shinozaki, T. 2021. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in neural information processing systems*, 34: 18408–18419.
- Zhao, X.; Chen, F.; Hu, S.; and Cho, J.-H. 2020. Uncertainty aware semi-supervised learning on graph data. *Advances in neural information processing systems*, 33: 12827–12836.
- Zheng, M.; You, S.; Huang, L.; Wang, F.; Qian, C.; and Xu, C. 2022. Simmatch: Semi-supervised learning with similarity matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 14471–14481.