Unlocking the Power of Open Set: A New Perspective for Open-Set Noisy Label Learning

Wenhai Wan*, Xinrui Wang*, Ming-Kun Xie, Shao-Yuan Li[†], Sheng-Jun Huang, Songcan Chen

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics MIIT Key Laboratory of Pattern Analysis and Machine Intelligence

Abstract

Learning from noisy data has attracted much attention, where most methods focus on closed-set label noise. However, a more common scenario in the real world is the presence of both open-set and closed-set noise. Existing methods typically identify and handle these two types of label noise separately by designing a specific strategy for each type. However, in many real-world scenarios, it would be challenging to identify open-set examples, especially when the dataset has been severely corrupted. Unlike the previous works, we explore how models behave when faced with open-set examples, and find that a part of open-set examples gradually get integrated into certain known classes, which is beneficial for the separation among known classes. Motivated by the phenomenon, we propose a novel two-step contrastive learning method CECL (Class Expansion Contrastive Learning) which aims to deal with both types of label noise by exploiting the useful information of open-set examples. Specifically, we incorporate some open-set examples into closed-set classes to enhance performance while treating others as delimiters to improve representative ability. Extensive experiments on synthetic and real-world datasets with diverse label noise demonstrate the effectiveness of CECL.

Introduction

Deep neural networks (DNNs) have achieved remarkable success in various tasks. The great success is primarily attributed to large amounts of data with high-quality annotations, which are expensive or even inaccessible in practice. Actually, datasets collected via search engines or crowd-sourcing platforms inevitably involve noisy labels (Xiao et al. 2015; Li et al. 2018; Li, Huang, and Chen 2021; Shi, Li, and Huang 2023). Given the powerful learning capacities of DNNs, the model will ultimately overfit the label noise and lead to poor generalization performance (Arpit et al. 2017; Zhang et al. 2021a). To mitigate this issue, it is significant to develop robust models for learning from noisy labels.

As shown in Figure 1, we divide the mislabeled examples into two types: *closed-set* and *open-set*. More specifically, a *closed-set* mislabeled example occurs when its true class fall within the set of known classes {*cat*, *dog*, *elephant*}, while



Figure 1: An example of open-set noisy label learning problem. {*cat, dog, elephant*} is the concerned *known classes*. The left, middle, and right columns respectively show images that are correctly labeled, wrongly labeled with *closedset* and *open-set* noise.

an *open-set* mislabeled example occurs when its true class does not fall within the set of known classes in the training data. To our knowledge, existing works mainly focus on closed-set scenarios (Li, Socher, and Hoi 2020; Li et al. 2022a), while a more common scenario in the real world is the presence of both *closed-set* and *open-set*.

The problem has been formalized as a learning framework called Open-Set Noisy Label Learning (OSNLL) (Wang et al. 2018), which is also the main focus of our work. The most challenging issue in OSNLL arises due to the presence of open-set examples. The detection and handling of these examples pose a significant challenge, especially when their distribution is unknown. Several papers (Wang et al. 2018; Yao et al. 2021; Sun et al. 2022) have proposed specific methods and frameworks by identifying mislabeled open-set examples and minimizing their impact. However, empirical findings (Morteza and Li 2022) and theoretical results (Fang et al. 2022) suggest that some open-set examples become increasingly difficult for the model to distinguish. In some cases, it is even impossible to recognize open-set examples

^{*}Equal contribution.

[†]Corresponding author.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

from clean datasets, making learning with both closed-set and open-set examples particularly challenging.

In this paper, we give a more in-depth study of the effect of open-set examples on OSNLL. Specifically, we first investigate the model's training behavior when facing openset examples, observing that some open-set classes get integrated with closed-set classes. By treating these open-set examples as closed-set ones, the models surprisingly acquire substantial enhancement in the classification ability. We call this the **Class Expansion** phenomenon. An intuitive explanation is, by introducing these "false positive" examples from the unknown class data, the representation of different known classes becomes more generalized.

Based on the above observations, we propose a novel twostep Class Expansion Contrastive Learning (CECL) framework to make full use of open-set examples while minimizing the negative impact of label noise. In the first step, we roughly tackle the data noise and maintain the basic concept of clean and noisy examples. In the second step, we adopt a contrastive learning scheme and selectively incorporate certain indistinct open-set examples into their corresponding similar known classes. In fact, the concept of known classes is broadened to include these similar, but previously unknown examples. Additionally, to enhance the discrimination between different classes, we use the remaining distinguishable open-set examples as delimiters.

In summary, our contributions are threefold:

- We explore the relationship between open-set and closedset classes and highlight the phenomenon of Class Expansion in OSNLL, demonstrating the potential of openset examples to facilitate known class learning.
- We propose a novel CECL framework that incorporates contrastive learning to fully utilize open-set examples, which better generalizes and discriminates the representation of different known classes.
- We provide theoretical guarantees for better representation capability of CECL, and conduct extensive experiments on both synthetic and real-world noisy datasets to validate its effectiveness.

Related Work

Noisy Label Learning Most of the methods in literature mitigate the label noise by robust loss functions(Wang et al. 2019b,a; Xu et al. 2019; Liu et al. 2020; Wei et al. 2021), noise transition matrix(Patrini et al. 2017; Tanno et al. 2019; Xia et al. 2019; Li et al. 2022b), sample selection(Han et al. 2018; Yu et al. 2019; Wei et al. 2020; Li et al. 2022a; Huang, Zhang, and Shan 2023; Xia et al. 2023), and label correction(Li, Socher, and Hoi 2020; Li, Xiong, and Hoi 2021; Ortego et al. 2021; Zhang et al. 2023).

Open-set Noisy Label Learning Open-set examples in the real-world are ubiquitous(Yang et al. 2022, 2023). Open-set noisy label learning aims to develop a well-performing model from real-world datasets that contain both closed-set and open-set noisy labels. (Wang et al. 2018) proposes an iterative learning framework called pcLOF to detect open-set noisy labels and then uses a re-weighting strategy in the

objective function. (Yao et al. 2021; Sun et al. 2022) introduces an approach based on consistency to identify openset examples, and then minimize the impact they bring, (Li, Xiong, and Hoi 2020) proposes momentum prototypes for webly-supervised representation learning which lacks consideration for the diversity of open-set examples, (Wu et al. 2021) introduces a noisy graph cleaning framework that simultaneously performs noise correction and clean data selection, (Xia et al. 2022) designs an extended T-estimator to estimate the cluster-dependent transition matrix by exploiting the noisy data. These methods rest on the assumption that the open-set examples are harmful to known class learning, which is questionable as we will show in this work.

The Class Expansion Phenomenon

In real-world scenarios, open-set examples are pervasive, posing significant challenges to many problems. Many researchers strive to distinguish open-set examples and mitigate their influence. In fact, each inclusion of an open-set example in the dataset has a reason behind it, either due to misjudgment, inherent attributes that are easily confused, or a lack of clear definition of the open-set, and so on. It might be unreasonable to simply treat all open-set examples caused by various reasons as a single type of outlier! Furthermore, some open-set examples may share similar features with closed-set examples, which can be confusing even for humans. Simply eliminating the impact of open-set examples may potentially exert negative effects on the learning of closed-set examples. So, is there room for us to explore a different avenue?



Figure 2: The distribution of open-set examples among different known classes on CIFAR10 and MNIST, respectively.

In the pursuit of an answer to the above question, we investigate the behavior of DNNs when encountering unseen open-set examples. Specifically, we conducted a series of experiments on the CIFAR10 (Krizhevsky 2009), MNIST (Le-Cun 1998) and Tiny-Imagenet (Le and Yang 2015) datasets. We randomly select 80% of the total categories as known classes and treat the remaining classes as unknown for all datasets. We train the model on labeled closed-set examples and generated pseudo-labels on open-set examples. Figure 2 illustrates the class distribution of pseudo-labels generated by the model. On the CIFAR10 dataset, it can be observed that open-set *dog* examples are almost uniformly classified as *cat*. This phenomenon occurs mainly due to the fact that

dog and cat share high-level similarity, making the model hard to distinguish between these two semantic objects and merge them into a generalized concept of cat. We use the term "class expansion" to describe this phenomenon. The class expansion phenomenon can also be found on MNIST (e.g. open-set digit 7 are consistently being classified as closed-set class 2), however, it is noteworthy that not all open-set class concepts can be treated as class expansion to a specific closed-set class concept (e.g. open-set digit 3 are relatively evenly distributed among multiple closed-set classes). Our observations indicate that when encountering open-set examples, the model tends to make reasoned judgments rather than reckless ones.¹



(b) Training process on Tiny-Imagenet.

Figure 3: Experimental results of incorporating examples from open-set classes into known classes for learning on CI-FAR10 and Tiny-Imagenet. 'CS', 'CS and OS' respectively denote training only on closed-set class examples and a mixture of closed-set and open-set class examples.

To further explore the effectiveness of class expansion. we perform simple experiments by selecting those open-set examples with high prediction scores on closed-set classes into the training set and retraining the model. Figure 3 (a) presents the results obtained by training using only closedset examples (CS) and training using both closed-set examples and open-set examples simultaneously (CS and OS). It is surprising that the incorporation of open-set examples significantly improves the model's performance on closedset classes, which contradicts our common sense. It is also worth noting that the improvement becomes more obvious as the dataset's scale increases, see the training process for Tiny-Imagenet shown in Figure 3 (b). The above findings indicate the potential power of open-set examples to boost known class recognition. Below, we amplify the usefulness of open-set examples by incorporating the contrastive learning into model training, leading to our Class Expansion Contrastive Learning (CECL) method.

Methodology

Let \mathcal{X} denotes the instance space and $\mathcal{Y} = \{1, 2, ..., c\}$ the label space with c distinct classes. We use $\mathcal{D} = \{(x_i, y_i) | 1 \le i \le n\}$ to denote the training dataset, with $y_i \in [0, 1]^c$ denoting the one-hot label vector over c classes. For datasets with noisy labels, $\{y_i\}$ are not guaranteed to be correct. Formally, we represent the ground-truth label of example x_i with a one-hot label vector $\hat{y}_i \in [0, 1]^{c+1}$ over c + 1 classes, where the (c + 1)-th class indicates the open-set class. It is noteworthy that \hat{y}_i is unknown throughout the entire training process. Due to the presence of closed-set and open-set label noise, directly training a model on \mathcal{D} fails to achieve favorable generalization performance.

In this paper, building on the comprehensive success of contrastive learning in extensive tasks, we propose a novel class expansion contrastive learning (CECL) framework to better leverage and unlock the power of open-set examples. Figure 4 shows the intuition of CECL. Provided that we have detected representative examples for the known classes, then the *distinguishable* open-set examples that possess very different features from the known class examples can be reliably detected. The remaining examples can be either known class examples or class expansion contributing open-set examples. From the contrastive learning perspective, we can wisely use these two types of open-set examples to push away the classes and better generalize their concept boundary. In the following, we introduce the details.



Figure 4: The intuition of CECL. CECL incorporates certain indistinct open-set examples into the known classes, which are expected to contribute to class expansion with better generalization. Additionally, the distinguishable open-set examples are used as delimiters, which are expected to push away between the known classes with better discrimination.

Class Expansion Contrastive Learning

Although supervised contrastive learning has been extensively studied, its application in OSNLL has not been explored due to two main challenges. First, it is difficult to construct the positive example set, second, it is challenging to determine whether an open-set example should be included in the positive example set.

To solve the above challenges, we introduce an efficient and robust two-step framework. In the first step, we per-

¹A similar phenomenon is observed on Tiny-Imagenet, and we provide the transition matrix in the Appendix.



Figure 5: Illustration of CECL. According to the information obtained in the first step, clean examples are used to generate prototypes for each class, certain open-set examples are incorporated into known classes in the form of class expansion, and remaining are perceived as delimiters. The momentum embeddings are maintained by a queue structure. '//' means stop gradient.

form a pretraining phase to obtain coarse labels for the entire dataset and identify whether each example is clean or not.

In the second step, we leverage the predictions from the first step to generate cluster prototypes for each class, which we then use to perform label noise filter and open-set decision. The identified distinguishable open-set examples are used to enhance the model's representation learning ability in our improved contrastive learning framework. Overall, our proposed framework provides an effective approach for training models on datasets with noisy labels in open-set scenarios. Our approach is illustrated in Figure 5.

Step 1: Clean Example Identification. To ensure more efficient training in the subsequent step, we conduct the first step in our proposed method. Specifically, we adopted one effective off-the-shelf closed-set noisy label learning method Promix (Wang et al. 2022) to identify a subset of clean examples. Promix adopts a dual network co-teaching learning style, we further ameliorate it with one correction record Tindicating whether one instance is mislabeled. $T(\cdot)$ is a 0/1valued indicator that records whether the model's prediction is the same as the instance's label during the whole training process. Then, we can divide the original noisy dataset \mathcal{D} into two subsets \mathcal{D}_{clean} and \mathcal{D}_{noisy} . \mathcal{D}_{clean} includes the examples that are highly believed to be clean, i.e., $T(\cdot) = 0$. \mathcal{D}_{noisy} contains the rest unclear examples might be noisy. For \mathcal{D}_{noisy} , Promix re-labels them with the most similar known class, thus, we obtain coarse labels Y' for \mathcal{D} .

Step 2: Contrastive Learning We adopt the most popular contrastive learning setups following SupCon (Khosla et al. 2020) and MoCo (He et al. 2020). Given each example (x, y), we generate its query and key view through randomized data augmentation $\operatorname{Aug}(x)$, then feed them into the query and key network $g(\cdot), g'(\cdot)$, yielding a pair of L_2 -normalized embeddings $q = g(\operatorname{Aug}_q(x)), k = g'(\operatorname{Aug}_k(x))$. In implementations, the query network shares the same convolutional blocks as the classifier, followed by a prediction head. Following MoCo, the key network uses a

momentum update with the query network. We additionally maintain a queue storing the most current key embeddings kand update it chronologically. To this end, we have the contrastive embedding pool $A = B_q \cup B_k \cup queue$, where B_q and B_k are vectorial embeddings corresponding to the query and key views of the current mini-batch. Given an example x, the per-example contrastive loss is defined by contrasting its query embedding with the remainder of the pool A:

$$\mathcal{L}_{\text{cont}}(g; \boldsymbol{x}, t, A) = -\frac{1}{|P(\boldsymbol{x})|} \sum_{\boldsymbol{k}_{+} \in P(\boldsymbol{x})} \log \frac{\exp\left(\boldsymbol{q}^{\top} \boldsymbol{k}_{+}/t\right)}{\sum_{\boldsymbol{k}' \in A(\boldsymbol{x})} \exp\left(\boldsymbol{q}^{\top} \boldsymbol{k}'/t\right)}.$$
 (1)

Here $P(\mathbf{x})$ is the set of positive examples for \mathbf{x} , and $A(\mathbf{x}) = A \setminus \{q\}, \tau$ is the temperature. Conventionally, the positive examples are from the same class and the negative examples are from different classes. However, in the OSNLL problem, the existence of label noise and open-set examples makes constructing the positive example set $P(\mathbf{x})$ particularly challenging. Below, we explain how we perform prototype-based contrastive learning to select positive example set and make use of open-set examples.

Positive Set Selection The positive example set is entirely derived from A(x), where A(x) consists of subsets from \mathcal{D}_{clean} and \mathcal{D}_{noisy} , which are referred to as the clean part and the noisy part, respectively. For the clean part, we can directly construct the positive examples based on its label:

$$P_{clean}(\boldsymbol{x}) = \{ \boldsymbol{k} \mid \boldsymbol{k} \in A(\boldsymbol{x}) \cap \mathcal{D}_{clean}, y = y' \}, \quad (2)$$

where y' and y are the coarse label of example x and k.

However, we cannot directly construct the positive example set from the noisy part due to the existence of noise. Therefore, we utilize \mathcal{D}_{clean} to guide the construction of positive example set within the noisy part, by generating an initialized L_2 -normalized prototype for each class:

$$Q_i = \frac{1}{n_i} \sum_{j=1}^{n_i} q_{ij}, \quad i \in \mathcal{Y},$$
(3)

where n_i denote the total number of examples for class i in \mathcal{D}_{clean} and q_{ij} represents the L_2 -normalized embedding of j-th example in class i.

Subsequently, we compute the distances between the examples in noisy part and their corresponding class prototypes and then determine whether they belong to the positive example set based on a threshold τ :

$$P_{noisy}(\boldsymbol{x}) = \{ \boldsymbol{k} \mid \boldsymbol{k} \in A(\boldsymbol{x}) \cap \mathcal{D}_{noisy}, y = y', \\ Distance(\boldsymbol{k}, Q_y) < \tau \},$$
(4)

where $Distance(a, b) = 1 - (a \cdot b)/(||a||||b||)$. Then, we combine with P_{clean} to obtain a complete positive example set: $P(x) = P_{clean}(x) \cup P_{noisy}(x)$. Correspondingly, we can also formalized the incorporated examples from \mathcal{D}_{noisy} as: $F_q = \mathbb{I}(q \in \mathcal{D}_{noisy} \text{ and } Distance(q, Q_{y'}) < \tau)$. When $F_q = 0$, we treat q as a distinguishable open-set example and refer to this process as the **open set decision**. When $F_q = 1$, we consider q as a clean example. During the subsequent training process, we update the prototype in a moving-average style: $Q_i = \text{Normalize}(\gamma Q_i + (1-\gamma)q)$, if i = y'. Here the prototype Q_i of class i is defined by the moving average of the normalized query embeddings q whose coarse label conforms to i. γ is a tunable hyperparameter.

Also, our classification term can be represented as:

$$\mathcal{L}_{CLS} = -\frac{1}{|\mathcal{D}_{clean}|} \sum_{i \in \mathcal{D}_{clean}} \sum_{j=1}^{c} y_i^j \log\left(p_i^j\right) - \frac{1}{\sum_{k=1}^{n} \mathbb{I}(F_k = 1)} \sum_{i \in \mathcal{D}_{noisy}} \sum_{j=1}^{c} \mathbb{I}(F_i = 1) y_i'^j \log\left(p_i^j\right),$$
(5)

where y_i^j and p_i^j denote values of the one-hot label and softmax output of the network of example x_i in the *j*-th class. **Training Objective** For examples that are distinct from any known classes, their embeddings tend to project toward the inter-class rather than the intra-class. Therefore, by leveraging the nature of these examples, we treat them as delimiters, continuously pushing the known classes away from them, i.e., they are used as negative examples for all known classes. Thus, our contrastive term can be represented as:

$$\mathcal{L}_{\text{CONT}} = \frac{1}{|\mathcal{D}_{clean}|} \sum_{i \in \mathcal{D}_{clean}} \mathcal{L}_{\text{cont}}(g; \boldsymbol{x}_{i}, t, A) + \frac{1}{\sum_{i=1}^{n} \mathbb{I}(F_{i}=1)} \sum_{i \in \mathcal{D}_{noisy}} \mathbb{I}(F_{i}=1) \mathcal{L}_{\text{cont}}(g; \boldsymbol{x}_{i}, t, A).$$
(6)

Finally, we put it all together and jointly train the classifier as well as the contrastive network with the overall loss function as:

$$\mathcal{L} = \mathcal{L}_{\text{CLS}} + \beta \mathcal{L}_{\text{CONT}}, \qquad (7)$$

in which β is a trade-off parameter.

Theoretical Analysis

In this section, we theoretically demonstrate that distinguishable open-set examples can contribute to enhancing the discriminative capabilities of contrastive learning. **Definition 1** ((σ , δ)-Augmentation). The augmentation set A is called a (σ , δ)-augmentation, if for each class C_k , there exists a subset $C_k^0 \subseteq C_k$ (called a main part of C_k), such that both $\mathbb{P}[\boldsymbol{x} \in C_k^0] \ge \sigma \mathbb{P}[\boldsymbol{x} \in C_k]$ where $\sigma \in (0, 1]$ and $\sup_{x_1, x_2 \in C_k^0}, d_A(x_1, x_2) \le \delta$ hold, where $d_A(\boldsymbol{x}_1, \boldsymbol{x}_2) = \min_{\boldsymbol{x}'_1 \in A(\boldsymbol{x}_1), \boldsymbol{x}'_2 \in A(\boldsymbol{x}_2)} \|\boldsymbol{x}'_1 - \boldsymbol{x}'_2\|$ It represents the augmented distance. (Wang and Isola 2020; Khosla et al. 2020; Duchi and Namkoong 2021).

With Definition 1,our analysis will focus on the examples in the main parts with good alignment, i.e., $(C_1^0 \cup \cdots \cup C_K^0) \cap S_{\varepsilon}$, where $S_{\varepsilon} := \{ \boldsymbol{x} \in \bigcup_{k=1}^K C_k : \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in A(\boldsymbol{x}), \|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)\| \leq \varepsilon \}$ is the set of examples with ε -close representations among augmented data. Furthermore we let $R_{\varepsilon} := \mathbb{P}[\overline{S_{\varepsilon}}]$.

We transform Eq.(6) to the following formulation:

$$\mathcal{L}_{CONT} = a \sum_{i \in \mathcal{Y}} \sum_{m \in D_i} \mathcal{L}_{align} + b \sum_{\substack{i,j \in \mathcal{Y} \\ i \neq j}} \sum_{\substack{m \in D_i \\ t \in D_j}} \mathcal{L}_{uniform},$$

where D_y denotes index set corresponding to true class y, a and b are fixed constant related to the example quantity, and $\mathcal{L}_{align} = \mathbb{E}_{u_m^+ \in Pos(u_m)} [||f(u_m) - f(u_m^+)||^2],$ $\mathcal{L}_{uniform} = \mathbb{E}_{u_t \in A(u_m)} [-||f(u_m) - f(u_t)||^2],$ where $u_t \notin Pos(u_m)$

 $u_m \sim P(u|x_m), Pos(\cdot)$ denotes positive example set. **Theorem 1** We assume f is normalized by |f| = r, and it is L-Lipschitz continuity, i.e., for any $x_1, x_2, ||f(x_1) - f(x_2)|| \le L ||x_1 - x_2||$. We let $p_k := \mathbb{P}[x \in C_k]$ for any $k \in [K]$. Let $\mu_k = \sum_{i \in D_k} \frac{f(u_i)}{|D_k|}, \ \mu_\ell = \sum_{i \in D_\ell} \frac{f(u_i)}{|D_\ell|}$ be centroids of cluster k and cluster ℓ with $k \neq \ell$. If the augmented data is (σ, δ) -Augmented, then for any $\varepsilon \ge 0$, we have

$$\mu_k^{\top} \mu_{\ell} \leq \log(\exp\{\frac{\mathcal{L}_{\text{uniform}} + \tau(\sigma, \delta, \varepsilon, R_{\varepsilon})}{p_k p_{\ell}}\} - \exp(1 - \varepsilon)),$$

where $\tau(\sigma, \delta, \varepsilon, R_{\varepsilon})$ is a non-negative term, decreasing with smaller $\varepsilon, R_{\varepsilon}$ or sharper concentration of augmented data. The specific formulation of $\tau(\sigma, \delta, \varepsilon, R_{\varepsilon})$ and the proof are deferred to the appendix.

Remark. Theorem 1 indicates that the distance between the cluster k and the cluster ℓ can be lower bounded by $-\mathcal{L}_{uniform}$. With the introduction of distinguishable openset examples, a higher lower bound is ensured, thereby enhancing the discriminative nature of different class.

Experiments

Experiment Setting

Datasets We evaluate our CECL approach using the setup from (Sun et al. 2022) on several datasets. CIFAR80N and CIFAR100N, generated from CIFAR100 (Krizhevsky 2009), offer distinct challenges. CIFAR80N contains an open-set noisy dataset with 20 unknown classes, while CI-FAR100N is a closed-set noisy dataset examining methods' versatility with different label noise scenarios (*Sym-20%*, *Sym-80%*, *Asym-40%*). Real-world datasets (Web-Aircraft, Web-Bird, Web-Car) tackle fine-grained vision categorization, presenting over 25% training examples with ambiguous noisy labels and lacking label verification. Additionally,

The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

Methods	CIFAR80N			CIFAR100N		
	Sym - 20%	Sym - 80%	Asym - 40%	Sym - 20%	Sym - 80%	Asym - 40%
Standard	29.37 ± 0.09	4.20 ± 0.07	22.25 ± 0.08	35.14 ± 0.44	4.41 ± 0.14	27.29 ± 0.25
Decoupling	43.49 ± 0.39	10.01 ± 0.29	33.74 ± 0.26	33.10 ± 0.12	3.89 ± 0.16	26.11 ± 0.39
Co-teaching	60.38 ± 0.22	16.59 ± 0.27	42.42 ± 0.30	43.73 ± 0.16	15.15 ± 0.46	28.35 ± 0.25
Co-teaching+	53.97 ± 0.26	12.29 ± 0.09	43.01 ± 0.59	49.27 ± 0.03	13.44 ± 0.37	33.62 ± 0.39
JoCoR	59.99 ± 0.13	12.85 ± 0.05	39.37 ± 0.16	53.01 ± 0.04	15.49 ± 0.98	32.70 ± 0.35
MoPro	65.60 ± 0.34	30.29 ± 0.21	60.22 ± 0.12	54.22 ± 0.26	28.32 ± 0.34	49.69 ± 0.45
NGC	74.26 ± 0.23	36.36 ± 0.48	65.73 ± 0.44	68.47 ± 0.28	$\textbf{37.17} \pm 0.41$	64.79 ± 0.38
Jo-SRC	65.83 ± 0.13	29.76 ± 0.09	53.03 ± 0.25	58.15 ± 0.14	23.80 ± 0.05	38.52 ± 0.20
PNP	67.00 ± 0.18	34.36 ± 0.18	61.23 ± 0.17	64.25 ± 0.12	31.32 ± 0.19	60.25 ± 0.21
CECL	$\textbf{77.23} \pm 0.26$	37.21 ± 0.11	68.48 ± 0.14	69.20 ± 0.09	36.37 ± 0.12	$\textbf{65.49} \pm 0.24$

Table 1: Test accuracy (%) comparison on synthetic noisy datasets CIFAR80N and CIFAR100N. The average mean and standard deviation results over the last 10 epochs are recorded. Bold values represent the best methods.

Methods	Web-Aircraft	Web-Bird	Web-Car	Food101N
Standard	60.80	64.40	60.60	84.51
Decoupling (Malach and Shalev-Shwartz 2017)	75.91	71.61	79.41	-
Co-teaching (Han et al. 2018)	79.54	76.68	84.95	-
CleanNet-hard (Lee et al. 2018)	-	-	-	83.47
CleanNet-soft (Lee et al. 2018)	-	-	-	83.95
SELFEI (Song, Kim, and Lee 2019)	79.27	77.20	82.90	-
PENCIL (Yi and Wu 2019)	78.82	75.09	81.68	-
Co-teaching+ (Yu et al. 2019)	74.80	70.12	76.77	-
Deep-Self (Han, Luo, and Wang 2019)	-	-	-	85.11
JoCoR (Wei et al. 2020)	80.11	79.19	85.10	-
AFM (Peng et al. 2020)	81.04	76.35	83.48	-
CRSSC (Sun et al. 2020)	82.51	81.31	87.68	-
Self-adaptive (Huang, Zhang, and Zhang 2020)	77.92	78.49	78.19	-
DivideMix (Li, Socher, and Hoi 2020)	82.48	74.40	84.27	-
Jo-SRC (Yao et al. 2021)	82.73	81.22	88.13	86.66
Sel-CL (Li et al. 2022a)	86.79	83.61	90.40	-
PLC (Zhang et al. 2021b)	79.24	76.22	81.87	-
NGC (Wu et al. 2021)	85.94	83.12	91.83	89.64
Peer-learning (Sun et al. 2021)	78.64	75.37	82.48	-
PNP (Sun et al. 2022)	85.54	81.93	90.11	87.50
CECL	87.46	83.87	92.61	90.24

Table 2: Test accuracy(%) comparison on four real-world noisy datasets. Bold values represent the best methods.

Food101N (Lee et al. 2018) showcases 101 food categories with more than 310k examples, posing a challenge due to its unspecified noise ratio and diverse noise types.

Comparison methods On CIFAR80N and CIFAR100N, referring (Sun et al. 2022), we compare with the following baselines: **Standard** which conducts cross entropy loss minimization, **Decoupling** (Malach and Shalev-Shwartz 2017), **Co-teaching** (Han et al. 2018), **Co-teaching+** (Yu et al. 2019), **JoCoR**, which are commonly used denoise approaches in learning with noisy labels, and recently proposed methods **MoPro**, **NGC**, **Jo-Src**, **PNP** customized for OSNLL. On the four real-word datasets, we compare with **Standard**, **NGC**, **Jo-Src**, **PNP** and **16** other closed-set noisy label learning approaches as shown in Table 2.

For baselines, their experimental results are directly adopted from (Sun et al. 2022). For our CECL method, we maintain the same backbone as PNP. For optimization, we uniformly adopt SGD with a momentum of 0.9, a learning rate decay strategy of CosineAnnealingLR, and a batch size of 256, 16, 96 for CIFAR, Web, Food101N datasets.

Classification Comparison

Table 1 and Table 2 respectively show results for the synthesized and real-world datasets. Regarding the error bars in Table 1, we have maintained the same setting as PNP, and reported the average and standard deviation of the last 10 epochs, under the condition of a random seed value 0. Bold values represent the best results among all methods.

Table 1 shows results for the 20%, 80% symmetric noise case and 40% asymmetric noise case. By varying the structure and ratio of label noise, we try to provide a more comprehensive view of our proposed method. Compared with existing OSNLL methods which minimize the impact of open-set examples, CECL wisely leverages them to learn

	CONT	OSD	sym 20%	asym 40%
CECL w/o CONT & OSD	X	×	70.54	62.41
CECL w/o CONT	×	~	73.17	64.57
CECL w/o OSD	~	×	72.87	64.26
CECL with RDOS	~	~	75.86	66.73
CECL	~	~	77.23	68.48

Table 3: Ablation study on CIFAR80N with *sym 20%* and *asym 40%*, *OSD* denotes *open-set decision*, *RDOS* denotes *removing distinguishable open-set examples*.

better representation space for the known classes, leading to significant performance gains. Surprisingly, on the closedset label noise dataset CIFAR100N, our proposed CECL still achieves better performance than the closed-set baselines. We hypothesize that this may be due to the models' tendency to treat the difficult and ambiguous examples as open-set examples, which is more reasonable than forcing them to some incorrect classes and resulting in negative effects.

Table 2 displays the experimental results on the four realworld noisy datasets. Compared with the synthetic datasets, the types of label noise in these datasets become more complicated with relatively lower noise ratios. We can observe similar superiority of our CECL method, which outperforms other methods. Such results validate our method works robustly on complex scenarios in learning with noise.

Ablation Study

In this subsection, we present our ablation results to show the effectiveness of CECL. We ablate two key components of CECL: contrastive learning (CONT) and open-set decision (OSD).

We compare CECL from four views: 1) To validate the effectiveness of Step 2: CECL w/o CONT & OSD, which uses neither CONT nor OSD, that is, our Step 1, Promix; 2) To validate the effectiveness of contrastive component: CECL w/o CONT, which removes the contrastive learning and only trains a classifier with clean examples and closed-set examples with corrected label identified by the prototype-based open-set decision; 3) To validate the effectiveness of OSD: CECL w/o OSD, specifically, we regard all the examples identified as noise in Step 1 as known classes examples and let them participate in contrastive learning, that is, the conventional supervised contrastive learning; 4) To validate the benefits brought by the distinguishable open-set examples: CECL with removing distinguishable open-set examples (RDOS), this means that we ignore the distinguishable open-set examples. As shown in Table 3, the combination of CONT and OSD can achieve the best performance.

Further Analysis

Feature Space Visualization The CECL framework consistently emphasizes improving representation learning. To evaluate its progress in representation capabilities, we compared it with PNP and visualized the results using t-SNE(Maaten and Hinton 2008), as shown in Figure 6.



Figure 6: The t-SNE on CIFAR80N with Sym 20%.

We show the t-SNE of the randomly selected 20 classes on CIFAR80N with symmetric 20% noise. The features learned by our method are more discriminative in every class compared with PNP, which verifies our method can learn better features under different noise levels and types.

Parameter Sensitivity We explore the impact of varying τ on the performance of our method. We conduct experiments on the CIFAR80N with *Sym20%* and *Asym40%*, and the results are shown in Figure 7. The results indicate that the performance significantly higher than PNP at different values of τ , demonstrating the robustness of our method.



Figure 7: Sensitivity analysis of hyper-parameter τ on CI-FAR80N with Sym 20% and Asym 40%.

Conclusion

This paper tackles the task of learning from real-world openset noisy labels. Unlike conventional methods that merely identify and minimize the impact of open-set examples, we reveal the Class Expansion phenomenon, demonstrating their ability to enhance learning for known classes. Our proposed CECL approach integrates appropriate open-set examples into known classes, treating the remaining distinguishable open-set examples as delimiters between known classes. This strategy, coupled with a modified contrastive learning scheme, boosts the model's representation learning.

Acknowledgments

This work was supported by the National Key R&D Program of China (2022ZD0114801), National Natural Science Foundation of China (61906089), Jiangsu Province Basic Research Program (BK20190408).

References

Arpit, D.; Jastrzebski, S.; Ballas, N.; Krueger, D.; Bengio, E.; Kanwal, M. S.; Maharaj, T.; Fischer, A.; Courville, A.; Bengio, Y.; and Lacoste-Julien, S. 2017. A Closer Look at Memorization in Deep Networks. In Precup, D.; and Teh, Y. W., eds., *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, 233–242. PMLR.

Duchi, J. C.; and Namkoong, H. 2021. Learning models with uniform performance via distributionally robust optimization. *The Annals of Statistics*, 49(3): 1378–1406.

Fang, Z.; Li, Y.; Lu, J.; Dong, J.; Han, B.; and Liu, F. 2022. Is Out-of-Distribution Detection Learnable? In Oh, A. H.; Agarwal, A.; Belgrave, D.; and Cho, K., eds., *Advances in Neural Information Processing Systems*.

Han, B.; Yao, Q.; Yu, X.; Niu, G.; Xu, M.; Hu, W.; Tsang, I.; and Sugiyama, M. 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31.

Han, J.; Luo, P.; and Wang, X. 2019. Deep self-learning from noisy labels. In *Proceedings of the IEEE/CVF international conference on computer vision*, 5138–5147.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9729–9738.

Huang, L.; Zhang, C.; and Zhang, H. 2020. Self-adaptive training: beyond empirical risk minimization. *Advances in neural information processing systems*, 33: 19365–19376.

Huang, Z.; Zhang, J.; and Shan, H. 2023. Twin contrastive learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11661–11670.

Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 33: 18661–18673.

Krizhevsky, A. 2009. Learning Multiple Layers of Features from Tiny Images. 32–33.

Le, Y.; and Yang, X. 2015. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7): 3.

LeCun, Y. 1998. The MNIST database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*.

Lee, K.-H.; He, X.; Zhang, L.; and Yang, L. 2018. Cleannet: Transfer learning for scalable image classifier training with label noise. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5447–5456.

Li, J.; Socher, R.; and Hoi, S. C. 2020. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv* preprint arXiv:2002.07394.

Li, J.; Xiong, C.; and Hoi, S. C. 2020. Mopro: Webly supervised learning with momentum prototypes. *arXiv preprint arXiv:2009.07995*.

Li, J.; Xiong, C.; and Hoi, S. C. 2021. Learning from noisy data with robust representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9485–9494.

Li, S.; Xia, X.; Ge, S.; and Liu, T. 2022a. Selectivesupervised contrastive learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 316–325.

Li, S.; Xia, X.; Zhang, H.; Zhan, Y.; Ge, S.; and Liu, T. 2022b. Estimating noise transition matrix with label correlations for noisy multi-label learning. *Advances in Neural Information Processing Systems*, 35: 24184–24198.

Li, S.-Y.; Huang, S.-J.; and Chen, S. 2021. Crowdsourcing aggregation with deep Bayesian learning. *Science China Information Sciences*, 64: 1–11.

Li, S.-Y.; Jiang, Y.; Chawla, N. V.; and Zhou, Z.-H. 2018. Multi-label learning from crowds. *IEEE Transactions on Knowledge and Data Engineering*, 31(7): 1369–1382.

Liu, S.; Niles-Weed, J.; Razavian, N.; and Fernandez-Granda, C. 2020. Early-learning regularization prevents memorization of noisy labels. *Advances in neural information processing systems*, 33: 20331–20342.

Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Malach, E.; and Shalev-Shwartz, S. 2017. Decoupling" when to update" from" how to update". *Advances in neural information processing systems*, 30.

Morteza, P.; and Li, Y. 2022. Provable guarantees for understanding out-of-distribution detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 8.

Ortego, D.; Arazo, E.; Albert, P.; O'Connor, N. E.; and McGuinness, K. 2021. Multi-objective interpolation training for robustness to label noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6606–6615.

Patrini, G.; Rozza, A.; Krishna Menon, A.; Nock, R.; and Qu, L. 2017. Making deep neural networks robust to label noise: A loss correction approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1944–1952.

Peng, X.; Wang, K.; Zeng, Z.; Li, Q.; Yang, J.; and Qiao, Y. 2020. Suppressing mislabeled data via grouping and self-attention. In *European Conference on Computer Vision*, 786–802. Springer.

Shi, Y.; Li, S.-Y.; and Huang, S.-J. 2023. Learning from crowds with sparse and imbalanced annotations. *Machine Learning*, 112(6): 1823–1845.

Song, H.; Kim, M.; and Lee, J.-G. 2019. Selfie: Refurbishing unclean samples for robust deep learning. In *International Conference on Machine Learning*, 5907–5915. PMLR.

Sun, Z.; Hua, X.-S.; Yao, Y.; Wei, X.-S.; Hu, G.; and Zhang, J. 2020. Crssc: salvage reusable samples from noisy data for

robust learning. In *Proceedings of the 28th ACM International Conference on Multimedia*, 92–101.

Sun, Z.; Shen, F.; Huang, D.; Wang, Q.; Shu, X.; Yao, Y.; and Tang, J. 2022. PNP: Robust Learning From Noisy Labels by Probabilistic Noise Prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5311–5320.

Sun, Z.; Yao, Y.; Wei, X.-S.; Zhang, Y.; Shen, F.; Wu, J.; Zhang, J.; and Shen, H. T. 2021. Webly supervised finegrained recognition: Benchmark datasets and an approach. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10602–10611.

Tanno, R.; Saeedi, A.; Sankaranarayanan, S.; Alexander, D. C.; and Silberman, N. 2019. Learning from noisy labels by regularized estimation of annotator confusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11244–11253.

Wang, H.; Xiao, R.; Dong, Y.; Feng, L.; and Zhao, J. 2022. ProMix: Combating Label Noise via Maximizing Clean Sample Utility. *arXiv preprint arXiv:2207.10276*.

Wang, T.; and Isola, P. 2020. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International Conference on Machine Learning*, 9929–9939. PMLR.

Wang, X.; Hua, Y.; Kodirov, E.; and Robertson, N. M. 2019a. Imae for noise-robust learning: Mean absolute error does not treat examples equally and gradient magnitude's variance matters. *arXiv preprint arXiv:1903.12141*.

Wang, Y.; Liu, W.; Ma, X.; Bailey, J.; Zha, H.; Song, L.; and Xia, S.-T. 2018. Iterative learning with open-set noisy labels. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8688–8696.

Wang, Y.; Ma, X.; Chen, Z.; Luo, Y.; Yi, J.; and Bailey, J. 2019b. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 322–330.

Wei, H.; Feng, L.; Chen, X.; and An, B. 2020. Combating noisy labels by agreement: A joint training method with coregularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13726–13735.

Wei, H.; Tao, L.; Xie, R.; and An, B. 2021. Open-set label noise can improve robustness against inherent label noise. *Advances in Neural Information Processing Systems*, 34: 7978–7992.

Wu, Z.-F.; Wei, T.; Jiang, J.; Mao, C.; Tang, M.; and Li, Y.-F. 2021. Ngc: A unified framework for learning with openworld noisy data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 62–71.

Xia, X.; Han, B.; Wang, N.; Deng, J.; Li, J.; Mao, Y.; and Liu, T. 2022. Extended *T* T: Learning With Mixed Closed-Set and Open-Set Noisy Labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3): 3047–3058.

Xia, X.; Han, B.; Zhan, Y.; Yu, J.; Gong, M.; Gong, C.; and Liu, T. 2023. Combating noisy labels with sample selection by mining high-discrepancy examples. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1833–1843. Xia, X.; Liu, T.; Wang, N.; Han, B.; Gong, C.; Niu, G.; and Sugiyama, M. 2019. Are anchor points really indispensable in label-noise learning? *Advances in neural information processing systems*, 32.

Xiao, T.; Xia, T.; Yang, Y.; Huang, C.; and Wang, X. 2015. Learning from massive noisy labeled data for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2691–2699.

Xu, Y.; Cao, P.; Kong, Y.; and Wang, Y. 2019. L_dmi: A novel information-theoretic loss function for training deep nets robust to label noise. *Advances in neural information processing systems*, 32.

Yang, Y.; Wei, H.; Sun, Z.; Li, G.; Zhou, Y.; Xiong, H.; and Yang, J. 2022. S2OSC: A Holistic Semi-Supervised Approach for Open Set Classification. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 16: 34:1–34:27.

Yang, Y.; Zhang, Y.; Song, X.; and Xu, Y. 2023. Not All Out-of-Distribution Data Are Harmful to Open-Set Active Learning. In *NeurIPS*. New Orleans, Louisiana.

Yao, Y.; Sun, Z.; Zhang, C.; Shen, F.; Wu, Q.; Zhang, J.; and Tang, Z. 2021. Jo-src: A contrastive approach for combating noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5192–5201.

Yi, K.; and Wu, J. 2019. Probabilistic end-to-end noise correction for learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7017–7025.

Yu, X.; Han, B.; Yao, J.; Niu, G.; Tsang, I.; and Sugiyama, M. 2019. How does disagreement help generalization against label corruption? In *International Conference on Machine Learning*, 7164–7173. PMLR.

Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; and Vinyals, O. 2021a. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3): 107–115.

Zhang, Y.; Zheng, S.; Wu, P.; Goswami, M.; and Chen, C. 2021b. Learning with feature-dependent label noise: A progressive approach. *arXiv preprint arXiv:2103.07756*.

Zhang, Z.; Chen, W.; Fang, C.; Li, Z.; Chen, L.; Lin, L.; and Li, G. 2023. RankMatch: Fostering Confidence and Consistency in Learning with Noisy Labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1644–1654.