Unveiling Open-set Noise: Theoretical Insights into Label Noise

Chen Feng
University College London
London, UK
Queen's University Belfast
Belfast, UK
chen.feng@ucl.ac.uk

Nicu Sebe
University of Trento
Trento, Italy
niculae.sebe@unitn.it

Georgios Tzimiropoulos
Queen Mary University of London
London, UK
g.tzimiropoulos@qmul.ac.uk

Miguel R.D. Rodrigues
University College London
London, UK
m.rodrigues@ucl.ac.uk

Ioannis Patras
Queen Mary University of London
London, UK
i.patras@qmul.ac.uk

Abstract

Learning with Noisy Labels (LNL) reduces reliance on high-quality labeled data but often overlooks *open-set noise*, where noisy samples belong to unknown classes, unlike *closed-set noise* within known categories.This paper advances LNL by reformulating the problem to incorporate open-set noise through a complete noise transition matrix, enabling a theoretical comparison of its impact on classification error rates against closed-set noise. Our analysis reveals that open-set noise induces smaller error increases, with distinct effects from 'hard' (semantically similar to inliers) and 'easy' (dissimilar) variants. We evaluate entropy-based detection, finding it effective only for easy open-set noise, and propose solutions leveraging vision-language models and self-supervised learning to address hard noise challenges. For empirical validation, we introduce *CIFAR100-O*, *ImageNet-O*, and a *WebVision* open-set test set, enabling robust benchmarking of LNL methods under open-set noise conditions. Recognizing classification accuracy's limitations in capturing model robustness, we advocate out-of-distribution (OOD) detection as a complementary metric. Our theoretical and empirical results highlight the unique challenges of open-set noise, offering new tools and evaluation frameworks to enhance LNL robustness in real-world scenarios.

CCS Concepts

• **Computing methodologies → Computer vision representations**; **Supervised learning by classification**.

Keywords

Noisy Labels; Open-set Noise; OOD Detection; Sample Selection; Robust Loss Function

ACM Reference Format:
Chen Feng, Nicu Sebe, Georgios Tzimiropoulos, Miguel R.D. Rodrigues, and Ioannis Patras. 2025. Unveiling Open-set Noise: Theoretical Insights into Label Noise. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3746027.3755040

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3755040

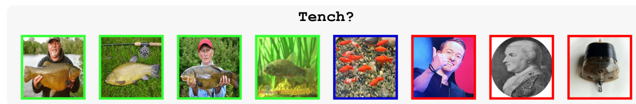


Figure 1: Example images of the class 'Tench' from the WebVision dataset. Clean samples are marked in green, closed-set noise is marked in blue, and open-set noise is marked in red. See Appendix H for further discussion.

1 Introduction

The remarkable success of machine learning in recent years has largely depended on the assumption that data labels are accurate and noise-free. Yet, in real-world scenarios, label noise-stemming from factors such as annotation errors and label ambiguity-prevails widely, presenting substantial obstacles to model performance and generalization. To tackle this issue, various approaches have been developed for learning with noisy labels (LNL), encompassing noise transition matrix estimation [22, 53, 56], noisy label correction [3, 40], robust loss functions [20, 46, 61], and, more recently, dominant sample selection-based methods [2, 11–13, 23, 30, 54].

Despite progress in LNL, most existing efforts predominantly focus on closed-set noise, where the true labels of noisy samples correspond to another known class. Common noise models include symmetric noise, where sample labels are randomly flipped to any other known class with a certain probability, and asymmetric noise, where label confusion arises from class similarity (e.g., a 'cat' is more likely to be mislabeled as a 'dog' than as an 'airplane'). Recent research has further investigated instance-dependent noise models [5, 8–10, 14–19, 41, 44, 55, 57, 63, 64], where label confusion hinges on the semantic properties of individual instances.

In contrast to the extensive research on closed-set noise, open-set noise-where the true labels of noisy samples do not align with any known category-has garnered significantly less attention. This gap is particularly pressing, given that one of the primary motivations for LNL is to handle datasets collected via web crawling, where open-set noise is commonplace. An analysis of the widely used WebVision dataset [32] reinforces the frequent presence of open-set noise, as demonstrated in Figure 1. The 'open-world' assumption, which accounts for the existence of unknown classes, has been thoroughly explored in other weakly supervised learning paradigms, such as open-set recognition and outlier detection. However, its implications for LNL remain largely uncharted. To address this gap, this paper delivers a comprehensive analysis of open-set noise. Our main contributions are outlined as follows:

- **Reformulation**: A complete noise transition matrix to incorporate open-set noise in LNL.
- **Impact Analysis**: Comparing open-set and closed-set noise, distinguishing 'hard' and 'easy' open-set noise, and proposing out-of-distribution (OOD) detection as a metric.
- **Detection**: Evaluating entropy-based detection's limitations for hard open-set noise and proposing vision-language and self-supervised solutions.
- **Benchmarks**: Introducing *CIFAR100-O*, *ImageNet-O*, and a WebVision open-set test set, with evaluations of LNL methods.

2 Related Work

This section reviews key approaches in learning with noisy labels and discusses research related to our study.

Learning with Noisy Labels. Research in learning with noisy labels (LNL) primarily involves two types of methods: statistical-consistent and statistical-approximate. Statistical-consistent methods aim to estimate the noise transition matrix [5, 22, 33, 44, 56, 57] or design robust loss functions [4, 20, 34, 36, 46, 50, 61] to create models that are theoretically risk-consistent or probabilistically sound. However, these approaches often rely on the idealized assumption that models can perfectly fit the data distribution, ignoring practical challenges such as overfitting due to excessive capacity or limited data.

By contrast, statistical-approximate methods focus on practical performance by selecting reliable samples, commonly using the 'small loss' strategy [2, 23, 28, 30, 37, 42, 47, 49, 59] to identify low-loss samples as clean, or leveraging advanced feature-based techniques [11, 35, 51, 52] like kNN and graph-based clustering. These methods are often strengthened by hybrid approaches that combine regularization and contrastive learning [7, 21, 27, 29, 31, 35, 58, 62] with semi-supervised learning and co-training for greater robustness. However, a key limitation of these approaches is their lack of theoretical guarantees, as they rely heavily on empirical heuristics rather than a formal probabilistic foundation, potentially leading to inconsistent performance across diverse noise scenarios.

Exploration of Open-set Noise. Research on open-set noise remains relatively scarce. Some studies concentrate on detecting open-set noise using methods like the Local Outlier Factor algorithm [45] or subgraph connectivity [52]. Entropy-based techniques have also been investigated [1, 39]. Rather than detecting open-set noise, Feng et al. [11] suggest mitigating its impact by refraining from relabeling. He et al. [25] propose a method for visual out-of-distribution detection in open-set noisy environments. Zhang et al. [60] employ a multi-prototype modeling approach to address the challenge of learning from noisy labels in open-set scenarios. Wan et al. [43] explore leveraging the properties of open sets to enhance noisy label learning. However, most of these works do not focus on providing a solid analysis of open-set noise or comparing it with closed-set noise.

More closely aligned with our work, Xia et al. [53] propose a noise transition matrix for open-set noise but assume all such noise belongs to a single class. We extend this framework by accounting for multiple open-set noise classes and examining their distinct effects. Additionally, Wei et al. [48] argue that open-set noise may alleviate overfitting caused by closed-set noise. In contrast, our study seeks to provide a deeper theoretical understanding of various types of open-set noise and their influence on model performance.

3 Theoretical Foundations on Open-set Noise

In this section, we perform a comparative analysis of open-set and closed-set noise by examining their effects on classification performance. We begin by reformulating the LNL problem, introducing the concept of open-set noise, and extending the label space to include both inlier and outlier classes. To facilitate a fair comparison, we use proxy samples with controlled noise ratios and offer theoretical insights into how open-set and closed-set noise affect model performance under different scenarios. Additionally, we examine the impact of 'hard' and 'easy' open-set noise, highlighting their contrasting effects.

3.1 Revisiting LNL Considering Open-set Noise

In supervised classification learning, it is typically assumed that a set of K independently and identically distributed (i.i.d.) training samples $\{x_k, y_k\}_{k=1}^K$ is drawn from a joint distribution $P(\mathbf{x}, \mathrm{y}; \mathrm{y} \in \mathcal{Y}^{in})$. The discrete label space, $\mathcal{Y}^{in} = \{1, 2, \ldots, \mathrm{A}\}$, referred to as the *inlier classes*, is known in advance. For a given loss function, the goal is to train a model $f : \mathcal{X} \to \mathcal{Y}$ that minimizes the classification error over the entire joint distribution.

In the LNL setting, it is assumed that the clean conditional distribution $P(\mathrm{y}|\mathbf{x}; \mathrm{y} \in \mathcal{Y}^{in})$ is perturbed to a noisy version $P^n(\mathrm{y}|\mathbf{x}; \mathrm{y} \in \mathcal{Y}^{in})$, resulting in noisy labels y_k^n in the training set $\{x_k, y_k^n\}_{k=1}^K$ that no longer conform to the clean distribution.

This work extends the standard LNL assumption by considering the presence of unknown *outlier classes*, denoted as $\mathcal{Y}^{out} = \{\mathrm{A+1}, \mathrm{A+2}, \ldots, \mathrm{A+B}\}$, where B is the number of outlier classes[1]. Thus, we extend the joint distribution to include both inlier and outlier classes, denoted as $P(\mathrm{y}|\mathbf{x}; \mathrm{y} \in \mathcal{Y}^{all})$ and its noisy counterpart $P^n(\mathrm{y}|\mathbf{x}; \mathrm{y} \in \mathcal{Y}^{all})$, where $\mathcal{Y}^{all} \triangleq \mathcal{Y}^{in} \cup \mathcal{Y}^{out}$.

3.1.1 Complete Noise Transition Matrix. We then define the complete noise transition matrix as follows:

Definition 3.1 (Complete noise transition matrix)**.** For a given sample x, the complete noise transition matrix T is defined as:

$$T = \left[\begin{array}{c|c} T_{in_{\mathrm{A}\times\mathrm{A}}} & \mathbf{0}_{\mathrm{A}\times\mathrm{B}} \\ \hline T_{out_{\mathrm{B}\times\mathrm{A}}} & \mathbf{0}_{\mathrm{B}\times\mathrm{B}} \end{array}\right]_{(\mathrm{A+B})\times(\mathrm{A+B})}$$

Here, $T_{ij} \triangleq P(\mathrm{y}^n = j | \mathrm{y} = i, \mathbf{x} = x; \mathrm{y}^n, \mathrm{y} \in \mathcal{Y}^{all})$. The matrix T_{in} captures noise within inlier classes, while T_{out} represents the transition from outlier classes to inlier classes[2].

[1] The following analysis is independent of the specific values of A and B, though typically B > A.

[2] Please note, unlike most statistical-consistent works, we do not assume that the noise transition matrix is class-dependent for all samples.

We explicitly define T_{out} as the *open-set noise mode*. Notably, since all collected samples are assumed to be labeled as inlier classes, the transitions within outlier classes are set to zero (marked in gray). For a specific sample $\boldsymbol{x}$, the relationship between the clean and noisy distributions can be expressed as:

$$P^n(\mathrm{y}=j|\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{all})=\sum_{i=1}^{A+B}P(\mathrm{y}=i|\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{all})\cdot T_{ij} \quad (1)$$

3.1.2 Label Noise. Label noise is commonly categorized into closed-set and open-set noise. Most existing studies define open-set noise as cases where a sample's true label belongs to an unknown outlier class but is mislabeled as an inlier class. To distinguish these noise types more clearly, we formally define label noise as follows:

Definition 3.2 (Label noise)**.** For a sample $\boldsymbol{x}$ with clean label y and noisy label y^n:

- If $y=y^n$, $(\boldsymbol{x},y,y^n)$ is a clean sample;
- If $y\neq y^n$ and $y\in\mathcal{Y}^{in}$, $(\boldsymbol{x},y,y^n)$ is closed-set noise;
- If $y\neq y^n$ and $y\in\mathcal{Y}^{out}$, $(\boldsymbol{x},y,y^n)$ is open-set noise.

Given that the clean label y is unknown (agnostic) in LNL setting, identifying the noise type requires sampling from the unknown clean distribution $P(\mathrm{y}|\mathbf{x};\mathrm{y}\in\mathcal{Y}^{all})$. To enable instance-level analysis, we define the $(O_{\boldsymbol{x}},C_{\boldsymbol{x}})$ label noise as follows:

Definition 3.3 ($(O_{\boldsymbol{x}},C_{\boldsymbol{x}})$ label noise)**.** For a sample $\boldsymbol{x}$ with clean distribution $P(\mathrm{y}|\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{all})$ and noise transition matrix T:

$$\begin{aligned}
O_{\boldsymbol{x}}&=\sum_{i=A+1}^{A+B}\sum_{j=1}^{A}T_{ij}P(\mathrm{y}=i|\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{all})\\
&=\sum_{i=A+1}^{A+B}P(\mathrm{y}=i|\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{all}),\\
C_{\boldsymbol{x}}&=\sum_{i=1}^{A}\sum_{j=1,j\neq i}^{A}T_{ij}P(\mathrm{y}=i|\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{all})\\
&=\sum_{i=1}^{A}(1-T_{ii})P(\mathrm{y}=i|\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{all}).
\end{aligned} \quad (2)$$

$O_{\boldsymbol{x}}$ denotes the expected open-set noise ratio for instance $\boldsymbol{x}$, $C_{\boldsymbol{x}}$ denotes the expected closed-set noise ratio, and their sum, $O_{\boldsymbol{x}}+C_{\boldsymbol{x}}$, represents the sample-wise noise ratio.

3.2 Comparative Analysis of Different Label Noise

To explore the impact of label noise, we adopt a simplified scenario where the conditional distribution over $\mathcal{X}$ remains constant, except for a single sample, denoted as $\boldsymbol{x}$, which is influenced by specific label noise. This simplification is motivated by the complexity of real-world datasets, where noise often intertwines with fluctuating distributions, obscuring its isolated effects. By focusing on a single sample $\boldsymbol{x}$, we can disentangle label noise from confounding factors, gaining critical insights into how it propagates and shapes the broader learning dynamics. This controlled approach establishes a foundation for understanding noise behavior, paving the way for robust strategies to mitigate its adverse effects in practical settings.

To dissect the nuances of different label noise types, we introduce two proxy samples, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, each representing a distinct noise scenario. Specifically, their clean conditional probabilities are specified as follows:

$$P(\mathrm{y}|\mathbf{x}=\boldsymbol{x}_1;\mathrm{y}\in\mathcal{Y}^{all})=[p_1^1,\ldots,p_A^1,\ldots,p_{A+B}^1],$$
$$P(\mathrm{y}|\mathbf{x}=\boldsymbol{x}_2;\mathrm{y}\in\mathcal{Y}^{all})=[p_1^2,\ldots,p_A^2,\ldots,p_{A+B}^2].$$

Their associated noise transition matrices are represented as $T^1=\{T_{ij}^1\}_{i,j=1}^{A+B}$ and $T^2=\{T_{ij}^2\}_{i,j=1}^{A+B}$.

3.2.1 Assumptions for Fair Comparison. A fair comparison is crucial to isolate the distinct characteristics of each noise type, unhindered by variables such as noise magnitude or sampling disparities. For this purpose, we establish two assumptions for $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to ensure fairness and precision, facilitating a thorough analysis of how noise type shapes learning outcomes.

Assumption 3.4 (Equal Noise Ratio)**.** To fairly assess open-set and closed-set noise, we must ensure that model performance differences arise from noise type rather than quantity. Thus, we assume both proxy samples share the same overall noise ratio:

$$O_{\boldsymbol{x}_1}+C_{\boldsymbol{x}_1}=O_{\boldsymbol{x}_2}+C_{\boldsymbol{x}_2}, \quad (3)$$

where $O_{\boldsymbol{x}}$ and $C_{\boldsymbol{x}}$ denote the open-set and closed-set noise ratios, respectively, as defined in Definition 3.3.

Assumption 3.5 (Equal Sampling Probability)**.** In real-world settings, unequal sampling frequencies can skew the learning process by amplifying the impact of overrepresented samples. To address this, we assume equal sampling probability, ensuring that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are treated equivalently by the model-that is, both samples have the same prior probability during dataset sampling:

$$P(\mathbf{x}=\boldsymbol{x}_1;\mathrm{y}\in\mathcal{Y}^{all})=P(\mathbf{x}=\boldsymbol{x}_2;\mathrm{y}\in\mathcal{Y}^{all}). \quad (4)$$

3.2.2 Quantifying Error Due to Noisy Labels. In our reformulated LNL framework, open-set noise from outlier classes $\mathcal{Y}^{out}$ is present in the training set, yet remains undetected and unaccounted for during training and testing. As a standard practice in machine learning, the classifier f is trained and evaluated exclusively on the inlier classes $\mathcal{Y}^{in}$, with performance metrics focusing solely on their error rates.

Let $P^f(\mathrm{y}|\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{in})$ denote the learned *inlier conditional probability* of model f. The predicted label for a sample $\boldsymbol{x}$ is then determined as:

$$y^f=\arg\max_k P^f(\mathrm{y}=k|\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{in})\in\mathcal{Y}^{in}.$$

The expected classification error rate for sample $\boldsymbol{x}$ is thus expressed as:

$$\begin{aligned}
E_{\boldsymbol{x}}&=\sum_{y\neq y^f}P(\mathrm{y}=y,\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{in})\\
&=(1-P(\mathrm{y}=y^f|\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{in}))\cdot P(\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{in}),
\end{aligned}$$

while the Bayes error rate for the optimal model f^* is given by:

$$E_{\boldsymbol{x}}^*=(1-\max_k P(\mathrm{y}=k|\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{in}))\cdot P(\mathbf{x}=\boldsymbol{x};\mathrm{y}\in\mathcal{Y}^{in}).$$

To measure the effect of noisy labels, we introduce the concept of *error rate inflation*, defined as the additional error induced by noise:

Definition 3.6 (Error Rate Inflation). Given the Bayes error rate $E^*_{\boldsymbol{x}}$, the error rate inflation for a sample $\boldsymbol{x}$ is defined as:

$$\Delta E_{\boldsymbol{x}} = E_{\boldsymbol{x}} - E^*_{\boldsymbol{x}}. \tag{5}$$

3.2.3 Fitted and Overfitted Cases Under Noise. The learned inlier conditional probability $P^f(\mathrm{y}|\mathbf{x} = \boldsymbol{x}; \mathrm{y} \in \mathcal{Y}^{in})$ depends on factors such as model capacity, dataset size, and training parameters, making its precise estimation at any given training stage difficult. This variability complicates direct analysis of noise effects across the training spectrum.

To tackle this, we focus on two practical cases-fitted (to the noisy distribution) and overfitted (memorize noisy labels)-representing the extreme ends of model behavior under noisy conditions. We choose these two scenarios because they effectively bound the range of possible outcomes of a classification model in real-world training. By limiting our analysis to these extremes, we simplify the problem while retaining key insights into how noise influences learning, laying a foundation for broader understanding and mitigation strategies.

- **Fitted Case**: The model aligns with the noisy distribution:

$$P^f(\mathrm{y}|\mathbf{x} = \boldsymbol{x}; \mathrm{y} \in \mathcal{Y}^{in}) = P^n(\mathrm{y}|\mathbf{x} = \boldsymbol{x}; \mathrm{y} \in \mathcal{Y}^{in}). \tag{6}$$

  Here, the model captures the noisy data distribution without overfitting to individual labels. This typically occurs when fine-tuning a linear classifier on a frozen pretrained model (e.g., ResNet pretrained on ImageNet). The pretrained backbone offers robust feature representations, while the linear classifier's limited capacity adapts to the noisy data without memorization.
- **Overfitted Case**: The model memorizes the noisy labels:

$$P^f(\mathrm{y}|\mathbf{x} = \boldsymbol{x}; \mathrm{y} \in \mathcal{Y}^{in}) = P^{y^n}(\mathrm{y}|\mathbf{x} = \boldsymbol{x}; \mathrm{y} \in \mathcal{Y}^{in}). \tag{7}$$

  In this case, P^{y^n} denotes the one-hot encoded noisy label y^n, with the model assigning full probability to the observed label for each sample. This arises when training a high-capacity deep network (e.g., a CNN or transformer) from scratch on a small, noisy dataset, where ample parameters—like those in ResNet-50—enable memorization rather than generalization.

3.3 Comparison of Open-set and Closed-set Noise

We begin by examining the differences between open-set and closed-set noise, focusing on how increasing the proportion of open-set noise affects model performance while keeping the total noise ratio fixed. To facilitate this analysis without loss of generality, we assume:

$$O_{\boldsymbol{x}_1} > O_{\boldsymbol{x}_2}, \quad C_{\boldsymbol{x}_1} < C_{\boldsymbol{x}_2}, \tag{8}$$

where sample $\boldsymbol{x}_1$ exhibits a higher susceptibility to open-set noise than $\boldsymbol{x}_2$, representing a scenario dominated by open-set noise.

However, these constraints alone-combined with Equation (3)-permit infinitely many noise transition matrices T^1 and T^2 that satisfy both Equation (3) and Equation (8). This non-uniqueness arises because, for given clean conditional probabilities $P(\mathrm{y}|\mathbf{x} = \boldsymbol{x}_1; \mathrm{y} \in \mathcal{Y}^{all})$ and $P(\mathrm{y}|\mathbf{x} = \boldsymbol{x}_2; \mathrm{y} \in \mathcal{Y}^{all})$, multiple T^1 and T^2 configurations can produce the same noise ratios (see the toy example below). As a result, directly comparing the error rate inflation $\Delta E_{\boldsymbol{x}_1}$ versus $\Delta E_{\boldsymbol{x}_2}$ becomes challenging, as detailed in Appendix C.

To address this ambiguity, we introduce a practical assumption commonly observed in classification tasks, which enables us to derive a definitive comparison between open-set and closed-set noise:

Assumption 3.7 (Dominant Class Probability). In typical classification settings, for samples $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, let $a = \arg\max_i P(\mathrm{y} = i|\mathbf{x} = \boldsymbol{x}_1; \mathrm{y} \in \mathcal{Y}^{all})$ and $b = \arg\max_i P(\mathrm{y} = i|\mathbf{x} = \boldsymbol{x}_2; \mathrm{y} \in \mathcal{Y}^{all})$. We assume that, with high probability, $p^1_a \to 1$ and $p^2_b \to 1$, meaning the clean conditional probability concentrates predominantly on a single class for each sample.

Toy Example: Ambiguity in T
Consider a ternary classification task with two inlier classes ("0" and "1") and one outlier class ("2"). Suppose sample $\boldsymbol{x}_1$ has a clean conditional probability distribution of $[0.1, 0.2, 0.7]$. Below are two possible noise transition matrices for T^1:

$$[0.55, 0.45, 0.0] = [0.1, 0.2, 0.7]\begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.75 & 0.25 & 0 \\ \hline 0.5 & 0.5 & 0 \end{bmatrix},$$

$$[0.45, 0.55, 0.0] = [0.1, 0.2, 0.7]\begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0.5 & 0 \\ \hline 0.5 & 0.5 & 0 \end{bmatrix}.$$

In both instances, $O_{\boldsymbol{x}_1} = 0.7$ and $C_{\boldsymbol{x}_1} = 0.2$, yet the resulting noisy distributions differ ($[0.55, 0.45, 0.0]$ vs. $[0.45, 0.55, 0.0]$), illustrating the ambiguity inherent in the noise transition process.

This assumption has been widely adopted in prior theoretical studies [6]. Building on this, we establish the following result:

Theorem 3.8 (Comparison of Open-set and Closed-set Noise). *Consider samples $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ satisfying Equation* (3) *and Equation* (8)*, where $\boldsymbol{x}_1$ is more prone to open-set noise than $\boldsymbol{x}_2$. Under Assumption 3.7, it holds that, in both the **fitted case** and the **overfitted case**:*

$$\Delta E_{\boldsymbol{x}_1} < \Delta E_{\boldsymbol{x}_2}, \tag{9}$$

where $\Delta E_{\boldsymbol{x}}$ denotes the increase in classification error due to noise.

A detailed proof is provided in Appendix D.1. This theorem demonstrates that open-set noise consistently induces a smaller error increase than closed-set noise across training scenarios. **In essence, under typical classification conditions, open-set noise proves less detrimental to model classification performance than closed-set noise.**

Remark 3.9 (Enhanced Evaluation via OOD Detection). Our analysis in Theorem 3.8 suggests that the milder effect of open-set noise on classification accuracy may limit the utility of accuracy alone in fully assessing a model's robustness under noisy conditions. To address this and enrich our evaluation framework, we integrate out-of-distribution (OOD) detection as a complementary metric tailored to our study of noise impacts. Unlike traditional accuracy-focused evaluations, OOD detection targets the model's ability to discern samples deviating from the training distribution-particularly those affected by open-set noise-offering critical insights into its behavior beyond inlier classification performance. For this purpose, we apply the maximum softmax probability method, originally proposed by Hendrycks and Gimpel [26], within the context of model evaluation in learning with noisy labels. For complete details on the OOD detection task, please refer to Appendix B.

3.4 Comparison of Different Open-set Noise Modes

We further extend our analysis to examine how distinct open-set noise modes influence model performance. Specifically, we investigate the effects of varying open-set noise transition matrices (T_{out}) while maintaining a constant open-set noise ratio:

$$O_{\boldsymbol{x}_1} = O_{\boldsymbol{x}_2}. \tag{10}$$

To enhance clarity and isolate the impact of open-set noise, we exclude closed-set noise in this section by setting:

$$C_{\boldsymbol{x}_1} = C_{\boldsymbol{x}_2} = 0. \tag{11}$$

This simplification allows us to focus solely on open-set noise variations. For a complementary analysis including consistent closed-set noise, refer to Appendix D.3. We further assume that the clean conditional probabilities for samples $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are identical:

$$P(\mathrm{y}|\mathbf{x}=\boldsymbol{x}_1;\mathrm{y}\in\mathcal{Y}^{all}) = P(\mathrm{y}|\mathbf{x}=\boldsymbol{x}_2;\mathrm{y}\in\mathcal{Y}^{all}) = [p_1,\ldots,p_A,\ldots,p_{A+B}]. \tag{12}$$

This ensures that any observed differences stem solely from the open-set noise modes (T_{out}).

Inspired by symmetric and asymmetric noise in closed-set settings, we define two distinct open-set noise modes:

- **'Easy' Open-set Noise**: Outlier classes are uniformly mapped to all inlier classes, distributing errors evenly across $\mathcal{Y}^{in}$.
- **'Hard' Open-set Noise**: Each outlier class is mapped to its most semantically similar inlier class, concentrating errors on specific inliers.

We assign $\boldsymbol{x}_1$ to the 'easy' mode with transition matrix T^{easy} and $\boldsymbol{x}_2$ to the 'hard' mode with T^{hard}, defined as:

$$T^1_{out} = T^{easy} = \begin{bmatrix} \frac{1}{A} & \cdots & \frac{1}{A} \\ \vdots & \ddots & \vdots \\ \frac{1}{A} & \cdots & \frac{1}{A} \end{bmatrix}_{\mathrm{B}\times\mathrm{A}}, \tag{13}$$

$$T^2_{out} = T^{hard} = \begin{bmatrix} 0 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 0 \end{bmatrix}_{\mathrm{B}\times\mathrm{A}}, \tag{14}$$

and

$$T^1_{in} = T^2_{in} = \mathbf{I}. \tag{15}$$

In T^{easy}, each entry $T_{ij} = \frac{1}{A}$ ensures uniform error distribution from outlier to inlier classes. In T^{hard}, we denote $H_i = \{j \mid T^{hard}_{ji} = 1\}^A_{i=1}$ as the set of outlier classes $j \in \mathcal{Y}^{out}$ mapped to inlier class $i \in \mathcal{Y}^{in}$, capturing targeted semantic similarity. These definitions enable the following analysis:

Theorem 3.10 (Comparison of 'Hard' and 'Easy' Open-set Noise). *Consider samples $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ satisfying Equation (10) and Equation (11), with noise transition matrices $T^1_{out} = T^{easy}$ and $T^2_{out} = T^{hard}$. Given identical clean conditional probability distributions, the following holds:*

$$\begin{aligned} &\textbf{\textit{Fitted Case:}} && \Delta E_{\boldsymbol{x}_1} \le \Delta E_{\boldsymbol{x}_2}, \\ &\textbf{\textit{Overfitted Case:}} && \Delta E_{\boldsymbol{x}_1} \ge \Delta E_{\boldsymbol{x}_2}. \end{aligned} \tag{16}$$

A detailed proof is available in Appendix D.2. This theorem highlights distinct behaviors of open-set noise under different training regimes. **In essence, 'easy' open-set noise, with its uniform dispersion, proves less detrimental in the fitted case, while 'hard' open-set noise, due to its concentrated mapping, exhibits a milder impact in the overfitted case.**

4 Rethinking Open-set Noise Detection

Building on our prior analysis of how open-set noise affects model generalization, we now shift our focus to detecting open-set noise within training data. As introduced, most LNL methods are tailored to mitigate closed-set noise, leaving open-set noise largely underexplored. Although a handful of approaches extend their scope to open-set noise, they typically target simpler 'easy' scenarios only. In this section, we critically assess the efficacy of a prevalent entropy-based detection method across diverse open-set noise types and propose innovative strategies to enhance detection for 'hard' open-set noise.

4.1 Entropy-Based Open-set Noise Detection

We begin by examining the widely adopted *entropy-based detection* approach for identifying open-set noise. Previous studies [1, 39] have adapted closed-set noise detection techniques to open-set scenarios, leveraging the insight that samples with low-confidence, uniformly distributed predictions often indicate open-set instances. Such samples typically exhibit elevated entropy in the model's output probability distribution.

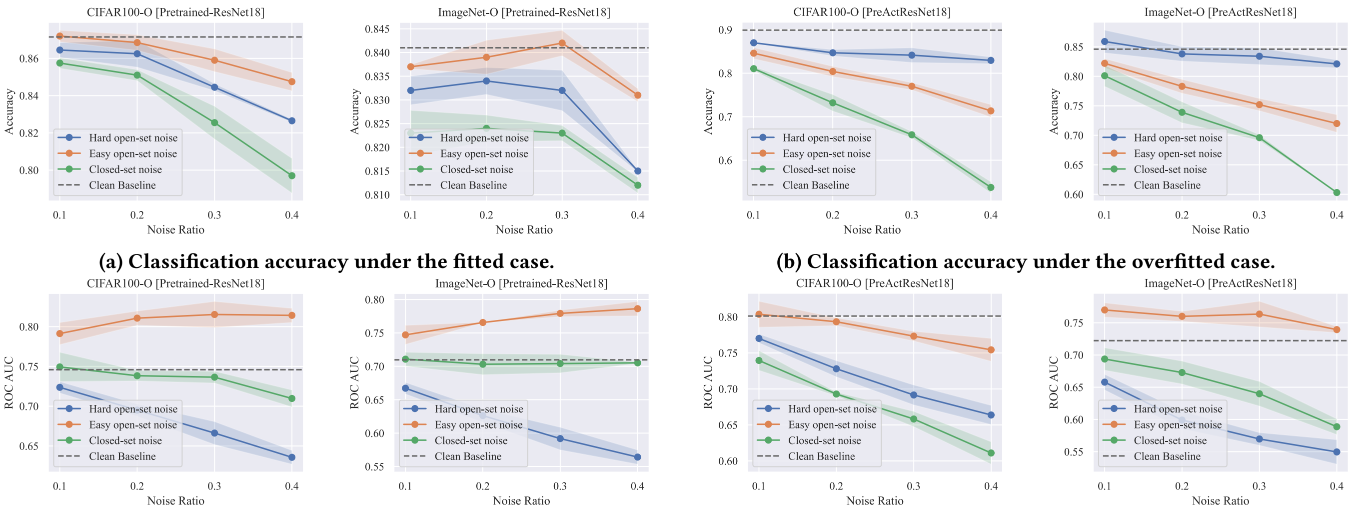
Entropy-based methods are commonly deployed after a warm-up phase, an early training stage where the model is expcted to capture meaningful patterns before severely overfitting to noisy labels. In this section, we build on this observation to formalize the entropy-based detection process and assess its effectiveness in distinguishing clean samples from open-set noise instances. Notably, we draw upon results from the *fitted case* analyzed in Section 3.3, as the model's state more closely resembles the fitted condition after warm-up phase. The following theorem provides a mathematical formulation of entropy values for clean samples and two distinct open-set noise types.

Theorem 4.1 (Entropy-Based Open-set Noise Detection)**.** *For a given sample $\boldsymbol{x}$, the entropy values across three cases-clean samples, 'easy' open-set noise, and 'hard' open-set noise-are defined as:*

$$\begin{aligned} \mathcal{H}_{clean} &= \textsc{Ent}\left(\left[p_1 + \frac{p_1}{\sum_{i=1}^A p_i}\sum_{i=A+1}^{A+B} p_i, \ldots, p_A + \frac{p_A}{\sum_{i=1}^A p_i}\sum_{i=A+1}^{A+B} p_i\right]\right), \\ \mathcal{H}_{easy} &= \textsc{Ent}\left(\left[p_1 + \frac{1}{A}\sum_{i=A+1}^{A+B} p_i, \ldots, p_A + \frac{1}{A}\sum_{i=A+1}^{A+B} p_i\right]\right), \\ \mathcal{H}_{hard} &= \textsc{Ent}\left(\left[p_1 + \sum_{j\in H_1} p_j, \ldots, p_A + \sum_{j\in H_A} p_j\right]\right), \end{aligned} \tag{17}$$

where $H_i = \{j|T^{hard}_{ji} = 1\}^A_{i=1}$ denotes the set of outlier classes $j \in \mathcal{Y}^{out}$ mapped to inlier class $i \in \mathcal{Y}^{in}$, reflecting the targeted semantic confusion of 'hard' noise. It follows that:

$$\mathcal{H}_{easy} \ge \mathcal{H}_{clean}, \tag{18}$$



(a) Classification accuracy under the fitted case.

(b) Classification accuracy under the overfitted case.

(c) OOD detection under the fitted case.

(d) OOD detection under the overfitted case.

Figure 2: Performance of supervised training across noise modes and ratios. The PreActResNet18 model represents the overfitted case, while the Pretrained-ResNet18 model represents the fitted case.

whereas the relationship between $\mathcal{H}_{hard}$ *and* $\mathcal{H}_{clean}$ *remains indeterminate without specific probability values.*

Please refer to Appendix E for the detailed proof. **Notably, entropy-based methods excel at identifying 'easy' open-set noise due to its consistently higher entropy, yet they falter in detecting 'hard' open-set noise, underscoring the need for methods that leverage richer contextual cues beyond mere distributional uniformity.**

4.2 Towards New Detection Methods

To overcome the possible limitations of existing open-set detection methods, we explore advanced techniques aimed at enhancing open-set noise detection, particularly for tackling the subtle yet critical challenges posed by 'hard' noise in real-world scenarios.

4.2.1 Utilization of Pretrained Models for Entropy-Based Detection. We first examine whether pretrained models can strengthen entropy-based open-set noise detection. Unlike randomly initialized feature spaces, which lack discriminative structure, pretrained encoders leverage prior knowledge to yield robust, structured representations. These enriched features may better disentangle challenging open-set samples, including those obscured by semantic overlap with inliers. To this end, we investigate two pretraining strategies tailored to refine feature extraction for improved noise detection.

Self-Supervised Pretraining. We first consider utilizing the MoCo framework [24] to pretrain the model's encoder. This self-supervised pretraining yields refined representations that may enhance the model's capacity to differentiate open-set noise-particularly 'hard' noise-from inlier samples, offering a potential boost to entropy-based detection in complex settings.

Vision-Language Contrastive Pretraining. We further explore vision encoders pretrained with the CLIP model [38], a multi-modal framework that jointly optimizes visual and textual representations. CLIP consists of: a visual encoder v to extract image features and a textual encoder t to encode textual descriptions. In our experiments, we substitute the randomly initialized encoder with the pretrained visual encoder v, leveraging its enriched representations to refine entropy-based detection.

4.2.2 Zero-Shot Detection with CLIP. Leveraging CLIP's multi-modal capabilities, we explore its efficacy in zero-shot open-set noise detection by aligning visual features with textual descriptions in a shared embedding space. We propose an algorithm that computes an open-set indicator score, $D_{\boldsymbol{x}}$, to identify noisy samples.

Algorithm 1 Zero-Shot Open-set Noise Detection

Require: Sample $\boldsymbol{x}$ with label y, candidate class set C_y, CLIP model (visual encoder v, text encoder t), similarity function $\mathcal{S}$

Ensure: Open-set indicator score $D_{\boldsymbol{x}}$

1: Generate target prompt: $t_y = t(\texttt{"A photo of class } y\texttt{"})$

2: Generate non-target prompts for all $i \in C_y$:

$$t_i = t(\texttt{"A photo of class } i\texttt{"})$$

3: Extract visual features: $v_{\boldsymbol{x}} = v(\boldsymbol{x})$

4: Compute target similarity: $S_y = \mathcal{S}(v_{\boldsymbol{x}}, t_y)$

5: Compute maximum non-target similarity:

$$S_{\text{other}} = \max\{\mathcal{S}(v_{\boldsymbol{x}}, t_i) \mid i \in C_y\}$$

6: **Output:** $D_{\boldsymbol{x}} = S_y - S_{\text{other}}$

The candidate class set C_y may be defined as:

- A curated set of suspected true classes for $\boldsymbol{x}$, based on prior knowledge or model predictions.
- A comprehensive set of categories from a large-scale dataset (e.g., ImageNet-1k), offering a broad reference for comparison.

The indicator $D_{\boldsymbol{x}}$ measures the alignment of $\boldsymbol{x}$'s visual features with its assigned label y relative to alternative classes in C_y, with lower scores suggesting potential open-set noise.

5 Experiments

We validate our theoretical findings through empirical evaluation in this section. Section 5.1 compares various label noise types based on our theoretical analysis, while Section 5.2 investigates entropy dynamics across different open-set noise conditions. In Section 5.3 we checked the performance of existing LNL methods on newly-proposed open-set noise modes and benchmarks.

Dataset Details. For controlled and unbiased experiments, we introduce two synthetic open-set noisy datasets-*CIFAR100-O* and *ImageNet-O*-derived from CIFAR100 and ImageNet, respectively. Unlike prior studies that construct noisy datasets by using separate datasets as mutual open-set noise sources [39, 52], which may introduce domain gaps and complicate analysis, we select inlier and outlier classes from the same dataset to ensure consistency.

Previous work often modeled open-set noise as random flipping of outlier samples across all inlier classes, aligning with our 'easy' open-set noise setting. However, our analysis reveals distinct behaviors between 'easy' and 'hard' open-set noise, suggesting that focusing solely on 'easy' noise is inadequate for a thorough evaluation. To address this, we also propose an open-set test set for WebVision. Complete dataset construction and implementation details are available in Appendix A.

5.1 Empirical Impact Analysis of Open-set Noise

In this section, we conduct experiments to assess the impact of open-set noise. Since deep models typically possess high capacity, we perform supervised learning from scratch on the noisy dataset, treating the final model as the *overfitted case*. Conversely, achieving a perfect fit to the data distribution is challenging; thus, we train a single-layer linear classifier on a frozen pretrained encoder. Due to its limited capacity, this setup approximates the *fitted case*.

We evaluate classification accuracy-computed as 1− classification error rate-on the CIFAR100-O and ImageNet-O datasets under different noise ratios, as illustrated in Figure 2(a/b). Our key observations are as follows: (1) Open-set noise has a significantly smaller impact on classification accuracy compared to closed-set noise in both the fitted and overfitted cases. (2) The behavior of 'hard' and 'easy' open-set noise varies across the two cases. These findings are consistent with our theoretical analysis.

Furthermore, we analyze OOD detection performance in Figure 2(c/d). The results indicate that 'hard' open-set noise degrades OOD detection performance in both cases, whereas 'easy' open-set noise can sometimes enhance it. For instance, in the fitted case, OOD detection performance improves progressively with increasing open-set noise in both datasets. We leave further investigation to future work.

However, these contrasting trends suggest that, beyond standard closed-set classification, incorporating alternative evaluation frameworks such as OOD detection can provide a more comprehensive assessment of LNL methods.

5.2 Evaluating Open-set Noise Detection

In Section 4, we analyze the entropy-based open-set noise detection mechanism and observe that it is primarily effective in identifying 'easy' open-set noise. In this section, we empirically validate this observation across varying open-set noise ratios. Specifically, we present the 0-1 normalized entropy values in Figure 3 following a warm-up training stage, during which the model is trained on the entire dataset for a fixed number of epochs (20 in our experiments).

Our results confirm that entropy values serve as a more reliable indicator for detecting 'easy' open-set noise compared to 'hard' open-set noise, as demonstrated by the comparisons between (a) and (e), (b) and (f), and (c) and (g) in Figure 3.

However, an examination of (e), (f), and (g) in Figure 3 reveals that neither self-supervised pretraining nor CLIP pretraining significantly improves the detection of 'hard' open-set noise. Entropy-based detection remains effective primarily for 'easy' open-set noise, even when pretrained encoders are employed, but exhibits limited sensitivity to 'hard' cases. This finding underscores the challenge of distinguishing open-set noise samples that exhibit strong semantic similarity to inlier classes.

Notably, compared to entropy-based detection methods, the CLIP zero-shot open-set detection approach consistently outperforms. It provides a more robust solution for detecting both 'easy' and 'hard' open-set noise scenarios, highlighting its potential in overcoming the limitations of entropy-based approaches.

5.3 Benchmarking Existing LNL Methods

We also assess the performance of existing Learning with Noisy Labels (LNL) methods on our newly proposed benchmarks.

5.3.1 Robust Loss Functions. To this end, we first evaluate robust loss functions across various open-set noise scenarios in this section. Specifically, we examine two widely adopted robust loss functions-Symmetric Cross Entropy (SCE) [46] and Generalized Cross Entropy (GCE) [61]-and report their classification accuracy and OOD detection AUC scores on the CIFAR100-O dataset:

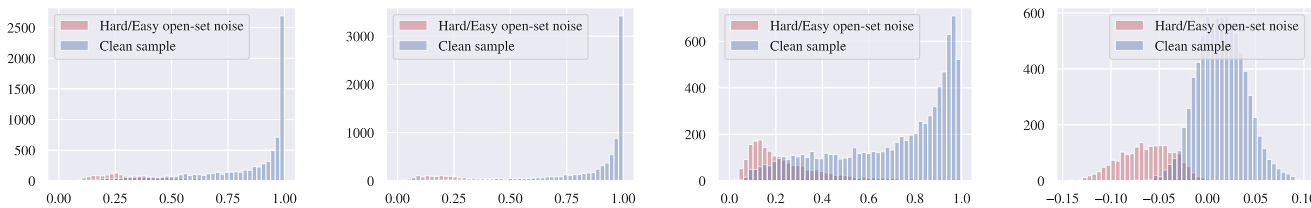
Table 1: Classification accuracy with robust loss functions on CIFAR100-O dataset.

Noise Mode	Easy				Hard			
Noise Ratio	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
CE	0.846	0.804	0.770	0.714	**0.872**	0.847	**0.842**	**0.829**
GCE	**0.854**	0.810	0.763	0.708	0.864	0.840	0.813	0.800
SCE	0.846	**0.822**	**0.787**	**0.729**	0.871	**0.854**	0.840	0.814

Table 2: OOD detection AUC with robust loss functions on CIFAR100-O dataset.

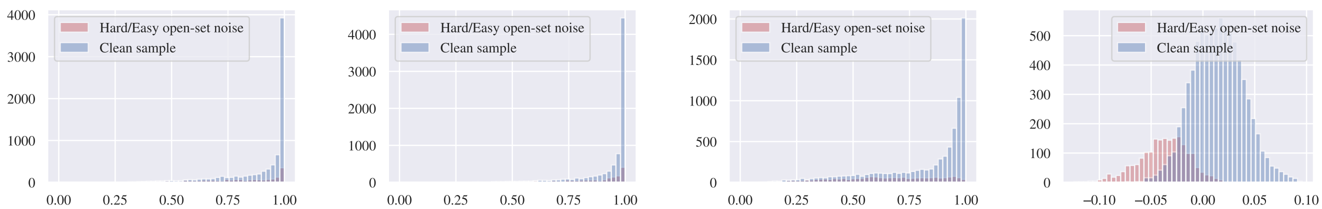
Noise Mode	Easy				Hard			
Noise Ratio	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
CE	**0.804**	0.793	0.773	0.754	**0.770**	**0.728**	**0.692**	**0.664**
GCE	0.782	0.771	0.752	0.719	0.759	0.718	0.679	0.639
SCE	0.794	**0.799**	**0.784**	**0.756**	0.749	0.718	0.682	0.651

ImageNet-O with 40% 'Easy' Open-set Noise



(a) Entropy-based detection (b) Self-supervised pretraining (c) CLIP pretraining (d) CLIP zero-shot

ImageNet-O with 40% 'Hard' Open-set Noise



(e) Entropy-based detection (f) Self-supervised pretraining (g) CLIP pretraining (h) CLIP zero-shot

Figure 3: Performance of open-set noise detection. The first three columns depict 0-1 normalized entropy values, while the rightmost column represets the open-set indicator score D_x.

We observe that the performance differences between the two robust loss functions under consideration and the standard cross-entropy (CE) loss are not substantial, which is understandable given that existing robust loss function frameworks do not account for the presence of open-set noise. For a more detailed analysis of these results, please refer to Appendix F.

5.3.2 Sample Selection Methods. Additionally, we evaluate the performance of another prominent technique in learning with noisy labels (LNL)—sample selection methods—including their effectiveness on the real-world WebVision dataset. We benchmark these methods on the CIFAR100-O dataset under varying open-set noise ratios and modes (see Table 3). For more results and experiment details please refer to Appendix G. Our analysis includes four representative methods: SSR [11] and DivideMix [30], exemplifying loss-based and feature-based sample selection approaches, respectively, alongside EvidentialMix [39] and DSOS [1], which further incorporate mechanisms tailored to mitigate open-set noise.

Table 3: Benchmarking results on the CIFAR100-O dataset under different noise modes and ratios.

Noise Ratio & Noise Mode	0.2 Easy	0.4 Easy	0.2 Hard	0.4 Hard
SSR [11]	**0.889**	**0.875**	0.895	**0.871**
DivideMix [30]	0.783	0.754	0.738	0.675
EvidentialMix [39]	0.884	0.827	**0.898**	0.872
DSOS [1]	0.846	0.765	0.854	0.832

The results suggest that different methods exhibit varying levels of sensitivity to open-set noise. Notably, the specialized open-set noise handling techniques employed in EvidentialMix and DSOS do not consistently outperform standard methods such as SSR across all noise conditions. This may be partially attributed to the influence of hyperparameter settings, which require further fine-tuning for optimal performance.

Despite these observations, evaluating the performance of these methods remains a complex task, as they often involve multiple interdependent components and regularization strategies. A thorough investigation of these factors is beyond the scope of this paper and warrants future exploration.

6 Conclusions

This paper explores the impact of open-set label noise on model performance, an overlooked aspect of Learning with Noisy Labels (LNL). Although the 'open world' concept is well-established in other weakly supervised learning fields, its relevance to LNL has received little attention. To address this, we redefine the LNL problem by introducing a noise transition matrix that explicitly accounts for open-set noise.

Our theoretical and empirical results show that open-set noise disrupts classification performance less than closed-set noise. However, the differing effects of 'easy' and 'hard' open-set noise underscore the need for more refined evaluation methods. To better assess models under open-set noise, we propose incorporating out-of-distribution (OOD) detection as a metric and introduce synthetic open-set noise datasets, to enable thorough robustness testing.

Additionally, our analysis reveals that entropy-based detection effectively identifies 'easy' open-set noise but struggles with more challenging cases. To address this limitation, we investigate pre-trained vision-language models and self-supervised learning techniques to enhance open-set noise detection. Overall, This work sheds light on open-set noise, offers practical solutions, and opens doors for future study.

Acknowledgements. We acknowledge support from Leverhulme Trust via research grant RPG-2022-198.

References

[1] Paul Albert, Diego Ortego, Eric Arazo, Noel E O'Connor, and Kevin McGuinness. 2022. Addressing out-of-distribution label noise in webly-labelled data. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 392–401.

[2] Eric Arazo, Diego Ortego, Paul Albert, Noel O'Connor, and Kevin McGuinness. 2019. Unsupervised label noise modeling and loss correction. In *International Conference on Machine Learning*. PMLR, 312–321.

[3] Paola Cascante-Bonilla, Fuwen Tan, Yanjun Qi, and Vicente Ordonez. 2021. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 6912–6920.

[4] Mingcai Chen, Hao Cheng, Yuntao Du, Ming Xu, Wenyu Jiang, and Chongjun Wang. 2023. Two wrongs don't make a right: Combating confirmation bias in learning with label noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 14765–14773.

[5] Pengfei Chen, Junjie Ye, Guangyong Chen, Jingwei Zhao, and Pheng-Ann Heng. 2021. Beyond class-conditional assumption: A primary attempt to combat instance-dependent label noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 11442–11450.

[6] Pengfei Chen, Junjie Ye, Guangyong Chen, Jingwei Zhao, and Pheng-Ann Heng. 2021. Robustness of accuracy metric and its inspirations in learning with noisy labels. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 11451–11461.

[7] Filipe R Cordeiro, Ragav Sachdeva, Vasileios Belagiannis, Ian Reid, and Gustavo Carneiro. 2021. Longremix: Robust learning with high confidence samples in a noisy label environment. *arXiv preprint arXiv:2103.04173* (2021).

[8] Chen Feng, Ziquan Liu, Zhuo Zhi, Ilija Bogunovic, Carsten Gerner-Beuerle, and Miguel Rodrigues. 2025. PROSAC: Provably Safe Certification for Machine Learning Models under Adversarial Attacks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 2933–2941.

[9] Chen Feng and Ioannis Patras. 2022. Adaptive soft contrastive learning. In *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 2721–2727.

[10] Chen Feng and Ioannis Patras. 2023. Maskcon: Masked contrastive learning for coarse-labelled dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 19913–19922.

[11] Chen Feng, Georgios Tzimiropoulos, and Ioannis Patras. 2022. SSR: An Efficient and Robust Framework for Learning with Unknown Label Noise. In *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*. BMVA Press.

[12] Chen Feng, Georgios Tzimiropoulos, and Ioannis Patras. 2024. CLIPCleaner: Cleaning noisy labels with CLIP. In *Proceedings of the 32nd ACM International Conference on Multimedia*. 876–885.

[13] Chen Feng, Georgios Tzimiropoulos, and Ioannis Patras. 2024. NoiseBox: towards more efficient and effective learning with noisy labels. *IEEE Transactions on Circuits and Systems for Video Technology* (2024).

[14] Zheng Gao, Chen Feng, and Ioannis Patras. 2024. Self-supervised representation learning with cross-context learning between global and hypercolumn features. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1773–1783.

[15] Arpit Garg, Cuong Nguyen, Rafael Felix, Thanh-Toan Do, and Gustavo Carneiro. 2023. Instance-Dependent Noisy Label Learning via Graphical Modelling. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2288–2298.

[16] Jiawei Ge, Jiuxin Cao, Yingxing Bao, Biwei Cao, and Bo Liu. 2024. GAL: combining global and local contexts for interpersonal relation extraction toward document-level Chinese text. *Neural Computing and Applications* 36, 11 (2024), 5715–5731.

[17] Jiawei Ge, Jiuxin Cao, Xiangmei Chen, Xuelin Zhu, Weijia Liu, Chang Liu, Kun Wang, and Bo Liu. 2025. Beyond visual cues: Synchronously exploring target-centric semantics for vision-language tracking. *ACM Transactions on Multimedia Computing, Communications and Applications* 21, 5 (2025), 1–21.

[18] Jiawei Ge, Jiuxin Cao, Xuelin Zhu, Xinyu Zhang, Chang Liu, Kun Wang, and Bo Liu. 2024. Consistencies are all you need for semi-supervised vision-language tracking. In *Proceedings of the 32nd ACM International Conference on Multimedia*. 1895–1904.

[19] Jia-Wei Ge, Jiu-Xin Cao, Zhi-Xiang Zhao, and Bo Liu. 2025. FSD-GAN: Generative Adversarial Training for Face Swap Detection via the Latent Noise Fingerprint. *Journal of Computer Science and Technology* 40, 2 (2025), 397–412.

[20] Aritra Ghosh, Himanshu Kumar, and PS Sastry. 2017. Robust loss functions under label noise for deep neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31.

[21] Aritra Ghosh and Andrew Lan. 2021. Contrastive Learning Improves Model Robustness Under Label Noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2703–2708.

[22] Jacob Goldberger and Ehud Ben-Reuven. 2017. Training deep neural-networks using a noise adaptation layer. In *International conference on learning representations*.

[23] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *arXiv preprint arXiv:1804.06872* (2018).

[24] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9729–9738.

[25] Rundong He, Zhongyi Han, Xiushan Nie, Yilong Yin, and Xiaojun Chang. 2024. Visual Out-of-Distribution Detection in Open-Set Noisy Environments. *International Journal of Computer Vision* 132, 11 (2024), 5453–5470.

[26] Dan Hendrycks and Kevin Gimpel. 2016. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. In *International Conference on Learning Representations*.

[27] Zhizhong Huang, Junping Zhang, and Hongming Shan. 2023. Twin Contrastive Learning with Noisy Labels. In *CVPR*.

[28] Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. 2018. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *International Conference on Machine Learning*. PMLR, 2304–2313.

[29] Nazmul Karim, Mamshad Nayeem Rizve, Nazanin Rahnavard, Ajmal Mian, and Mubarak Shah. 2022. Unicon: Combating label noise through uniform selection and contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9676–9686.

[30] Junnan Li, Richard Socher, and Steven CH Hoi. 2020. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv preprint arXiv:2002.07394* (2020).

[31] Shikun Li, Xiaobo Xia, Shiming Ge, and Tongliang Liu. 2022. Selective-supervised contrastive learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 316–325.

[32] Wen Li, Limin Wang, Wei Li, Eirikur Agustsson, and Luc Van Gool. 2017. Webvision database: Visual learning and understanding from web data. *arXiv preprint arXiv:1708.02862* (2017).

[33] Yang Liu, Hao Cheng, and Kun Zhang. 2023. Identifiability of label noise transition matrix. In *International Conference on Machine Learning*. PMLR, 21475–21496.

[34] Anqi Mao, Mehryar Mohri, and Yutao Zhong. 2023. Cross-entropy loss functions: Theoretical analysis and applications. In *International conference on Machine learning*. PMLR, 23803–23828.

[35] Diego Ortego, Eric Arazo, Paul Albert, Noel E O'Connor, and Kevin McGuinness. 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.

[36] Deep Patel and PS Sastry. 2023. Adaptive sample selection for robust learning under label noise. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3932–3942.

[37] Deep Patel and PS Sastry. 2023. Adaptive sample selection for robust learning under label noise. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3932–3942.

[38] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[39] Ragav Sachdeva, Filipe R Cordeiro, Vasileios Belagiannis, Ian Reid, and Gustavo Carneiro. 2021. Evidentialmix: Learning with combined open-set and closed-set noisy labels. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3607–3615.

[40] Hwanjun Song, Minseok Kim, and Jae-Gil Lee. 2019. Selfie: Refurbishing unclean samples for robust deep learning. In *International Conference on Machine Learning*. PMLR, 5907–5915.

[41] Zhonglin Sun, Chen Feng, Ioannis Patras, and Georgios Tzimiropoulos. 2024. LAFS: Landmark-based facial self-supervised learning for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1639–1649.

[42] Zeren Sun, Fumin Shen, Dan Huang, Qiong Wang, Xiangbo Shu, Yazhou Yao, and Jinhui Tang. 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.

[43] Wenhai Wan, Xinrui Wang, Ming-Kun Xie, Shao-Yuan Li, Sheng-Jun Huang, and Songcan Chen. 2024. Unlocking the power of open set: A new perspective for open-set noisy label learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 38. 15438–15446.

[44] Lei Wang, Jieming Bian, and Jie Xu. 2024. Federated Learning with Instance-Dependent Noisy Label. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 8916–8920.

[45] Yisen Wang, Weiyang Liu, Xingjun Ma, James Bailey, Hongyuan Zha, Le Song, and Shu-Tao Xia. 2018. Iterative learning with open-set noisy labels. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8688–8696.

[46] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. 2019. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of*

the IEEE/CVF International Conference on Computer Vision. 322–330.

[47] Hongxin Wei, Lei Feng, Xiangyu Chen, and Bo An. 2020. Combating noisy labels by agreement: A joint training method with co-regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13726–13735.

[48] Hongxin Wei, Lue Tao, Renchunzi Xie, and Bo An. 2021. Open-set label noise can improve robustness against inherent label noise. *Advances in Neural Information Processing Systems* 34 (2021), 7978–7992.

[49] Qi Wei, Haoliang Sun, Xiankai Lu, and Yilong Yin. 2022. Self-filtering: A noise-aware sample selection for label noise with confidence penalization. In *European Conference on Computer Vision*. Springer, 516–532.

[50] Jonathan Wilton and Nan Ye. 2024. Robust Loss Functions for Training Decision Trees with Noisy Labels. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 15859–15867.

[51] Pengxiang Wu, Songzhu Zheng, Mayank Goswami, Dimitris Metaxas, and Chao Chen. 2020. A topological filter for learning with label noise. *Advances in neural information processing systems* 33 (2020), 21382–21393.

[52] Zhi-Fan Wu, Tong Wei, Jianwen Jiang, Chaojie Mao, Mingqian Tang, and Yu-Feng Li. 2021. NGC: A Unified Framework for Learning with Open-World Noisy Data. *arXiv preprint arXiv:2108.11035* (2021).

[53] Xiaobo Xia, Bo Han, Nannan Wang, Jiankang Deng, Jiatong Li, Yinian Mao, and Tongliang Liu. 2022. Extended T: Learning With Mixed Closed-Set and Open-Set Noisy Labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 3 (2022), 3047–3058.

[54] Xiaobo Xia, Tongliang Liu, Bo Han, Mingming Gong, Jun Yu, Gang Niu, and Masashi Sugiyama. 2021. Sample selection with uncertainty of losses for learning with noisy labels. *arXiv preprint arXiv:2106.00445* (2021).

[55] Xiaobo Xia, Tongliang Liu, Bo Han, Nannan Wang, Mingming Gong, Haifeng Liu, Gang Niu, Dacheng Tao, and Masashi Sugiyama. 2020. Part-dependent label noise: Towards instance-dependent label noise. *Advances in Neural Information Processing Systems* 33 (2020), 7597–7610.

[56] Xiaobo Xia, Tongliang Liu, Nannan Wang, Bo Han, Chen Gong, Gang Niu, and Masashi Sugiyama. 2019. Are anchor points really indispensable in label-noise learning? *Advances in neural information processing systems* 32 (2019).

[57] Shuo Yang, Erkun Yang, Bo Han, Yang Liu, Min Xu, Gang Niu, and Tongliang Liu. 2022. Estimating instance-dependent bayes-label transition matrix using a deep neural network. In *International Conference on Machine Learning*. PMLR, 25302–25312.

[58] Yazhou Yao, Zeren Sun, Chuanyi Zhang, Fumin Shen, Qi Wu, Jian Zhang, and Zhenmin Tang. 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.

[59] Xingrui Yu, Bo Han, Jiangchao Yao, Gang Niu, Ivor Tsang, and Masashi Sugiyama. 2019. How does disagreement help generalization against label corruption?. In *International Conference on Machine Learning*. PMLR, 7164–7173.

[60] Yue Zhang, Yiyi Chen, Chaowei Fang, Qian Wang, Jiayi Wu, and Jingmin Xin. 2025. Learning from open-set noisy labels based on multi-prototype modeling. *Pattern Recognition* 157 (2025), 110902.

[61] Zhilu Zhang and Mert R Sabuncu. 2018. Generalized cross entropy loss for training deep neural networks with noisy labels. *arXiv preprint arXiv:1805.07836* (2018).

[62] Evgenii Zheltonozhskii, Chaim Baskin, Avi Mendelson, Alex M Bronstein, and Or Litany. 2021. Contrast to Divide: Self-Supervised Pre-Training for Learning with Noisy Labels. *arXiv preprint arXiv:2103.13646* (2021).

[63] Xuelin Zhu, Jian Liu, Weijia Liu, Jiawei Ge, Bo Liu, and Jiuxin Cao. 2023. Scene-aware label graph learning for multi-label image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1473–1482.

[64] Zhaowei Zhu, Tongliang Liu, and Yang Liu. 2021. A second-order approach to learning with instance-dependent label noise. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10113–10123.