

# Open-Set Recognition with Scalable Rejection under Large-Class Scenarios<sup>\*</sup>

Jing Yang<sup>1</sup>, Bingbing Wu<sup>2</sup>, Qi Li<sup>3</sup> and Bang Li<sup>1,\*</sup>

<sup>1</sup>Key Laboratory of Oracle Bone Inscriptions Information Processing, Anyang Normal University, Anyang, China

<sup>2</sup>School of computer & information engineering, Anyang Normal University, Anyang, China

<sup>3</sup>Graduate School of Science and Engineering, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga, 525-8577, Japan

## Abstract

In real-world scenarios, classification systems often encounter previously unseen categories unavailable during training, posing a significant challenge to conventional closed-world models. To address this issue, Open-set Recognition has emerged with the goal of enabling models to not only accurately classify known categories but also effectively reject out-of-distribution samples. In recent years, various studies have attempted to balance these two objectives within unified frameworks. However, as the number of classes increases, such frameworks often suffer from blurred decision boundaries and diminished rejection capability. To mitigate this problem, we propose a rejection rule equipped with a logarithmic scaling mechanism, which dynamically adjusts the rejection boundary to maintain its stability and enhance the model's discriminative power in large-class scenarios. Experiments conducted on the CIFAR-100 benchmark with the WideResNet-28-10 (WRN-28-10) architecture show that our method achieves the highest AUROC in the Class Prototype Network (CPN) group for OOD detection, reaching 83.43%, representing a 2.98% improvement over the previous best method. Additionally, it improves AUPR by 1.03% and reduces FPR95 by 4.92%, while maintaining classification accuracy. These results highlight the method's strong capability in rejecting unknown samples and its robustness to class expansion.

## Keywords

open-set recognition, one-vs-all architecture, rejection rule, confidence calibration, out-of-distribution detection

## 1. Introduction

With the rapid advancement of deep learning technologies, neural networks have been widely adopted in tasks such as image classification [1, 2]. Under the closed-world assumption, these models have achieved continuously improving recognition accuracy on known categories, thereby enabling the practical deployment of various intelligent systems. However, real-world

---

*The 7th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT 2025), September 10-11, 2025, Kusatsu, Japan*

<sup>\*</sup> You can use this document as the template for preparing your publication. We recommend using the latest version of the ceurart style.

<sup>\*</sup>Corresponding author.

✉ JingYang@stu.aynu.edu.cn (J. Yang); 230901034@stu.aynu.edu.cn (B. Wu); liqi24@fc.ritsumei.ac.jp (Q. Li); libang@aynu.edu.cn (B. Li)

🆔 0009-0007-6593-7987 (J. Yang); 0009-0008-7795-3754 (B. Wu); 0000-0002-1963-5263 (Q. Li); 0000-0002-7254-6569 (B. Li)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

environments are inherently open, where deployed systems inevitably encounter inputs from previously unseen, unknown categories [3, 4]. Conventional classifiers lack mechanisms to handle such inputs and tend to misclassify them as the most similar known categories, severely undermining the system’s reliability and safety. As a result, Open-set Recognition (OSR) has emerged as a critical research focus in computer vision, particularly in safety-sensitive domains such as medical diagnosis and autonomous driving [5, 6]. The goal of OSR is to ensure that a model trained solely on known-class data can not only maintain reliable recognition of known inputs but also robustly reject unknown samples.

To better evaluate model performance under open-world conditions, recent studies have gradually adopted two complementary evaluation sub-tasks [7, 8]: Misclassification Detection (MisD) and Out-of-Distribution Detection (OOD). MisD focuses on identifying incorrect predictions within known classes, while OOD assesses the model’s ability to reject inputs that lie outside the training distribution. Although this task decomposition is not yet a universal standard, it has been widely employed in recent OSR research as a practical framework for jointly measuring classification robustness and rejection capability. This in turn has inspired the development of unified modeling strategies that simultaneously address both objectives.

Against this backdrop, several approaches have proposed integrated frameworks that jointly optimize classification and rejection goals [9]. A representative method, Unified Classification and Rejection, builds on the One-vs-All (OVA) strategy [10], incorporating class prototype constraints and end-to-end training to achieve balanced classification and rejection performance on small- and medium-scale tasks. However, as the number of classes increases, the model must accommodate more known-class clusters within a limited feature space. This leads to the erosion of necessary decision boundary margins, making it easier for OOD samples to fall into the regions of known classes. Consequently, the rejection boundaries become unstable, and the model’s discriminative reliability degrades.

To alleviate this issue, we propose a Log-scaled Rejection Calibration mechanism as a lightweight extension to the One-Versus-All with Prototype Learning (OVA-PL) framework. By introducing a category-aware scaling factor into the rejection score computation during inference, the method dynamically adjusts the boundary sensitivity under varying class scales. Unlike traditional post-hoc thresholding methods, our approach is tightly integrated with the structure of OVA-PL and maintains compatibility with the trained classifier logits. It requires no additional parameters or loss modifications, making it simple, interpretable, and robust in high-class-count settings.

Extensive experiments conducted on the CIFAR-100 dataset using the WRN-28-10 backbone show that our method achieves leading OOD detection performance, surpassing all other CPN-based approaches. Meanwhile, it maintains competitive MisD rejection compared to the standard baseline. These results demonstrate the proposed method’s superior robustness and boundary stability under large-class open-set conditions. Overall, our main contributions can be summarized below:

- We propose a scalable **log-scaled rejection calibration** mechanism that explicitly accounts for the impact of class expansion on rejection confidence. By incorporating a class-count-aware scaling term into the rejection score computation, the method adaptively calibrates decision boundaries without altering the model structure or loss function.

It enhances the robustness of unified classification–rejection systems under large-class open-set scenarios while maintaining compatibility with existing OVA-based frameworks.

- We empirically validate its effectiveness under challenging settings. On the CIFAR-100 dataset, our method achieves the best AUROC (83.43%) and lowest MisD FPR95 (43.26%) in the CPN group, verifying its practical value in open-set recognition under class expansion.

The remainder of this article is organized as follows. Section 2 reviews the background of open-set recognition and the development of unified classification–rejection strategies. Section 3 details the proposed log-scaled rejection calibration mechanism. Section 4 presents the experimental setup, evaluation metrics, and result analysis. Section 5 concludes the paper and outlines potential future research directions.

## 2. Related Work

### 2.1. Open-set recognition and out-of-distribution detection

Open-set Recognition, first proposed by Scheirer et al. in 2013 [3], aims to break the limitations of the “closed-world” assumption inherent in traditional classification models. It enables systems to identify and reject unknown class samples that do not appear in the training set during multi-class tasks, while maintaining accurate classification for known categories. With the increasing demand for model deployment in open environments, OSR has become a critical research direction in safety-aware learning. Against this backdrop, Out-of-Distribution Detection, systematically introduced by Hendrycks and Gimpel in 2017 [4], focuses on determining whether an input deviates from the overall distribution of the training data, emphasizing model robustness under distributional shifts. Although OSR and OOD differ in focus—OSR emphasizes openness in label space while OOD emphasizes externality in data distribution—they share high consistency in the core subtask of identifying and rejecting inputs beyond the training data distribution. Both require models to rely primarily on known class information and respond to anomalous or unknown inputs with uncertainty, thereby avoiding overconfident mispredictions.

### 2.2. Post-hoc rejection methods

Consequently, a growing body of research in recent years has treated OSR and OOD as two complementary perspectives that can be jointly modeled, showing significant overlap in evaluation metrics, model design, and system architecture [10, 11]. Comparatively, the OOD task features more standardized evaluation protocols, richer public benchmarks, and stronger reproducibility, and has gradually become an important reference for assessing the performance of open-world recognition systems. It has also driven the development of numerous post-processing rejection mechanisms. These methods typically do not alter the main classifier’s structure, but instead construct additional rejection scoring functions based on its output. Representative approaches include: ODIN [12], which enhances Softmax output discrimination via temperature scaling and input perturbation; the Mahalanobis distance method [13], which builds Gaussian models in feature space for anomaly detection; and energy-based OOD method [14], which map logits to energy scores to better respond to low-confidence inputs.

These methods offer clear advantages in deployment flexibility and generalizability. However, since the rejection mechanism in post-processing approaches is not jointly optimized with the classification boundary, their scoring functions often misalign with the original decision surface, leading to the mistaken rejection of known samples that could otherwise be correctly classified. This degrades performance in both Misclassification Detection and In-Distribution Classification. Despite notable progress in enhancing OOD detection, such methods often come at the cost of undermining the confidence in known class predictions [15].

### 2.3. Towards unified classification–rejection frameworks

To address the disconnect in rejection modeling introduced by post-processing, researchers have gradually shifted toward unified training frameworks that integrate classification and rejection. A representative work by Yang et al. constructs a prototype space by combining discriminative and generative losses, and applies distance-based and probability-based rules to reject unknown samples [16]. This method demonstrates significant advantages in both closed-set classification and open-set unknown detection, simultaneously improving the accuracy of known class predictions and the effectiveness of unknown sample detection, thus better meeting the requirements of OSR tasks. Building upon this, Cheng et al. proposed the OVA-PL framework [10], which integrates the One-vs-All strategy into the unified training process. By jointly optimizing OVA loss and multi-class cross-entropy loss, the method effectively combines the decision boundary control of OVA learning with the representational power of multi-class discrimination. It achieves notable improvements in OOD detection while maintaining MisD performance and offering structural simplicity. However, as the number of classes continues to grow, the rejection signal becomes increasingly diluted in feature space, making it difficult to maintain sufficient rejection margins and causing the rejection boundaries to contract—thereby increasing the risk of misclassifying OOD samples as known classes [10]. Overall, although current research has made some progress, the joint optimization of classification, OOD detection, and MisD remains an open challenge. Developing robust and scalable classification–rejection collaboration mechanisms—especially for high-class-count and complex scenarios—remains a key frontier in OSR and OOD research [17, 18].

## 3. Methodology

In this study, we adopt a unified classification–rejection architecture built upon the OVA-PL framework, which has demonstrated superior performance in open-set recognition tasks [10]. Our approach retains the standard OVA-based binary classifiers and prototype learning structure, while introducing a novel log-scaled rejection calibration mechanism to enhance robustness under class expansion. This mechanism operates directly on classifier outputs without modifying the network structure or training objectives, ensuring compatibility with existing OVA-PL systems. All components are jointly evaluated on open-set benchmarks to validate the effectiveness of our proposed design.

### 3.1. Unified Modeling with OVA-PL Architecture

We adopt a unified modeling framework based on the One-vs-All (OVA) classification strategy and Prototype Learning (PL), which has proven effective for open-set recognition (OSR). In this architecture, each binary classifier is trained to distinguish a specific known class from all others. Formally, for an input sample  $x$ , the posterior probability of class  $i$  from the  $i$ -th binary classifier is given by:

$$p_i(x) = \sigma(g_i(x)) = \frac{1}{1 + \exp(-g_i(x))} \quad (1)$$

where  $g_i(x)$  is the logit output of the classifier. The OVA classification probability is then converted to a multi-class prediction by computing the softmax over the positive responses.

To enhance representation learning, the Prototype Learning loss  $L_{PL}$  is integrated into the training objective. Additionally, a regularization term  $L_{reg}$ , defined as the cross-entropy between the predicted softmax and the ground-truth class label, is used to stabilize learning. The total training loss for the hybrid OVA-PL model is defined as:

$$L = \beta \cdot L_{OVA} + (1 - \beta) \cdot L_{reg} + \lambda \cdot L_{PL} \quad (2)$$

where  $\beta \in [0, 1]$  controls the trade-off between binary classification and softmax-based regularization, and  $\lambda$  is the weight for the prototype loss.

### 3.2. Challenges of Rejection under Class Expansion

In the OVA-PL framework, rejection is performed by estimating the probability that an input sample does not belong to any known class. This is defined as:

$$p_{\text{OOD}}(x) = 1 - \sum_{i=1}^K p_i(x) \quad (3)$$

where  $p_i(x)$  denotes the output probability of the  $i$ -th binary classifier for class  $i$ . Based on this, the vanilla rejection rule is defined as:

$$\phi_{K+1}(x) = \min \left\{ 1 - p_{\text{OOD}}(x), \max_i p_i(x) + \epsilon \right\} \quad (4)$$

where  $\epsilon$  is a small calibration constant.

This rule jointly enables the handling of out-of-distribution detection (via  $p_{\text{OOD}}(x)$ ), misclassification detection (via maximum confidence thresholding), and standard classification (via  $\arg \max$ ). It serves as a unified decision criterion in open-set recognition.

However, under class expansion, this rule exhibits notable instability. As the number of known classes  $K$  increases, the summation  $\sum_{i=1}^K p_i(x)$  includes more terms, which causes  $p_{\text{OOD}}(x)$  to shrink. As a result, the rejection signal is weakened, making it harder to distinguish OOD samples. Moreover,  $\phi_{K+1}(x)$  becomes dominated by the confidence term  $\max_i p_i(x)$ , which leads to overly conservative rejection and increases false acceptance of unknowns.

To improve robustness under such conditions, we propose a log-scaled rejection calibration mechanism that adaptively adjusts the rejection score based on the number of known classes. This mechanism is presented in the following section.

### 3.3. Log-scaled Rejection Calibration

As discussed in Section 3.2, the rejection rule in Eq. (4) fundamentally relies on the unknown class probability in Eq. (3). This design interprets the remaining confidence after summing all known class probabilities as the likelihood that a sample belongs to an unknown class. While effective in small-class scenarios, this formulation becomes unstable as the number of known classes  $K$  increases. Specifically, the summation  $\sum_{i=1}^K p_i(x)$  naturally grows with  $K$ , leading to a systematic shrinkage of  $p_{\text{OOD}}(x)$ . As a result, the rejection signal weakens and becomes increasingly dominated by high-confidence known class predictions, hindering the detection of unknown samples.

To alleviate this issue, we propose a *Log-scaled Rejection Calibration* mechanism. This method is based on the normalization reformulation of OOD probability using Dempster-Shafer Theory of Evidence (DSTE) as adopted in the Unified framework, expressed as:

$$p_{\text{OOD}}(x) = \frac{1}{1 + \sum_{j=1}^K \exp(g_j(x))} \quad (5)$$

where the unknown class is modeled as a virtual node with logit zero and a fixed prior mass of 1. However, this constant becomes insufficient under class expansion; as  $K$  increases, the relative contribution of the unknown class diminishes rapidly.

To address this, we replace the constant term with a logarithmic prior that grows with the number of classes, leading to the following adaptive formulation:

$$p_{\text{OOD}}^\alpha(x) = \frac{\alpha \cdot \log(K + 1)}{\alpha \cdot \log(K + 1) + \sum_{j=1}^K \exp(g_j(x))} \quad (6)$$

where  $\alpha > 0$  is a tunable hyperparameter that controls the compensation strength for class expansion. This design is functionally equivalent to introducing a virtual unknown class with a class-dependent prior weight, allowing its representation to remain distinguishable as  $K$  increases. It effectively prevents the OOD probability from being overwhelmed by the logits of known classes in high-class-count scenarios.

The final rejection rule retains the same structure as Eq. (5), replacing only the OOD estimation term with  $p_{\text{OOD}}^\alpha(x)$ . The mechanism adds no extra training parameters or architectural changes, and can be seamlessly integrated into the OVA-PL framework, significantly improving rejection robustness in large-class OSR tasks without increasing computational overhead.

## 4. Experiment

In this chapter, we systematically evaluate the effectiveness of the proposed rejection mechanism, focusing on two core tasks: Out-of-Distribution detection and Misclassification Detection.

We formally introduce the term **(K+extra) rejection rule** to refer to the log-scaled rejection calibration method proposed earlier in the paper.

Experiments are conducted on the CIFAR-100 benchmark [19], using a medium-scale backbone WRN-28-10 [20] to verify the generalizability of our method. We compare our approach with a variety of widely-used post-hoc rejection techniques, as well as rejection rules under the Unified framework. Results demonstrate that our method achieves consistent improvements across multiple evaluation metrics, showing superior rejection robustness and detection stability, particularly in scenarios with a large number of known classes.

#### 4.1. Dataset

To construct a challenging open-set recognition scenario and ensure the comparability of experimental results, we strictly follow the CIFAR benchmark settings and data preprocessing procedures adopted in previous studies [10], as detailed below.

In this study, CIFAR-100 is used as the in-distribution dataset for training and evaluation. It contains 100 distinct classes, with 500 training images and 100 test images per class. All images have a resolution of  $32 \times 32$ .

We adopt a standard set of out-of-distribution test datasets from the CIFAR benchmark, including:

- **Textures** [21]: 47 classes of natural texture images, totaling 5,640 images;
- **SVHN** [22]: 10 digit classes of street view house number images, with 26,032 test images;
- **Places365** [23]: A large-scale scene dataset, from which a subset of categories is randomly sampled for OOD testing;
- **LSUN-Crop / LSUN-Resize** [24]: Derived from the LSUN scene classification dataset (10,000 images in total), where **LSUN-Crop** applies random cropping and **LSUN-Resize** applies global downscaling;
- **iSUN** [25]: A natural image dataset collected via eye-tracking, containing 2,000 test images.

During testing, all OOD images are downsampled to  $32 \times 32$  to match the InD input size. For each OOD dataset, 2,000 images are randomly sampled and the experiment is repeated multiple times to obtain the average performance.

#### 4.2. Implementation details

The experiments were conducted on an NVIDIA® GeForce RTX 4090-based platform. We adopted PyTorch 2.4.1 with CUDA 12.4 as the deep learning framework. To ensure fair comparison with prior work [10], we strictly followed their training configuration.

For the CIFAR-100 benchmark, the WRN-28-10 model was trained for 200 epochs using the AdamW optimizer, with momentum set to 0.9 and weight decay of  $2 \times 10^{-4}$ . The learning rate was initialized at 0.1 and decayed to 0.01, 0.001, and 0.0001 at epochs 100, 150, and 200, respectively. Cosine annealing with a 40-epoch warm-up phase was applied. The batch size was fixed at 64 for both training and evaluation. The loss coefficients were set to  $\lambda = 0.05$  and  $\beta = 0.95$ ; the temperature scaling factor  $\lambda_1$  was set to 2.0.



The proposed Log-scaled Rejection Calibration introduces a tunable parameter  $\alpha$ , applied only during inference. We conducted a grid search over  $\alpha \in [0, 2.0]$  with a step size of 0.01 to balance OOD detection and MisD rejection. Larger  $\alpha$  values generally improved OOD metrics but degraded MisD performance due to excessive penalization of low-confidence predictions. The best trade-off was found at  $\alpha = 0.21$  on CIFAR-100 with WRN-28-10.

### 4.3. Evaluation metrics

To comprehensively evaluate the effectiveness of the proposed (**K+extra**) rejection mechanism, we adopt several commonly used metrics in OSR, covering three main aspects: OOD detection performance, MisD rejection ability, and in-distribution classification accuracy. The specific metrics are as follows:

- **AUROC (Area Under the Receiver Operating Characteristic Curve):** For OOD detection, AUROC measures the model’s ability to distinguish between InD and OOD samples by adjusting the confidence score threshold used for rejecting OOD inputs. For MisD detection, it evaluates whether the model can accurately separate correctly and incorrectly predicted InD samples by tuning the threshold used to reject InD inputs. A higher AUROC indicates stronger discrimination ability across thresholds and is one of the core metrics in evaluating rejection mechanisms.
- **AUPR (Area Under the Precision-Recall Curve):** When the number of InD and OOD samples is highly imbalanced, AUPR places more emphasis on the precision of identified OOD instances. It is particularly suitable for safety-critical applications.
- **FPR95 (False Positive Rate at 95% TPR):** For OOD evaluation, FPR95 quantifies the proportion of InD samples incorrectly identified as OOD when the model detects 95% of OOD inputs. For MisD evaluation, it measures how many incorrectly predicted InD samples are wrongly retained, while 95% of the correctly predicted samples are preserved. A lower FPR95 indicates more robust rejection behavior.
- **Acc (In-Distribution Accuracy):** This is the classification accuracy over InD test samples. In OSR tasks, the model is expected to maintain high accuracy on known classes while identifying unknown ones. We retain this metric under OOD settings to ensure that the proposed rejection rule does not degrade the original classification performance.
- **AURC (Area Under the Risk-Coverage Curve) and E-AURC (Expected AURC):** These metrics evaluate how prediction risk changes with coverage. A lower AURC means high-risk samples are rejected earlier, reducing error on the retained set. E-AURC normalizes for accuracy and confidence distribution, enabling fairer model comparisons.

### 4.4. Performance Comparison

To verify the effectiveness of our proposed *Log-scaled Rejection Calibration*, we evaluate its performance against existing baselines on the CIFAR-100 benchmark using the WRN-28-10 backbone. As shown in Table 1, our method—Hybrid+PL with (*K+extra*) rejection—achieves the best overall performance within the CPN group. Compared to the standard Hybrid+PL with (*K+1*) rejection, it improves OOD detection with an AUROC gain of **2.98%** (from 80.45% to



83.43%), an AUPR increase of **1.03%**, and a reduction in FPR95 by **4.92%**, while maintaining the same in-distribution classification accuracy.

For MisD rejection, our method further lowers FPR95 by **10.42%** (from 53.68% to 43.26%) and achieves a reduced EAURC (33.74 vs. 35.26), reflecting better ranking quality for misclassified samples. Although AURC increases slightly (from 57.10 to 59.20), the drop in EAURC suggests more stable and equitable rejection across varying confidence distributions. Overall, these results confirm that the proposed *log-scaled calibration mechanism* enhances both OOD detection and MisD rejection without introducing any additional training cost.

**Table 1**

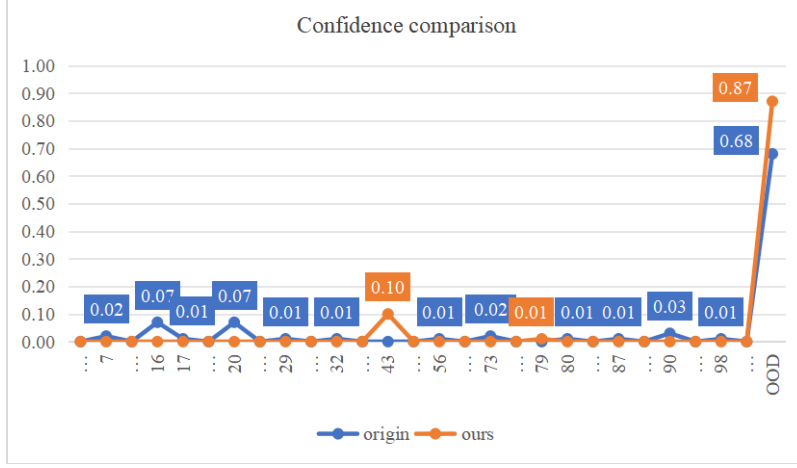
Performance of OOD detection on CIFAR-100 benchmark with WRN-28-10 backbone. OOD metrics are averaged across six datasets. All values are in percentage (%). Best results within CNN and CPN groups are in **bold**.

Model	Loss	Rejection Rule	OOD: CIFAR-100				MisD: CIFAR-100			
			AUROC↑	AUPR↑	FPR95↓	ACC↑	AUROC↑	FPR95↓	AURC↓	EAURC↓
CNN	CE	ODIN [12]	81.25	95.42	74.28	<b>79.72</b>	84.23	57.42	63.22	41.35
		Maha [13]	63.32	88.83	79.36	<b>79.72</b>	73.49	69.02	91.43	69.56
		EBO [14]	78.03	94.47	78.65	<b>79.72</b>	83.56	58.04	64.81	42.94
		GradNorm [26]	64.90	89.18	81.82	<b>79.72</b>	55.54	91.61	170.43	148.56
		ReAct [27]	84.09	<b>96.33</b>	68.46	<b>79.72</b>	70.31	74.75	104.38	82.51
		MaxLogit [28]	77.82	94.41	79.40	<b>79.72</b>	84.23	57.42	63.22	41.36
		KNN [29]	80.50	95.03	69.68	<b>79.72</b>	84.84	53.04	60.46	38.59
		ASH [15]	<b>85.22</b>	96.24	<b>54.97</b>	79.60	79.20	68.76	80.83	58.43
		Relation [30]	82.53	95.42	69.19	<b>79.72</b>	<b>88.50</b>	<b>40.52</b>	58.00	36.14
		K probs(MSP) [4]	75.77	93.92	81.53	<b>79.72</b>	<b>88.50</b>	<b>40.52</b>	<b>49.45</b>	<b>27.59</b>
CPN	DCE+PL	Distance-based [11]	80.06	94.93	73.06	<b>80.66</b>	84.90	62.25	60.51	40.61
	DCE+PL	Probs.-based [11]	79.40	94.70	75.27	<b>80.66</b>	<b>87.96</b>	43.63	<b>49.00</b>	<b>28.96</b>
	OVA+PL	(K+1) probs [10]	80.15	95.00	74.89	80.29	86.80	52.13	55.79	34.75
	hybrid+PL	(K+1) probs [10]	80.45	94.97	72.38	79.84	86.86	53.68	57.10	35.26
	hybrid+PL	(K+extra) probs	<b>83.43</b>	<b>96.00</b>	<b>67.46</b>	79.84	86.90	<b>43.26</b>	59.20	33.74

To further illustrate the effectiveness of our proposed *log-scaled calibration*, Figure 1 visualizes the predicted confidence scores across all known classes and the OOD class. Compared with the baseline Hybrid+PL using  $(K+1)$  rejection, our method with  $(K+extra)$  produces a significantly sharper and more discriminative confidence peak on the OOD dimension, while effectively suppressing spurious high scores on known categories. This clearer separation reinforces the improved OOD detection performance reported in Table 1.

## 5. Conclusion

This paper presents a log-scaled rejection calibration mechanism designed to enhance rejection reliability in open-set scenarios with many of known classes. By adaptively scaling the



**Figure 1:** Confidence distribution comparison between *Hybrid+PL (K+1)* and our *Hybrid+PL (K+extra)* on an OOD input. Our method produces significantly higher confidence for the OOD class while suppressing spurious high-confidence predictions on known classes.

prior of the virtual unknown class according to class count, the method effectively mitigates the degradation in OOD confidence estimation caused by class expansion. It integrates seamlessly with existing OVA-PL frameworks without requiring any structural or training modifications.

Extensive experiments on the CIFAR-100 benchmark with the WRN-28-10 backbone demonstrate the effectiveness and scalability of the proposed method. It achieves the highest AUROC (83.43%) and the lowest FPR95 (67.46%) among all CPN-group methods for OOD detection. For MisD performance, it yields the lowest FPR95 (43.26%) and a favorable EAURC (33.74). Meanwhile, the in-distribution classification accuracy remains unchanged, confirming that the enhanced OOD rejection capability does not compromise base classification or MisD reliability. Compared to all CNN-based baselines, our method also maintains competitive performance across key metrics, showing its comprehensive advantage in both groups. These results validate the practical value of the proposed approach in building robust and unified classification–rejection systems under large-class open-set conditions.

## Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 62506007), the Natural Science Foundation of Henan Province (Grant No. 242300420680), the Paleography and Chinese Civilization Inheritance and Development Program (Grant Nos. G1807, G1806, G2821), the Henan Province Science and Technology Research Project (Grant Nos. 242102210116, 252102321071), the Major Science and Technology Project of Anyang (Grant No. 2025A02SF007), and the Henan Province High-Level Talents International Training Program (Grant No. GCC2025028).

## Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

## References

- [1] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [2] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Communications of the ACM* 60 (2017) 84–90.
- [3] W. J. Scheirer, A. Rocha, A. Sapkota, T. E. Boult, Toward open set recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (2013) 1757–1772.
- [4] D. Hendrycks, K. Gimpel, A baseline for detecting misclassified and out-of-distribution examples in neural networks, *International Conference on Learning Representations (ICLR)* (2017).
- [5] Y. Xu, R. Wang, R.-W. Zhao, X. Xiao, R. Feng, Semi-supervised and class-imbalanced open set medical image recognition, *IEEE Access* (2024).
- [6] L. Balasubramanian, F. Kruber, M. Botsch, K. Deng, Open-set recognition based on the combination of deep learning and ensemble method for detecting unknown traffic scenarios, in: 2021 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2021, pp. 674–681.
- [7] S. Vaze, K. Han, A. Vedaldi, A. Zisserman, Open-set recognition: A good closed-set classifier is all you need?, in: *International Conference on Learning Representations (ICLR)*, 2022.
- [8] J. Yang, K. Zhou, Y. Li, Z. Liu, Generalized out-of-distribution detection: A survey, *Advances in Neural Information Processing Systems (NeurIPS)* 34 (2021) 1–15.
- [9] F. Zhu, Z. Cheng, X.-Y. Zhang, C.-L. Liu, Openmix: Exploring outlier samples for misclassification detection, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 12074–12083.
- [10] Z. Cheng, X. Zhang, C. Liu, Unified classification and rejection: A one-versus-all framework, *Machine Intelligence Research* 21 (2024) 870–887. doi:10.1007/s11633-024-1514-4.
- [11] H.-M. Yang, X.-Y. Zhang, F. Yin, Q. Yang, C.-L. Liu, Convolutional prototype network for open set recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44 (2022) 2358–2370. doi:10.1109/TPAMI.2020.3045079.
- [12] S. Liang, Y. Li, R. Srikant, Enhancing the reliability of out-of-distribution image detection in neural networks, in: *International Conference on Learning Representations (ICLR)*, 2018.
- [13] K. Lee, K. Lee, H. Lee, J. Shin, A simple unified framework for detecting out-of-distribution samples and adversarial attacks, *Advances in Neural Information Processing Systems (NeurIPS)* (2018).
- [14] W. Liu, X. Wang, J. D. Owens, Y. Li, Energy-based out-of-distribution detection, in: *Advances in Neural Information Processing Systems*, volume 34, Vancouver, Canada, 2020.
- [15] A. Djurisic, N. Bozanic, A. Ashok, R. Liu, Extremely simple activation shaping for out-of-

- distribution detection, in: International Conference on Learning Representations, Kigali, Rwanda, 2023.
- [16] H.-M. Yang, X.-Y. Zhang, F. Yin, C.-L. Liu, Robust classification with convolutional prototype learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 3474–3482.
  - [17] H. Wang, S. Vaze, K. Han, Dissecting out-of-distribution detection and open-set recognition: A critical analysis of methods and benchmarks, *International Journal of Computer Vision* (2024). doi:10.1007/s11263-024-02222-4.
  - [18] F. Zhu, S. Ma, Z. Cheng, X.-Y. Zhang, Z. Zhang, C.-L. Liu, Open-world machine learning: A systematic review and future directions, *arXiv preprint arXiv:2403.01759* (2024).
  - [19] A. Krizhevsky, Learning multiple layers of features from tiny images, Technical Report, University of Toronto, 2009.
  - [20] S. Zagoruyko, N. Komodakis, Wide residual networks, in: British Machine Vision Conference (BMVC), 2016.
  - [21] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, A. Vedaldi, Describing textures in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 3606–3613.
  - [22] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, A. Y. Ng, Reading digits in natural images with unsupervised feature learning, in: Advances in Neural Information Processing Systems (NIPS), 2011, pp. 1–9.
  - [23] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, A. Torralba, Places: A 10 million image database for scene recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (2017) 1452–1464.
  - [24] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, J. Xiao, Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop, *CoRR* abs/1506.03365 (2015).
  - [25] P. Xu, K. A. Ehinger, Y. Zhang, A. Finkelstein, S. R. Kulkarni, J. Xiao, Turkergaze: Crowdsourcing saliency with webcam based eye tracking, *arXiv preprint arXiv:1504.06755* (2015).
  - [26] R. Huang, A. Geng, Y. Li, On the importance of gradients for detecting distributional shifts in the wild, in: Advances in Neural Information Processing Systems (NeurIPS), volume 34, Virtual, 2021.
  - [27] Y. Sun, C. Guo, Y. Li, React: Out-of-distribution detection with rectified activations, in: Advances in Neural Information Processing Systems (NeurIPS), volume 34, Virtual, 2021.
  - [28] D. Hendrycks, S. Basart, M. Mazeika, A. Zou, J. Kwon, M. Mostajabi, J. Steinhardt, D. Song, Scaling out-of-distribution detection for real-world settings, in: Proceedings of the 39th International Conference on Machine Learning (ICML), Baltimore, USA, 2022, pp. 8759–8773.
  - [29] Y. Sun, Y. Ming, X. Zhu, Y. Li, Out-of-distribution detection with deep nearest neighbors, in: Proceedings of the 39th International Conference on Machine Learning (ICML), Baltimore, USA, 2022, pp. 20827–20840.
  - [30] J.-H. Kim, S. Yun, H. O. Song, Neural relation graph: A unified framework for identifying label noise and outlier data, in: Advances in Neural Information Processing Systems (NeurIPS), New Orleans, USA, 2023. Article No. 1898.