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

Towards Evidential and Class Separable Open Set Object Detection

Ruofan Wang[1], Rui-Wei Zhao[2], Xiaobo Zhang[3*], Rui Feng[1,2,3,4*]

[1]School of Computer Science, Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai
[2]Academy for Engineering and Technology, Fudan University, Shanghai
[3]Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China
[4]Shanghai Collaborative Innovation Center of Intelligent Visual Computing
21212010034@m.fudan.edu.cn, rwzhao@fudan.edu.cn, zhangxiaobo0307@163.com, fengrui@fudan.edu.cn

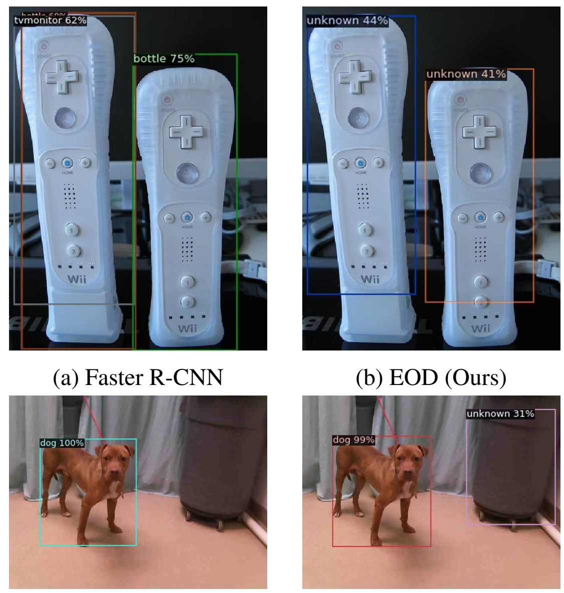
Abstract

Detecting in open-world scenarios poses a formidable challenge for models intended for real-world deployment. The advanced closed set object detectors achieve impressive performance under the closed set setting, but often produce overconfident misprediction on unknown objects due to the lack of supervision. In this paper, we propose a novel Evidential Object Detector (EOD) to formulate the Open Set Object Detection (OSOD) problem from the perspective of Evidential Deep Learning (EDL) theory, which quantifies classification uncertainty by placing the Dirichlet Prior over the categorical distribution parameters. The task-specific customized evidential framework, equipped with meticulously designed model architecture and loss function, effectively bridges the gap between EDL theory and detection tasks. Moreover, we utilize contrastive learning as an implicit means of evidential regularization and to encourage the class separation in the latent space. Alongside, we innovatively model the background uncertainty to further improve the unknown discovery ability. Extensive experiments on benchmark datasets demonstrate the outperformance of the proposed method over existing ones.

Introduction

Thanks to the powerful learning representation ability, deep neural networks have achieved great success in object detection (Girshick 2015; Ren et al. 2015; Redmon et al. 2016) under the closed world assumption, i.e., training and testing sets follow the same class distribution. However, in the deployment of real-world detector, unknown objects that were not seen during training (even in the background) may appear. Although object detector is trained to reject objects of no interest as background, deep neural networks tend to misclassify some unknown objects into one of the known classes since they have not been presented during training (Nguyen, Yosinski, and Clune 2015; Dhamija et al. 2020). This could lead to disastrous results in safety-critical applications such as autopilot. Therefore, the concept of Open Set Object Detection (OSOD) is proposed, which aims to detect the objects of interest correctly and recognize the unknown objects during testing.

*Corresponding author.



(a) Faster R-CNN (b) EOD (Ours)

(c) Faster R-CNN (d) EOD (Ours)

Figure 1: Compared to Faster R-CNN, which may misclassify unknown as incorrect known (a) or completely ignore unknown (c) in the open set problem, the proposed EOD can well locate and identify the unknown objects (b, d).

Compared with previous OSOD works (Miller et al. 2018; Du et al. 2022; Han et al. 2022), e.g., simply augmenting the K-way classifier to the K+1-way classifier based on comparatively simple and heuristic decision rules, thus causing to suffer from mistaking high probability scores for model certainty. Our method draws inspiration from the Evidential Deep Learning (EDL) theory and estimates epistemic uncertainty from the perspective of model confidence, which has better mathematical basis and interpretability. However, as the EDL theory was initially validated in relatively simple experiments, its direct application to real-world tasks often faces challenges of sensitivity to initialization and approximation difficulties. To address these problems, we further introduce task-specific customized modifications for OSOD.

In this paper, we propose Evidential Object Detector (EOD) for the OSOD task. More specifically, we estimate the Dirichlet Prior of the categorical distribution parameters based on the evidential theory and predict objects with high uncertainty as unknown. Furthermore, our customized framework introduces a meticulously designed loss function to train the detector for better evidence acquisition. It is worth noting that since the newly introduced EDL head can be regarded as a natural replacement of classification head with different optimization targets, it can plug and play on existing closed set object detectors without any adjustment to the model structure. And by encouraging class separation in the feature space and implicit regularization of evidenital learning, the auxiliary contrastive learning module provides the ideal characteristic for the open set problem. Figure 1 shows an example of our proposed method. The main contributions of our work can be summarized as:

- We establish an Evidential Object Detector (EOD) framework to formulate the OSOD task as an uncertainty estimation problem based on the evidential deep learning theory.
- The task-specific customized framework introduces a novel hybrid evidential loss and weak regularization term to accelerate early-stage evidence extraction and alleviate later-stage over-fitting for achieving stable OSOD performance.
- We adopt contrastive learning as an implicit evidence regularization method and encourage class separation in the feature space.
- Extensive experiments on several benchmark datasets show that the proposed EOD outperforms existing methods and achieves new state-of-the-art results.

Related Work

Open Set Recognition. The task of Open Set Recognition (OSR) was formalized in (Scheirer et al. 2012), which requires a classifier not only to accurately classify the images of known classes but also to reject unknown ones during testing. A benchmark OSR method is OpenMax (Bendale and Boult 2016), which replaced the SoftMax layer in DNNs with an OpenMax layer. This method redistributed the probability distribution of Softmax and obtained the class probability of unknown samples. More recently, (Sun et al. 2020) proposed Conditional Gaussian Distribution Learning (CGDL) model, which used Conditional VAE with the probabilistic ladder architecture (Rasmus et al. 2015; Sønderby et al. 2016) to model the probability distribution of each known class in the latent space. (Zhou, Ye, and Zhan 2021) proposed PROSER, which prepared for the unknown classes by learning placeholders for both data and classifier. Despite significant progress in OSR, OSOD remains challenging due to the complexity of various OSR methods, which cannot be readily applied to the classification sub-network of object detectors.

Deep Learning Uncertainty. To ascertain the reliability of deep learning outputs, numerous researchers have focused on modeling and quantifying uncertainty within the neural network. (Hüllermeier and Waegeman 2021) proposed the concepts of aleatoric uncertainty (the uncertainty caused by the noisy input data) and epistemic uncertainty (the uncertainty caused by missing training data). Bayesian Neural Networks (BNNs) (Denker et al. 1987; Tishby, Levin, and Solla 1989) introduces uncertainty to the weight of neural network through Bayesian learning, but is limited by the intractability of posterior inference and the difficulty of specifying a meaningful prior. Ensemble methods (Sagi and Rokach 2018; Hansen and Salamon 1990) estimate uncertainty and derive predictions based on multiple decision makers. Test-time data augmentation methods (Molchanov et al. 2020) quantify uncertainty based on several inference results by using different data augmentations in the testing phase. However, both the ensemble methods and the test-time data augmentation methods suffer from the drawbacks of high computation effort and memory consumption.

Evidential Deep Learning (EDL) was proposed by (Sensoy, Kaplan, and Kandemir 2018) for uncertainty quantification, and can be used both for the classification and regression tasks (Amini et al. 2020). In classification, EDL assumes that class probability follows the Dirichlet distribution and each parameter of the Dirichlet distribution accounts for the evidence of each class. As a single deterministic network method, it boasts higher computational efficiency and lower memory cost compared with other uncertainty estimation methods. Recently, evidential deep learning has shown great potential in real-world applications, e.g., (Bao, Yu, and Kong 2021) incorporated the evidential theory with video action recognition tasks and achieved impressive results. However, EDL faces challenges related to sensitivity to initialization and training process parameters (Gawlikowski et al. 2021), which are rarely addressed. In contrast, we have meticulously designed a customized evidential framework specifically for OSOD, mitigating this phenomenon from the perspective of loss design.

Open Set Object Detection. Open Set Object Detection (OSOD) can be viewed as an extension of Open Set Recognition (OSR). (Miller et al. 2018) proposed Dropout Sampling to approximate Bayesian neural network by keeping the Dropout layers activated during testing, to produce different prediction results and extract label uncertainty to improve the robustness of object detectors in open set condition. (Dhamija et al. 2020) first formalized the problem of open set object detection and proposed open set evaluation protocol Average Wilderness Impact (AWI) to quantify the sensitivity of an algorithm to open set. (Joseph et al. 2021) proposed ORE, leveraging contrastive clustering, Unknown Aware RPN and Energy Based Unknown Identifier, to address the challenges of unknown detection and incremental learning. (Han et al. 2022) proposed OpenDet with Contrastive Feature Learner (CFL) and Unknown Probability Learner (UPL) to separate high/low-density regions in the latent space to detect unknown classes. Compared with the prior works, we are the first to introduce Evidential deep learning (EDL) theory in OSOD tasks. Our approach explicitly formulates the open set problem from the perspective of Bayesian Prior and achieves state-of-the-art results.

Methodology

Without loss of generality and to ensure fair comparison with existing algorithms, we use Faster R-CNN with Fea-

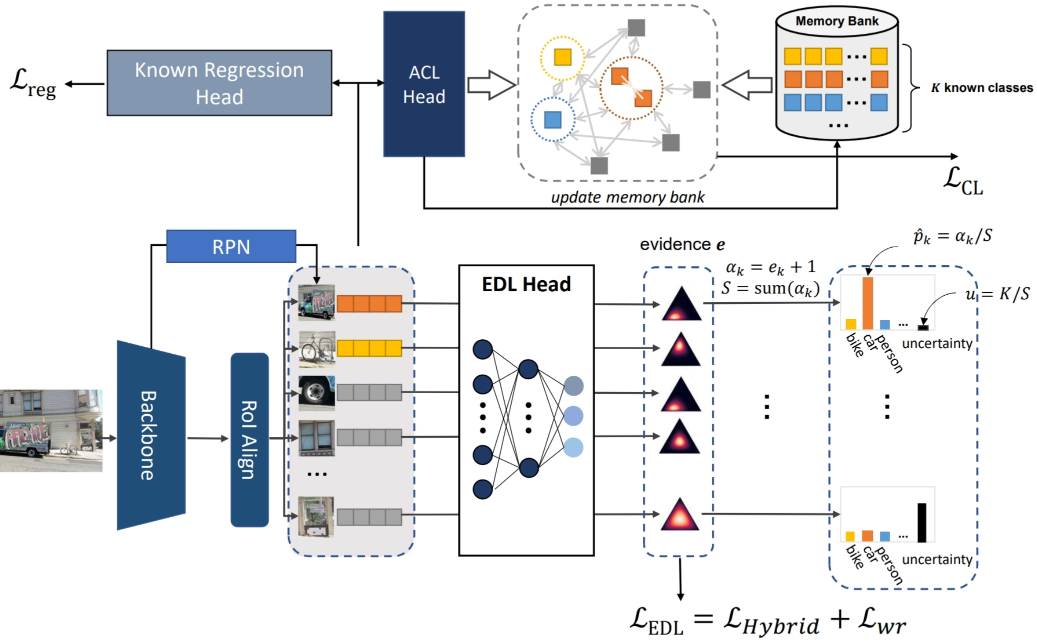


Figure 2: Overview of our proposed method. EDL head predicts evidence e to model the Dirichlet distribution of class probability. ACL head encourages class separation in the latent space. The expected probability $\hat{p}$ and uncertainty u derived from the Dirichlet distribution are used for known classification and unknown identification.

ture Pyramid Network (FPN) (Lin et al. 2017) as the base detector. The proposed Evidential Object Detector (EOD) is illustrated in Figure 2.

Customized Evidential Framework

Evidence Formation Evidenital deep learning is essentially a K-way classification system based on the Dirichlet hypothesis, where class probability follows the Dirichlet Prior. It utilizes deep neural network to directly learn the parameters $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_K] \in R_K$ of the Dirichlet posterior distribution $Dir(\mathbf{p}|\boldsymbol{\alpha})$. The class probability of K categories $\boldsymbol{p} = [p_1, \cdots, p_K] \in R_K$ is regarded as sampling from $Dir(\mathbf{p}|\boldsymbol{\alpha})$. It provides a heuristic way to quantify classification uncertainty without using Softmax to model the class probability.

In our proposed method, the traditional Faster R-CNN's Softmax-based classification head is replaced with a novel Evidential Deep Learning (EDL) head. For the plug-and-play purpose, the EDL head shares the same Multilayer Perceptron (MLP) structure as the previous classification head.

Following the theoretical framework in (Sensoy, Kaplan, and Kandemir 2018), for each proposal feature x, we use the neural network to directly learn evidence $\boldsymbol{e} = [e_1, \cdots, e_K] \geq 0$ for each class k. We employ exponent on the neural network output to ensure tractable training and non-negative evidence, which has been demonstrated to be more efficient than using ReLU or Softplus:

$$\boldsymbol{e} = \exp(\mathbf{NN}(x)) \tag{1}$$

The corresponding Dirichlet distribution parameters $\boldsymbol{\alpha}$ and Dirichlet strength S can be calculated as:

$$\boldsymbol{\alpha} = \boldsymbol{e} + 1\,,\; S = \sum_{i=1}^{K} \alpha_i \tag{2}$$

The uncertainty u and expected probability $\hat{p}$ for class k are computed as:

$$u = \frac{K}{S}\,,\; \hat{p}_k = \frac{\alpha_k}{S} \tag{3}$$

By introducing evidential theory and contrastive learning, the loss function in Evidenital Object Detector (EOD) is defined as:

$$\begin{aligned} \mathcal{L}_{total} &= \mathcal{L}_{rpn} + \mathcal{L}_{reg} + \mathcal{L}_{EDL} + \mathcal{L}_{CL} \\ &= \mathcal{L}_{rpn} + \mathcal{L}_{reg} + (\mathcal{L}_{Hybrid} + \gamma_t \mathcal{L}_{wr}) + \mathcal{L}_{CL} \end{aligned} \tag{4}$$

where $\mathcal{L}_{rpn}$ denotes the total loss of RPN; $\mathcal{L}_{reg}$ is the smooth L1 loss for box regression; $\mathcal{L}_{CL}$ stands for contrastive learning loss; $\mathcal{L}_{EDL}$ for evidence formation is composed of the proposed hybrid evidential loss $\mathcal{L}_{Hybrid}$ and weak regularization term $\mathcal{L}_{wr}$; and γ_t serves as annealing coefficient, which is increased gradually with iteration t.

Hybrid Evidential Loss EDL uses Type II Maximum Likelihood or Bayes risk as evidential loss function to maximize the collection of evidence to support accurate classification. For the cross-entropy loss, the Bayes risk $\mathcal{L}_{CE_{br}}$ is defined as:

$$\begin{aligned}\mathcal{L}_{CE_{br}} &= \int \left[\sum_{j=1}^{K} -y_j \log\left(p_j\right)\right] \frac{1}{B\left(\boldsymbol{\alpha}\right)} \prod_{j=1}^{K} p_j^{\alpha_j - 1} d\mathbf{p} \\ &= \sum_{j=1}^{K} y_j \left(\psi\left(S\right) - \psi\left(\alpha_j\right)\right) \end{aligned} \tag{5}$$

where y is the ground truth label; $\psi(\cdot)$ is the *digamma* function; S and $\boldsymbol{\alpha}$ are the Dirichlet strength and parameters.

For the Mean Square Error (MSE), the Bayes risk $\mathcal{L}_{MSE_{br}}$ is defined as:

$$\begin{aligned}\mathcal{L}_{MSE_{br}} &= \int \|\mathbf{y}-\mathbf{p}\|_2^2 \frac{1}{B\left(\boldsymbol{\alpha}\right)} \prod^{K} p_j^{\alpha_j - 1} d\mathbf{p} \\ &= \sum_{j=1}^{K} \left(y_j^2 - 2y_j E\left[p_j\right] + E\left[p_j^2\right]\right) \\ &= \sum_{j=1}^{K} \left(y_j - \hat{p}_j\right)^2 + \frac{\hat{p}_j\left(1-\hat{p}_j\right)}{(S+1)} \end{aligned} \tag{6}$$

where y is the ground truth label; S and $\hat{p}$ are the Dirichlet strength and expected probability.

Due to the increased difficulty in approximating the Dirichlet Prior compared to the categorical distribution, selecting an appropriate loss function becomes crucial in effectively training the model. However, both above losses have deficiencies. $\mathcal{L}_{MSE_{br}}$ is more likely to fall into local minima, and suffers from slow weight update during early training. $\mathcal{L}_{CE_{br}}$ is hard to optimize and exhibits relatively less stable performance in practical applications (Sensoy, Kaplan, and Kandemir 2018).

To solve such problem, we propose a novel Hybrid Evidential Loss $\mathcal{L}_{Hybrid}$, which is formulated as:

$$\begin{aligned}\mathcal{L}_{Hybrid} &= (1-\lambda_t)\mathcal{L}_{CE_{br}} + \lambda_t \mathcal{L}_{MSE_{br}} \\ &= (1-\lambda_t)\sum_{j=1}^{K} y_j\left(\psi\left(S\right) - \psi\left(\alpha_j\right)\right) + \\ &\quad \lambda_t\left(\sum_{j=1}^{K}\left(y_j - \hat{p}_j\right)^2 + \frac{\hat{p}_j\left(1-\hat{p}_j\right)}{(S+1)}\right) \end{aligned} \tag{7}$$

Inspired by the fact that $\mathcal{L}_{CE_{br}}$ tends to generate excessively high expected probability for classes compared with $\mathcal{L}_{MSE_{br}}$ (Sensoy, Kaplan, and Kandemir 2018), we leverage $\mathcal{L}_{CE_{br}}$ to accelerate the early evidence learning and gradually increase the proportion of $\mathcal{L}_{MSE_{br}}$ with annealing coefficient λ_t as a means to prevent overfitting and ensure stable OSOD performance.

Weak Regularization Term EDL introduced a Kullback-Leibler (KL) divergence term $\mathcal{L}_r$ to regularize the predictive distribution by penalizing evidence for the incorrect labels. However, this regularization term often leads to performance degradation in real-world applications. Especially in the detection tasks, the number of *background* proposals far exceeds each foreground class. Penalizing all non-ground-truth labels hinders the learning of foreground objects. To solve the above problem, we introduce a novel weak regularization term $\mathcal{L}_{wr}$, which selectively penalizes excessively high evidence for incorrect labels:

$$\mathcal{L}_{wr} = \beta KL\left[D\left(\mathbf{p} \mid \hat{\boldsymbol{\alpha}}\right) \| D\left(\mathbf{p} \mid \mathbf{1}\right)\right] \tag{8}$$

where $\mathbf{1}$ represents K ones vector; β serves as weighting factor; and $\hat{\boldsymbol{\alpha}}$ can be calculated from the following equations:

$$\tilde{\boldsymbol{\alpha}} = \mathbf{y} + (1-\mathbf{y}) \odot \boldsymbol{\alpha} \tag{9}$$

$$\hat{\boldsymbol{\alpha}} = \mathbf{1}_{\tilde{\boldsymbol{\alpha}}<\tau}(\tilde{\boldsymbol{\alpha}}) + (1 - \mathbf{1}_{\tilde{\boldsymbol{\alpha}}<\tau}(\tilde{\boldsymbol{\alpha}})) \odot \tilde{\boldsymbol{\alpha}} \tag{10}$$

where τ is the pre-defined threshold; $\mathbf{1}_{\tilde{\boldsymbol{\alpha}}<\tau}$ is the indicator function to filter excessive evidence for the incorrect labels with threshold τ.

Further, the KL divergence term can be calculated as:

$$\begin{aligned} KL[D(\mathbf{p} \mid \hat{\boldsymbol{\alpha}}) \| D(\mathbf{p} \mid \mathbf{1})] &= \log\left(\frac{\Gamma\left(\sum_{k=1}^{K} \hat{\alpha}_k\right)}{\Gamma(K)\prod_{k=1}^{K}\Gamma\left(\hat{\alpha}_k\right)}\right) \\ &+ \sum_{k=1}^{K}\left(\hat{\alpha}_k - 1\right)\left[\psi\left(\hat{\alpha}_k\right) - \psi\left(\sum_{j=1}^{K}\hat{\alpha}_j\right)\right] \end{aligned} \tag{11}$$

where $\Gamma(\cdot)$ and $\psi(\cdot)$ refer to the *gamma* function and the *digamma* function, respectively.

Auxiliary Contrastive Learning

Learning more discriminative features for each category would further help in the open set object detection task. Inspired by (Joseph et al. 2021; Han et al. 2022), we model it as a contrastive clustering problem, where the characteristic of inter-class separation and intra-class aggregation is encouraged. Moreover, contrastive learning can also be viewed as an implicit regularization method for eliminating shared evidence, aligning with the same objective of the weak regularization term.

The adopted Auxiliary Contrastive Learning (ACL) module consists of an ACL head (Multilayer Perceptron structured network) and a dynamic memory bank. In each iteration, a batch of proposal feature maps is compressed into low-dimensional embeddings by the ACL head. A queue-based memory bank performs an enqueue/dequeue operation and updates the model based on the following contrastive learning loss $\mathcal{L}_{CL}$ (Han et al. 2022):

$$\mathcal{L}_{CL} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_{CL}\left(\mathbf{z}_i\right) \tag{12}$$

$$\mathcal{L}_{CL}\left(\mathbf{z}_i\right) = \frac{-\nu}{\left|M_{\mathbf{c}_i}\right|}\sum_{\mathbf{z}_j \in M_{\mathbf{c}_i}} \log\frac{\exp\left(\frac{\mathbf{z}_i \cdot \mathbf{z}_j}{\tau}\right)}{\sum_{\mathbf{z}_k \in M \setminus M_{\mathbf{c}_i}} \exp\left(\frac{\mathbf{z}_i \cdot \mathbf{z}_k}{\tau}\right)} \tag{13}$$

where N denotes the batch size, ν and τ serve as weighting factor and temperature hyperparameter, c_i is the class label of i-th proposal, M is the memory bank and $M_{\mathbf{c}_i}$ is the memory queue of class c_i.

By optimizing $\mathcal{L}_{CL}$, the network is trained to maximize the similarity within the same class while minimizing the similarity across different classes.

Unknown Identification Procedure

Compared with other uncertainty-based OSOD methods (Miller et al. 2018) that exclusively model the uncertainty of foreground objects, we observe that the background proposals with higher uncertainty are likely to contain unknown objects. The joint modeling of uncertainty for both foreground objects and background proposals has greatly enhanced the unknown discovery ability.

To avoid redundant computation, we first follow the standard post-processing protocol of the closed set object detector, i.e., Non-Maximum Suppression (NMS), to get the reserved foreground box $\mathbf{B} = \{b_1, \cdots, b_N\}$ and the predicted class $\mathbf{C} = \{c_1, \cdots, c_N\}$. We then calculate the uncertainty $\mathbf{U} = \{u_1, \cdots, u_N, u_1^*, \cdots, u_M^*\}$ for each reserved foreground box and background proposal $\mathbf{P} = \{p_1, \cdots, p_M\}$. Recognizing that object detectors typically assign higher confidence scores to background proposals, we apply distinct uncertainty thresholds for foreground objects and background proposals. Details of the unknown identification procedure are summarized in Algorithm 1.

Algorithm 1: Testing procedure of EOD

Require: $\mathbf{B} = \{b_1, \cdots, b_N\}$; $\mathbf{C} = \{c_1, \cdots, c_N\}$; $\mathbf{P} = \{p_1, \cdots, p_M\}$; $\mathbf{U} = \{u_1, \cdots, u_{N+M}\}$; Uncertainty threshold τ_f, τ_b for foreground objects and background proposals;

for $i = 1, \cdots, N$ **do**
    if $u_i > \tau_f$ **then**
        modify c_i as *unknown*;
    else
        continue;
    end if
end for
for $j = 1, \cdots, M$ **do**
    if $u_j^* > \tau_b$ **then**
        predict p_j as *unknown*;
    else
        continue;
    end if
end for
filter with NMS in the *unknown* class;

Experiments and Results

Datasets

Following the OSOD benchmark in (Han et al. 2022), the $trainval$ set and the $test$ set of **PASCAL VOC** (Everingham et al. 2010) are used for training and closed set evaluation. **VOC-COCO-$\{$20, 40, 60$\}$** and **VOC-COCO-$\{$2500, 5000, 20000$\}$** are used to evaluate the performance under different open set settings.

VOC-COCO-$\{$20, 40, 60$\}$ consist of 5000 **PASCAL VOC** testing images and $\{5000, 10000, 15000\}$ **MS COCO** images containing $\{20, 40, 60\}$ non-VOC classes. This setting is more similar to the real-world applications where objects of non-VOC classes and VOC classes could both appear in the **COCO** images.

VOC-COCO-$\{$2500, 5000, 20000$\}$ consist of 5000 **PASCAL VOC** testing images and $\{2500, 5000, 20000\}$ **MS COCO** (Lin et al. 2014) images disjointing with VOC classes. This setting aims to evaluate the model under extreme open set condition, where only objects of non-VOC classes appear in the **COCO** images.

Evaluation Metrics

1) **Wilderness Impact (WI)** (Dhamija et al. 2020) measures the open set impact on precision.

$$\begin{aligned} \text{WI} &= \frac{\text{Precision in closed set}}{\text{Precision in open set}} - 1 \\ &= \frac{FP_o}{TP_c + FP_c} \end{aligned} \tag{14}$$

where FP_o is the number of false positive open set detections, i.e., unknown objects are misclassified to one of known classes; TP_c and FP_c are the number of true positive and false positive closed set detections.
2) **Absolute Open Set Error (AOSE)** (Miller et al. 2018) measures the total number of unknown objects misclassified into known classes.
3) **Mean Average Precision** of known classes ($\mathbf{mAP}_{\mathcal{K}}$). An ideal open set object detector is expected to identify unknown classes without compromising performance on $\text{mAP}_{\mathcal{K}}$.
4) **Average Precision** of *unknown* class ($\mathbf{AP}_{\boldsymbol{u}}$) (Han et al. 2022) measures the unknown discovery ability.

Implementation Details

For the fair comparison, we follow the settings in (Han et al. 2022). ResNet-50 with Feature Pyramid Network is used as the backbone of the detector. We adopt the default learning rate schedule of Detectron2 (Wu et al. 2019), and use the SGD optimizer with an initial learning rate of 0.02, momentum of 0.9, and weight decay of 0.0001. The max iteration is set to 35000. All models are trained on 8 GPUs with a batch size of 16. Annealing coefficients γ_t, λ_t are set as $\min(1.0, \max(0.0, (t-20000)/10000))$ and $\min(1.0, t/25000)$, where t is the index of the current iteration. Weighting factors ν for contrastive learning is set to 0.4 and β for regularization term is set to 0.05. The uncertainty threshold τ_f for unknown identification is set to 0.02 by default.

State-of-the-art Comparison

In this subsection, we compare EOD with state-of-the-art methods under different open set settings. We reproduce OpenDet with the official code and report results of other methods directly from (Han et al. 2022).

Method	VOC	VOC-COCO-20				VOC-COCO-40				VOC-COCO-60			
	$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{WI}_{\downarrow}$	$\mathrm{AOSE}_{\downarrow}$	$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{AP}_{u\uparrow}$	$\mathrm{WI}_{\downarrow}$	$\mathrm{AOSE}_{\downarrow}$	$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{AP}_{u\uparrow}$	$\mathrm{WI}_{\downarrow}$	$\mathrm{AOSE}_{\downarrow}$	$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{AP}_{u\uparrow}$
Faster R-CNN	**80.10**	18.39	15118	58.45	0.00	22.74	23391	55.26	0.00	18.49	25472	55.83	0.00
Faster R-CNN*	80.01	18.83	11941	57.91	0.00	23.24	18257	54.77	0.00	18.72	19566	55.34	0.00
PROSER	79.68	19.16	13035	57.66	10.92	24.15	19831	54.66	7.62	19.64	21322	55.20	3.25
ORE	79.80	18.18	12811	58.25	2.60	22.40	19752	55.30	1.70	18.35	21415	55.47	0.53
DS	80.04	16.98	12868	58.35	5.13	20.86	19775	55.31	3.39	17.22	21921	55.77	1.25
OpenDet	79.23	11.85	11737	57.72	14.07	14.58	17754	54.50	10.46	12.25	21892	55.18	4.32
EOD (Ours)	79.96	**11.57**	**8989**	**58.97**	**16.55**	**14.05**	**13292**	**55.71**	**11.53**	**12.08**	**16914**	**56.29**	**4.75**

Table 1: Comparison with State-of-the-Art Methods on VOC and VOC-COCO-$\{20, 40, 60\}$ datasets. * means a higher score threshold (i.e., 0.1) for testing.

Method	VOC-COCO-2500				VOC-COCO-5000				VOC-COCO-20000			
	$\mathrm{WI}_{\downarrow}$	$\mathrm{AOSE}_{\downarrow}$	$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{AP}_{u\uparrow}$	$\mathrm{WI}_{\downarrow}$	$\mathrm{AOSE}_{\downarrow}$	$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{AP}_{u\uparrow}$	$\mathrm{WI}_{\downarrow}$	$\mathrm{AOSE}_{\downarrow}$	$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{AP}_{u\uparrow}$
Faster R-CNN	9.25	6015	**77.97**	0.00	16.14	12409	**74.52**	0.00	32.89	48618	63.92	0.00
Faster R-CNN*	9.01	4599	77.66	0.00	16.00	9477	74.17	0.00	33.11	37012	63.80	0.00
PROSER	9.32	5105	77.35	7.48	16.65	10601	73.55	8.88	34.60	41569	63.09	11.15
ORE	8.39	4945	77.84	1.75	15.36	10568	74.34	1.81	32.40	40865	**64.59**	2.14
DS	8.30	4862	77.78	2.89	15.43	10136	73.67	4.11	31.79	39388	63.12	5.64
OpenDet	**6.10**	4233	77.38	8.91	11.10	8735	73.86	12.03	24.28	34111	63.43	16.54
EOD (Ours)	**6.10**	**3547**	77.44	**11.50**	**10.81**	**7357**	74.38	**14.06**	**23.06**	**28653**	64.29	**16.76**

Table 2: Comparison with State-of-the-Art Methods on VOC-COCO-$\{2500, 5000, 20000\}$ datasets.

Table 1 presents the results on the *test* set of **PASCAL VOC** and **VOC-COCO-$\{20, 40, 60\}$**. We first compare the closed set $\mathrm{mAP}_{\mathcal{K}}$ performance of different methods. Although our proposed method is applied to open set condition, its $\mathrm{mAP}_{\mathcal{K}}$ on closed **VOC** *test* set can achieve comparable results to the most advanced closed detector Faster R-CNN (79.96 vs. 80.10). Furthermore, on the more real-world-like **VOC-COCO-$\{20, 40, 60\}$** test sets, EOD achieves significant performance improvements compared with existing open set methods. Taking **VOC-COCO-20** as an example, the WI, AOSE, $\mathrm{mAP}_{\mathcal{K}}$ and AP_u of EOD reach 11.57, 8989, 58.97, 16.55, respectively, which are remarkably superior to the previous best OpenDet. We also note that although EOD deals with the open set problem, it reaches the highest $\mathrm{mAP}_{\mathcal{K}}$ on **VOC-COCO-$\{20, 40, 60\}$** even compared with the closed detector Faster R-CNN by eliminating open set errors. Methods are also evaluated under the extreme wilderness condition. Table 2 presents the results on **VOC-COCO-$\{2500, 5000, 20000\}$**. On the lower openness test sets **VOC-COCO-$\{2500, 5000\}$**, closed detector Faster-RCNN shows the highest $\mathrm{mAP}_{\mathcal{K}}$, reaching 77.97 and 74.52, respectively. In comparison, EOD achieves the second highest $\mathrm{mAP}_{\mathcal{K}}$ on **VOC-COCO-$\{5000, 20000\}$** (74.38 vs. 74.52 and 64.29 vs. 64.59). Furthermore, EOD exhibits the best open set performance, measured by WI, AOSE, and AP_u, outperforming all other methods on **VOC-COCO-$\{2500, 5000, 20000\}$**. Taking **VOC-COCO-5000** as an example, the WI, AOSE and AP_u of EOD reach 10.81, 7357 and 14.06, respectively, which are remarkably superior to the previous best OpenDet. The experimental results demonstrate that even in extreme situations, EOD can also achieve comparable $\mathrm{mAP}_{\mathcal{K}}$ with the state-of-the-art methods while having the optimal open set performance.

Ablation Studies

In this subsection, We perform ablation stuides on **VOC-COCO-20** to analyze each part's contribution of the proposed framework.

Hybrid Evidential Loss. We first conduct ablation experiments on different evidential losses. As shown in Table 3, $\mathcal{L}_{CE_{br}}$ exhibits less stable performance and proves challenging to optimize as mentioned before. It is also found that $\mathcal{L}_{CE_{br}}$ is more sensitive to the random initialization and training process. By comparison, $\mathcal{L}_{MSE_{br}}$ has similar WI (11.36 vs. 11.57) with $\mathcal{L}_{Hybrid}$, but its other open set performance is considerably worse than $\mathcal{L}_{Hybrid}$, e.g., AOSE increases from 8989 to 13064 and AP_u decreases from 16.55 to 15.57. Conversely, $\mathrm{mAP}_{\mathcal{K}}$ of $\mathcal{L}_{Hybrid}$ reach 58.97, which is remarkably superior to $\mathcal{L}_{MSE_{br}}$ ($\mathrm{mAP}_{\mathcal{K}}$=57.88).

Loss	VOC-COCO-20			
	$\mathrm{WI}_{\downarrow}$	$\mathrm{AOSE}_{\downarrow}$	$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{AP}_{u\uparrow}$
$\mathcal{L}_{CE_{br}}$	15.14	14157	48.72	3.64
$\mathcal{L}_{MSE_{br}}$	**11.36**	13064	57.88	15.57
$\mathcal{L}_{Hybrid}$	11.57	**8989**	**58.97**	**16.55**

Table 3: Comparison results on VOC-COCO-20 with different evidential loss functions.

Weak Regularization Term. We then perform ablation studies to explore different regularization strategies. As demonstrated in Table 4, the introduction of the original regularization term $\mathcal{L}_r$ (Sensoy, Kaplan, and Kandemir 2018) results in performance degradation for EOD. Specifically, $\mathrm{mAP}_{\mathcal{K}}$ decreases from 58.45 to 56.90, and AOSE increases from 11072 to 14999. As a comparison, the proposed weak regularization term $\mathcal{L}_{wr}$ further increases $\mathrm{mAP}_{\mathcal{K}}$ and AP_u

Figure 3: Visualization of the detection results with comparison between the baseline (top) and EOD (bottom). Both models are trained on PASCAL VOC dataset and tested on MS COCO dataset. EOD can detect the objects of interest correctly and recognize the unknown objects during testing.

to 58.97 and 16.55, respectively, while reducing AOSE to 8989.

$\mathcal{L}_r$	$\mathcal{L}_{wr}$	VOC-COCO-20			
		WI$_\downarrow$	AOSE$_\downarrow$	mAP$_{\mathcal{K}\uparrow}$	AP$_{u\uparrow}$
✗	✗	12.58	11072	58.45	14.55
✓	✗	13.79	14999	56.90	15.49
✗	✓	**11.57**	**8989**	**58.97**	**16.55**

Table 4: Comparison results on VOC-COCO-20 with different regularization strategies.

Contrastive Learning Module. The effectiveness of the Auxiliary Contrastive Learning (ACL) module is also verified, as presented in Table 5, the performance of all evaluation metrics is significantly improved after adopting the ACL module, e.g., AOSE decreases from 10515 to 8989 and $\mathrm{mAP}_{\mathcal{K}}$ increases from 58.84 to 58.97. The experimental results demonstrate that the ACL module can help EOD to learn better feature representation by implicitly regularizing evidential learning and encouraging class separation in the latent space.

ACL Module	VOC-COCO-20			
	WI$_\downarrow$	AOSE$_\downarrow$	mAP$_{\mathcal{K}\uparrow}$	AP$_{u\uparrow}$
w/o	12.15	10515	58.84	15.97
w/	**11.57**	**8989**	**58.97**	**16.55**

Table 5: Comparison results on VOC-COCO-20 with/without auxiliary contrastive learning module.

Background Uncertainty Modeling. We further analyze the design of background uncertainty modeling in Table 6. The first row exclusively models uncertainty on foreground objects. The second row identifies unknown objects in the *background* class using the same threshold as that for the foreground classes τ_f, while the third row uses a lower background threshold τ_b, which is set to half of τ_f by default. As demonstrated in Table 6, the unknown discovery ability can be further improved by modeling the background uncertainty without affecting WI, AOSE and $\mathrm{mAP}_{\mathcal{K}}$, which only focus on the foreground detection performance. After setting the same background uncertainty threshold as the foreground classes τ_f, the AP_u significantly rises from 12.68 to 14.40. Conversely, since the *background* class tends to have higher evidence than foreground objects, the AP_u can be further increased from 14.40 to 16.55 by using a lower uncertainty threshold τ_b.

τ_f	τ_b	VOC-COCO-20			
		WI$_\downarrow$	AOSE$_\downarrow$	mAP$_{\mathcal{K}\uparrow}$	AP$_{u\uparrow}$
✗	✗	11.57	8989	58.97	12.68
✓	✗	11.57	8989	58.97	14.40
✗	✓	11.57	8989	58.97	**16.55**

Table 6: Comparison results on VOC-COCO-20 with different uncertainty modeling procedures.

Conclusions

In this paper, we propose Evidential Object Detector (EOD) for the challenging Open Set Object Detection task. Our method leverages Evidential Deep Learning theory to approximate the Bayesian Prior of the classification distribution parameters, utilizing a task-specific customized framework to enhance real-world performance. Moreover, the auxiliary contrastive learning module implicitly regularizes the evidence acquisition and promotes class separation in the feature space. By modeling the background uncertainty, we further improve the unknown discovery capability. Extensive experiments on benchmark datasets demonstrate that the proposed method exhibits a substantial improvement over existing ones and achieves new state-of-the-art results. Codes are available at https://github.com/roywang021/EOD.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 62172101), and the Science and Technology Commission of Shanghai Municipality (No. 21511100500), Regional social experimentation with the "Clinical Decision Support System for Paediatric Outpatient Clinics" (No. 21002411800); Auxiliary Diagnosis and Rare Disease Screening System for Children's Pneumonia (No. yg2022-7).

References

Amini, A.; Schwarting, W.; Soleimany, A.; and Rus, D. 2020. Deep evidential regression. *Advances in Neural Information Processing Systems*, 33: 14927–14937.

Bao, W.; Yu, Q.; and Kong, Y. 2021. Evidential deep learning for open set action recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 13349–13358.

Bendale, A.; and Boult, T. E. 2016. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1563–1572.

Denker, J.; Schwartz, D.; Wittner, B.; Solla, S.; Howard, R.; Jackel, L.; and Hopfield, J. 1987. Large automatic learning, rule extraction, and generalization. *Complex systems*, 1(5): 877–922.

Dhamija, A.; Gunther, M.; Ventura, J.; and Boult, T. 2020. The overlooked elephant of object detection: Open set. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1021–1030.

Du, X.; Wang, Z.; Cai, M.; and Li, Y. 2022. VOS: Learning What You Don't Know by Virtual Outlier Synthesis. *arXiv preprint arXiv:2202.01197*.

Everingham, M.; Van Gool, L.; Williams, C. K.; Winn, J.; and Zisserman, A. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2): 303–338.

Gawlikowski, J.; Tassi, C. R. N.; Ali, M.; Lee, J.; Humt, M.; Feng, J.; Kruspe, A.; Triebel, R.; Jung, P.; Roscher, R.; et al. 2021. A survey of uncertainty in deep neural networks. *arXiv preprint arXiv:2107.03342*.

Girshick, R. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 1440–1448.

Han, J.; Ren, Y.; Ding, J.; Pan, X.; Yan, K.; and Xia, G.-S. 2022. Expanding Low-Density Latent Regions for Open-Set Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9591–9600.

Hansen, L. K.; and Salamon, P. 1990. Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, 12(10): 993–1001.

Hüllermeier, E.; and Waegeman, W. 2021. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110(3): 457–506.

Joseph, K.; Khan, S.; Khan, F. S.; and Balasubramanian, V. N. 2021. Towards open world object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5830–5840.

Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*, 740–755. Springer.

Miller, D.; Nicholson, L.; Dayoub, F.; and Sünderhauf, N. 2018. Dropout sampling for robust object detection in open-set conditions. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 3243–3249. IEEE.

Molchanov, D.; Lyzhov, A.; Molchanova, Y.; Ashukha, A.; and Vetrov, D. 2020. Greedy policy search: A simple baseline for learnable test-time augmentation. *arXiv preprint arXiv:2002.09103*.

Nguyen, A.; Yosinski, J.; and Clune, J. 2015. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 427–436.

Rasmus, A.; Berglund, M.; Honkala, M.; Valpola, H.; and Raiko, T. 2015. Semi-supervised learning with ladder networks. *Advances in neural information processing systems*, 28.

Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

Sagi, O.; and Rokach, L. 2018. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4): e1249.

Scheirer, W. J.; de Rezende Rocha, A.; Sapkota, A.; and Boult, T. E. 2012. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7): 1757–1772.

Sensoy, M.; Kaplan, L.; and Kandemir, M. 2018. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31.

Sønderby, C. K.; Raiko, T.; Maaløe, L.; Sønderby, S. K.; and Winther, O. 2016. How to train deep variational autoencoders and probabilistic ladder networks. *arXiv preprint arXiv:1602.02282*, 3(2).

Sun, X.; Yang, Z.; Zhang, C.; Ling, K.-V.; and Peng, G. 2020. Conditional gaussian distribution learning for open set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13480–13489.

Tishby, N.; Levin, E.; and Solla, S. A. 1989. Consistent inference of probabilities in layered networks: Predictions and generalization. In *International Joint Conference on Neural Networks*, volume 2, 403–409. IEEE New York.

Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.-Y.; and Girshick, R. 2019. Detectron2. https://github.com/facebookresearch/detectron2.

Zhou, D.-W.; Ye, H.-J.; and Zhan, D.-C. 2021. Learning placeholders for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4401–4410.