Towards Evidential and Class Separable Open Set Object Detection

Ruofan Wang¹, Rui-Wei Zhao², Xiaobo Zhang³*, Rui Feng¹,²,³,⁴*

¹School of Computer Science, Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai
²Academy for Engineering and Technology, Fudan University, Shanghai
³Children’s Hospital of Fudan University, National Children’s Medical Center, Shanghai, China
⁴Shanghai 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.

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 evidential 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 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.

Customized Evidential Framework

Evidence Formation Evidential 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 $\alpha = [\alpha_1, \cdots, \alpha_K] \in \mathbb{R}_+^K$ of the Dirichlet posterior distribution $\text{Dir}(p|\alpha)$. The class probability of $K$ categories $p = [p_1, \cdots, p_K] \in \mathbb{R}_+^K$ is regarded as sampling from $\text{Dir}(p|\alpha)$. 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 $e = [e_1, \cdots, e_K] \geq 0$ 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:

$$e = \exp(\text{NN}(x)) \quad (1)$$

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

$$\alpha = e + 1, \quad S = \sum_{i=1}^{K} \alpha_i \quad (2)$$

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

$$u = \frac{K}{S}, \quad \hat{p}_k = \frac{\alpha_k}{S} \quad (3)$$

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

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{rpn}} + \mathcal{L}_{\text{reg}} + \mathcal{L}_{\text{EDL}} + \mathcal{L}_{\text{CL}} = \mathcal{L}_{\text{rpn}} + \mathcal{L}_{\text{reg}} + (\mathcal{L}_{\text{Hybrid}} + \gamma_t \mathcal{L}_{\text{wr}}) + \mathcal{L}_{\text{CL}} \quad (4)$$

where $\mathcal{L}_{\text{rpn}}$ denotes the total loss of RPN; $\mathcal{L}_{\text{reg}}$ is the smooth L1 loss for box regression; $\mathcal{L}_{\text{CL}}$ stands for contrastive learning loss; $\mathcal{L}_{\text{EDL}}$ for evidence formation is composed of the proposed hybrid evidential loss $\mathcal{L}_{\text{Hybrid}}$ and weak regularization term $\mathcal{L}_{\text{wr}}$; and $\gamma_t$ serves as annealing coefficient, which is increased gradually with iteration $t$.
defined as:
\[
\mathcal{L}_{CEwr} = \int \left[ \sum_{j=1}^{K} -y_j \log(p_j) \right] \frac{1}{B(\alpha)} \prod_{j=1}^{K} p_j^{\alpha_j - 1} dp \\
= \sum_{j=1}^{K} y_j (\psi(S) - \psi(\alpha_j))
\]
(5)
where \(y\) is the ground truth label; \(\psi(\cdot)\) is the digamma function; \(S\) and \(\alpha\) are the Dirichlet strength and parameters.

For the Mean Square Error (MSE), the Bayes risk \(\mathcal{L}_{MSEbr}\) is defined as:
\[
\mathcal{L}_{MSEbr} = \int \|y - \hat{p}\|^2 \frac{1}{B(\alpha)} \prod_{j=1}^{K} p_j^{\alpha_j - 1} dp \\
= \sum_{j=1}^{K} (y_j^2 - 2y_j E[p_j] + E[p_j^2]) \\
= \sum_{j=1}^{K} (y_j - \hat{p}_j)^2 + \hat{p}_j (1 - \hat{p}_j) \frac{1}{(S + 1)}
\]
(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. \(\mathcal{L}_{MSEbr}\) is more likely to fall into local minima, and suffers from slow weight update during early training. \(\mathcal{L}_{CEwr}\) 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:
\[
\mathcal{L}_{Hybrid} = (1 - \lambda_t)\mathcal{L}_{CEwr} + \lambda_t\mathcal{L}_{MSEbr} \\
= (1 - \lambda_t) \sum_{j=1}^{K} y_j (\psi(S) - \psi(\alpha_j)) + \\
\lambda_t \sum_{j=1}^{K} (y_j - \hat{p}_j)^2 + \hat{p}_j (1 - \hat{p}_j) \frac{1}{(S + 1)}
\]
(7)

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

**Weak Regularization Term** EDL introduced a Kullback-Leibler (KL) divergence term \(\mathcal{L}_{KL}\) 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:
\[
\mathcal{L}_{wr} = \beta KL \left[ D (p \mid \hat{\alpha}) \| D (p \mid 1) \right]
\]
(8)
where 1 represents K ones vector; \(\beta\) serves as weighting factor; and \(\hat{\alpha}\) can be calculated from the following equations:
\[
\hat{\alpha} = y + (1 - y) \odot \alpha
\]
(9)
\[
\hat{\alpha} = \mathbf{1}_{\alpha < \tau} (\hat{\alpha}) + (1 - \mathbf{1}_{\alpha < \tau}) \odot \hat{\alpha}
\]
(10)
where \(\tau\) is the pre-defined threshold; \(\mathbf{1}_{\alpha < \tau}\) is the indicator function to filter excessive evidence for the incorrect labels with threshold \(\tau\).

Further, the KL divergence term can be calculated as:
\[
KL[D(p \mid \hat{\alpha})\|D(p \mid 1)] = \log \left( \frac{\Gamma \left( \sum_{k=1}^{K} \hat{\alpha}_k \right)}{\Gamma(K) \prod_{k=1}^{K} \Gamma(\hat{\alpha}_k)} \right) \\
+ \sum_{k=1}^{K} (\hat{\alpha}_k - 1) \left[ \psi(\hat{\alpha}_k) - \psi \left( \sum_{j=1}^{K} \hat{\alpha}_j \right) \right]
\]
(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):
\[
\mathcal{L}_{CL} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_{CL} (z_i)
\]
(12)
\[
\mathcal{L}_{CL} (z_i) = -\frac{\nu}{|M_{cl}|} \sum_{z_j \in M_{cl}} \log \frac{\exp \left( \frac{z_i \cdot z_j}{\tau} \right)}{\sum_{z_k \in M_{cl}} \exp \left( \frac{z_i \cdot z_k}{\tau} \right)}
\]
(13)
where \(N\) denotes the batch size, \(\nu\) and \(\tau\) serve as weighting factor and temperature hyperparameter, \(c_i\) is the class
label of i-th proposal, \( M \) is the memory bank and \( M_c \) is the memory queue of class \( c_i \).

By optimizing \( L_{CL} \), 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 \( B = \{b_1, \cdots , b_N \} \) and the predicted class \( C = \{c_1, \cdots , c_N \} \). We then calculate the uncertainty \( U = \{u_1, \cdots , u_N, u_1^*, \cdots , u_M^* \} \) for each reserved foreground box and background proposal \( P = \{p_1, \cdots , p_M \} \).

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](image)

**Experiments and Results**

**Datasets**

Following the OSOD benchmark in (Han et al. 2022), the \textit{trainval} set and the \textit{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.

\[
WI = \frac{\text{Precision in closed set}}{\text{Precision in open set}} - 1
\]

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 (m\(\text{AP}_K\)). An ideal open set object detector is expected to identify unknown classes without compromising performance on m\(\text{AP}_K\).

4) **Average Precision** of unknown class (AP\(_u\)) (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 \( \gamma_t, \lambda_t \) 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 \( \nu \) for contrastive learning is set to 0.4 and \( \beta \) for regularization term is set to 0.05. The uncertainty threshold \( \tau_f \) 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).
Table 1 presents the results on the test set of PASCAL VOC and VOC-COCO-\{20, 40, 60\}. We first compare the closed set \( mAP_K \) performance of different methods. Although our proposed method is applied to open set condition, its \( mAP_K \) 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, \( mAP_K \), 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 \( mAP_K \) 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 \( mAP_K \), reaching 77.97 and 74.52, respectively. In comparison, EOD achieves the second highest \( mAP_K \) 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 \( mAP_K \) with the state-of-the-art methods while having the optimal open set performance.

### Ablation Studies

In this subsection, We perform ablation studies 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, \( \mathcal{L}_{CEbr} \) exhibits less stable performance and proves challenging to optimize as mentioned before. It is also found that \( \mathcal{L}_{CEbr} \) is more sensitive to the random initialization and training process. By comparison, \( \mathcal{L}_{MSEbr} \) 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, \( mAP_K \) of \( \mathcal{L}_{Hybrid} \) reach 58.97, which is remarkably superior to \( \mathcal{L}_{MSEbr} \) (mAP\(_K\)=57.88).

#### 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, \( mAP_K \) 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 \( mAP_K \) and \( AP_u \).
to 58.97 and 16.55, respectively, while reducing AOSE to 8989.

<table>
<thead>
<tr>
<th>( \mathcal{L}_{\tau} )</th>
<th>( \mathcal{L}_{\tau \text{wr}} )</th>
<th>VOC-COCO-20</th>
<th>WI</th>
<th>AOSE</th>
<th>mAP(_K)</th>
<th>AP(_u)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>12.58</td>
<td>11072</td>
<td>58.45</td>
<td>14.55</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>13.79</td>
<td>14999</td>
<td>56.90</td>
<td>15.49</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>11.57</td>
<td>8989</td>
<td>58.97</td>
<td>16.55</td>
<td></td>
</tr>
</tbody>
</table>

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 mAP\(_K\) 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.

<table>
<thead>
<tr>
<th>ACL Module</th>
<th>VOC-COCO-20</th>
<th>WI</th>
<th>AOSE</th>
<th>mAP(_K)</th>
<th>AP(_u)</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o</td>
<td>12.15</td>
<td>10515</td>
<td>58.84</td>
<td>15.97</td>
<td></td>
</tr>
<tr>
<td>w/</td>
<td>11.57</td>
<td>8989</td>
<td>58.97</td>
<td>16.55</td>
<td></td>
</tr>
</tbody>
</table>

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 \( \tau_f \), while the third row uses a lower background threshold \( \tau_b \), which is set to half of \( \tau_f \) 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 mAP\(_K\), which only focus on the foreground detection performance. After setting the same background uncertainty threshold as the foreground classes \( \tau_f \), 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 \( \tau_b \).

<table>
<thead>
<tr>
<th>( \tau_f )</th>
<th>( \tau_b )</th>
<th>VOC-COCO-20</th>
<th>WI</th>
<th>AOSE</th>
<th>mAP(_K)</th>
<th>AP(_u)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>11.57</td>
<td>8989</td>
<td>58.97</td>
<td>12.68</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>11.57</td>
<td>8989</td>
<td>58.97</td>
<td>14.40</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>11.57</td>
<td>8989</td>
<td>58.97</td>
<td>16.55</td>
<td></td>
</tr>
</tbody>
</table>

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


Tishby, N.; Levin, E.; and Solla, S. A. 1989. Consistent inference of probabilities in layered networks: Predictions and
