Meta Evidential Transformer for Few-Shot Open-Set Recognition

Hitesh Sapkota [1 2] **Krishna Prasad Neupane** [1 2] **Qi Yu** [2]

Abstract

Few-shot open-set recognition (FSOSR) aims to detect instances from unseen classes by utilizing a small set of labeled instances from closed-set classes. Accurately rejecting instances from open-set classes in the few-shot setting is fundamentally more challenging due to the weaker supervised signals resulting from fewer labels. Transformer-based few-shot methods exploit attention mapping to achieve a consistent representation. However, the softmax-generated attention map normalizes all the instances that assign unnecessary high attentive weights to those instances not close to the closed-set classes that negatively impact the detection performance. In addition, open-set samples that are similar to a certain closed-set class also pose a significant challenge to most existing FSOSR models. To address these challenges, we propose a novel Meta Evidential Transformer (MET) based FSOSR model that uses an evidential open-set loss to learn more compact closed-set class representations by effectively leveraging similar closed-set classes. MET further integrates an evidence-to-variance ratio to detect fundamentally challenging tasks and uses an evidence-guided cross-attention mechanism to better separate the difficult open-set samples. Experiments on real-world datasets demonstrate consistent improvement over existing competitive methods in unseen class recognition without deteriorating closed-set performance.

1. Introduction

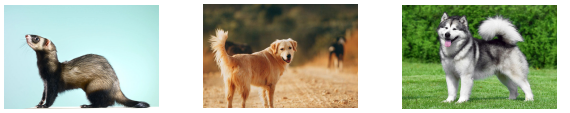
Various learning strategies have been explored to reduce label dependency, including semi-supervised learning (Oliver et al., 2018; Chapelle et al., 2006) and weakly supervised learning (Ilse et al., 2018; Sapkota et al., 2021). Few shot learning (FSL) offers another promising direction by assuming that only limited labeled data samples are available for model training (Jeong et al., 2021). Once trained, the model is expected to perform well on unseen data samples. While existing FSL models achieve promising results, most of them primarily focus on the closed-set setting, where both the training and test samples are assumed to be from the same data distribution over a common set of classes (Luo et al., 2019). Nevertheless, when deployed in a practical setting, the model may very likely be exposed to samples from unknown classes, which are not part of the training distribution. In this case, it is ideal that the model can detect these samples as unknown.

The open-set recognition (OSR) problem has been studied in the general setting with ample training data samples (Bendale & Boult, 2016; Ge et al., 2017; Yoshihashi et al., 2019; Sun et al., 2020). However, the few-shot setting poses unique challenges, making existing solutions inadequate. There have been few attempts to address few-shot open-set recognition (FSOSR). For example, PEELER is designed to learn a high-entropy posterior distribution for samples from the open-set classes (Liu et al., 2020). SnaTCHer further improves PEELER by leveraging transformer consistency (Jeong et al., 2021). It considers a set as a whole that includes all the prototypes of closed-set classes to detect the open-set ones. Because of the attention-based transformation of the entire set, this approach can provide a compact representation for the entire closed-set classes, leading to improved detection performance. However, when facing more challenging scenarios, where open-set classes share some similarities with closed-set ones, existing techniques become less effective.

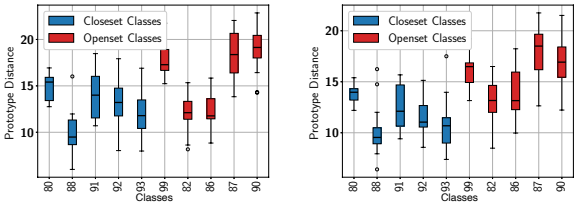
As shown in Figure 1 (a), golden retriever (Class ID: 82) shares some feature similarities (*e.g.,* body structure, whiskers, and tail) with the closed-set class ferrets (Class ID: 88). As such, when a golden retriever is evaluated, it may be predicted as a ferrets. In this case, the distance between the altered prototype (where the class ferrets prototype is replaced by a golden retriever sample) and the original prototype will be very small, resulting in the mis-classification of an open-set sample as a closed-set one with high confidence. As illustrated in Figure 1 (b), the mean prototype distance from this open-set class (*i.e.,* 82) is smaller than most other closed-set classes (*e.g.,* 80, 91, 92), leading

[1]Amazon Inc. (Work was done at RIT, which is not relate to the position at Amazon) [2]Rochester Institute of Technology (RIT). Correspondence to: Qi Yu <qi.yu@rit.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.



(a) Open-set sample (golden retriever) shares similar features with closed-set samples including ferrets and malamute.



(b) SnaTCHer (c) MET

Figure 1. OSR performance (AUROC) of a difficult task consisting of similar closed- and open-set images with examples shown in (a). (b) SnaTCHer (72.84%) and (c) MET (83.34%) that uses Class mamalute (83) serving as the opponent class for better separation between Ferrets (closed) and Golden Retriever (open).

to a relatively low detection rate with a 72.84% AUROC score. It is noted in the figure, prototype distance is the distance that tells how far the original prototype is from the altered prototype, where the closest class prototype is replaced by a test sample (see Eq. (9) for a definition).

Similar cases as described above may commonly occur in an open world. This makes it inherently challenging to detect open-set classes similar to certain closed-set ones but with subtle and important differences (*e.g.,* golden retriever and ferrets as described above). Since recognizing open-set samples that are very different from their closed-set counterparts is relatively trivial with promising results achieved by existing methods, we will focus on attacking the more challenging cases. Furthermore, due to the limited control over the open-set samples, the goal is to *learn a more compact representation of closed-set classes*. To this end, we propose a novel **M**eta **E**vidential **T**ransformer (MET) that integrates uniquely designed training and inference modules to address the central challenges in hard FSOSR problems.

During training, MET leverages the power of similar closed-set classes playing a role as open-set samples (referred to as opponent classes) for improved model training. MET assigns a high uncertainty to the *opponent classes* that serve as training-time open-set samples. This will help the model make (relatively) more confident predictions on the closed-set samples while being uncertain of unseen open-set samples that may share similar features as the opponent classes. To achieve this, a straightforward way would be enforcing the model to produce a high uncertainty on the opponent class samples through the entropy maximization technique (Liu et al., 2020). However, a high entropy cannot tell whether a sample is close to multiple closed-set classes or far away from all of them (Shi et al., 2020), where the former corresponds to a confusing closed-set sample and the latter is a true open-set one. To address this issue, we propose to integrate evidential learning (Sensoy et al., 2018b), which allows us to design an evidence-based loss function to guide model training. Intuitively, a data sample with a small sum of evidence from all closed-set classes is more likely to be from the open-set while one with strong conflicting evidence from multiple classes should be a confusing closed-set sample. Figure 1 (b) shows an improved OSR performance by MET as compared with SnaTCHer.

While the use of a transformer coupled with the special training process allows us to improve the overall compactness of entire closed-set samples, one challenging issue still remains when a certain class is very different from others in the closed-set. Given an open-set sample that is relatively similar to this special closed-set class, it will also be very different from other classes (and their prototypes). Due to the normalization effect during transformation, this data sample will likely be assigned to the special closed-set class. We propose a novel *evidence-to-variance ratio (EVR)* to identify such cases during inference time. The inference module then conducts evidence-guided cross-attention in the transformer to improve detection performance with theoretical guarantees. Our main contribution is threefold:

- an MET model that uses an evidential open-set loss to learn more compact closed-set representations by leveraging similar closed-set classes as opponent open-set classes,
- a novel evidence-to-variance ratio (EVR) to identify challenging open-set samples by collectively considering both the predicted evidence and their distribution over all closed-set classes,
- a uniquely designed evidence-based cross-attention mechanism to form a more accurate representation of the prototypes for improved OSR performance,

We conduct extensive experiments on real-world datasets and the results show that MET achieves SOTA OSR performance as compared with the competitive baselines.

2. Related Work

We discuss representative works most relevant to ours. Additional related works are presented in the appendix.

Few-shot Open-set Recognition. There are recent OSR models specifically developed for few-shot learning under the meta-learning setting. Liu et al. propose an oPen set mEta LEaRning (PEELER) model that leverages ProtoNet for few shot open-set recognition (Liu et al., 2020), which makes an assumption that the unknown samples are available during the training process. The key limitation of this approach is the learned embedding representation is not

task adaptive and the open-set detection process heavily depends on the used open-set samples during the training process. To address those limitations, Jeong et al. propose the SnaTCHer model based on FEAT (Ye et al., 2020), which makes the embedding task specific by leveraging different transformer functions (Jeong et al., 2021). While the training paradigm is very similar to FEAT (not requiring unknown samples), SnaTCHer proposes a unique process to detect unknown samples during testing by leveraging the transformed set of prototypes to represent all closed-set classes. However, SnaTCHer may suffer from more challenging open-set samples and the normalization effect may miss detecting open-set samples with a strong confidence. Similarly, Huang et al. leverage task-adaptive negative class prototypes to learn dynamic rejection boundaries for FSOSR tasks (Huang et al., 2022). However, learning from negative samples generated from closed-set prototypes may not help to deal with challenging open-set samples. In their recent work, Boudiaf et al. (Boudiaf et al., 2023) introduce the Open-Set Likelihood Optimization (OSLO) technique to address the FSOSR task in a transductive setting. OSLO requires that all unlabeled query samples from the test set are available altogether, which may affect its applicability in practice. Additionally, the OSLO technique does not incorporate the transformer architecture to achieve a more concise prototype representation, which may result in the model struggling to identify challenging open-set samples. Wang et al. also propose the Glocal framework to tackle FSOSR that consists of two branches: a closed-set classification branch aimed at improving closed-set accuracy, and an energy-based open-set recognition branch to enhance FSOSR (Wang et al., 2023). The classification branch utilizes class-wise similarity between query samples and prototypes, while the open-set recognition model considers both pixel-wise and class-wise similarity between query samples and prototypes. Unlike our technique, the Glocal approach does not explicitly handle challenging open-set samples.

Uncertainty-aware Open-set Recognition. Multiple approaches have been developed that explicitly consider uncertainty during model training (Sensoy et al., 2018a; Malinin & Gales, 2018; Charpentier et al., 2020). For instance, Sensoy et al. propose an evidential deep learning (EDL) model that leverages the subjective logic principle to learn the evidence and uncertainty explicitly based on the training data samples (Sensoy et al., 2018a). Similarly, Malinin et al. propose a Prior Network (PN) that uses an explicit mechanism to quantify the distributional uncertainty coming from the distributional mismatch (Malinin & Gales, 2018). However, this approach requires unknown-class samples during the training time and therefore limits its applicability in practical settings. Considering this limitation, Charpentier et al. propose the posterior network that leverages the normalizing flows to estimate the density in the latent space in order to predict the posterior distribution based upon the in-distribution samples (Charpentier et al., 2020). The proposed MET model extends these approaches to the few-shot setting through seamless and novel integration with a transformer architecture for effective open-set recognition.

3. Preliminaries

3.1. Meta Learning

A meta-learner learns a learning algorithm by exploiting a pool of learning tasks. Meta learning splits data into two sets: *meta-train* and *meta-test* consisting of distinct training and test classes. Meta-train $\mathcal{MS}$=$\{(\mathcal{S}_i^{tr}, \mathcal{Q}_i^{tr})\}_{i=1}^{N^{tr}}$ includes support ($\mathcal{S}_i^{tr}$) and query ($\mathcal{Q}_i^{tr}$) sets for the i^{th} task and N^{tr} is the number of training tasks. Similarly, meta-test $\mathcal{MT} = \{(\mathcal{S}_i^{te}, \mathcal{Q}_i^{te})\}_{i=1}^{N^{te}}$ includes support ($\mathcal{S}_i^{te}$) and query ($\mathcal{Q}_i^{te}$) sets for the i^{th} task and N^{te} is a number of test tasks. Meta-learning performs training by minimizing the error of label prediction for the query set $\mathcal{Q}^{tr}$ conditioned on the support set $\mathcal{S}^{tr}$. Specifically, the meta-training objective is

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \sum_{(\mathbf{x}_j, y_j) \in T_i^{tr}} \mathcal{L}(y_j, P_{\boldsymbol{\theta}}(\cdot|\mathbf{x}_j, S_i^{tr})) \quad (1)$$

where $\mathcal{L}$ is a loss function (*e.g.,* cross-entropy or mean-square error), which is suitable for the optimization procedure, and $P_{\boldsymbol{\theta}}(\cdot)$ is a parametric neural network or other models to make predictions. Meta-learning is a popular approach for few-shot learning. It forms support and query sets by sampling N-classes from the pool of classes with few training samples (*e.g.,* K-shot examples per class) commonly referred to as a N-way K-shot problem.

3.2. Evidential Learning

Theory of evidence and subjective logic (SL) (Dempster, 1968; Jøsang, 2016) are utilized to address inexact and expensive posterior inference of Bayesian and Monte-Carlo approximation. It also provides predictive uncertainty, including both aleatoric and epistemic uncertainty. In particular, *evidence* provides a measure of the number of supportive observations from data for each class and let e_k denote the evidence for a class k. Then, the Dirichlet concentration parameter α_k for each class $k \in \mathbb{Y}$ can be calculated as: $\alpha_k = e_k + a_k W$, where $e_k \geq 0$. The belief mass and uncertainty mass (*a.k.a.,* vacuity) is computed as:

$$b_k = \frac{e_k}{S}, \quad u = \frac{K}{S} \text{ with } S = \sum_{k=1}^{K}(e_k + 1) \quad (2)$$

Evidential learning essentially places a Dirichlet prior $\text{Dir}(p_i|\alpha_i)$ on a multinomial likelihood $\text{Mult}(y_i|p_i)$ and then uses the negative log-likelihood to train the model:

$$\mathcal{L}_{\text{EDL}} = \sum_{k=1}^{K} y_{ik} \left(\log S_i - \log \alpha_{ik}\right) \quad (3)$$

where y_{ik} is an one-hot encoding of ground truth label y_i of a data sample $\mathbf{x}_i$, α_{ik} is a corresponding Dirichlet parameter and S_i is the total Dirichlet strength.

4. Methodology

4.1. Transformer based FSOSR

Transformers leverage the similarity among the closed-set classes through the attention mechanism, which results in a more compact representation of the entire closed-set classes. As such, the open-set sample representation can stay away from all of the closed-set class representations, improving the openset detection capability. Let $F(\cdot)$ be the feature extractor and we can define the class-representation (*i.e.,* prototype) of closed-set class n as follow:

$$\mathbf{p}_n = \frac{1}{K} \sum_{\mathbf{x} \in \text{ class } n} F(\mathbf{x}; \boldsymbol{\theta}_f) \tag{4}$$

where K is the total number of samples belonging to class n in the support set, $\boldsymbol{\theta}_f$ denotes the parameters associated with feature extractor F, $\mathbf{x}$ represents a data sample belonging to class n, and N is the total number of closed-set classes for a given task. The overall prototype representation can then be formed as a concatenation of N closed-set class prototypes:

$$\mathcal{P} = \{\mathbf{p}_n\}_{n=1}^{N} \tag{5}$$

The above prototype representation does not leverage the similarity among closed-set classes. For a more compact representation, we can transform the prototype using the transformation function $\mathtt{T}(\cdot)$. Specifically, we transform the prototype ($\mathcal{P}$) in the form of a triplet $[key\ (\mathcal{K}), value\ (\mathcal{V}), query\ (\mathcal{Q})]$ with trainable transformer weight matrices and then the transformed prototype is achieved by

$$\begin{aligned}\mathcal{P}' = \mathtt{T}(\mathcal{P}; \boldsymbol{\theta}_t) = &\mathtt{LayerNorm}(\mathcal{P} + \frac{1}{N}(W_{\mathcal{V}}\mathcal{P})) \\ &\left[\mathtt{softmax}\left(\frac{(W_{\mathcal{K}}\mathcal{P})^\top (W_{\mathcal{Q}}\mathcal{P})}{\sqrt{d}}\right)^\top\right]\end{aligned} \tag{6}$$

where $\boldsymbol{\theta}_t = \{W_{\mathcal{K}}, W_{\mathcal{V}}, W_{\mathcal{Q}}\} \in \mathbb{R}^{d \times d}$ are learnable transformation matrices and d is the feature dimension.

During the training process, we aim to minimize the distance between a closed-set query sample feature representation with its corresponding transformed prototype class while maximizing with the rest. To achieve this, we can leverage cross-entropy loss with the following inverse of distance as the logits for that loss.

$$\begin{aligned} o_{jn} &= [\mathrm{d}(F(\mathbf{x}_j), \mathbf{p}'_n)]^{-1} \\ &= \left[\sqrt{(F(\mathbf{x}_j) - \mathbf{p}'_n)^\top (F(\mathbf{x}_j) - \mathbf{p}'_n)}\right]^{-1} \end{aligned} \tag{7}$$

where $\mathbf{x}_j$ is the j^{th} query sample, $\mathbf{p}'_n$ is the transformed prototype of class n, and $d(\cdot, \cdot)$ is the Euclidean distance.

During the inference process, for open-set detection, one straightforward process would be computing the distance using (7) and deciding the sample as open-set or closed-set based on its value. Compared to this, SnaTCHer considers a set as a whole that includes all the prototypes of the closed-set classes to detect the open-set samples which results in better open-set detection performance. Specifically, let $\mathbf{x}_j$ be a query sample, and c be the closest closed-set class with this sample, then we alter the prototype in (5) as

$$\mathcal{P}_a = \mathcal{P} - \{\mathbf{p}_c\} + F(\mathbf{x}_j) \tag{8}$$

Next, the altered prototype is passed through the transformer using (6). This yields the transformed representation of altered prototype represented as $\mathcal{P}'_a$. Finally, we compute the distance between transformed prototype and the altered transformed prototype which is given as

$$\delta(\mathbf{x}_j) = d(\mathcal{P}'_a, \mathcal{P}') \tag{9}$$

For an open-set sample, the transformed $\mathcal{P}'_a$ is expected to be very different from the original $\mathcal{P}'$ which has a compact representation, leading to an improved OSR performance.

Remarks. As described in the introduction, in case of more challenging scenarios where open-set classes share some similarities with closed-set classes, the existing techniques like SnaTCHer become less effective. Because of the feature similarity between the open-set class samples with one of the closed-set class samples, $\mathcal{P}'_a$ can become similar to $\mathcal{P}'$, which compromises the open-set detection ability.

4.2. Meta Evidential Transformer (MET)

MET is designed to attack the most challenging few-shot open-set detection tasks that the state-of-the-art FSOSR techniques are less effective to handle. It integrates uniquely designed training and inference modules to achieve significantly improved OSR performance in these challenging settings. Specifically, training of MET is guided by a novel *evidential open-set loss* that learns a more compact closed-set representation by leveraging similar closed-set classes playing the role as open-set classes (referred to as opponent classes). As a result, open-set samples can be more easily separated from the closed-set ones falling into this compact representation. Another difficulty arises when the closed-set involves classes that are very different from all other closed-set classes. In this case, learning a compact representation that covers all the closed-set classes becomes challenging due to the large difference within the set. We propose a novel evidence-to-variance ratio (EVR) to identify such cases during the inference time using the predicted evidence by the trained evidential transformer. The inference module then conducts evidential cross-attention in the transformer to improve the detection performance

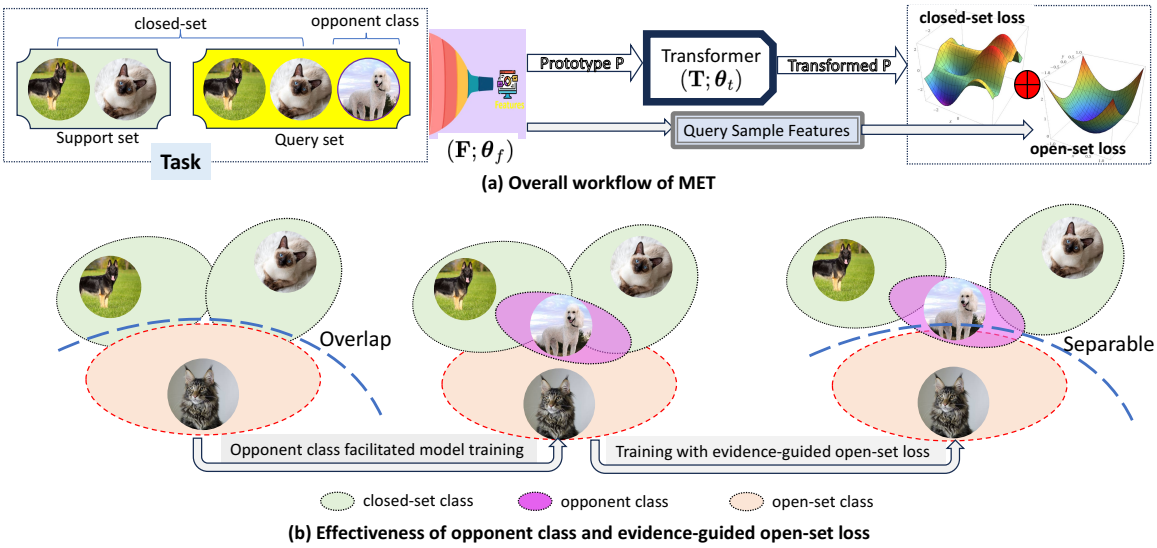


Figure 2. Overview of Meta Evidential Transformer: (a) training pipeline and (b) effectiveness of evidence-guided loss-set loss

MET **Training via Evidence-Guided Open-Set Loss.** We first construct a meta-training ($\mathcal{MS}$) set consisting of only training classes so there is no overlap with samples from meta-test ($\mathcal{MT}$) classes. Furthermore, we choose a set of opponent classes from the existing known closed-set classes to serve as open-set classes, aiming to learn a more compact representation of the known classes. Figure 2 (b) and (c) provides an illustrative example. We develop a unique mechanism to select opponent classes that are similar to some other closed-set classes. Specifically, within a training set, we perform semantic analysis at the class level to identify groups of semantically relevant classes (*e.g.,* different categories of dogs). For datasets with a relatively small number of classes, this introduces minimal overhead. For larger datasets, we can usually benefit from some existing hierarchical structure among the classes. If a hierarchical structure is unavailable, similar classes can be identified based on their semantic similarity. A neural network trained on a training dataset with a cross-entropy loss can be used to form a feature representation for each sample that could be used to pick similar classes.

Evidence Guided Evidential Loss. The overall training pipeline of MET is shown in Figure 2 (a) and the respective training algorithm is presented in the Appendix. We proceed to conduct episodic training, where the test procedure mimics the training procedure. In training, we sample $(K+M)$ instances from N closed-set classes and form a support set (S_i^{tr}) utilizing K instances from each closed-set classes and a query set (Q_i^{tr}) with the M closed-set samples as well as O samples from open-set classes (*i.e.,* the chosen opponent classes) for the i^{th} training task T_i^{tr}=(S_i^{tr},Q_i^{tr}). In this way, MET has a similar procedure as standard meta-learning and we propose the following learning process:

$$\boldsymbol{\theta}^* = \arg\min_{\theta} \Bigg\{ \sum_{(\mathbf{x}_j,y_j)\in T_i^{tr} | y_j \in C^s} \mathcal{L}_{close}\left(y_j, P_{\boldsymbol{\theta}}(.|\mathbf{x}_j, S_i^{tr})\right) + \lambda \sum_{(\mathbf{x}_j,y_j)\in T_i^{tr} | y_j \in C^u} \mathcal{L}_{open}\left(P_{\boldsymbol{\theta}}(.|\mathbf{x}_j, S_i^{tr})\right) \Bigg\} \tag{10}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_f, \boldsymbol{\theta}_t\}$ indicates total parameters consisting of both feature extractor parameter $\boldsymbol{\theta}_f$ and transformer parameter $\boldsymbol{\theta}_t$. $\mathcal{L}_{close}$ is a closed-set loss and suitable loss functions include cross-entropy and mean-square error. However, our method is based on evidential learning so it leverages the evidential loss given in (3). Similarly, $\mathcal{L}_{open}$ is an open-set loss applied to the open-set classes introduced into the training process and we will discuss our novel approach next. Also, C^s and C^u are the sets of closed-set classes and open-set classes of few-shot training task $\mathcal{T}_i^{tr}$.

During the meta-update, the model uses the query set, which includes samples from those opponent classes chosen from the closed-set playing the role of challenging open-set classes. Our goal is to shrink the total evidence towards zero for these samples that effectively learn a more compact representation for the closed-set classes. To this end, we utilize KL-divergence between the predictive distribution on these open-set classes and a uniform distribution that indicates a maximum uncertainty mass (*i.e.,* $u=1$):

$$\mathcal{L}_{open}(\cdot) = \sum_{j|y_j\in C^u} KL[\text{Dir}(\mathbf{p}_j|\boldsymbol{\alpha}_j)||\text{Dir}(\mathbf{p}_j|(1,...,1)^\top)] \tag{11}$$

where $\mathbf{p}_j$ represents the class probabilities of sample $\mathbf{x}_j$,

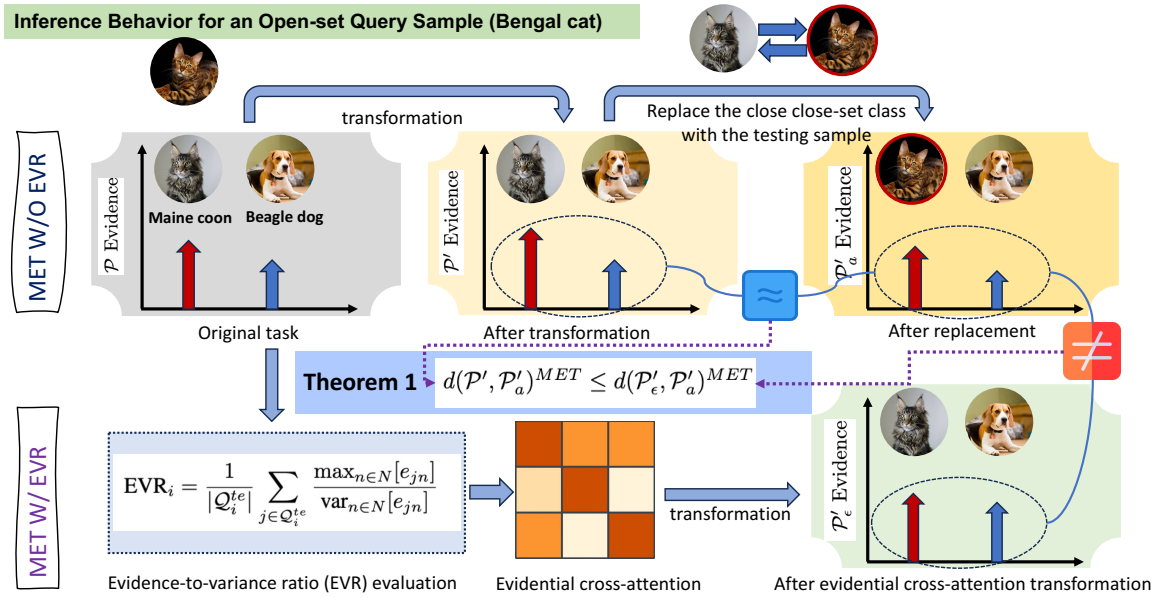


Figure 3. A model trained using evidential loss fails to identify the Bengal Cat as open-set because it shares feature similarity with Maine Coon and is very distinct from the Dog class. In contrast, EVR and evidential cross-attention help to recognize the open-set sample.

Dir represents Dirichlet distribution, $\boldsymbol{\alpha}_j$ is the Dirichlet parameter given as follow

$$\alpha_{jn} = \{e_{jn} + 1\}, \quad e_{jn} = o_{jn} \tag{12}$$

where o_{jn} is defined in (7), which is non-negative.

Evidential Cross-Attention. When the closed-set involves classes that are inherently different from other classes, learning a compact closed-set representation is more difficult. Using a loose representation may cause trouble during inference especially when evaluating a test open-set sample that is similar to one of the closed-set classes. Figure 3 provides an illustrative example on this scenario. Due to the similarity between the open-set Bengal Cat and the original closed-set Maine Coon, after the latter is replaced by the former using (8), the transformed representation of the altered prototype ($\mathcal{P}'_a$) and that of the original prototype ($\mathcal{P}'$) remains similar to each other, resulting in a low distance $d(\mathcal{P}', \mathcal{P}'_a)$. As a result, the model fails to recognize the open-set Bengal Cat. To improve the detection performance, the proposed inference module leverages the predicted evidence by MET to detect the challenging FSOSR tasks using a uniquely designed *Evidence-to-Variance Ration (EVR)* metric. Once detected, it further performs *evidential cross-attention*, leading to a transformed representation ($\mathcal{P}'_\epsilon$) of the original prototypes that is more compact than $\mathcal{P}'$ as shown in Figure 3. Thus, the distance $d(\mathcal{P}'_\epsilon, \mathcal{P}'_a)$ becomes much larger, which results in a successful detection of the open-set Bengal Cat.

Let e_{jn} denote the predicted evidence on class n for the j-th sample in the query set of a meta-test task, *i.e.*, $j \in \mathcal{Q}_i^{te}$. The EVR metric is designed based on the following two properties of the predicted evidence:

- $\mathbf{P}_1$: If j is an open-set sample, $\max_n[e_{jn}]$ is not high (since the model has not learned from the same class); if j is a closed-set sample, $\max_n[e_{jn}]$ is high.
- $\mathbf{P}_2$: For a challenging FSOSR task, if j is an open-set sample similar to some closed-set class n', $\text{var}_{n \in N}[e_{jn}]$ is high (because a very low evidence for all other classes while a relative higher evidence for n'); if j is a closed-set sample, $\text{var}_{n \in N}[e_{jn}]$ is even higher (because a high evidence to the true closed-set class).

Guided by these properties, EVR is defined as

$$\text{EVR}_i = \frac{1}{|\mathcal{Q}_i^{te}|} \sum_{j \in \mathcal{Q}_i^{te}} \frac{\max_{n \in N}[e_{jn}]}{\text{var}_{n \in N}[e_{jn}]} \tag{13}$$

Remark. Consider a properly trained MET model that can predict evidence satisfying the two properties ($\mathbf{P}_1$, $\mathbf{P}_2$). Given two FSOSR test tasks a and b, with a being a challenging task having a loose transformed closed-set representation $\mathcal{P}'$, and b being a regular task, we have $\text{EVR}_a < \text{EVR}_b$. Leveraging this key result, we propose an evidential cross-attention mechanism to improve the OSR performance. Let c be the class nearest to the given query point $\mathcal{Q}_{ij}^{te}$, and $A_i \in \mathbb{R}^{N \times N}$ be the attention matrix obtained from the transformer network then, we update the attention

$$A_i[c_1, c_2] = \left\{ \begin{array}{ll} A_i[c_1, c_2] \times \frac{\epsilon}{\text{EVR}_i} & \text{if } cond == true \\ A_i[c_1, c_2] & \text{else} \end{array} \right\}$$
$$cond = \{(c_1 == c || c_2 == c) \;\&\; c_1 \neq c_2\} \tag{14}$$

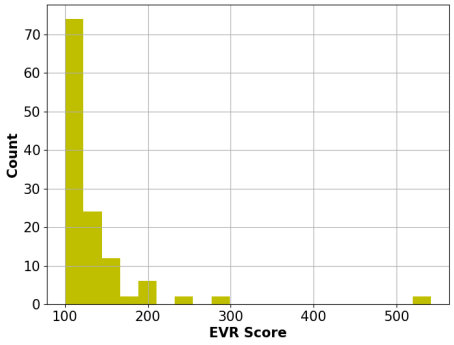


Figure 4. EVR score distribution of 1-Shot MiniImageNet

where ϵ is the threshold which is chosen in a way that $\epsilon > \text{EVR}_i$ for all tasks. The ratio $\frac{\epsilon}{\text{EVR}_i}$ should be large for challenging FSOSR tasks whereas small for easier tasks. As such, the ratio $\frac{\epsilon}{\text{EVR}_i}$ will have a minimal impact on the easier tasks but a drastic effect on challenging tasks.

Theorem 1. *Consider a challenging FSOSR testing task and a properly trained* MET *model that can predict evidence satisfying the two properties* (**P1, P2**)*. Let $\mathcal{P}', \mathcal{P}'_a$ denote the transformed representations of the original prototypes and altered prototypes, respectively; let $\mathcal{P}'_\epsilon$ be the transformed representation of the original prototypes augmented through the evidential cross-attention. Then, when $\frac{\epsilon}{EVR_i} > 1$ holds true, we have following inequality:*

$$d(\mathcal{P}'_a, \mathcal{P}') \leq d(\mathcal{P}'_a, \mathcal{P}'_\epsilon) \tag{15}$$

Remark. The evidential cross-attention essentially forces the (originally) different closed-set classes to attend to each other through (14). This leads to a more compact representation $\mathcal{P}'_\epsilon$. On the other hand, the representation of the altered prototypes $\mathcal{P}'_a$ is much less compact as it involves an open-set sample, leading to a large distance $d(\mathcal{P}'_a, \mathcal{P}'_\epsilon)$ that helps to improve the detection performance. Figure 4 demonstrates the EVR scores of query samples in 1-Shot tasks on MiniImageNet. As shown, by choosing a relatively large ϵ (*e.g.,* 500), we have $\frac{\epsilon}{EVR} \leq 1$ for most of the samples. This indicates that the above inequality holds true most of the times especially for the challenging FSOSR tasks. Figure 3 shows the impact of evidential cross-attention. The overall EVR augmented inference process is also illustrated in the figure, which integrates the EVR metric and the evidential cross-attention. For our inference algorithm along with the proof of Theorem 1, please refer to the Appendix.

5. Experiments

We conduct experiments to evaluate the effectiveness of the proposed MET model. Through these experiments, we aim to demonstrate: (i) state-of-the-art open-set detection performance in comparison to existing competitive baselines, and (ii) deeper insights on better detection performance through a qualitative and ablation study.

Datasets and Evaluation Metrics. We conduct experimentation on multiple datasets, including MiniImageNet (Vinyals et al., 2016), TiredImageNet (Ren et al., 2018), Cifar100 (Krizhevsky et al., 2009), and Caltech101 (Fei-Fei et al., 2004). Table 5 in the Appendix shows the data split for each dataset.As our goal is open-set recognition, we use the Area Under the ROC curve (AUROC) as the detection performance metric. AUROC measures unseen class instance detection capability using both seen and unseen class samples. We set five classes as known classes and the other non-overlapped five classes as unknown classes to compose a single 5-way classification problem during the experiments. We collected 15 instances for each class as queries, which leads to 75 known queries and 75 unknown queries for a 5-way classification problem. We use 1 shot and 5 shot indicating the number of examples per class in the support set.

Comparison Baselines. We compare MET with the state-of-the-art few-shot learning and open-set recognition methods. For the few-shot learning methods, we consider a metric-based meta-learning method FEAT (Ye et al., 2020) as it can be applied to a similar setting as ours. Additionally, we have included two standard metric-based few-shot learning models: Prototypical Network (Snell et al., 2017) and Relation Network (Sung et al., 2018). We also include a set of representative FSOSR methods: PEELER (Liu et al., 2020), SnaTCHer (Jeong et al., 2021), TANE (Huang et al., 2022), and Glocal (Wang et al., 2023) as the state-of-the-art comparison baselines. We furthermore include a classical open-set detection baseline, *i.e.,* OpenMax (Bendale & Boult, 2016). Comparison with some additional baselines are also included in the Appendix.

Implementation details. We used the ResNet-12 as a backbone architecture for the feature extractor followed by the transformer network. For good initialization, the feature extractor is connected to a fully connected layer (with output nodes equal to a number of classes present in the training set) and trained using the cross entropy (CE) classification loss by treating it as a multi-class classification problem. Once the model is trained, the last layer is removed and the transformer network is connected. Finally, the model is trained in the FSL open-set detection setting using the training loss defined in (10). For the training, stochastic gradient descent (SGD) is used with a total of 200 epochs. The initial learning rate of 0.002 is set and is decreased by 10% at an interval of every 20 epochs. The weight decay is set to 0.005 and λ is set to 1 throughout the experimentation.

5.1. Results and Discussion

The comparison results are obtained through 1,000 evaluation episodes with 1,000 tasks per episode and computed

Table 1. OSD (AUROC) performance comparison

Approaches	MiniImageNet 5-way		TieredImagenet 5-way		Cifar100 5-way		Caltech101 5-way	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
ProtoNet (Snell et al., 2017)	51.63 ± 0.47	60.26 ± 0.56	58.48 ± 0.50	63.46 ± 0.24	48.12 ± 0.23	50.63 ± 0.35	47.34 ± 0.26	51.35 ± 0.54
RelationNet (Sung et al., 2018)	53.14 ± 0.67	62.22 ± 0.78	60.85 ± 0.68	64.42 ± 0.57	48.76 ± 0.65	51.54 ± 0.56	47.95 ± 0.46	52.62 ± 0.35
OpenMAX (Bendale & Boult, 2016)	71.67 ± 0.87	76.75 ± 0.80	62.27 ± 0.55	70.92 ± 0.52	51.42 ± 0.54	54.45 ± 0.60	49.18 ± 0.52	49.77 ± 0.52
FEAT (Probability) (Ye et al., 2020)	45.00 ± 0.70	53.82 ± 0.78	57.14 ± 0.57	63.94 ± 0.52	49.25 ± 0.60	52.30 ± 0.59	48.99 ± 0.53	51.08 ± 0.52
FEAT (Distance) (Ye et al., 2020)	67.71 ± 0.92	75.32 ± 0.84	61.52 ± 0.58	70.77 ± 0.52	54.69 ± 0.46	59.86 ± 0.46	59.19 ± 0.52	65.30 ± 0.49
PEELER (Liu et al., 2020)	60.36 ± 0.72	68.45 ± 0.78	58.24 ± 0.65	66.14 ± 0.74	52.46 ± 0.43	56.11 ± 0.15	51.10 ± 0.72	55.96 ± 0.22
SnaTCHer (Jeong et al., 2021)	67.37 ± 0.91	77.99 ± 0.76	71.00 ± 0.66	79.49 ± 0.47	57.60 ± 0.57	62.06 ± 0.52	62.37 ± 0.63	67.35 ± 0.53
TANE (Huang et al., 2022)	73.23 ± 0.25	81.15 ± 0.18	74.89 ± 0.64	80.45 ± 0.49	55.14 ± 0.64	63.08 ± 0.58	52.71 ± 0.19	56.65 ± 0.18
Glocal (Wang et al., 2023)	73.41 ± 0.83	83.45 ± 0.57	72.23 ± 0.82	80.16 ± 0.81	57.71 ± 0.79	$60.36 \pm 0..74$	60.81 ± 0.78	66.06 ± 0.65
MET	$\mathbf{76.93 \pm 0.59}$	$\mathbf{84.90 \pm 0.41}$	$\mathbf{78.77 \pm 0.46}$	$\mathbf{84.37 \pm 0.35}$	$\mathbf{61.76 \pm 0.60}$	$\mathbf{66.17 \pm 0.54}$	$\mathbf{64.85 \pm 0.55}$	$\mathbf{72.12 \pm 0.47}$

Table 2. MiniImageNet performance with: (a) Different backbones, (b) Original data split.

Approach	ResNet12	ResNet18
PEELER	60.36 ± 0.72	61.52 ± 0.64
SnaTCHer	67.37 ± 0.91	68.11 ± 0.94
MET	76.93 ± 0.59	77.11 ± 0.62

(a) Performance on Different Backbones.

Approach	1-shot	5-shot
PEELER	60.12 ± 0.72	68.23 ± 0.66
SnaTCHer	72.98 ± 0.61	79.57 ± 0.49
MET	73.20 ± 0.45	81.19 ± 0.47

(b) Original Data Split Performance.

the mean over 1,000 episodes. It is worth mentioning that we achieve a comparable closed-set accuracy with regard to competitive baselines as demonstrated in the Table 6 in the Appendix. Table 1 shows the open-set performance comparison between different competitive models and the proposed MET. As demonstrated, our approach has a much more superior performance compared to these competitive baselines. As shown compared to SnaTCHer, our approach has more than 9% performance improvement in 1-shot and more than 6% in case of 5-shot setting.

Similarly for TieredImagenet, there is a substantial performance gain over the second best baselines for both 5-shot and 1-shot settings. Compared to SnaTCHer, the performance improvement is more than 7% in the 1-shot and around 5% in the 5-shot setting. This justifies the effectiveness of our proposed technique. For other standard few-shot learning baselines, we leverage distance and relation scores between the prototype and the query sample to perform open-set recognition in prototypical and relational networks, respectively. Also, they don't have information about opponent classes and we compute the naive distance between prototypes and the query sample to provide its corresponding score. This results in poor performance of those models in all datasets. Similarly, another standard open-set recognition baseline OpenMax has better results than some other baselines in 1-shot miniImageNet but has poor performance compared to the proposed MET model in all cases.

5.2. Ablation Study

In this section, we first demonstrate the robustness of our technique with respect to different backbones. We then show performance of the proposed technique under the original data split. Next, we demonstrate the effectiveness of each key component in MET. In the next subsection, we perform a qualitative analysis to further justify the effectiveness of our proposed technique. Because of the limited space, we present additional ablation studies and qualitative analysis in the Appendix.

Different backbones. Table 2 (a) demonstrates the performance comparison between different backbones for the MiniImageNet dataset (1-Shot setting). As shown, for multiple backbones, our technique has a superior performance compared to the competitive baselines.

Original data split. Our technique requires similar classes in open-set as well as closed-set to demonstrate the effectiveness of our technique. Therefore, in the main evaluation, we altered the data split. In this section, we show that even for the original data split, MET is able to outperform the competitive baselines. Table 2 (b) shows the performance for different baselines for the MiniImageNet dataset in the original data split. As shown, MET has superior performance compared to the other baselines. It should be noted that because of the evidential open-set loss along with the novel cross-attention technique, MET demonstrates a clear advantage over other baselines. Different from the evaluation strategy used in other baselines (*e.g.,* SnaTCHer), we fixed open-set and closed-set samples in both training as well as testing processes while keeping the original split, *i.e.,* training, validation, and testing identical. To achieve a fair comparison, we consider the identical setting (*i.e.,* backbone transformer) for PEELER and SnaTCHer and rerun them.

Table 3. Ablation study results on MiniImageNet

Transformer	Evidential Loss	EVR	AUROC	
			1-shot	5-shot
✓	✗	✗	67.37	77.99
✓	✓	✗	74.35	81.47
✓	✓	✓	**76.93**	**84.90**

Impact of key components. We conduct an ablation study to justify the effectiveness of each key component, including the evidential loss and EVR augmented inference process. Table 3 shows the effectiveness of each component by using the MiniImageNet dataset in the 5-way 1-shot setting as an example. Similar results are achieved on other datasets. As can be seen, the performance using the proposed evidential loss yields better performance compared to without using it (first row). Further combining both the evidential loss and EVR significantly boosts the performance as demonstrated in the third row of the table.

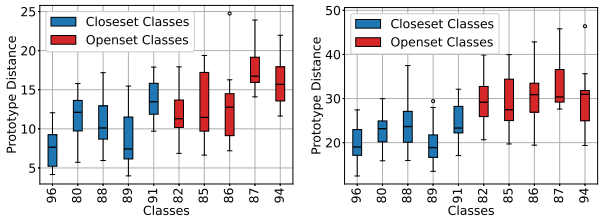
5.3. Qualitative Analysis

In addition to the ablation study, we perform a qualitative analysis to show the effectiveness of each proposed component (evidential loss and EVR).

It should be noted that using SnaTCHer, challenging open-set classes `82` and `85` have low prototype distances making them overlap with many closed-set classes. This is because, `82` and `85` share similarities with closed-set class `88` (see the Appendix for more details about these classes and their data samples). The proposed evidential loss helps to increase the separation between open-set and closed-set class samples. Specifically, as shown in Figure 5 (a), the loss helps to enlarge the prototpye distances of the difficult open-set classes, including `82` and `85`. As demonstrated in Figure 5 (b), by leveraging EVR, we can further push those samples upward and thereby creating a bigger separation between open-set and closed-set samples. In terms of detection performance, SnaTCHer, MET (w/o EVR), and MET (w/ EVR) achieve 65.16%, 73.40%, and 85.53% AUROC, respectively.

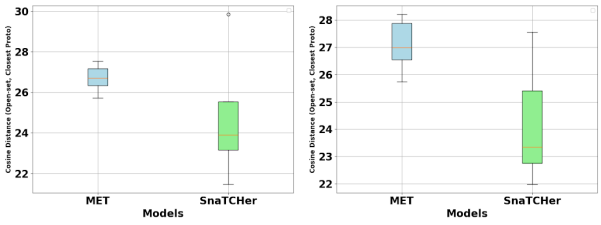
Apart from theoretical result as presented in Theorem 1, we also empirically show that MET learns more compact representation compared to that of SnaTCHER, leading to the better OSD. To show this, we compute the cosine distance (*i.e.,* 1/cosine similarity) between the prototype from the closest closed-set class and open-set query sample in the transformed space. Figure 6 show the boxplot for the distribution of open-set query samples (1-shot tasks on MiniImageNet) in terms of their cosine distance with the closest prototype. As shown in the figure, our approach is being able to push the open-set query away from the closest prototypes resulting in more compact representation with respect to open-set samples. Furthermore, as shown by the t-SNE plot of prototype and open-set query embedding representation distribution in Figure 7, our approach is able to push open-set samples away from the close-set samples, whereas in case of the SnaTCHer the distribution is much less compact.

For more detailed qualitative analysis along with visualization, please refer to the Appendix.



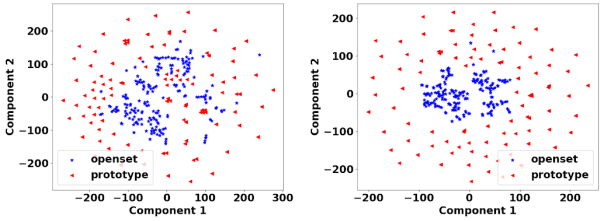
(a) MET (w/o EVR) (b) MET (w/ EVR)

Figure 5. OSR performance comparison on MiniImageNet



(a) MET (w/o EVR) (b) MET (w/ EVR)

Figure 6. Boxplot demonstrating the compactness behavior



(a) SnaTCHer (b) MET

Figure 7. t-SNE plot of the open-set and prototype representations

6. Conclusion

To tackle the FSOSR task, we propose a novel meta evidential transformer (MET) that uses an evidential open-set loss during training to learn more compact closed-set representation by leveraging similar closed-set classes. Furthermore, MET integrates an evidence-to-variance ratio to detect fundamentally challenging open-set samples by using an evidence-guided cross-attention mechanism. Experimental results on multiple real-world datasets demonstrate the effectiveness of the proposed technique over existing competitive methods in terms of better recognizing unseen class samples without deteriorating closed-set performance.

Acknowledgement

This research was supported in part by an NSF IIS award IIS-1814450. The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing any funding agency.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. By facilitating the model training with the limited data, our proposed MET model significantly reduces the cost associated with model training with state-of-the-art open-set detection ability thereby helping to reduce the annotation expenses and save energy. Furthermore, MET offers a novel way *i.e.,* cross-attention mechanism to capture the uncertainty information in the transformer-based architecture. This technique can be leveraged in the Generative AI techniques to quantify the uncertainty associated with a given response.

References

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

Boudiaf, M., Bennequin, E., Tami, M., Toubhans, A., Piantanida, P., Hudelot, C., and Ayed, I. B. Open-set likelihood maximization for few-shot learning, 2023.

Cevikalp, H. and Yavuz, H. S. Fast and accurate face recognition with image sets. *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 1564–1572, 2017.

Chapelle, O., Schölkopf, B., and Zien, A. *Semi-Supervised Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2006. ISBN 0262033585.

Charpentier, B., Zügner, D., and Günnemann, S. Posterior network: Uncertainty estimation without ood samples via density-based pseudo-counts. *ArXiv*, abs/2006.09239, 2020.

Dempster, A. P. A generalization of bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):205–232, 1968.

Fei-Fei, L., Fergus, R., and Perona, P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *2004 conference on computer vision and pattern recognition workshop*, pp. 178–178. IEEE, 2004.

Finn, C., Abbeel, P., and Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pp. 1126–1135. PMLR, 2017.

Finn, C., Xu, K., and Levine, S. Probabilistic model-agnostic meta-learning. *arXiv preprint arXiv:1806.02817*, 2018.

Ge, Z., Demyanov, S., Chen, Z., and Garnavi, R. Generative openmax for multi-class open set classification. *arXiv preprint arXiv:1707.07418*, 2017.

Grant, E., Finn, C., Levine, S., Darrell, T., and Griffiths, T. Recasting gradient-based meta-learning as hierarchical bayes. *arXiv preprint arXiv:1801.08930*, 2018.

Huang, S., Ma, J., Han, G., and Chang, S.-F. Task-adaptive negative envision for few-shot open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7171–7180, 2022.

Ilse, M., Tomczak, J. M., and Welling, M. Attention-based deep multiple instance learning. In *ICML*, 2018.

Jain, L. P., Scheirer, W. J., and Boult, T. E. Multi-class open set recognition using probability of inclusion. In *ECCV*, 2014.

Jeong, M., Choi, S., and Kim, C. Few-shot open-set recognition by transformation consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12566–12575, 2021.

Jøsang, A. *Subjective logic*. Springer, 2016.

Júnior, P. R. M., de Souza, R. M., de Oliveira Werneck, R., Stein, B. V., Pazinato, D. V., de Almeida, W. R., Penatti, O. A. B., da Silva Torres, R., and Rocha, A. Nearest neighbors distance ratio open-set classifier. *Machine Learning*, 106:359–386, 2016.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.

Liu, B., Kang, H., Li, H., Hua, G., and Vasconcelos, N. Few-shot open-set recognition using meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8798–8807, 2020.

Luo, T., Li, A., Xiang, T., Huang, W., and Wang, L. Few-shot learning with global class representations, 2019.

Malinin, A. and Gales, M. Predictive uncertainty estimation via prior networks. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 7047–7058, Red Hook, NY, USA, 2018. Curran Associates Inc.

Munkhdalai, T. and Yu, H. Meta networks. In *International Conference on Machine Learning*, pp. 2554–2563. PMLR, 2017.

Oliver, A., Odena, A., Raffel, C., Cubuk, E. D., and Goodfellow, I. J. Realistic evaluation of deep semi-supervised learning algorithms. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 3239–3250, Red Hook, NY, USA, 2018. Curran Associates Inc.

Ravi, S. and Larochelle, H. Optimization as a model for few-shot learning. 2016.

Ren, M., Triantafillou, E., Ravi, S., Snell, J., Swersky, K., Tenenbaum, J. B., Larochelle, H., and Zemel, R. S. Meta-learning for semi-supervised few-shot classification. *ArXiv*, abs/1803.00676, 2018.

Rusu, A. A., Rao, D., Sygnowski, J., Vinyals, O., Pascanu, R., Osindero, S., and Hadsell, R. Meta-learning with latent embedding optimization. *arXiv preprint arXiv:1807.05960*, 2018.

Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., and Lillicrap, T. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pp. 1842–1850. PMLR, 2016.

Sapkota, H., Ying, Y., Chen, F., and Yu, Q. Distributionally robust optimization for deep kernel multiple instance learning. In Banerjee, A. and Fukumizu, K. (eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 2188–2196. PMLR, 13–15 Apr 2021.

Scheirer, W. J., Rocha, A., Sapkota, A., and Boult, T. E. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:1757–1772, 2013.

Sensoy, M., Kandemir, M., and Kaplan, L. M. Evidential deep learning to quantify classification uncertainty. In *NeurIPS*, 2018a.

Sensoy, M., Kaplan, L., and Kandemir, M. Evidential deep learning to quantify classification uncertainty. *Advances in Neural Information Processing Systems*, 31, 2018b.

Shi, W., Zhao, X., Chen, F., and Yu, Q. Multifaceted uncertainty estimation for label-efficient deep learning. *Advances in Neural Information Processing Systems*, 33: 17247–17257, 2020.

Snell, J., Swersky, K., and Zemel, R. S. Prototypical networks for few-shot learning. *arXiv preprint arXiv:1703.05175*, 2017.

Sun, X., Yang, Z., Zhang, C., Peng, G., and Ling, K.-V. Conditional gaussian distribution learning for open set recognition. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13477–13486, 2020.

Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H., and Hospedales, T. M. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1199–1208, 2018.

Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29:3630–3638, 2016.

Wang, H., Pang, G., Wang, P., Zhang, L., Wei, W., and Zhang, Y. Glocal energy-based learning for few-shot open-set recognition, 2023.

Ye, H.-J., Hu, H., Zhan, D.-C., and Sha, F. Few-shot learning via embedding adaptation with set-to-set functions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8808–8817, 2020.

Yoshihashi, R., Shao, W., Kawakami, R., You, S., Iida, M., and Naemura, T. Classification-reconstruction learning for open-set recognition. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4011–4020, 2019.

Zhang, H. and Patel, V. M. Sparse representation-based open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:1690–1696, 2017.

Appendix

Organization of Appendix In this Appendix, first, we summarize all notations used in our paper. After that, we discuss other related works in additional to those reviewed in the related work section of the main paper. Next, we provide the theoretical proof for our theoretical results. Followed by that, we provide algorithms for the training and inference process. Then, we provide experimental details along with additional results. Finally, we provide a link to the source code.

A. Summary of Notations

Table 4 organizes all the major notations into three groups and describes their meanings.

Table 4. Notations with Descriptions

Symbol Group	Notation	Description
Meta Learning	N^{tr}	Number of training tasks
	$\mathcal{S}_i^{tr}$	Support set for i^{th} task in Meta-train
	$\mathcal{Q}_i^{tr}$	Query set for i^{th} task in Meta-train
	$\mathcal{S}_i^{te}$	Support set for i^{th} task in Meta-test
	$\mathcal{Q}_i^{te}$	Query set for i^{th} task in Meta-test
	K	Number of examples in support set
	N	Number of Classes in support set
	C^s	Set of Closed-set Classes
	C^u	Set of Open-set Classes
Evidential Loss	$\boldsymbol{\theta}$	Neural Network Parameter
	h	Hidden dimensionality of feature extractor
	e_k	Evidence belonging to class k
	S	Total Dirichlet Strength
	u	Uncertainty (vacuity) mass associated with a given data point
	α_{ik}	Dirichlet parameter for the i^{th} data point in the k^{th} class
	$KL(P\|\|Q)$	Kullback–Leibler divergence between two distributions P and Q
Transformer	A	Square $N \times N$ matrix with attention weights
	F	Backbone feature extractor
	T	Transformer
	$\mathcal{P}$	Prototype obtained from backbone
	$\mathcal{P}'$	Transformed prototype representation
	$\mathbf{p}_n$	Prototype corresponding to the n^{th} class
	$\mathbf{p}'_n$	Transformed prototype representation corresponding to class n
	$\mathcal{P}_a$	Altered prototype by replacing nearest class prototype by query instance
	$\mathcal{P}'_a$	Transformed prototype representation of altered prototype
	$\mathcal{P}'_\epsilon$	Transformed prototype representation of original prototype using cross-attention

B. Additional Related Work

In this section, we review some additional related works, including few-shot learning and open-set recognition.

Few-shot Learning. Few-shot learning is becoming a popular method due to its ability to quickly generalize to new tasks containing only a few examples. These methods are grouped into three categories: *model-based*, *optimization-based*, and *metric-based*. Model-based methods largely depend on a model design for the fast adaptation (Santoro et al., 2016; Munkhdalai & Yu, 2017), which are less frequently used in recent years. Optimization-based methods back-propagate the gradients to deal with generalization problems. Ravi et al. (Ravi & Larochelle, 2016) model a meta-learner as an LSTM so that knowing historical gradients can benefit current gradient updates. MAML (Finn et al., 2017) and its variants (Grant et al., 2018; Rusu et al., 2018; Finn et al., 2018) learn meta parameters with outer updates utilizing query samples and task-specific parameters via support samples. Metric-based methods learn a good distance function to compare feature similarity between support and query set samples. Cosine distance is learned in (Vinyals et al., 2016) with a recurrent network to measure similarities between samples. Prototypical network (Snell et al., 2017) represents each class as a prototype utilizing support

set samples and then computes its similarity with the query set ones. Relation network (Sung et al., 2018) predicts a relation score between a pair of support and query set samples rather utilizing metrics directly on the feature space. FEAT (Ye et al., 2020) transforms each class prototype via transformer functions and results in a richer representation. Since feature-based metrics are also useful for open-set recognition, we largely focus on those approaches. While those methods show promising results in closed-set settings, few attempts have explored whether they can be effectively adapted for open-set recognition.

Open-set Recognition. Various support vector machines (SVMs) and reconstruction-based approaches have been proposed to tackle the OSR problem in existing literature (Scheirer et al., 2013; Jain et al., 2014). For instance, Scheirer et al. (Scheirer et al., 2013) propose a Weibull-calibrated SVM (W-SVM) technique by leveraging the Extreme Value Theory (EVT). Zhang & Patel (Zhang & Patel, 2017) propose a reconstruction-based approach, where a threshold defined over the reconstruction error is used to distinguish known-class samples from unknown classes. Additionally, various traditional models such as nearest neighbor (Júnior et al., 2016) and quasi-linear function (Cevikalp & Yavuz, 2017), have also been used in the open-set detection tasks. More recently, deep learning models have been adopted for open-set detection and multiple approaches have been proposed. For instance, Scheirer et al. propose OpenMax (Scheirer et al., 2013), in which the probability output from a softmax function is redistributed in order to produce the probability of being unknown. VAE-based approaches have also been proposed for the open-set detection (Yoshihashi et al., 2019), (Sun et al., 2020). For example, Yoshihashi et al. propose a reconstruction-based approach that performs open-set detection similar to OpenMax by leveraging the effective latent representation trained using VAE (Yoshihashi et al., 2019).

C. Theoretical Proof

In this section, we first show why $EVR_a < EVR_b$ (which is a key ingredient in our methodology). We then provide the detailed proof of Theorem 1.

C.1. Proof of $EVR_a < EVR_b$

Proof. Based on $\mathbf{P1}$, we can write the following

$$\left\{\max_{n\in N}[e_{jn}]\right\}_{\text{easy-closed}} \geq \left\{\max_{n\in N}[e_{jn}]\right\}_{\text{ch-open}} \tag{16}$$

where easy-closed indicates the easy closed-set sample whereas ch-open indicates a challenging open-set sample. According to this equation, for the easy-closed set sample as $\max_{n\in N}[e_{jn}]$ is high, the EVR_i will be high *i.e.,* evidence dominates EVR_i to make it high. Based on $\mathbf{P2}$, we can write the following

$$\{\text{var}_{n\in N}[e_{jn}]\}_{\text{easy-open}} \leq \{\text{var}_{n\in N}[e_{jn}]\}_{\text{ch-open}} \tag{17}$$

where easy-open indicates an easy open-set sample. In this case, for the easy open-set sample the output evidence will remain low (closed to 0) with respect to all closed-set classes making $\text{var}_{n\in N}[e_{jn}]$ low. This low variance will dominate EVR_i to make it high. In case of a challenging open-set sample, the maximum evidence $\max_{n\in N}[e_{jn}]$ is relatively low while variance $\text{var}_{n\in N}[e_{jn}]$ is high, making EVR_i low. Therefore, we can say that with a being a challenging task and b being a regular task *i.e.,* easy closed-set or easy open-set, we have $\text{EVR}_a < \text{EVR}_b$. This completes the proof. □

C.2. Proof of Theorem 1

Proof. Specifically, we need to show the following:

$$d(\mathcal{P}'_a, \mathcal{P}') \leq d(\mathcal{P}'_a, \mathcal{P}'_\epsilon) \tag{18}$$

In the above equation, the transformed representation for $\mathcal{P}'$ on j^{th} prototype on the l^{th} dimension can be represented as

$$\{\mathcal{P}'\}_{jl} = \sum_{n=1}^{N} a_{jn} f_{nl}; \forall j \in [N], \forall l \in [h] \tag{19}$$

where a_{jn} is the attention for j^{th} row and n^{th} column and f_{nl} be the associated feature obtained with value $value$ $(\mathcal{V})$ in the transformer network, h is the feature dimensionality. Let c be the closest class for a given query sample then each element

of the altered transformed prototype *i.e.,* $\mathcal{P}'_a$ for this sample can be represented as

$$\{\mathcal{P}'_a\}_{jl} = \sum_{n=1}^{N} a'_{jn} f'_{nl}; \forall j \in [N], \forall l \in [h] \tag{20}$$

where a'_{jn} is the attention weight for altered transformed prototype for j^{th} row and n^{th} column and f'_{jl} being associated feature. It should be noted that $a'_{jn} = a_{jn}$ if $j \neq c$ or $n \neq c$ and $f'_{nl} = f_{nl}$ if $n \neq c$. The transformed prototype obtained using our proposed cross-attention mechanism can be represented as

$$\{\mathcal{P}'_\epsilon\}_{jl} = \left\{ \begin{array}{ll} a_{cc} f_{cl} + \frac{\epsilon}{EVR} \sum_{n=1, n\neq c}^{N} a_{jn} f_{nl} & \text{if } j == c \\ \frac{\epsilon}{EVR} a_{jc} f_{cl} + \sum_{n=1, n\neq c}^{N} a_{jn} f_{nl}, \text{otherwise} & \end{array} \right\} \tag{21}$$

Considering d being the Euclidean distance, we compute $d(\mathcal{P}'_a, \mathcal{P}')$ as:

$$d(\mathcal{P}'_a, \mathcal{P}') = \frac{1}{N} \sum_{j=1}^{N} \sqrt{\sum_{l=1}^{h} \left(\{\mathcal{P}'_a\}_{jl} - \{\mathcal{P}'\}_{jl}\right)^2} \tag{22}$$

Similarity, we can compute $d(\mathcal{P}'_a, \mathcal{P}'_\epsilon)$ term as:

$$d(\mathcal{P}'_a, \mathcal{P}'_\epsilon) = \frac{1}{N} \sum_{j=1}^{N} \sqrt{\sum_{l=1}^{h} \left(\{\mathcal{P}'_a\}_{jl} - \{\mathcal{P}'_\epsilon\}_{jl}\right)^2} \tag{23}$$

It it is noted that $\forall j \in [N], \forall l \in [h]$, if we show $\left(\{\mathcal{P}'_a\}_{jl} - \{\mathcal{P}'\}_{jl}\right)^2 \leq \left(\{\mathcal{P}'_a\}_{jl} - \{\mathcal{P}'_\epsilon\}_{jl}\right)^2$ then the inequality in Eq (19) becomes valid. For simplicity, let us assume that $\mathbf{U}_{jl} = \{\mathcal{P}'_a\}_{jl}$, $\mathbf{V}_{jl} = \{\mathcal{P}'\}_{jl}$, $\mathbf{W_{jl}} = \{\mathcal{P}'_\epsilon\}_{\mathbf{jl}}$. Then, $\forall j \in [N], l \in [h]$, we need to prove:

$$(\mathbf{U}_{jl} - \mathbf{V}_{jl})^2 \leq (\mathbf{U}_{jl} - \mathbf{W_{jl}})^{\mathbf{2}} \tag{24}$$

Let us write write both sides where we seek to find the inequality relation between them

$$(\mathbf{U}_{jl} - \mathbf{V}_{jl})^2 \textcircled{?} (\mathbf{U}_{jl} - \mathbf{W_{jl}})^{\mathbf{2}} \tag{25}$$

Expanding both sides and canceling common terms we have the following

$$\mathbf{V}_{jl}^2 - 2\mathbf{U}_{jl}\mathbf{V}_{jl} \textcircled{?} \mathbf{W}_{jl}^2 - 2\mathbf{U}_{jl}\mathbf{W}_{jl} \tag{26}$$

Further simplification leads to the following

$$2\mathbf{U}_{jl}(\mathbf{W}_{jl} - \mathbf{V}_{jl}) \textcircled{?} (\mathbf{W}_{jl} - \mathbf{V}_{jl})(\mathbf{W}_{jl} + \mathbf{V}_{jl}) \tag{27}$$

As $\frac{\epsilon}{EVR} > 1$, $\mathbf{W}_{jl} > \mathbf{V}_{jl}$. As such $(\mathbf{W}_{jl} - \mathbf{V}_{jl})$ is non-negative and therefore, we can cancel $(\mathbf{W}_{jl} - \mathbf{V}_{jl})$ on both sides without changing their inequality sign. This leads to the following:

$$2\mathbf{U}_{jl} \textcircled{?} (\mathbf{W}_{jl} + \mathbf{V}_{jl}) \tag{28}$$

In the inequality, since $\frac{\epsilon}{EVR} > 1$, there exists a constant $k > 1$ that makes $\mathbf{W}_{jl} = k\mathbf{V}_{jl}$. Substituting this in the above equation, we have the following:

$$2\mathbf{U}_{jl} \textcircled{?} (1 + k)\mathbf{V}_{jl} \tag{29}$$

It is noted that attention weights of the altered prototype in $\mathbf{U}_{jl}$ are likely to be similar to $\mathbf{V}_{jl}$ in case of a challenging query sample. This is because the challenging sample may be very similar to one of the prototypes making the output representation almost identical. This makes $\mathcal{U}_{jl}$ to be similar to $\mathcal{V}_{jl}$. However, on the right-hand side, we have $(1 + k) > 2$, which therefore makes the right term bigger than the left term. It should be noted that k is the term dependent on the ratio $\frac{\epsilon}{EVR}$ where the higher the ratio, the higher the k term would be. Therefore, under the non-negativity assumption of the feature representation (which can be achieved simply using a non-negative transformation function in the output), with high probability following holds for the challenging samples.

$$2\mathbf{U}_{jl} \leq (1 + k)\mathbf{V}_{jl} \tag{30}$$

This completes the proof of Theorem 1. □

D. Training and Inference Algorithms

Algorithm 1 shows the overall training process of our proposed MET technique. As shown we use both open-set as well as closed-set loss to optimize the network parameters. Algorithm 2 shows the corresponding inference algorithm that leverages the cross-attention mechanism.

Algorithm 1 MET Training

Require: Hyperparameters: λ
Require: Task distribution: $P(\mathcal{T})$, Feature extractor $F(.)$, Transformer network $\mathtt{T}(.)$
  Initialize meta evidential transformer, $\boldsymbol{\theta}$
  while not converge **do**
    Sample tasks $\mathcal{T}^{tr} \sim P(\mathcal{T})$
    for all $\mathcal{T}^{tr}$ **do**
      Sample support set, $\mathcal{S}^{tr} \in \mathcal{T}^{tr}$
      Compute prototype for each seen class $\mathbf{p}_n$ with $\mathcal{S}^{tr}$ using Equation 4
      Compute transformed prototype $\mathcal{P}' = \mathtt{T}(\mathcal{P})$ using Equation 6
      Sample query set $\mathcal{Q}_{tr} \in \mathcal{T}^{tr}$ for the meta update
      Compute evidence for each query sample using Equation 12
    end for
    Perform meta update using Equation (10) with $\mathcal{Q}_{tr}$ leveraging evidence of each query
  end while

Algorithm 2 MET Inference

Require: Trained MET model $\boldsymbol{\theta}$
Require: Threshold: ϵ
  for each test task $\mathcal{T}^{te}$ **do**
    Sample support set, $\mathcal{S}^{te} \in \mathcal{T}^{te}$
    Compute prototype for each seen class $\mathbf{p}_n$ with $\mathcal{S}^{te}$ using Equation (4)
    Compute transformed prototype $\mathcal{P}' = \mathtt{T}(\mathcal{P})$
    Sample query set $\mathcal{Q}_{te} \in \mathcal{T}^{te}$ for performance evaluation
    Compute EVR threshold for a query set using Equation (13)
    for each query sample q_i **do**
      Compute evidence using Equation (12)
      Predict c as a predicted query class
      Construct altered prototype $\mathcal{P}_a = \mathcal{P} - \{\mathbf{p}_c\} + \{\mathbf{p}_i\}$
      Pass through the transformer to get attention weights A_i
      Update the attention weights of transformer using Equation (14) yielding $\mathtt{T}'$
      Computed the transformed representation of original prototype $P'_\epsilon = \mathtt{T}'(P_a)$
      Compute distance between two sets $\mathcal{P}'$ and $\mathcal{P}'_\epsilon$ and store the distance
    end for
    Perform open-set recognition by passing stored distances through selected recognition metric
  end for

E. Experimental Details and Additional Results

In this section, we first provide the dataset distribution of all datasets used in the experimentation. Next, we provide the closed-set performance of those datasets with respect to competitive baselines. Next, we explain the ROC curves generated for the same set of datasets. After that, we conduct an additional ablation study. Finally, we conduct an in-depth qualitative analysis to show the effectiveness of our proposed technique.

E.1. Dataset Distribution

Table 5 shows the dataset splits for four datasets: MiniImageNet, TieredImageNet, Cifar100, and Caltech101. It should be noted that to serve our purpose we have considered the whole data distribution and divided it into closed-set and open-set.

E.2. Experimentation Details

In this section, we describe the way we split the data along with the implementation details.

Table 5. Train/Evaluation/Test partition on different datasets.

Split	MiniImageNet			TieredImagenet			Cifar100			Caltech101		
	Train	**Eval**	**Test**	**Train**	**Eval**	**Test**	**Train**	**Eval**	**Test**	**Train**	**Eval**	**Test**
Open-set	21	6	8	116	24	41	27	5	8	20	10	10
Closed-set	46	10	9	259	73	95	35	10	15	40	10	10

Dataset Split. According to Table 5, we first partition the entire dataset into training, validation, and testing. Within the training set, we perform semantic analysis at the class level to identify groups of semantically relevant classes (*e.g.,* different categories of dogs). For datasets with a relatively small number of classes (*e.g.,* MiniImageNet), this introduces minimal overhead. For larger datasets, we can usually benefit from some existing hierarchical structure among the classes. For instance, in MiniImageNet, training classes Ferrets (`88`) and Malamute dog (`83`) are semantically similar. Instead of using both of them as closed-set samples during training like all existing approaches, we assign Malamute dog (`83`) as one of the opponent classes, which are used as part of the evidential open-set loss. As demonstrated in our experiments, this arrangement clearly improves the detection of some similar open-set classes, such as Golden Retriever Golden Retriever (`82`) and African Hunting dog (`85`).

E.3. Closed-Set Performance

Table 6 shows the closed-set performance of MET with respect to the competitive baselines. As shown our approach generates comparable closed-set performance while having a much better OSR performance as demonstrated in Table 1.

E.4. ROC Curves

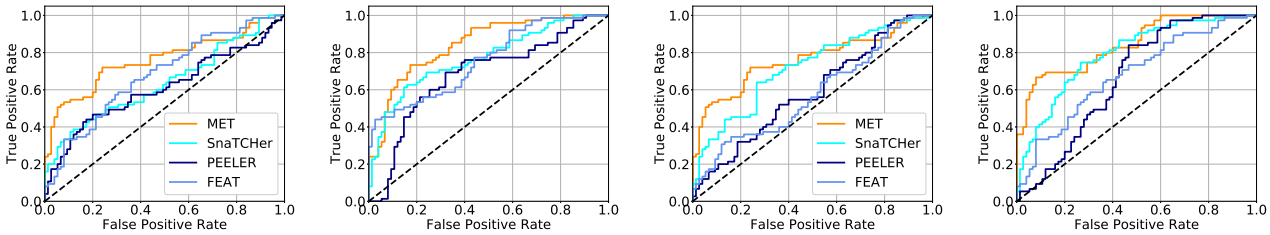
To provide a detailed view of AUROC, we further show the ROC curves for the 1-shot and 5-shot scenarios in the miniImageNet and TieredImagenet datasets as shown in Figure 8. The ROC plot has a similar pattern in the other two datasets. It is worth mentioning from the ROC curves that the proposed technique stays on the top, especially for the lower false positive rate (FPR) region. For example, in the case of Figure 8 (a), we can achieve True Positive Rate (TPR) around 70% while maintaining FPR below 30% which is more than 20% higher than the second best competitive model. This concludes that the proposed approach can correctly identify far more open-set samples compared to other baselines while being able to maintain a low FPR.

E.5. Ablation Study

In this section, we conduct an ablation study with regard to the hyperparameters λ and ϵ. Next, we explain the effectiveness of our proposed technique using different backbones. After that, we report the performance in the original data split. Finally,

Table 6. Closed set performance (ACC) on different datasets.

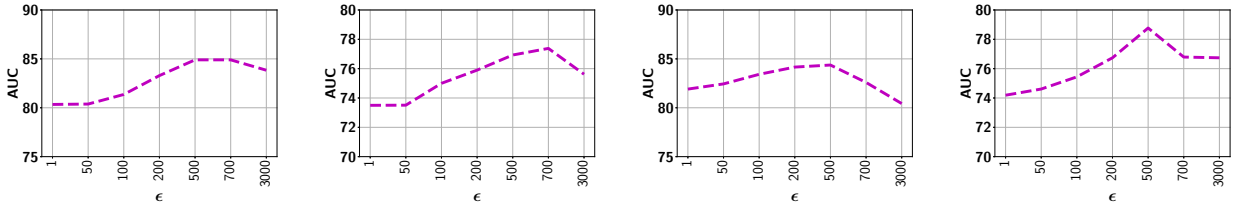
Approach	**MiniImageNet**		**TieredImagenet**	
	1-Shot	5-Shot	1-Shot	5-Shot
PEELER	61.24 ± 0.46	76.46 ± 0.50	45.64 ± 0.44	60.22 ± 0.23
SnaTCHer	63.91 ± 0.63	79.93 ± 0.43	47.75 ± 0.72	64.04 ± 0.65
MET	63.13 ± 0.62	79.00 ± 0.43	47.91 ± 0.73	63.33 ± 0.48



(a) MiniImageNet (b) TieredImagenet

Figure 8. ROC curves on both 5-way-1-shot and 5-way-5-shot tasks

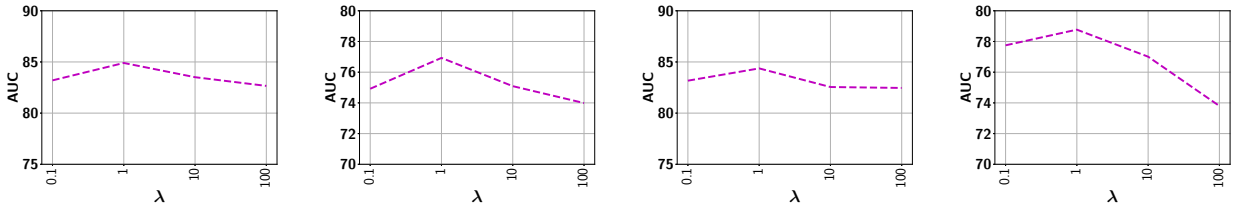
we also conduct an additional experimentation showcasing effectiveness of our proposed technique with additional baselines.



(a) 5-Shot (b) 1-Shot (c) 5-Shot (d) 1-Shot

Figure 9. OSR performance with respect to hyperparameter ϵ: (a-b) MiniImageNet, (c-d) TieredImageNet.

Sensitivity to λ and ϵ: Figure 9 shows the impact of hyperparameter ϵ on the model performance. As can be seen, a very small ϵ value is not beneficial because the model fails to shift the attention by assigning a larger weight to the predicted class. Having a higher ϵ value helps the model to change the attention weight according to EVR, *i.e.,* a higher EVR leads to a lower change. But, having a very high ϵ leads to degradation in performance as it dramatically changes the representation irrespective of the EVR value. In general, the model performance is quite stable as long as ϵ is not set to very high or very low values. With the middle range of ϵ as shown by Figure 9, the model automatically calibrates the change in accordance with EVR leading to better performance.



(a) 5-Shot (b) 1-Shot (c) 5-Shot (d) 1-Shot

Figure 10. OSR performance with respect to hyperparameter λ: (a-b) MiniImageNet, (c-d) TieredImageNet.

Figure 10 shows the impact of the open-set weight λ on the performance. As shown, having a low λ value, the model puts less emphasis on opponent open-set classes, leading to a less compact representation of closed-set classes that can benefit open-set detection. On the other hand, having a very high λ value may be problematic as the model puts too much emphasis on the opponent open-set classes without paying much attention on learning from the closed-set classes resulting in performance degradation as well. In general, the λ value in the middle range (*e.g.,* $\lambda = 1$) gives a good balance between open-set and closed-set losses resulting in the best performance.

Experimental Results on Additional Baselines. As discussed in the related work section, Open-Set Likelihood Optimization (OSLO) works in a transductive setting that requires the access to all unlabeled query samples from the test set (Boudiaf et al., 2023). In contrast, our proposed technique has not such constraint and can be applied in more general settings. Despite this difference, we still make a comparison by giving OSLO access to all unlabeled test samples. We

(a) Images: Ferrets (88), African Hunting dog (85)

Image	SnaTCHer	MET (w/o EVR)	MET (w/ EVR)
Image (a) 88	32	32	15
Image (b) 85	1	2	17
Image (c) 85	7	15	26

(b) Ranking

Figure 11. Examples of difficult images with the corresponding ranking

show the results in Table 7 on the two more challenging datasets: MiniImageNet and TinyImageNet. As can be seen, MET achieves a clearly better OSD performance on different few-shot tasks.

Table 7. OSD (AUROC) performance with a transductive baseline

Approaches	**MiniImageNet 5-way**		**TieredImagenet 5-way**	
	1-shot	5-shot	1-shot	5-shot
OSLO	75.52 ± 0.63	83.80 ± 0.44	62.82 ± 0.82	77.37 ± 0.46
MET	$\mathbf{76.93 \pm 0.59}$	$\mathbf{84.90 \pm 0.41}$	$\mathbf{78.77 \pm 0.46}$	$\mathbf{84.37 \pm 0.35}$

E.6. Qualitative Analysis

In this section, we perform an in-depth quantitative analysis to justify the effectiveness of our proposed technique. Figure 11 (a) demonstrates some difficult examples from closed-set class Ferrets (88) and open-set class African Hunting Dog (85). As shown in the first image *i.e.,* leftmost of Figure 11 (a), the image looks different from many others from class Ferrets (88) because of the different color and camera angle. As a result, SnaTCHer incorrectly classifies it as an open-set sample and assigns the highest distance among all samples in the same task. Although MET (w/o EVR) helps to decrease the distance by leveraging the opponent open-set classes in training, it is not sufficient to correctly identify it as open-set. With the help of the novel EVR detection, MET (w/ EVR) is able to correctly identify it as a closed-set sample. In the case of the second image *i.e.,* middle image in Figure 11 (a), because of its similarity with the first image, due to color (and possible other low-level image features), both SnaTCHer and MET (w/o EVR) incorrectly classify it as a closed-set sample. In contrast, MET (w/ EVR) is able to correctly identify it as an open-set sample. In the case of the third image i.e., the rightmost image in Figure 11 (a), it shares some similarities (in terms of color and body pattern) with closed-set samples, SnaTCHer has trouble identifying it as an open-set sample. MET (w/o EVR) helps to increase the distance (*i.e.,* uncertainty) but it is not sufficient to classify confidently as an open-set sample. With further help from EVR-based detection, we are able to correctly identify it as an open-set sample.

Similarly, Table 11 (b) shows the relative ranking of images based on the output prototype distance. It should be noted that a ranking of 1 *i.e.,* highest ranked (or lowest prototype distance) means the given sample is closest to the prototype among all samples. For Figure 11 (a) leftmost image, we want the sample to be ranked close to the top so that it is closer to the prototype compared to most open-set samples. However, it ranks at 32 by both SnaTCHer and MET (w/o EVR), which will negatively impact the overall AUROC score. In contrast, MET (w/ EVR) ranks as 15, higher than $\frac{2}{3}$ of other samples, most of which are open-set one, leading to an improved AUROC score. In the case of the middle image in Figure 11 (a), both SnaTCHer and MET (w/o EVR) rank it very high *i.e.,*, better than most closed-set samples. In the case of the rightmost image in Figure 11 (a), there is some improvement using our novel evidential loss by MET (w/o EVR). With the help of EVR, MET (w/ EVR) is able to further push this sample down to the rank of 26. We provide some additional qualitative analysis in the appendix.

F. Source Code

For the source code, please click here.