# Boosting Few-Shot Open-Set Recognition with Multi-Relation Margin Loss

**Yongjuan Che**[1,2] , **Yuexuan An**[1,2] and **Hui Xue**[1,2*]
[1]School of Computer Science and Engineering, Southeast University, Nanjing, 210096, China
[2]MOE Key Laboratory of Computer Network and Information Integration (Southeast University), China
{yjche, yx_an, hxue}@seu.edu.cn

## Abstract

Few-shot open-set recognition (FSOSR) has become a great challenge, which requires classifying known classes and rejecting the unknown ones with only limited samples. Existing FSOSR methods mainly construct an ambiguous distribution of known classes from scarce known samples without considering the latent distribution information of unknowns, which degrades the performance of open-set recognition. To address this issue, we propose a novel loss function called multi-relation margin (MRM) loss that can plug in few-shot methods to boost the performance of FSOSR. MRM enlarges the margin between different classes by extracting the multi-relationship of paired samples to dynamically refine the decision boundary for known classes and implicitly delineate the distribution of unknowns. Specifically, MRM separates the classes by enforcing a margin while concentrating samples of the same class on a hypersphere with a learnable radius. In order to better capture the distribution information of each class, MRM extracts the similarity and correlations among paired samples, ameliorating the optimization of the margin and radius. Experiments on public benchmarks reveal that methods with MRM loss can improve the unknown detection of AUROC by a significant margin while correctly classifying the known classes.

## 1 Introduction

Few-shot learning (FSL) has emerged as a promising direction for tackling the challenge of recognizing new classes with few labeled samples [Chen *et al.*, 2019; An *et al.*, 2023]. Conventional few-shot learning methods mostly hold a closed-set assumption that testing samples belong to the predefined classes of training samples [Liu *et al.*, 2020b]. However, in real-world applications, there always exist unforeseen testing samples out of classes of training samples, which results in severe distribution shifts [Hsu *et al.*, 2020]. Hence, the learned model should not only classify known classes but also reject unknowns.

Open-set recognition (OSR) [Bendale and Boult, 2016; Oza and Patel, 2019; Perera *et al.*, 2020] considers the scenario where testing samples could come from unknown classes. The existing OSR methods mainly learn an unknown class detector by utilizing the characteristics of known classes from substantial datasets. But having a sufficient number of training samples is inconsistent with few-shot learning. Therefore, directly applying OSR methods in the few-shot setting could degrade the performance.

Recently, few-shot open-set recognition (FSOSR) is proposed to tackle the new problems that classify the known classes and reject the unknown ones with a few samples. The main challenge of FSOSR is how to detect unknown class samples while maintaining the classification capability of known classes with limited labeled training samples. Existing methods for FSOSR mostly construct an ambiguous distribution of known classes while ignoring the latent distribution information of unknowns [Jeong *et al.*, 2021; Pal *et al.*, 2022]. These methods easily lead to confusion between unknown and known samples and further degrade the open-set performance. Limited methods attempt to learn the unknown information by directly utilizing or generating pseudo-unknown class samples [Huang *et al.*, 2022; Pal *et al.*, 2023]. However, these methods heavily rely on the quality of pseudo-unknown samples and the latent true distribution of unknown classes may not be adequately represented by these pseudo-unknown samples. Hence, how to properly delineate the distribution of unknowns is an essential problem for FSOSR.

In this paper, we propose multi-relation margin (MRM) loss, which seeks to dynamically refine the distinct decision boundary for known classes and implicitly delineate the distribution of unknown classes without relying on pseudo-unknown samples. Specifically, MRM enlarges the margin between different classes and adjusts a hypersphere for each known class with a learnable radius to form a finer boundary between the closed and open spaces. In order to capture the entire data distribution information and further refine the margin and radius, we excavate multiple relationships among paired samples, including self-relation and relative relation, where self-relation is the similarity relation inside a pair and relative relation is the correlation among different pairs. They

*Corresponding author

both make a significant impact on exploiting the potentiality of each pair and precisely measuring the distribution of paired samples. Therefore, the areas away from the boundaries of known classes implicitly delineate the distribution of unknowns, and open-set detection can be easily conducted by rejecting samples located in these areas. The effectiveness of MRM is demonstrated by experiments on public benchmark datasets.

In summary, our contributions are as follows:

- We propose a novel multi-relation margin (MRM) loss to boost the performance of FSOSR. Compared with previous methods, MRM is able to implicitly delineate the distribution of unknown classes without relying on additional pseudo-unknown class samples.
- We explore a new formulation of margin-based deep metric learning for FSOSR. Our method enforces the margin between different classes by extracting the multi-relationship of paired samples to learn a finer boundary between closed and open spaces.
- We perform adequate experiments to verify the effectiveness of MRM. Results reveal that few-shot methods with MRM loss can improve the unknown detection of AUROC by a significant margin while correctly classifying the closed-set.

## 2 Related Work

**Few-shot learning.** Few-shot learning focuses on training models with limited labeled samples to develop the capacity of recognizing novel classes [An *et al.*, 2021; Ma *et al.*, 2022a]. Representative few-shot learning literature can be typically organized into two branches: optimization-based methods and metric-based methods. Optimization-based methods can achieve generalization ability within a small number of update steps. Specifically, MAML [Finn *et al.*, 2017] adapts model parameters to novel tasks based on the loss computed by secondary gradients. Metric-based methods try to model an appropriate feature metric space and predict query samples based on the distance function. ProtoNet [Snell *et al.*, 2017] suggests measuring the Euclidean distance between the mean of the support samples, indicated as class prototypes, and the query samples for classification. On that basis, FEAT [Ye *et al.*, 2020] transforms class prototypes to task-adaptive prototypes with a set-to-set function. In our work, we mainly consider the classification combined with ProtoNet and FEAT. Although prior FSL studies have demonstrated promising results under the closed-set assumption that the query set and support set belong to the same classes, the performance in the open-set setting is not guaranteed.

**Open-set recognition.** Open-set recognition describes a scenario where new classes unseen in training occur in testing, thus classifiers must be able to properly identify seen samples while rejecting unseen ones [Geng *et al.*, 2020]. OpenMax [Bendale and Boult, 2016] was the first solution towards open-set deep network combined with Extreme Value Theory to detect unknowns by thresholding. Later researchers mostly use discriminative or generative approaches to tackle this issue. Specifically, C2AE [Oza and Patel, 2019] and CGDL [Sun *et al.*, 2020] utilize class conditioned autoencoders that have been trained on all of the train samples to define a threshold. Since it requires a large amount of training samples to train an unseen sample detector or a generative model, directly applying these OSR methods to the few-shot setting could cause over-fitting and lead to poor performance.

**Few-shot open-set recognition.** FSOSR has gained popularity recently, which addresses OSR in the FSL setting. Existing research on FSOSR mostly trains a model without considering the distribution of unseen classes, which could degrade the performance on unknown class detection. SnaTCHer [Jeong *et al.*, 2021] thresholds the distance between original prototypes and query replaced transformed prototypes to detect unseens. [Pal *et al.*, 2022] reinforce the feature extractor through a residual attention network. Limited methods focus on the distribution of unknowns by directly utilizing or generating additional pseudo-unknown samples. [Song *et al.*, 2022] utilize background features from seen classes as the pseudo-unseen classes for classifier training. ATT [Huang *et al.*, 2022] augments the classifier with additional negative prototypes via a negative generator. However, the pseudo-unknown based approach heavily depends on the quality of pseudo samples, and these pseudo-unknowns may not be adequate to represent the latent true distribution of unknown classes. In our work, we aim to implicitly delineate the distribution of unknowns through deep metric learning.

**Margin-based deep metric learning.** Deep metric learning is an important series of meta-learning methods for learning better features. In the metric learning area, the standard softmax loss is regarded as insufficient for discrimination on different training classes [Liu *et al.*, 2020a]. It can only construct an ambiguous distribution. According to previous studies [Sohn, 2016; Song *et al.*, 2016], integrating the large margin to the standard loss could help learn highly-discriminative features and efficiently improve the performance on visual recognition tasks. As the representative of margin-based methods, Triplet loss [Schroff *et al.*, 2015] enforces each input sample as an anchor and works on the triplet of samples $\{x_i^+, x_i^a, x_i^-\}$, where $y_i^+ = y_i^a$ and $y_i^- \neq y_i^a$. $x_i^a$ is usually called an anchor in the $i$-th triplet, $x_i^+$ is the positive sample and $x_i^-$ is the negative sample. It assumes the anchor is closer to the similar point than the dissimilar one by the margin. Center loss [Wen *et al.*, 2016] encourages samples to concentrate on their corresponding class center. However, these methods do not utilize all the samples contained in a batch, which could result in losing valuable information during training. On the contrary, Multi-similarity loss [Wang *et al.*, 2019b] considers each pair of samples in a batch that construct an informative structure among anchor-sample pairs and provide a new weighting scheme by calculating the derivative of the loss function.

Although previous margin-based losses have achieved promising results, they normally assume that sufficient samples can be provided for training. Recently, some researchers are devoted to applying margin-based deep metric learning in the few-shot setting [Li *et al.*, 2020]. For instance, [Liu *et al.*,

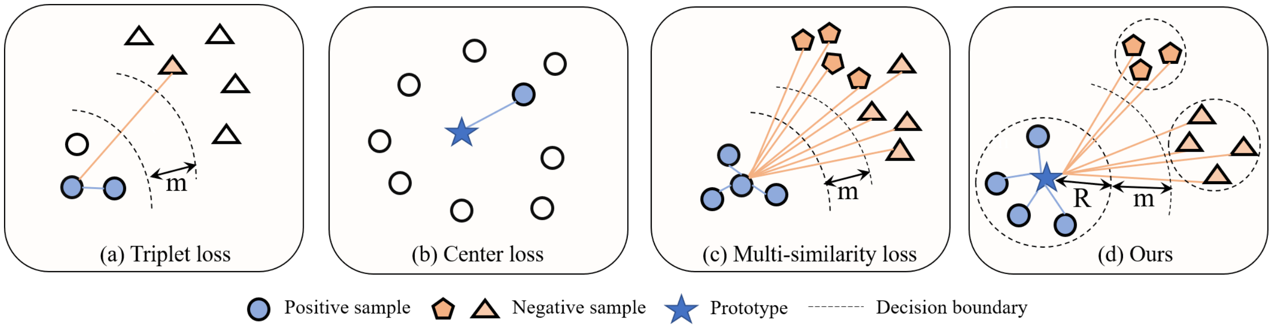



Figure 1: Comparison between popular margin-based metric learning losses. (a) Triplet loss operates on triplets of samples {anchor, positive sample, negative sample}, which associates each anchor with a positive and a negative sample. (b) Center loss simply focuses on the similarity relationship between a single sample and its class center. (c) Multi-similarity utilizes each pair of samples in a batch while ignoring the structure information inside the class. (d) Our MRM loss refines a hypersphere with a learnable radius for each class and effectively leverages the distribution information.

2020a] suggest using a negative margin to improve the novel class performance. However, the negative margin will simultaneously degrade the classification capacity for base classes. In our work, we focus on detecting unseen samples while maintaining the classification capability of known classes.

# 3 Method

## 3.1 Preliminary

Few-shot open-set recognition (FSOSR) aims to classify known classes and detect unknown ones with limited training samples. FSOSR follows the typical FSL episode training and testing strategy. Similarly to closed-set few-shot classification, for given $N$ classes with $K$ labeled supports per class, an episode is called as an $N$-way $K$-shot problem. Formally, a few-shot task can be represented as $\mathcal{T} = (\mathcal{D}^S, \mathcal{D}^Q)$. Specifically, for the support set $\mathcal{D}^S = \{x_i^S, y_i^S\}_{i=1}^{NK}$, where $x_i \in \mathcal{X}^S$ and $y_i \in \mathcal{Y}^S$ indicate sample in the support set $\mathcal{X}^S$ is from known classes with the known label space $\mathcal{Y}^S$. Different from closed-set few-shot learning, the query set includes unknown class queries which can be denoted as $\mathcal{D}^Q = \mathcal{D}^K \cup \mathcal{D}^U$, where $\mathcal{D}^K$ stands for the known set and $\mathcal{D}^U$ stands for the unknown set. The known query set is represented as $\mathcal{D}^K = \{x_i^K \in \mathcal{X}^K, y_i^K \in \mathcal{Y}^S\}_{i=1}^{NQ}$, which indicates the known query is sampled from classes with the same label space $\mathcal{Y}^S$ as support set, and $Q$ is the number of known queries in each class. The unknown query set is represented as $\mathcal{D}^U = \{x_i^U \in \mathcal{X}^U, y_i^U \in \mathcal{Y}^U\}_{i=1}^{N^U}$, where $N^U$ is the number of unknown queries and $\mathcal{Y}^S \cap \mathcal{Y}^U = \varnothing$. The goal of FSOSR is to train a model with small support set $\mathcal{D}^S$ that can classify known queries $\mathcal{D}^K$ and detect unknown queries $\mathcal{D}^U$.

## 3.2 Multi-Relation Margin Loss

Instead of generating pseudo-unknown class samples to approximate the distribution of unknowns, MRM aims to obtain a finer boundary between the closed and open spaces and further implicitly delineate the distribution of unknown classes. A question is then raised: how to learn discriminative spaces for closed and open sets from limited samples?

Inspired by margin-based methods, MRM tackles this question by optimizing margin spaces. Compared with existing methods, MRM divides the distribution of known classes into clearer regions and takes full advantage of the distribution information. The illustration and comparison of different margin-based losses and MRM loss are presented in Figure 1. Specifically, MRM mainly implements three strategies: (1) **Anchor generating**: utilize prototype as the anchor for each pair to represent global distribution in the class; (2) **Margin and radius refining**: force the margin between different classes while refining a hypersphere for each known class to preserve the distribution information; (3) **Pair-wise relation modeling**: extract multiple relationships among paired samples to cover the entire distribution information contained in a pair.

Based on these strategies, MRM loss can be formulated as:

$$\begin{aligned} L_{MRM} = \frac{1}{N}\sum_{c=1}^{N}\{\lambda R_c^2 + \frac{1}{\alpha}log[1 + \sum_{x_i^+ \in \mathcal{X}_c^+} e^{\alpha(d(p_c, f_\theta(x_i^+)) - R_c)}] \\ + \frac{1}{\beta}log[1 + \sum_{x_i^- \in \mathcal{X}_c^-} e^{-\beta(d(p_c, f_\theta(x_i^-)) - (R_c + m))}]\} \end{aligned} \tag{1}$$

Specifically, MRM works on pairs of samples that consist of positive pair $\{p_c, f_\theta(x_i^+)\}$ and negative pair $\{p_c, f_\theta(x_i^-)\}$, Where $x_i \in \mathcal{X}^S$ indicates the sample from the support set. For class $c$, the support set can be divided into positive and negative sets, including $\mathcal{X}_c^+ = \{x_i^+\}_{i=1}^K$ and $\mathcal{X}_c^- = \{x_i^-\}_{i=1}^{(N-1)K}$, where $y_i^+ = c$, $y_i^- \neq c$. $p_c$ represents the class prototype and $R_c$ denotes the radius of hypersphere. $f_\theta(\cdot)$ is the feature extractor with learnable parameters $\theta$. $d(\cdot)$ is the Euclidean distance function, $m$ stands for the margin. $\lambda$ is a constant trading off the regulation of radius and margin loss item, $\alpha$ and $\beta$ are the scaling factors for positive and negative sets respectively. The details of each strategy are as follows.

**Anchor Generating**

In this paper, we apply two types of anchor generation techniques. Initially, we utilize the class prototypes as the class-specific anchor. Moreover, we incorporate the anchor generation process with a transformation function to adapt and modify the task-specific anchor.

**Class-specific anchor.** The class prototype [Snell *et al.*, 2017] is proposed based on the assumption that there exists an embedding space where points cluster around a single prototype representation for each class. Therefore, MRM utilizes the class prototype as the anchor for each pair instead of an individual sample to achieve the representative features of known classes. The class prototype is formulated as the mean vector of support embedding vectors:

$$p_c = \frac{1}{|K|} \sum_{x_i \in \mathcal{X}_c^S} f_\theta(x_i) \tag{2}$$

where $\mathcal{X}_c^S$ is the support set labeled to class $c$. The central idea is to learn more centralized features by forcing the samples closer to their corresponding prototypes and further to samples from different classes. With the anchor $p_c$, MRM could obtain a finer boundary between the positive and negative set by a margin $m$. Specifically, MRM enforces the margin between the positive pair $\{p_c, f_\theta(x_i^+)\}$ and negative pair $\{p_c, f_\theta(x_i^-)\}$:

$$d(p_c, f_\theta(x_i^-)) > d(p_c, f_\theta(x_i^+)) + m \tag{3}$$

By utilizing the class-specific anchor $p$, we can derive the class-specific multi-relation margin loss, denoted as $L_{MRM_C}$, based on Eq.(1).

**Task-specific anchor.** According to previous studies, transformation functions are demonstrated to provide effective FSOSR performance [Jeong *et al.*, 2021; Ma *et al.*, 2022b]. Inspired by these methods, we apply a standard set-to-set function to modify more distinguishable prototypes as anchors to consider the entire distribution of the task [Ye *et al.*, 2020]. The set-to-set function transforms a set of original prototypes $P = \{p_c\}_{c=1}^N$ to a set of adapted prototypes $P' = \{p'_c\}_{c=1}^N$, which projects the input feature into query, key, and value space with learnable transformation matrices. The training process is as follows:

$$\begin{aligned} A(P) &= (W_v P)(s(\frac{(W_q P)^T (W_k P)}{\sqrt{M}})^T) \\ P' &= \sigma(P + A(P)) \end{aligned} \tag{4}$$

where $A$ is the attention weights matrix. $W_q, W_k$, and $W_v \in \mathbb{R}^{M \times M}$ are learnable projection kernels for query, key, and value space respectively. $\sigma$ indicates the layer normalization, and $s(\cdot)$ is the softmax function. By incorporating the modified prototype $p'$, we can obtain the task-specific multi-relation margin loss denoted as $L_{MRM_T}$, which adapts the original prototype $p$ as described in Eq.(1).

### Margin and Radius Refining

Previous algorithms mostly attempted to compress sample points of the same class as much as possible. Consequently, these samples will be shrunk into one point in the embedding space and lose their similarity structure inside the class, which degrades the recognition performance. Inspired by [Wang *et al.*, 2019a], instead of pulling intra-class samples as compact as possible, MRM learns a hypersphere for each known class with a learnable radius to preserve the distribution structure inside it. Specifically, MRM forces the distance between the class prototype and positive sample to be smaller than a threshold, which is the radius of each class's hypersphere. Mathematically,

$$d(p_c, f_\theta(x_i^+)) < R_c \tag{5}$$

where $x_i^+ \in \mathcal{X}_c^+$ and $R_c$ is the hypersphere radius of class $c$. Details about updating the hypersphere by optimizing the radius are explained in Section 4.2. Hence, we can obtain a distinct decision boundary for each class while taking the prototype as a center point. Based on the positive boundary, the margin between positive and negative sets in Eq.(3) can be refined as:

$$d(p_c, f_\theta(x_i^-)) > R_c + m \tag{6}$$

Since the number of negative pairs is much larger than that of positive pairs in a few-shot task, we adjust the radius based on the negative pairs which are more informative. According to Eq.(6), $R$ is limited to the distance between the anchor and negative samples, we can approximate it by calculating the negative distance. Besides, we add a trade-off hyperparameter that allows some points to be mapped outside the sphere. For class $c$, the hypersphere could be updated by adjusting the radius $R_c$:

$$R_c = h_\nu(\{d(p_c, f_\theta(x_i^-)) - m\}_{i=1}^{(N-1)K}) \tag{7}$$

Where $h(\cdot)$ is the quantile function and $\nu$ is the hyperparameter that controls violations of the boundary.

### Pair-Wise Relation Modeling

In order to utilize the entire distribution information contained in a pair, MRM extracts self-relations and relative relations among paired samples. Self-relation is denoted as the similarity relation between the anchor and positive or negative sample in a pair. Relative relation is computed by considering correlations of other pairs. According to previous research [Wang *et al.*, 2019b], they both have a significant impact on exploiting the potentiality of each pair and building a discriminative distribution structure, whereas most existing methods only partially explore this factor.

As Multi-similarity loss suggests, pair-wise relation could be modeled through sample weighting, and the partial derivative of the loss function with respect to each pair is its weight. Paired samples with higher weights are considered to be more informative in the optimizing process. For instance, negative pairs that are in close proximity to the anchor and closer to the anchor than other negative samples are assigned higher weights, which means such negative samples are more difficult to classify, and thus carry more valuable distribution information.

Since the main goal of MRM is to encourage positive pairs to be closer while pushing negative pairs away from each other, it is reasonable to assume $\frac{\partial L_{MRM}}{\partial d(p_c, f_\theta(x_i^+))} \geq 0$ for the positive pair $\{p_c, f_\theta(x_i^+)\}$, thus its weight $w_i^{c^+}$ can be computed as:

$$\begin{aligned} w_i^{c^+} &= \frac{\partial L_{MRM}}{\partial d(p_c, f_\theta(x_i^+))} \\ &= \frac{e^{\alpha(d(p_c, f_\theta(x_i^+)) - R_c)}}{1 + \sum\limits_{x_k^+ \in \mathcal{X}_c^+} e^{\alpha(d(p_c, f_\theta(x_k^+)) - R_c)}} \end{aligned} \tag{8}$$

where the weight is computed jointly from its self-relation by measuring the anchor and positive sample in $e^{\alpha(d(p_c, f_\theta(x_i^+))-R_c)}$, and relative relations by comparing to other positive pairs $\{p_c, f_\theta(x_k^+)\}$ with the same anchor $p_c$ in the item $d(p_c, f_\theta(x_+^k)) - R_c$. Similarly, for the negative pair $\{p_c, f_\theta(x_i^-)\}$, we assume $\frac{\partial L_{MRM}}{\partial d(p_c, f_\theta(x_i^-))} \leq 0$, and the weight $w_i^{c^-}$ is formulated as:

$$\begin{aligned} w_i^{c^-} &= -\frac{\partial L_{MRM}}{\partial d(p_c, f_\theta(x_i^-))} \\ &= \frac{e^{-\beta(d(p_c, f_\theta(x_i^-))-(R_c+m))}}{1 + \sum\limits_{x_k^- \in \mathcal{X}_c^-} e^{-\beta(d(p_c, f_\theta(x_k^-))-(R_c+m))}} \end{aligned} \tag{9}$$

**Remark.** In comparison to existing margin-based losses, MRM offers several key advantages: (1) MRM divides the distribution of known classes into distinct regions and employs a learnable radius centered on the prototype to refine the distribution of each class. This allows MRM to detect unknown samples that are located away from known regions, while existing methods only construct ambiguous decision boundaries. (2) MRM utilities both intra-class and inter-class distribution of samples to learn an appropriate radius for each known class, efficiently preventing overfitting and enhancing generalization in the few-shot setting, which is not addressed by existing methods.

# 4 Experiments

## 4.1 Datasets and Evaluation Protocols

Following previous studies, we conduct experiments on three public benchmark datasets CUB-200 [Wah *et al.*, 2011], *tiered*ImageNet [Ren *et al.*, 2018], and *mini*ImageNet [Vinyals *et al.*, 2016] to verify the effectiveness of MRM.

We evaluate the model with closed-set classification accuracy (Acc) and the area under ROC curve (AUROC) for unknown class detection. The Acc measures the classification capacity via known samples, and the AUROC evaluates the unknown detection capacity via both known and unknown samples. Since prior FSL studies have already demonstrated promising results under the closed-set assumption, the main challenge of FSOSR is how to obtain better AUROC while maintaining the Acc. Following [Liu *et al.*, 2020b], we set $N = 5$ and $K = 1, 5$ during meta-training and meta-testing. For each episode, we sample 5 known classes and 5 unknown classes, with each containing 15 queries. Code is available at https://github.com/Casie-che/MRM.

## 4.2 Implementation Details

**Backbone training.** Following previous FSOSR methods [Jeong *et al.*, 2021], we use ResNet12 [He *et al.*, 2016] based architecture as the feature extractor which creates 640-dimensional feature vectors before the classifier. As [Ye *et al.*, 2020] suggest, we first pre-train the backbone with a simple classifier to classify all known classes with the cross-entropy loss. The classification performance of sampled 1-shot tasks over the penultimate layer embedding is evaluated to identify the best pretrained model, whose weights are subsequently utilized to initialize the embedding function. Then we use a SGD optimizer to train the feature extractor and transformers over 200 epochs. The initial learning rate is set to 0.0002 for the feature extractor and 0.002 for transformers with a multi-step learning rate schedule. MRM is finetuning the feature extractor over 30 epochs with 0.0005 weight decay. During training, We use the validation set to select the best model. The comparison results are calculated over 600 evaluation episodes on the test set.

**Hypersphere updating.** To construct appropriate decision boundaries for each known class, we utilize dynamic radii instead of a fixed threshold during training. Since the network parameters $\theta$ and radius of hypersphere $R$ are generally on different scales, using a common optimal learning rate may be inefficient. Inspired by [Ruff *et al.*, 2018], we optimize $\theta$ and $R$ in an alternating minimization coordinate descent approach. We train the network parameters $\theta$ firstly while the radius $R$ is fixed, then after one epoch, we calculate the radius for each class based on the embeddings extracted from the network of the latest update.

**Classifier setting.** Inspired by [Jeong *et al.*, 2021], we utilize two classifiers to detect unknown samples. The first uses the distance between a query and its predicted class prototype, which is denoted as Distance. The other is SnaTCHerF, which measures the distance between the transformed prototypes and original class prototypes. To evaluate the efficiency of MRM, we first plug $L_{MRM_C}$ in different few-shot methods. We finetune the ProtoNet [Snell *et al.*, 2017] as the representative of class-specific methods with MRM loss and Distance classifier to get MRM$_C$-ProtoNet. Then plug $L_{MRM_T}$ in FEAT [Ye *et al.*, 2020] to get MRM$_T$-FEAT and MRM$_T$-SnaTCHerF, which respectively utilize Distance and SnaTCHer as the classifier.

**Experimental setting.** For simplicity, we maintain the same hyper-parameters of our methods for all the experiments which are obtained from the validation set. We $\lambda = 0.1, \alpha = 1, \beta = 3, m = 1$ for the loss function in Eq.(1), and $\nu = 0.1$ for hypersphere updating in Eq.(7).

## 4.3 Comparison with Prior Work

In order to validate the effectiveness of MRM, we conduct thorough comparisons of few-shot methods plugged in the MRM loss with state-of-the-art results in FSL. Additionally, we compare the performance of MRM with classical margin-based losses which are widely used in deep metric learning.

**Comparison with Few-Shot Methods**

We first compare with the standard FSL methods, including baseline [Chen *et al.*, 2019], ProtoNet [Snell *et al.*, 2017] and FEAT [Ye *et al.*, 2020]. Then we select conventional OSR methods, including NN [Mendes Júnior *et al.*, 2017], OpenMax [Bendale and Boult, 2016], and CounterFactual [Neal *et al.*, 2018]. Next, we choose PEELER [Liu *et al.*, 2020b], SnaTCHer [Jeong *et al.*, 2021], ATT [Huang *et al.*, 2022], and ProCAM [Song *et al.*, 2022] as representatives of FSOSR methods.

| Method | CUB-200 5-way | | | |
|---|---|---|---|---|
| | 1-shot | | 5-shot | |
| | Acc | AUROC | Acc | AUROC |
| baseline | 61.24 | 68.43 | 76.23 | 75.09 |
| PEELER | 62.62 | 57.26 | 82.44 | 65.01 |
| ProCAM | 65.88 | 75.88 | 81.14 | 83.70 |
| ProCAMsm | 68.54 | 76.05 | 82.22 | 82.34 |
| ProtoNet | 57.31 | 60.32 | 73.19 | 64.55 |
| MRM$_C$-ProtoNet (ours) | 60.44 ± 0.86 | 63.16 ± 0.86 | 76.58 ± 0.63 | 69.25 ± 0.68 |
| FEAT | 69.03 ± 0.84 | 63.22 ± 0.81 | 83.77 ± 0.55 | 69.74 ± 0.75 |
| MRM$_T$-FEAT (ours) | **70.00 ± 0.85** | 75.17 ± 0.74 | **84.51 ± 0.52** | 82.88 ± 0.60 |
| SnaTCHer-F | 69.03 ± 0.84 | 74.34 ± 0.73 | 83.77 ± 0.55 | 84.57 ± 0.57 |
| MRM$_T$-SnaTCHerF (ours) | **70.00 ± 0.85** | **76.30 ± 0.70** | **84.51 ± 0.52** | **85.91 ± 0.50** |

Table 1: Comparison with the state-of-the-art methods on CUB-200. Average closed-set classification accuracies (%) and average unknown detection AUROCs (%) over 600 test episodes with 95% confidence intervals. We cite the baseline results based on [Song *et al.*, 2022].

| Method | *tiered*ImageNet 5-way | | | |
|---|---|---|---|---|
| | 1-shot | | 5-shot | |
| | Acc | AUROC | Acc | AUROC |
| NN | 67.73 ± 0.96 | 62.70 ± 0.72 | 83.43 ± 0.66 | 69.77 ± 0.75 |
| OpenMax | 68.28 ± 0.95 | 60.13 ± 0.74 | 83.48 ± 0.66 | 65.51 ± 0.83 |
| CounterFactual | 70.08 ± 0.94 | 71.04 ± 0.80 | 85.36 ± 0.60 | 78.66 ± 0.62 |
| PEELER | 69.51 ± 0.92 | 65.20 ± 0.76 | 84.10 ± 0.66 | 73.27 ± 0.71 |
| SnaTCHer-T | 70.45 ± 0.95 | 74.84 ± 0.79 | 84.42 ± 0.68 | 82.03 ± 0.66 |
| SnaTCHer-L | 70.85 ± 0.99 | 74.95 ± 0.83 | 85.23 ± 0.64 | 80.81 ± 0.68 |
| ProCAMsm | 68.82 | 75.55 | **85.64** | 82.77 |
| ATT | 69.34 ± 0.95 | 72.74 ± 0.78 | 83.82 ± 0.63 | 78.66 ± 0.6 |
| ProtoNet | 68.26 ± 0.96 | 60.73 ± 0.80 | 83.40 ± 0.65 | 64.96 ± 0.83 |
| MRM$_C$-ProtoNet (ours) | 68.35 ± 0.94 | 72.99 ± 0.74 | 84.75 ± 0.62 | 78.18 ± 0.60 |
| FEAT | 70.52 ± 0.96 | 63.54 ± 0.76 | 84.74 ± 0.69 | 70.74 ± 0.75 |
| MRM$_T$-FEAT (ours) | **71.13 ± 0.91** | 74.30 ± 0.78 | 85.27 ± 0.62 | 80.83 ± 0.64 |
| SnaTCHer-F | 70.52 ± 0.96 | 74.28 ± 0.80 | 84.74 ± 0.69 | 82.02 ± 0.64 |
| MRM$_T$-SnaTCHerF (ours) | **71.13 ± 0.91** | **75.59 ± 0.77** | 85.27 ± 0.62 | **83.03 ± 0.63** |

Table 2: Comparison with the state-of-the-art methods on *tiered*ImageNet. Average closed-set classification accuracies (%) and average unknown detection AUROCs (%) over 600 test episodes with 95% confidence intervals. We cite the baseline results based on [Song *et al.*, 2022] and [Huang *et al.*, 2022].

We illustrate the average closed-set classification accuracies and unknown detection AUROCs over 600 test episodes with 95% confidence intervals under 5-way 5-shot and 1-shot settings respectively. We cite most of the baseline results based on [Huang *et al.*, 2022] and [Song *et al.*, 2022]. Table 1 demonstrates the results for CUB-200, Table 2 and Table 3 show comparisons on *tiered*ImageNet and *mini*ImageNet respectively. For each setting, the best results are highlighted in bold.

According to the results, we can find that OSR methods show unsatisfactory classification accuracies in the few-shot setting, whereas FSL methods perform poorly in unknown class detection. On the contrary, methods plugged in MRM loss can improve the unknown detection of AUROC by a significant margin and maintain the closed-set accuracy. For instance, on CUB-200, *tiered*ImageNet, and *mini*ImageNet, MRM$_C$-ProtoNet has around 13%, 13%, and 16% performance improvements compared with the vanilla ProtoNet under 1-shot setting, and 5%, 13%, 17% improvements under 5-shot setting, while correctly classifying the closed-set. Results further verify the effectiveness of the combination of MRM and transformation function. Compared with state-of-the-art FSOSR methods, methods with $L_{MRM_T}$ achieve better AUROC in most settings without relying on additional pseudo-unknown class samples.

| Method | *mini*ImageNet 5-way | | | |
|---|---|---|---|---|
| | 1-shot | | 5-shot | |
| | Acc | AUROC | Acc | AUROC |
| NN | 63.82 ± 0.85 | 56.96 ± 0.75 | 80.12 ± 0.57 | 63.43 ± 0.76 |
| OpenMax | 63.69 ± 0.84 | 62.64 ± 0.80 | 80.56 ± 0.58 | 62.27 ± 0.71 |
| CounterFactual | 63.7 ± 0.83 | 64.17 ± 0.88 | 81.44 ± 0.54 | 71.58 ± 0.76 |
| PEELER | 65.86 ± 0.85 | 60.57 ± 0.83 | 80.61 ± 0.59 | 67.35 ± 0.80 |
| SnaTCHer-T | 66.60 ± 0.80 | 70.17 ± 0.88 | 81.77 ± 0.53 | 76.66 ± 0.78 |
| SnaTCHer-L | 67.60 ± 0.83 | 69.40 ± 0.92 | 82.36 ± 0.58 | 76.15 ± 0.83 |
| ProCAMsm | **67.86** | 71.09 | 83.66 | 77.51 |
| ATT | 67.64 ± 0.81 | **71.35 ± 0.68** | 82.31 ± 0.49 | 79.85 ± 0.58 |
| ProtoNet | 64.01 ± 0.88 | 51.81 ± 0.93 | 80.09 ± 0.58 | 60.39 ± 0.92 |
| MRM$_C$-ProtoNet (ours) | 64.05 ± 0.82 | 68.11 ± 0.73 | **84.73 ± 0.51** | 77.07 ± 0.63 |
| FEAT | 67.02 ± 0.85 | 57.01 ± 0.84 | 82.02 ± 0.53 | 63.18 ± 0.78 |
| MRM$_T$-FEAT (ours) | 67.03 ± 0.83 | 71.22 ± 0.83 | 82.00 ± 0.55 | 78.99 ± 0.67 |
| SnaTCHer-F | 67.02 ± 0.85 | 68.27 ± 0.96 | 82.02 ± 0.53 | 77.42 ± 0.73 |
| MRM$_T$-SnaTCHerF (ours) | 67.03 ± 0.83 | 71.20 ± 0.80 | 82.00 ± 0.55 | **80.39 ± 0.59** |

Table 3: Comparison with the state-of-the-art methods on *mini*ImageNet. Average closed-set classification accuracies (%) and average unknown detection AUROCs (%) over 600 test episodes with 95% confidence intervals. We cite the baseline results based on [Song *et al.*, 2022] and [Huang *et al.*, 2022].

**Comparison with Margin-Based Methods**

To provide a more comprehensive explanation of the advantages of MRM loss, we conduct a comparative analysis with existing margin-based losses, including Triplet loss [Schroff *et al.*, 2015], Center loss [Wen *et al.*, 2016], and Multi-similarity loss [Wang *et al.*, 2019b]. We illustrate the average unknown detection AUROCs over 600 test episodes on 5-way 5-shot *mini*ImageNet tasks in Figure 2. Remarkably, results clearly demonstrate that MRM surpasses the performance of the compared margin-based losses. By dividing the distribution of known classes into clearer regions and refining the distribution of each class, MRM enables effective identification of unknown samples located away from known regions. Through the comparison of unknown detection AUROCs, MRM demonstrates its superiority in building a highly discriminative distribution structure.

## 4.4 Ablation Study

**Visualization of the effect of MRM loss.** We plot t-SNE [van der Maaten and Hinton, 2008] figure to visualize the distributions of images sampled from *mini*ImageNet in the feature space learned by standard softmax loss and MRM loss in Figure 3, respectively. In Figure 3a, the feature extractor learned by softmax loss can only construct ambiguous decision boundaries while ignoring the distribution of unknowns. It might easily lead to confusion between unknown and known samples as all features just follow one distribution. Figure 3b shows that the feature extractor trained by MRM loss can obtain a finer boundary between classes and a distinct decision boundary for each class, which enables the model to detect unknown samples by rejecting samples away from closed boundaries while maintaining the classification

| Method | Anchor generating | Radius refining | Relation modeling | 1-shot | 5-shot |
|---|---|---|---|---|---|
| MRM #1 | | | | $65.41 \pm 0.85$ | $74.80 \pm 0.84$ |
| MRM #2 | ✓ | | | $65.95 \pm 0.85$ | $75.42 \pm 0.57$ |
| MRM #3 | ✓ | ✓ | | $66.89 \pm 0.73$ | $76.35 \pm 0.57$ |
| Our model | ✓ | ✓ | ✓ | $68.11 \pm 0.73$ | $77.07 \pm 0.63$ |

Table 4: Ablation study on different strategies for MRM loss on 5-way *mini*ImageNet tasks. Average unknown detection AUROCs (%) over 600 test episodes with 95% confidence intervals.

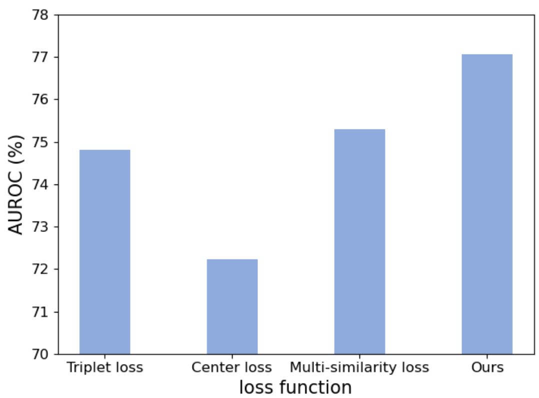
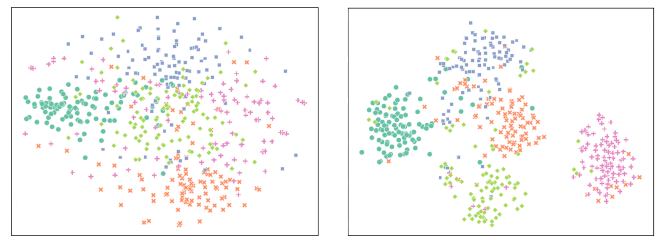
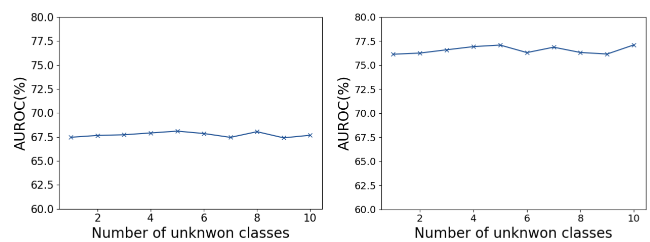



Figure 2: Comparisons of existing margin-based losses with MRM loss on 5-way 5-shot *mini*ImageNet tasks. Average unknown detection AUROCs (%) over 600 test episodes.



(a) Softmax loss (b) Ours: MRM loss

Figure 3: Visualization of distributions on images optimized with standard softmax loss (a) and MRM loss (b). The figure is plotted by applying the t-SNE of ConvNet4 features of samples from *mini*ImageNet. Different colors represent distinct classes in a few-shot task.




(a) 1-shot (b) 5-shot

Figure 4: Performances on different numbers of unknown classes on 5-way *mini*ImageNet tasks. Average unknown detection AUROCs (%) over 600 test episodes.

capability.

**Effect of different strategies for MRM loss.** We perform our ablation study on the 5-way 1-shot and 5-shot tasks on *mini*ImageNet to analyze the effectiveness of different strategies in Table 4. Since the main challenge of FSOSR is how to improve detection capability for unknown classes, we mainly focus on the performance of AUROC. We compare our full model with a number of stripped down versions by evaluating the AUROC of ProtoNet plugged in different margin losses and MRM$_C$-ProtoNet. Specifically, 'MRM #1' utilizes the individual sample as an anchor in a pair and forces the margin between positive and negative samples. 'MRM #2' utilizes class prototypes as anchors. On that basis, 'MRM #3' refines a hypersphere for each known class with a learnable radius. The results show an increasing trend in unknown class AUROC, which indicates that integrating a large margin to the loss function could improve the performance efficiently, and MRM could further build a more discriminative distribution structure to boost FSOSR performance.

**Performances on different numbers of unknown classes.** During the conducted experiments, we maintain a fixed number of 5 unknown classes. However, in real-world applications, the presence of unknown samples can vary. To account for this, we investigate the performance of AUROC for MRM$_C$-ProtoNet in different unknown class settings, specifically focusing on 5-way 1-shot and 5-shot tasks on *mini*ImageNet. We vary the number of unknown classes from 1 to 10, while keeping the loss function and classifier unchanged. As shown in Figure 4, MRM demonstrates robust performance across different unknown class configurations.

## 5 Conclusion

In this paper, we propose a novel loss function named multi-relation margin (MRM) that can plug in existing few-shot methods to boost the performance of few-shot open-set recognition. We first analyze the limitations of pseudo-unknown based approaches, then propose to enlarge the margin between different classes and refine the decision boundaries for known classes to implicitly delineate the distribution of unknowns without relying on pseudo-unknown samples. To better capture the distribution information, we further construct multiple relationships among paired samples. Detailed experiments reveal that MRM indeed helps to delineate the distribution of unknowns and leads to better unknown detection of AUROC while correctly classifying the known classes.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62076062) and the Social Development Science and Technology Project of Jiangsu Province ( No.BE2022811). Furthermore, the work was also supported by the Collaborative Innovation Center of Wireless Communications Technology and the Big Data Computing Center of Southeast University.

## References

[An *et al.*, 2021] Yuexuan An, Hui Xue, Xingyu Zhao, and Lu Zhang. Conditional Self-Supervised Learning for Few-Shot Classification. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2021.

[An *et al.*, 2023] Yuexuan An, Hui Xue, Xingyu Zhao, and Jing Wang. From Instance to Metric Calibration: A Unified Framework for Open-World Few-Shot Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Bendale and Boult, 2016] Abhijit Bendale and Terrance E. Boult. Towards Open Set Deep Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[Chen *et al.*, 2019] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A Closer Look at Few-shot Classification. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2019.

[Finn *et al.*, 2017] Chelsea Finn, P. Abbeel, and S. Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *Proceedings of International Conference on Machine Learning (ICML)*, 2017.

[Geng *et al.*, 2020] Chuanxing Geng, Sheng-jun Huang, and Songcan Chen. Recent Advances in Open Set Recognition: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[Hsu *et al.*, 2020] Yen-Chang Hsu, Yilin Shen, Hongxia Jin, and Zsolt Kira. Generalized ODIN: Detecting Out-of-Distribution Image Without Learning From Out-of-Distribution Data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[Huang *et al.*, 2022] Shiyuan Huang, Jiawei Ma, Guangxing Han, and Shih-Fu Chang. Task-Adaptive Negative Envision for Few-Shot Open-Set Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[Jeong *et al.*, 2021] Minki Jeong, Seokeon Choi, and Changick Kim. Few-shot Open-set Recognition by Transformation Consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[Li *et al.*, 2020] Aoxue Li, Weiran Huang, Xu Lan, Jiashi Feng, Zhenguo Li, and Liwei Wang. Boosting Few-Shot Learning With Adaptive Margin Loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[Liu *et al.*, 2020a] Bin Liu, Yue Cao, Yutong Lin, Qi Li, Zheng Zhang, Mingsheng Long, and Han Hu. Negative Margin Matters: Understanding Margin in Few-Shot Classification. In *Proceedings of the IEEE/CVF European Conference on Computer Vision (ECCV)*, 2020.

[Liu *et al.*, 2020b] Bo Liu, Hao Kang, Haoxiang Li, Gang Hua, and Nuno Vasconcelos. Few-Shot Open-Set Recognition Using Meta-Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[Ma *et al.*, 2022a] Rongkai Ma, Pengfei Fang, Gil Avraham, Yan Zuo, Tianyu Zhu, Tom Drummond, and Mehrtash Harandi. Learning instance and task-aware dynamic kernels for few-shot learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

[Ma *et al.*, 2022b] Rongkai Ma, Pengfei Fang, Tom Drummond, and Mehrtash Harandi. Adaptive poincaré point to set distance for few-shot classification. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2022.

[Mendes Júnior *et al.*, 2017] Pedro R. Mendes Júnior, Roberto M. de Souza, Rafael de O. Werneck, Bernardo V. Stein, Daniel V. Pazinato, Waldir R. de Almeida, Otávio A. B. Penatti, Ricardo da S. Torres, and Anderson Rocha. Nearest neighbors distance ratio open-set classifier. *Machine Learning*, 2017.

[Neal *et al.*, 2018] Lawrence Neal, Matthew Olson, Xiaoli Fern, Weng-Keen Wong, and Fuxin Li. Open Set Learning with Counterfactual Images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[Oza and Patel, 2019] Poojan Oza and Vishal M. Patel. C2AE: Class Conditioned Auto-Encoder for Open-Set Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[Pal *et al.*, 2022] Debabrata Pal, Valay Bundele, Renuka Sharma, Biplab Banerjee, and Yogananda Jeppu. Few-Shot Open-Set Recognition of Hyperspectral Images With Outlier Calibration Network. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022.

[Pal *et al.*, 2023] Debabrata Pal, Shirsha Bose, Biplab Banerjee, and Yogananda Jeppu. MORGAN: Meta-Learning-Based Few-Shot Open-Set Recognition via Generative Adversarial Network. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023.

[Perera *et al.*, 2020] Pramuditha Perera, Vlad I. Morariu, Rajiv Jain, Varun Manjunatha, Curtis Wigington, Vicente Ordonez, and Vishal M. Patel. Generative-Discriminative Feature Representations for Open-Set Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[Ren *et al.*, 2018] Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B. Tenenbaum, Hugo Larochelle, and Richard S. Zemel. Meta-Learning for Semi-Supervised Few-Shot Classification. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018.

[Ruff *et al.*, 2018] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep One-Class Classification. In *Proceedings of the International Conference on Machine Learning (ICML)*, July 2018.

[Schroff *et al.*, 2015] Florian Schroff, Dmitry Kalenichenko, and James Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[Snell *et al.*, 2017] Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical Networks for Few-shot Learning. In *Advancesof Neural Information Processing Systems (NeurIPS)*, 2017.

[Sohn, 2016] Kihyuk Sohn. Improved Deep Metric Learning with Multi-class N-pair Loss Objective. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

[Song *et al.*, 2016] Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese. Deep Metric Learning via Lifted Structured Feature Embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[Song *et al.*, 2022] Nan Song, Chi Zhang, and Guosheng Lin. Few-shot Open-set Recognition Using Background as Unknowns. In *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, 2022.

[Sun *et al.*, 2020] Xin Sun, Zhenning Yang, Chi Zhang, Keck-Voon Ling, and Guohao Peng. Conditional Gaussian Distribution Learning for Open Set Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[van der Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 2008.

[Vinyals *et al.*, 2016] Oriol Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and Daan Wierstra. Matching Networks for One Shot Learning. In *Advances of Neural Information Processing Systems (NeurIPS)*, 2016.

[Wah *et al.*, 2011] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. Technical Report CNS-TR-2010-001, California Institute of Technology, 2011.

[Wang *et al.*, 2019a] Xinshao Wang, Yang Hua, Elyor Kodirov, Guosheng Hu, Romain Garnier, and Neil M. Robertson. Ranked List Loss for Deep Metric Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[Wang *et al.*, 2019b] Xun Wang, Xintong Han, Weilin Huang, Dengke Dong, and Matthew R. Scott. Multi-Similarity Loss With General Pair Weighting for Deep Metric Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[Wen *et al.*, 2016] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A Discriminative Feature Learning Approach for Deep Face Recognition. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016.

[Ye *et al.*, 2020] Han-Jia Ye, Hexiang Hu, De-Chuan Zhan, and Fei Sha. Few-Shot Learning via Embedding Adaptation With Set-to-Set Functions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.