Ancestor Search: Generalized Open Set Recognition via Hyperbolic Side Information Learning

Xiwen Dengxiong

Golisano College of Computing and Information Sciences

Rochester Institute of Technology

sd6384@rit.edu

Yu Kong

Department of Computer Science and Engineering

Michigan State University

yukong@msu.edu

Abstract

Different from the open set recognition, generalized open set recognition learns the most similar known classes for unseen samples using known classes samples and side information of known classes. It is challenging because hierarchically structured side information is distorted when features are embedded in the Euclidean space in existing literature, which incurs the difficulty of identifying the unseen samples. In this paper, we introduce a side information learning algorithm for generalized open set recognition based on the hyperbolic space to alleviate the distortion and accurately identify the unknown samples. Specifically, we propose a hyperbolic side information learning framework to identify the unseen samples and an ancestor search algorithm to search the most similar ancestor from the taxonomy of selected known classes. Experiments on CUB-200 and AWA 2 datasets show that our method improves the performance of generalized open set recognition by a large margin.

1. Introduction

Generalized open set recognition (GOSR) is an important computer vision problem that recognizes unknown classes[1] and identifies new categories based on the side information of known classes[2] such as taxonomy and attribute annotations[16]. Unlike the open set recognition [41, 7, 8, 44] that simply rejecting novel samples as unknown, generalized open set recognition further classifies the super-classes [25]. For example, in Figure 1, GOSR recognizes the unknown animal (leopard) as a *Feline* or *Placental* based on the tree-structured taxonomy information rather than simply reject it as an unknown sample. This is beneficial if a system is able to provide more information about the unseen samples so that we can have a better un-

---

[1]Unknown classes refer to classes without information during training.

[2]Known classes refer to the classes given in training.
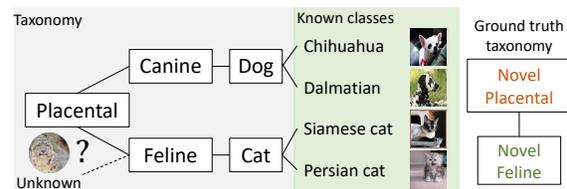
Figure 1. Generalized open set recognition further classifies unknown samples based on the side information (taxonomy). The goal of generalized open set recognition is to find the closest ancestor of the unseen samples from ancestor nodes. The unknown animal leopard is a novel feline. In GOSR problem, recognizing the leopard as novel placental is partially correct, and other predictions are wrong.

derstanding of the unknown samples.

GOSR is a challenging problem because the variability of unknown samples cannot be effectively captured by the existing approach [25]. In GOSR problem, training samples for unknown classes are not provided, solely capturing the variability of unknown samples from known training samples [25] degrades the recognition capability of the classifier. Although side information, including taxonomy and attribute annotations, is available in GOSR problem, directly embedding side information in the existing Euclidean classifiers [25] may lead to a high distortion for fine-grained classes. In addition, the number of classes in grows exponentially when the taxonomy becomes deeper, which impacts the recognition capability of the existing approaches. Therefore, how to effectively learn the side information is the key to recognizing the unknown samples.

To effectively learn the side information, this paper proposes a side information learning approach by generating a novel attribute feature in the hyperbolic space and an ancestor search algorithm to search the closest ancestor node for unseen samples in the taxonomy of known classes. Hyperbolic space is a non-Euclidean space, and its space capacity grows exponentially. According to its geometry property

[30], the hyperbolic space can be viewed as a continuous tree, and the tree-structured side information can be embedded in the hyperbolic space with minimal distortion [17]. Generating novel attribute features in hyperbolic space can better capture the variability of the unknown samples.

Specifically, we first propose hyperbolic placeholders to set aside a large area for unknown samples and an attribute generator in the hyperbolic space to generate novel attributes for unknown samples. Then we propose a novel attribute generator to generate novel attribute features based on the selected attribute features from known classes. Besides, we propose an ancestor node searching algorithm for recognizing the unknown on the taxonomy of similar known classes. Compared with the existing GOSR work [25], our novel attribute feature generator better captures the variability of the unknown classes. Compared with the existing Euclidean-based open set recognition work [44], our hyperbolic placeholder is more representative and could learn the side information from known classes, which enables hierarchical recognition of the unknown samples. Compared with the generative method [21], which is unable to further recognize the ancestor of the unknown, our ancestor search framework could achieve this and superior to [35, 3] because our method better describes the relationship between attributes and the known sample by the attribute generator and similarity in the hyperbolic space. Additionally, we develop hierarchical similarity indexes to evaluate the similarity between known and unknown samples for the GOSR problem because existing open set evaluation metrics do not consider the hierarchical distance of the predicted result.

Our contributions are summarized as follows:

- We propose a hyperbolic side information learning framework that better captures the variability of unknown samples.

- We propose an ancestor search algorithm to recognize the unseen classes via tree-structured taxonomy.

- We propose new measurements to evaluate the GOSR problem, and empirical results show the superiority of our method to state-of-the-art.

2. Related Work

Generalized Open Set Recognition GOSR detects a test sample that is significantly different from representative training data without prior knowledge. The GOSR can use the side information (semantic information in taxonomy) from the training classes. Some previous object recognition works incorporate the hierarchical taxonomy in the classification task. The music genre classification work[35] detects the novel class utilizing the hierarchical data. Zhao [43] estimates the predicted label and ground truth with hierarchy in the open set scene parsing framework.

Similar to GOSR problem, Mancini [27] proposes an open-world compositional framework to recognize unseen samples using attribute annotation under zero-shot learning scenario. Based on attribute annotations, the compositional framework first extracts concepts from images and generates novel images according to the concepts. Different from the open-world compositional framework, the GOSR framework only needs side information from known classes. The zero-shot learning framework cannot classify unknown classes without side information from unknown classes. Our task is similar to the novel detection task elucidated in [25]. Even though the proposed method considers the hierarchical relationships in recognizing the unseen samples, the top-down and flatten method [25] does not expand the gap between known and unknown classes, which impacts the novel class detection accuracy. Compared with another recognition work [3], the hierarchy leveraging method does not work when the probability of unknown classes is different from their assumption.

Open Set Recognition Pioneer work [2] in the open set recognition replaces the Softmax layer by adding the Weibull distribution fitting score to compute the pseudo-activation for unseen classes. The discriminative model and generative model are two mainstreams in the modern open set recognition model. Some applications [10, 19, 34, 32, 1] have applied discriminative models in image, video, and text recognition domains. Besides, the classification-reconstruction learning algorithm [40] improves the robustness of the unknown classes classification and decreases the misclassification in the known classes. The generative model improves the open set recognition task by providing more generated unseen samples [14, 41, 15, 44]. Zhou *et al*. [44] combines both discriminative and generative parts in their proposed method. The first part is to learn a multi-threshold schema to classify the known and the unknown. In the second part, the model generates unseen images from the known classes by randomly mixing-up samples from two different known classes. Compared to [44], our method embeds the placeholder in the hyperbolic space and uses the distance between two known classes to determine the attribute feature generating process. We also develop an approach to generating novel attributes for unknown objects.

Hyperbolic Embedding To mitigate the distortion of hierarchical class embedding, Riemannian optimization algorithms are developed in early pioneering literature [37, 4]. Following the Riemannian optimization, researchers leveraged hyperbolic geometry to naturally address the class hierarchy. The Poincaré ball model [36] embeds the hierarchical relation using a hyperbolic entailment cone. Based on the Poincaré ball model, Nickel *et al*. [30] projects the tree structure to the disk and proposes Poincaré disk model.
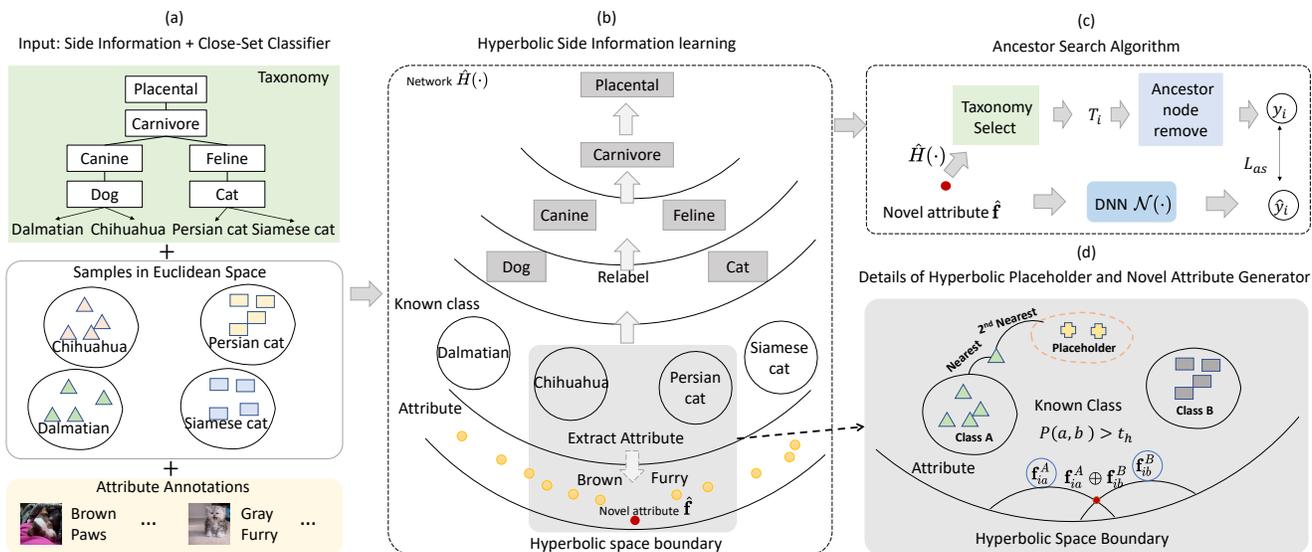
Figure 2. Overall framework of GOSR. In section a, the inputs are known samples and side information (taxonomy + attribute annotations). In (b), a neural network $\hat{H}(\cdot)$ is trained by hyperbolic side information learning process. The taxonomy is used for relabeling images, and attribute annotations are used to extract the attribute feature (yellow points). (d) illustrates details of hyperbolic placeholder learning and novel attribute generating process. In the ancestor search algorithm part (c), an ancestor search loss is designed to train a model $\mathcal{N}(\cdot)$ to recognize unseen samples.

Poincaré disk model learns better representation among hierarchically structured data, so that hyperbolic based models are utilized in solving natural language processing problems [11, 12]. After that, Ganea *et al.* [13] further proposes a feed-forward neural network and recurrent neural network based on the hyperbolic geometry. Furthermore, the graph convolutional network [6] is proposed using hyperbolic geometry features. Besides, hyperbolic embedding methods are widely used in other tasks with a hierarchical structure. Chami *et al.* [5] embeds a knowledge graph based on the Poincaré disk model. In video prediction, Compared with the hyperbolic embedding in few-shot learning [20, 9], side information for unseen classes like taxonomy is not available in our task. Feature representation in the hyperbolic space will assist us in using the taxonomy from the known classes. Compared with hyperbolic visual embedding in zero-shot learning [26], our method does not use the large text corpus to generate key words.

3. Approach

This section illustrates how the proposed framework recognizes unseen samples. Figure 2(a) introduces the input of the GOSR, including taxonomy, known samples, and attribute annotations. Known samples are training samples for known classes. To better describe the framework, we define known classes as leaf classes, and the ancestor of known classes are non-leaf classes. Figure 2(b) illustrates the hyperbolic side information learning process. To better recognize unseen samples, we propose an ancestor search
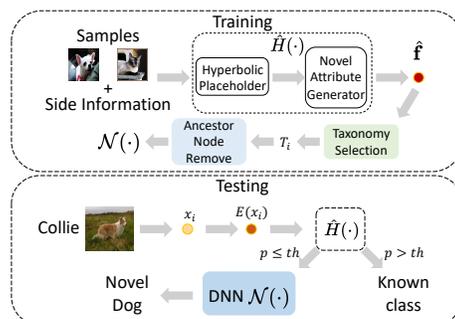
Figure 3. In the training process, novel attribute features $\hat{\mathbf{f}}$ are generated by the novel attribute generator. Taxonomy selection and ancestor node remove process are used for training the deep neural network $\mathcal{N}(\cdot)$. In the testing process, a threshold th is utilized to separate known and unknown. If the output probability $p < t_h$, $\mathcal{N}(\cdot)$ will classify the unseen sample as a novel dog.

algorithm in Figure 2(c). Figure 2(d) provides details of hyperbolic side information learning, including hyperbolic placeholder learning and novel attribute feature generator.

Training and Testing. In our proposed framework, the hyperbolic network $\hat{H}(\cdot)$ and deep neural network $\mathcal{N}(\cdot)$ are trainable to recognize unknown samples. Figure 3 illustrates the workflows of the training and testing. In testing, we first extract the feature of the input image as $\mathbf{x}$, e.g., *collie*. Then, we map the feature to the hyperbolic space by exponential mapping, e.g., $E(\mathbf{x})$. Taking the maximum

output probability of known class, i.e., $p = \max \hat{H}(\mathbf{x})$. If $p > t_h$ where t_h is the threshold of leaf classes, the image is classified as the known class taking the label by $\hat{y} = \arg\max \hat{H}(\mathbf{x})$. Otherwise, we use the hyperbolic embedding $E(\mathbf{x})$ as the input of the network $\mathcal{N}(\cdot)$ and search the closest ancestor of $\mathbf{x}$. The training process is slightly different from the testing process. In the hyperbolic side information learning part, the novel attribute features are only generated in the training process. Additionally, we use the generated attribute feature to train the deep neural network $\mathcal{N}(\cdot)$ after taxonomy selection and ancestor node remove process.

3.1. Hyperbolic Side Information Learning

Taxonomies and attribute annotations in the real world are naturally hierarchical. In this section, we design a placeholder learning process in the hyperbolic space to learn the side information because the number of nodes in taxonomies grows exponentially in the hierarchical structure. Additionally, unknown samples cannot be used in the training process for GOSR problem. We propose a novel attribute feature generator to capture the variability of the unknown samples based on the attribute annotations in hyperbolic space.

Hyperbolic Placeholder. The placeholder stands for the area that is set aside for the non-leaf classes. We propose hyperbolic placeholders to further classify non-leaf classes based on the closed-set classifier in the Euclidean space.

Given the closed-set classifier, matrix $\mathbf{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K]$ are the weights of the last linear layer of the classifier and K is the number of leaf classes (known classes). We transform the closed set classifier of known classes to the hyperbolic space using exponential mapping [31]. Assuming $E(\mathbf{x})$ is the exponential mapping transformation from Euclidean space to hyperbolic space and $\mathbf{x}$ is a feature extracted by the wide residual network [42]. The output logits of all leaf classes in the hyperbolic space is $h = \mathbf{W}^\top E(\mathbf{x})$. To enable the recognition of known and unknown classes, we introduce hyperbolic placeholders to classify samples that are not belong to the known classes. The output matrix of the hyperbolic placeholder is $h^p = \left[\mathbf{W}^\top E(\mathbf{x}), \hat{\boldsymbol{w}}^\top E(\mathbf{x})\right]$, where $\hat{\boldsymbol{w}}^\top$ represents the augmented parameters for classifying the hyperbolic placeholders. Different from the placeholder learning [44], we define the number of hyperbolic placeholder is L, where L represents the number of non-leaf classes. The size of the output is $K + L$. The first K outputs represent the leaf classes, and the latter L outputs denote non-leaf classes. The output probabilities of hyperbolic neural network are normalized logits by softmax function, i.e., $\hat{H}(\mathbf{x}) = softmax(h^p)$. Similar to [44], we propose a loss function for the hyperbolic placeholder process to

separate the leaf class samples and generated samples in novel attribute generating process. Figure 2(d) illustrates the hyperbolic placeholder. The hyperbolic placeholder loss on the training set $\mathcal{D}_{train}$ is shown as

$$\mathcal{L}_{hp} = \sum_{(\mathbf{x},y)\in\mathcal{D}_{train}} \ell(\hat{H}(\mathbf{x}),y) + \beta\ell(\hat{H}(\mathbf{x})\backslash y, K+1), \quad (1)$$

where ℓ denotes the cross-entropy loss, $y \in Y_l$ represents the label of leaf classes and $Y_l = \{1, 2, \ldots, K\}$. $\hat{H}(\mathbf{x})\backslash y$ means ignoring the probability of generated novel attribute features that belong to a known class by setting the $\mathbf{W}^\top E(\mathbf{x})$ to 0. β is the parameter that controls the loss of unknown samples.

For the GOSR problem, all training samples are labeled as leaf classes. Learning to separate non-leaf classes based on a close-set classifier is still a challenging problem. A simple way to solve the problem is to relabel some leaf class samples to non-leaf classes [25]. Specifically, relabel a portion (r) of image features to their ancestor nodes using the taxonomy information. For example in (Figure 2(b)), a *chihuahua* image can be randomly relabeled as one of the following non-leaf classes: *dog*, *canine*, *carnivore*, and *placental*. We denote the Y as the set of K leaf classes Y_l and L non-leaf classes, i.e., $Y = \{1, 2, \ldots, K, K+1, \ldots, K+L\}$. Then, we define a relabel loss given the new labels

$$L_{relabel} = -\sum_{\mathbf{x}\in\mathcal{D}_{train}} \sum_{i=1}^{K+L} t_i \log \hat{H}(\mathbf{x})_i, \quad (2)$$

where t_i is i-th element of the one-hot vector of the real-valued label $y \in Y$ and $\mathbf{x}$ is the input feature. The relabel loss enables hyperbolic placeholders to separate non-leaf classes.

Novel Attribute Feature Generator. In our framework, the most challenging part is how to separate the ancestor node in the hyperbolic space. Generating novel attributes based on the attributes from leaf classes assists the framework to capture the variability of the non-leaf classes and enable further classification ability in non-leaf classes. Attributes stand for tags of the image samples. Inspired by the Euclidean placeholder learning method [44] and open-world compositional method [27], we introduce the hyperbolic similarity constrain and Möbius transformations to generate novel attribute features.

Since unknown samples in the real world share similar attributes with the known samples to some extent [27], we can generate less noisy novel attribute features by combining the leaf class attribute features based on their similarity in the hyperbolic space. The similarity constraints of the generation mitigate the arbitrary combinations of leaf classes which could never exist in an open world. Formally,
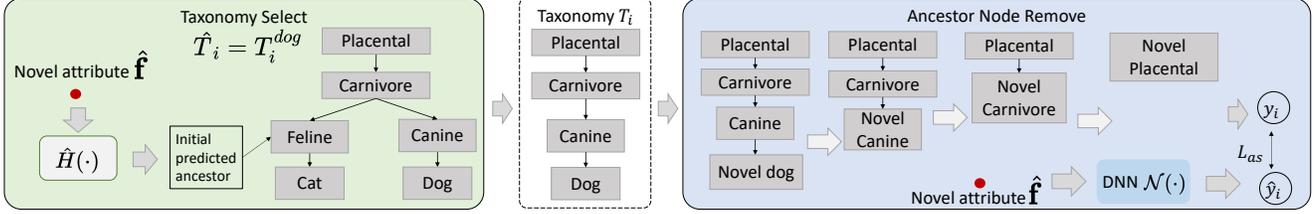
Figure 4. Ancestor search algorithm. The input is novel attribute $\hat{\mathbf{f}}$, the potential ground truth taxonomy $\hat{T}_i$. $\hat{H}(\cdot)$ represents the classifier in hyperbolic side information learning. The initial potential ancestor $\hat{n}_i = \arg\max \hat{H}(\mathbf{f})$. The selected taxonomy T_i is determined by $\hat{T}_i$ and $\hat{n}_i$. After that, an ancestor node remove process and an ancestor search loss is designed to train a model $\mathcal{N}(\cdot)$ for recognizing unseen samples.

we define $\mathbf{f}_{ia}^{(A)}$ and $\mathbf{f}_{jb}^{(B)}$ as the feature of the i-th sample $\mathbf{x}_i$ from class A on attribute a and the feature of the j-th sample $\mathbf{x}_j$ from class B on attribute b, respectively, their hyperbolic similarity $P(\mathbf{f}_{ia}^{(A)}, \mathbf{f}_{jb}^{(B)})$ is measured by

$$P(\mathbf{u}, \mathbf{v}) = \frac{\langle E(\mathbf{u}), E(\mathbf{v}) \rangle}{\langle E(\mathbf{u}), E(\mathbf{u}) \rangle \cdot \langle E(\mathbf{v}), E(\mathbf{v}) \rangle} \quad (3)$$

where $\langle \cdot, \cdot \rangle$ represents the Minkowski inner product and $E(\cdot)$ is the exponential mapping that transforms features from Euclidean space to hyperbolic space. Furthermore, the hyperbolic similarity between attribute a and attribute b can be defined as

$$P(a, b) = max\left\{ P(\mathbf{f}_{ia}^{(A)}, \mathbf{f}_{jb}^{(B)}) | i \in \Omega(A, a), j \in \Omega(B, b) \right\} \quad (4)$$

where the sample indices set $\Omega(A, a) = \{k | \mathbf{x}_k \in A, a \in Attr(\mathbf{x}_k)\}$ and similar definition for $\Omega(B, b)$. This indicates that the attribute similarity is defined as the maximum feature similarity among the samples within the subset of samples in class A (or B) that contain the same attribute a (or b).

To determine which two features are selected for generating the unknown samples, we define a threshold t_h that constrains the similarity between features from two different leaf classes. Given two distinct classes y_i and y_j where $y_i \neq y_j$ and they contain the attribute a and b respectively. In datasets with attribute annotations, if $P(a, b) > t_h$, a novel attribute feature can be generated in the hyperbolic space using the following transformation:

$$\hat{\mathbf{f}} = \alpha_i \otimes E\left(\mathbf{f}_{ia}^{(y_i)} \right) \oplus \alpha_j \otimes E\left(\mathbf{f}_{jb}^{(y_j)} \right) \quad (5)$$

where $y_i \neq y_j$, and α_i and α_j represent normalized confidence scores for leaf classes y_i and y_j, respectively. The notation $\oplus$ and $\otimes$ are Möbius addition and Möbius scalar multiplication [17], respectively. In this transformation, we add confidence scores from the side information to assist novel attribute feature generation. The novel attribute should be close to the attribute with a higher confidence score. Figure 2 section (d) shows the visualization of the Möbius addition.

For datasets without attribute annotations, we use a parameter λ_i as the confidence score for leaf class y_i and define the confidence score for leaf class y_j as $1 - \lambda_i$ in datasets without attribute annotations. The novel feature generating process in datasets without attribute annotations is $\hat{\mathbf{f}} = \lambda_i \otimes \mathbf{f}_i \oplus (1 - \lambda_i) \otimes \mathbf{f}_j$ where $y_i \neq y_j$. In this transformation, the novel attributes are determined by a parameter λ_i, which is sampled from a Beta distribution [44]. After generating the novel attributes, we design the loss to enlarge the feature distance between leaf classes:

$$\mathcal{L}_{\text{gen}} = -\sum_{\hat{\mathbf{f}} \in \mathcal{T}} \log \frac{e^{c \cdot \rho(\hat{\mathbf{f}}; i, j)}}{\sum_{i,j \in Y_l} e^{c \cdot \rho(\hat{\mathbf{f}}; i, j)}} \quad (6)$$

where $\rho(\hat{\mathbf{f}}; i, j)$ defines the total distance of $\hat{\mathbf{f}}$ to the attribute feature $\mathbf{f}_a^{(i)}$ and $\mathbf{f}_b^{(j)}$ in class i and j, i.e., $\rho(\hat{\mathbf{f}}, \mathbf{f}_a^{(i)}) + \rho(\hat{\mathbf{f}}, \mathbf{f}_b^{(j)})$. Here, the distance measure $\rho(\cdot, \cdot)$ is defined as the Euclidean distance in Euclidean space. The c is the temperature hyperparameter. $\mathcal{T}$ is the set of novel attribute features. The goal of $\mathcal{L}_{\text{gen}}$ is to maximize the distance from generated attribute to the original attributes. The loss $\mathcal{L}_h$ for hyperbolic side information learning shows as follows,

$$\mathcal{L}_h = \mathcal{L}_{hp} + \mathcal{L}_{gen} + c_r \mathcal{L}_{relabel}, \quad (7)$$

where c_r denotes the weight hyperparameter of relabel loss.

3.2. Ancestor Search

Motivation In the Section 3.1, the hyperbolic network $\hat{H}(\cdot)$ can only achieve the goal of the open set recognition without the capability to determine the specific semantic of the unknown samples in class taxonomy. Therefore, we propose the ancestor search algorithm to train a network $\mathcal{N}(\cdot)$ to classify unknown samples.

The ancestor search algorithm can be divided into two parts. The first part is to select taxonomy T_i using the generated novel attribute feature $\hat{\mathbf{f}}$. The T_i is defined as a taxonomy that contains all possible non-leaf nodes of the $\hat{\mathbf{f}}$. The second part is ancestor node remove process. This process recursively removes the node from T_i as the label of $\hat{\mathbf{f}}$ to train a deep neural network that classifies unseen samples.

Algorithm 1 Training Ancestor Search Algorithm
___
Require: Taxonomy for known classes T_k; Hyperbolic classifier $\hat{H}$; Generated novel attribute set I; Ground truth set y_i.
1: **for** each $i \in I$ **do**
2:     Find initial predicted ancestor node $\hat{n}_i = \arg\max \hat{H}(\hat{\mathbf{f}}_i)$.
3:     Select taxonomy T_i from T_k.
4:     Find LCA node n_{LCA} and put all ancestor nodes from T_i to y_i.
5:     Let start node $a = y_i(a)$
6:     Let $list$ be the linked list of $\{a \rightarrow n_{LCA}\}$
7:     **while** $a \in List$ **and** $a \neq n_{LCA}$ **and** $a \neq n_i$ **do**
8:         Remove a and update $T \backslash a$, $list \backslash a$
9:         Compute loss L_{as} for situation under $T \backslash a$.
10:        Move to the ancestor node of a
11:     **end while**
12: **end for**
___

Taxonomy Selection. The target of taxonomy selection is to find the taxonomy T_i for input attribute feature $\hat{\mathbf{f}}$. Since $\hat{\mathbf{f}}$ is generated by two attribute features from two leaf classes A and B, all non-leaf classes that have links to the leaf classes A and B should be considered. Here, we define n_p^A and n_q^B as the node of non-leaf classes that has linked path to the leaf classes A and B, respectively. $T_i^A = \{n_1^A, \cdots, n_p^A\}$ and $T_i^B = \{n_1^B, \cdots, n_q^B\}$ are two potential taxonomies of $\hat{\mathbf{f}}$. We define $\hat{T}_i$ as the potential ground truth taxonomy. $\{n_j^A | j = 1, \cdots, p\}$ are ancestor nodes of the leaf class A. However, during the Möbius transformation, confidence scores α_i and α_j in the attribute generator process considerably influence in the final generated result. Let α_i and α_j as the confidence score of attribute feature from leaf class A and B, respectively. If $\alpha_i > \alpha_j$, $\hat{T}_i = T_i^A$. Otherwise, $\hat{T}_i = T_i^B$. After that, we need the initial predicted node to finalize the taxonomy T_i. The initial predicted node $\hat{n}_i = \arg\max \hat{H}(\mathbf{f})$. The definition of T_i shows as follows,

$$T_i = \begin{cases} \hat{T}_i - \{\hat{n}_i, \cdots, \hat{n}_{root}\} & n_i \in \hat{T}_i \\ \hat{T}_i & n_i \notin \hat{T}_i \end{cases}. \quad (8)$$

where $\hat{n}_{root}$ is the root node of $\hat{T}_i$. Figure 4 illustrates the taxonomy select process, the initial ancestor label of $\hat{\mathbf{f}}$ is $\hat{n}_i = feline$ and the $\hat{T}_i = T_i^{dog}$ where the dog taxonomy $T_i^{dog} = \{placental, carnivore, canine, dog\}$. Since $feline$ is not in $\hat{T}_i$, the selected taxonomy $T_i = \hat{T}_i = T_i^{dog}$, which means that all the four nodes could be the ground truth label of the novel feature $\hat{\mathbf{f}}$.

Ancestor Node Remove Process. This process recursively removes ancestor nodes from the bottom of the tax-

onomy T_i to provide potential novel class y_i for novel attribute feature $\hat{\mathbf{f}}$. Let a denotes the novel class and n_{LCA} represents the lowest common ancestor of class A and $\hat{n}_i$. $T \backslash a$ represents deficient taxonomy. For example, in Figure 4, a is the ancestor node dog. Then, we train a deep neural network $\mathcal{N}$ to predict the novel class. To make sure the predicted novel class is close to the ground truth, the output of this neural network $\hat{y}_i = \mathcal{N}(\hat{H}(x))$ should be close to the potential novel class y_i (Figure 4). Let $a = y_i(a)$, where $y_i(a)$ denotes the ground truth node of $\hat{\mathbf{f}}$. $List$ is defined as the linked list from $\{a \rightarrow n_{LCA}\}$. The loss of ancestor search algorithm $\mathcal{L}_{as}$ shows as follows,

$$\mathcal{L}_{as} = \sum_{a \in List} \ell(\mathcal{N}(\hat{H}(\hat{\mathbf{f}})), y_i(a); T \backslash a) \quad (9)$$

where $y_i(a)$ is the ground truth in $T \backslash a$ situation, and $\mathcal{N}(\hat{H}(\hat{\mathbf{f}}))$ denotes the prediction of novel attribute i. The ancestor search loss add the cross entropy loss of all nodes a under $T \backslash a$ situation. Specific ancestor search algorithm can be found in Algorithm 1.

4. Experiments

4.1. Datasets

In GOSR experiments, we use the attributed image dataset Caltech-UCSD bird dataset (CUB-200) [38] and the Animal with attributes 2 dataset (AWA2) [23]. CUB-200 [38] consists of 11,788 images with accurate quantized attribute annotation from 200 bird species. AWA2 [23] contains 85 attributes for 50 animal species. We use the class taxonomy reported in previous hierarchical class detection work [25]. For basic open set recognition, we also report the result from some widely used datasets without taxonomy annotations using the same open set recognition setting in [7] including MNIST[24], CIFAR-10 [22], CIFAR-100 [22], SVHN [29], and Tiny Imagenet [33]. Details about datasets are shown in the Appendix.

4.2. Evaluation

The Area Under the Receiver Operating Characteristic (AUROC) curve [28] and Open Set Classification Rate (OSCR) [10] are chosen as evaluation metrics in open set recognition. The comprehensive evaluation of the generalized open set recognition consists of the average hierarchical distance [3] (AHD) and the hierarchical similarity index. AHD is the original evaluation metric for the hierarchical recognition tasks.

Hierarchical Similarity Index. Since AUROC and OSCR cannot illustrate the distance in a hierarchical structure, evaluation metrics in open set recognition are not appropriate to measure the difference of non-leaf classes.

Table 1. Generalized open set recognition results (%) on CUB-200 and AWA datasets. Best results are in bold.

Method	CUB-200[38]			AWA 2 [23]			AWA 1 [23]		
	AHD($\downarrow$)	HSI-b_1($\uparrow$)	HSI-b_2($\uparrow$)	AHD($\downarrow$)	HSI-b_1($\uparrow$)	HSI-b_2($\uparrow$)	AHD($\downarrow$)	HSI-b_1($\uparrow$)	HSI-b_2($\uparrow$)
Random guess	1.98	26.73	42.14	3.08	20.86	24.96	3.06	20.98	25.15
Clustering	1.82	33.56	54.28	2.81	45.58	50.29	2.78	45.77	50.96
TD-LOO [25]	1.73	36.48	60.91	2.67	63.23	62.49	2.64	64.05	63.11
Ours	**1.71**	**38.65**	**62.47**	**2.48**	**65.05**	**66.91**	**2.47**	**65.43**	**67.02**

We propose hierarchical similarity index (HSI) to evaluate the generalized open set recognition. The hierarchical similarity index is defined by the Lowest Common Ancestor (LCA) distance. The LCA means the lowest common ancestor between ground truth and the direct ancestor of the predicted class. In our experiments, we report two hierarchical similarity indexes (bottom-1, and bottom-2) and compare results with the average hierarchical distance. Both similarity indexes use the reciprocal of the LCA distance as the hierarchical similarity index. Higher hierarchical similarity index means better performance in the ancestor searching experiment. Bottom-1 and Bottom-2 hierarchical similarity metrics (HSI-b_1 and HSI-b_2) are defined as

$$\text{HSI-}b_1 = \frac{1}{m}\sum_{l=1}^{m}\frac{1}{d\left(y_{gt1}^l, y_{LCA1}^l\right)}$$

$$\text{HSI-}b_2 = \frac{1}{m}\sum_{l=1}^{m}\frac{1}{ln(d\left(y_{gt2}^l, y_{LCA2}^l\right)+1)e} \quad (10)$$

Here, m is the total number of testing data. In HSI-b_1, $d\left(y_{gt1}^l, y_{LCA1}^l\right)$ represent the distance between direct ground truth ancestor and the lowest common ancestor. In HSI-b_2, the $d\left(y_{gt2}^l, y_{LCA2}^l\right)$ is the distance between ground truth class and the lowest common ancestor. Since the meaning of hierarchical distance is the distance from the ground truth node to the LCA node, a lower hierarchical distance represents the better recognition result. HSI-b_1 and HSI-b_2 are different from AHD. A larger index means better performance in recognizing unknown classes.

4.3. Generalized Open Set Recognition

We first compare baseline methods, the Hierarchical novelty detection method (TD+LOO) [25], and our proposed method on CUB-200 and AWA 1&2 datasets in the generalized open set recognition task as shown in Table 1. Here, the **random guess** baseline is to guess the closest ancestor node based on the taxonomy from the known classes. The **clustering** baseline calculates the cosine similarity between the sample feature and the feature from all ancestor nodes and outputs the most similar node.

According to the result in Table 1, our method outperforms all comparing methods under three evaluation methods. Compared with TD-LOO [25], novel attribute features enhance the performance of ancestor search algorithm. Be-
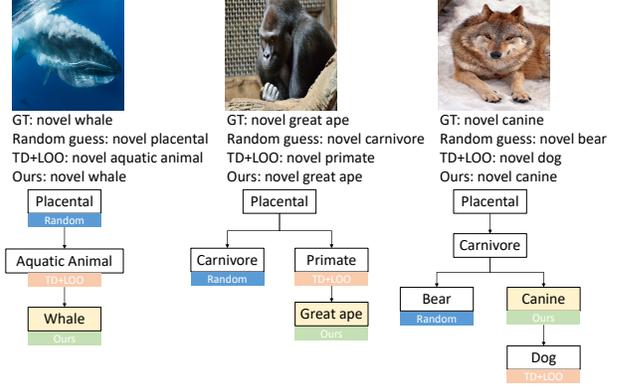
Figure 5. Result visualization on AWA 2 dataset[23]. The yellow boxes represent the ground truth of the generalized open set recognition. The taxonomy tree graph shows the relationship between ground truth nodes and the results from random guess (Blue), TD+LOO [25] (Orange), and our proposed method (Green).

sides, we find that HSI-b_1 tends to provide a better score for totally correct samples, which is more sensitive when the LCA is close to the leaf nodes. Figure 5 shows some generalized open set recognition results. The text under images illustrates the ground truth label and the result of comparing method. The taxonomy illustrates the relationship between ancestor nodes. From the visualization, ancestor node results from our method are more close to the ground truth node. For example, our method recognizes a *gorilla* as a novel *great ape*, while the result from TD+LOO[25] is a novel *primate*. This is because novel attribute features from hyperbolic side information learning process better capture the sample variability of the non-leaf classes.

Ablation Study. The ablation study results are reported in Table 3. In Section 3.1, we generate the unseen attribute features using Möbius transformation and control the generating process using the hyperbolic similarity. Here we design an ablation study on Möbius transformation, and the constraint of hyperbolic similarity to show how these components influence the result of generalized open set recognition. The evaluation methods are identical to the generalized open set recognition task. In our method (Case 4), two selected attribute features are determined by the hyperbolic similarity, and the novel attributes are generated by Möbius transformation using confidence score. In Table 3,

Table 2. Open set recognition results in five datasets, '-' represents result that is not reported in original paper. Numbers in bold represent the best performance. Tiny IN represents the Tiny ImageNet dataset. AUC denotes the AUROC result

Method	MNIST[24]		SVHN[29]		CIFAR10[22]		CIFAR+50[22]		Tiny IN[33]	
	AUC	OSCR	AUC	OSCR	AUC	OSCR	AUC	OSCR	AUC	OSCR
Softmax	97.83	99.26	88.62	92.84	67.76	88.94	84.32	81.93	57.71	59.93
Openmax [2]	98.13	-	89.41	-	69.51	-	79.63	-	57.64	-
Capsule [18]	99.22	-	94.60	-	83.53	-	88.98	-	71.51	-
GCPL [39]	99.21	99.11	94.30	**92.85**	84.69	82.42	85.45	88.30	69.43	49.47
RPL [8]	98.91	99.23	93.41	92.46	82.72	85.26	83.29	89.63	68.87	53.21
ARPL [7]	99.02	99.37	94.03	91.78	85.98	81.47	90.32	90.27	74.40	58.42
ARPL+cs [7]	**99.51**	**99.47**	94.61	91.85	86.17	81.30	**90.65**	86.80	78.07	65.58
Ours	99.43	99.40	**94.78**	92.71	**89.52**	**86.41**	90.45	**89.76**	**78.19**	**65.74**

Table 3. Ablation study on CUB-200 dataset. Similarity denotes the Hyperbolic similarity constraint, and Möbius represents Möbius transformation.

Method	Similarity	Möbius	AHD($\downarrow$)	HSI-$b_1(\uparrow)$	HSI-$b_2(\uparrow)$
Case1			1.75	35.23	59.83
Case2		✓	1.73	36.25	61.22
Case3	✓		1.73	36.78	60.54
Case4	✓	✓	**1.71**	**38.65**	**62.47**

Case 1 is randomly choosing attributes features from two different known classes without using the Möbius transformation. Case 2 only uses the Möbius transformation. Case 3 only uses the hyperbolic similarity constrain. From the results, both Möbius transformation and constraint of the hyperbolic similarity could improve the performance of generalized open set recognition. Besides, we provide analysis for hyperbolic feature mapping in the Appendix.

4.4. Open Set Recognition

Comparison with State-of-the-arts. We apply the AUROC and OSCR metrics to evaluate the performance of the basic open set recognition. Since there is no side information available in the open set recognition, the portion of relabeling r is set as zero. The number of hyperbolic placeholders is equal to the number of non-leaf classes. Results by AUROC and OSCR are reported in Table 2 in which the numbers are average scores from five randomized trials. In CIFAR+50 experiment, 50 non-overlapping classes from CIFAR-100 are randomly sampled as unknown classes. From Table 2, our method has better recognition capability in CIFAR10[22] and Tiny ImageNet[33] Dataset. Since the open set setting of comparing methods are not identical, we report the AUROC and OSCR performance based on our own open set setting. Some results may differ from the original paper. Details about the experiment setting can be found in the Appendix.

Discussion. Results in Table 2 are averaged by five independent experiment. Our proposed method shows better recognition ability in more complex datasets. Compared
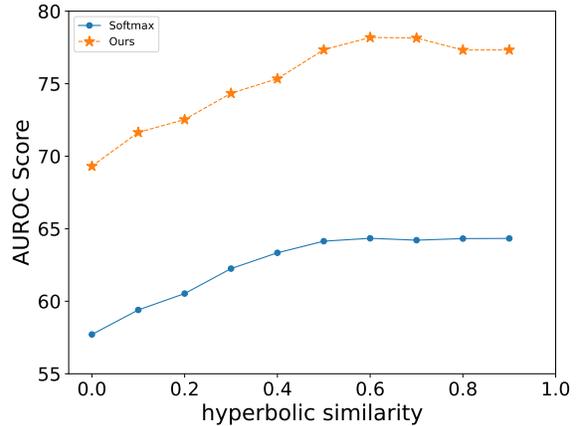
Figure 6. The AUROC results for different threshold t_h from random attribute generating to selective attribute generating.

with state-of-the-art method ARPL+cs [7], we do not generate confusing samples to ensure the performance in generalized open set recognition. To show that hyperbolic side information learning improves the recognition capability, we compare the AUROC results using different experiment settings. In Figure 6, we report the AUROC performance in Tiny ImageNet dataset using different hyperbolic similarity thresholds. Our method achieves the best AUROC score when the threshold t is equal to 0.6. More results can be found in the Appendix.

5. Conclusions

This paper proposes a hyperbolic side information learning framework and an ancestor search algorithm to capture the variability of unknown samples and solve the GOSR problem. In our framework, the generated attribute features better capture the variability of unknown samples. Additionally, we propose hierarchical similarity indexes to measure the performance of GOSR. From experiment results, our framework outperforms the state-of-the-art methods in the GOSR problem and the open set recognition.

References

[1] Wentao Bao, Qi Yu, and Yu Kong. Evidential deep learning for open set action recognition. In *International Conference on Computer Vision*, 2021.

[2] Abhijit Bendale and Terrance E Boult. Towards open set deep networks. In *Conference on Computer Vision and Pattern Recognition*, pages 1563–1572, 2016.

[3] Luca Bertinetto, Romain Mueller, Konstantinos Tertikas, Sina Samangooei, and Nicholas A Lord. Making better mistakes: Leveraging class hierarchies with deep networks. In *Conference on Computer Vision and Pattern Recognition*, pages 12506–12515, 2020.

[4] Silvere Bonnabel. Stochastic gradient descent on riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, 2013.

[5] Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. Low-dimensional hyperbolic knowledge graph embeddings. *arXiv preprint arXiv:2005.00545*, 2020.

[6] Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. *Advances in Neural Information Processing Systems*, 32:4868–4879, 2019.

[7] Guangyao Chen, Peixi Peng, Xiangqian Wang, and Yonghong Tian. Adversarial reciprocal points learning for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[8] Guangyao Chen, Limeng Qiao, Yemin Shi, Peixi Peng, Jia Li, Tiejun Huang, Shiliang Pu, and Yonghong Tian. Learning open set network with discriminative reciprocal points. In *European Conference on Computer Vision*, August 2020.

[9] Yawen Cui, Zitong Yu, Wei Peng, and Li Liu. Rethinking few-shot class-incremental learning with open-set hypothesis in hyperbolic geometry. *arXiv preprint arXiv:2207.09963*, 2022.

[10] Akshay Raj Dhamija, Manuel Günther, and Terrance E Boult. Reducing network agnostophobia. *arXiv preprint arXiv:1811.04110*, 2018.

[11] Bhuwan Dhingra, Christopher J Shallue, Mohammad Norouzi, Andrew M Dai, and George E Dahl. Embedding text in hyperbolic spaces. *arXiv preprint arXiv:1806.04313*, 2018.

[12] Tom Diethe and Oluwaseyi Feyisetan. Preserving privacy in analyses of textual data. In *PrivateNLP@ WSDM*, 2020.

[13] Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. *Advances in Neural Information Processing Systems*, 31, 2018.

[14] ZongYuan Ge, Sergey Demyanov, Zetao Chen, and Rahil Garnavi. Generative openmax for multi-class open set classification. *arXiv preprint arXiv:1707.07418*, 2017.

[15] Chuanxing Geng and Songcan Chen. Collective decision for open set recognition. *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[16] Chuanxing Geng, Sheng-jun Huang, and Songcan Chen. Recent advances in open set recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3614–3631, 2020.

[17] Mikhael Gromov. Hyperbolic groups. In *Essays in group theory*, pages 75–263. Springer, 1987.

[18] Yunrui Guo, Guglielmo Camporese, Wenjing Yang, Alessandro Sperduti, and Lamberto Ballan. Conditional variational capsule network for open set recognition, 2021.

[19] Mehadi Hassen and Philip K Chan. Learning a neural-network-based representation for open set recognition. In *International Conference on Data Mining*, pages 154–162. SIAM, 2020.

[20] Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. Hyperbolic image embeddings. In *Conference on Computer Vision and Pattern Recognition*, pages 6418–6428, 2020.

[21] Shu Kong and Deva Ramanan. Opengan: Open-set recognition via open data generation. In *ICCV*, 2021.

[22] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[23] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):453–465, 2013.

[24] Yann LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

[25] Kibok Lee, Kimin Lee, Kyle Min, Yuting Zhang, Jinwoo Shin, and Honglak Lee. Hierarchical novelty detection for visual object recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1034–1042, 2018.

[26] Shaoteng Liu, Jingjing Chen, Liangming Pan, Chong-Wah Ngo, Tat-Seng Chua, and Yu-Gang Jiang. Hyperbolic visual embedding learning for zero-shot recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[27] Massimiliano Mancini, Muhammad Ferjad Naeem, Yongqin Xian, and Zeynep Akata. Open world compositional zero-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5222–5230, 2021.

[28] Lawrence Neal, Matthew Olson, Xiaoli Fern, Weng-Keen Wong, and Fuxin Li. Open set learning with counterfactual images. In *European Conference on Computer Vision*, pages 613–628, 2018.

[29] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. 2011.

[30] Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. *Advances in Neural Information Processing Systems*, 30:6338–6347, 2017.

[31] Wei Peng, Tuomas Varanka, Abdelrahman Mostafa, Henglin Shi, and Guoying Zhao. Hyperbolic deep neural networks: A survey. *ArXiv*, abs/2101.04562, 2021.

[32] Sridhama Prakhya, Vinodini Venkataram, and Jugal Kalita. Open set text classification using convolutional neural networks. In *International Conference on Natural Language Processing, 2017*, 2017.

[33] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy,

Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[34] Lei Shu, Hu Xu, and Bing Liu. Unseen class discovery in open-world classification. *arXiv preprint arXiv:1801.05609*, 2018.

[35] Paolo Simeone, Raúl Santos-Rodríguez, Matt McVicar, Jefrey Lijffijt, and Tijl De Bie. Hierarchical novelty detection. In *International Symposium on Intelligent Data Analysis*, pages 310–321. Springer, 2017.

[36] Alexandru Tifrea, Gary Bécigneul, and Octavian-Eugen Ganea. Poincar\'e glove: Hyperbolic word embeddings. *arXiv preprint arXiv:1810.06546*, 2018.

[37] Abraham A Ungar. Hyperbolic trigonometry and its application in the poincaré ball model of hyperbolic geometry. *Computers & Mathematics with Applications*, 41(1-2):135–147, 2001.

[38] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

[39] Hong-Ming Yang, Xu-Yao Zhang, Fei Yin, and Cheng-Lin Liu. Robust classification with convolutional prototype learning. In *Conference on Computer Vision and Pattern Recognition*, pages 3474–3482, 2018.

[40] Ryota Yoshihashi, Wen Shao, Rei Kawakami, Shaodi You, Makoto Iida, and Takeshi Naemura. Classification-reconstruction learning for open-set recognition. In *Conference on Computer Vision and Pattern Recognition*, pages 4016–4025, 2019.

[41] Yang Yu, Wei-Yang Qu, Nan Li, and Zimin Guo. Open-category classification by adversarial sample generation. *arXiv preprint arXiv:1705.08722*, 2017.

[42] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.

[43] Hang Zhao, Xavier Puig, Bolei Zhou, Sanja Fidler, and Antonio Torralba. Open vocabulary scene parsing. In *International Conference on Computer Vision*, pages 2002–2010, 2017.

[44] Da-Wei Zhou, Han-Jia Ye, and De-Chuan Zhan. Learning placeholders for open-set recognition. In *Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2021.