Research Collection School Of Computing and Information Systems   School of Computing and Information Systems

5-2024

Learning adversarial semantic embeddings for zero-shot recognition in open worlds

Tianqi LI

Guansong PANG
Singapore Management University, gspang@smu.edu.sg

Xiao BAI

Jin ZHENG

Lei ZHOU

See next page for additional authors

Citation

Author

Tianqi LI, Guansong PANG, Xiao BAI, Jin ZHENG, Lei ZHOU, and Xin NING

Learning Adversarial Semantic Embeddings for Zero-Shot Recognition in Open Worlds

Tianqi Li[1]
tianqili@buaa.edu.cn

Guansong Pang[2*]
gspang@smu.edu.sg

Xiao Bai[1*]
baixiao@buaa.edu.cn

Jin Zheng[1]
JinZheng@buaa.edu.cn

Lei Zhou[3]
leizhou@whut.edu.cn

Xin Ning[4]
ningxin@semi.ac.cn

[1] School of Computer Science and Engineering, Beihang University
[2] School of Computing and Information Systems, Singapore Management University
[3] School of Computer Science and Artificial Intelligence, Wuhan University of Technology
[4] Institute of Semiconductors, Chinese Academy of Sciences

Abstract

Zero-Shot Learning (ZSL) focuses on classifying samples of **unseen classes** with only their side semantic information presented during training. It cannot handle real-life, open-world scenarios where there are test samples of **unknown classes** for which neither samples (*e.g*., images) nor their side semantic information is known during training. Open-Set Recognition (OSR) is dedicated to addressing the unknown class issue, but existing OSR methods are not designed to model the semantic information of the unseen classes. To tackle this combined ZSL and OSR problem, we consider the case of "Zero-Shot Open-Set Recognition" (ZS-OSR), where a model is trained under the ZSL setting but it is required to accurately classify samples from the unseen classes while being able to reject samples from the unknown classes during inference. We perform large experiments on combining existing state-of-the-art ZSL and OSR models for the ZS-OSR task on four widely used datasets adapted from the ZSL task, and reveal that ZS-OSR is a non-trivial task as the simply combined solutions perform badly in distinguishing the unseen-class and unknown-class samples. We further introduce a novel approach specifically designed for ZS-OSR, in which our model learns to generate adversarial semantic embeddings of the unknown classes to train an unknowns-informed ZS-OSR classifier. Extensive empirical results show that our method 1) substantially outperforms the combined solutions in detecting the unknown classes while retaining the classification accuracy on the unseen classes and 2) achieves similar superiority under generalized ZS-OSR settings. Our code is available at https://github.com/lhrst/ASE.

1   Introduction

'A zebra is a horse with black and white striped coats.' With this description, a child who has never seen a zebra can recognize it at first sight. Humans can recognize images of such unseen classes using the shared semantic knowledge learned from images of the classes previously seen. Inspired by this phenomenon, Zero-Shot Learning (ZSL) was proposed to learn a multi-modal cognition

---

*Corresponding author: G. Pang (gspang@smu.edu.sg) and X. Bai (baixiao@buaa.edu.cn)
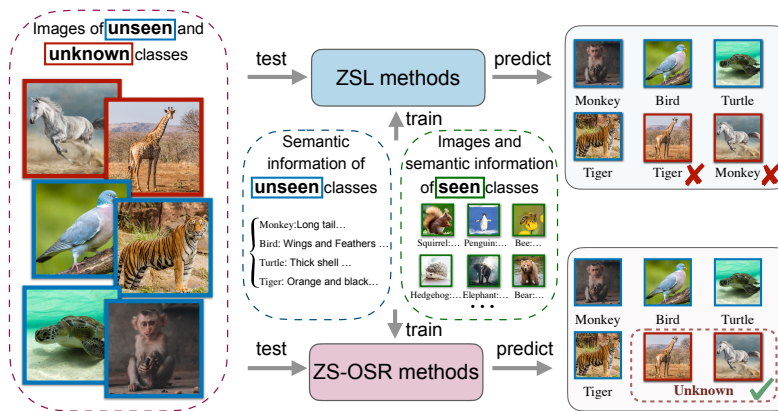
Figure 1: The ZS-OSR task vs. the ZSL task. ZS-OSR considers open-set scenarios where test data can also contain samples of unknown classes, unlike the closed-set setting in ZSL that presumes the presence of unseen-class samples only. As a result, ZSL methods would misclassify the unknown-class samples into one unseen class, whereas ZS-OSR methods can reject these unknown samples while correctly classifying the unseen-class samples.

ability in image classification [1]. Given only some side semantic information like attribute vectors or description text of a set of targeted classes (**unseen classes**), together with samples (*e.g.*, images) of another set of classes (**seen classes**) and their semantic information, ZSL aims to learn a model to recognize images of the unseen classes. Many ZSL methods have been introduced over the years, including embedding methods [2, 3] that learn mappings between the given semantic knowledge and images in a new feature space, and generative methods [4–6] that train a generator to synthesize training samples for the unseen classes, transforming the ZSL to a standard image recognition task.

However, current ZSL approaches cannot handle real-life, open-world scenarios, where there are test samples of **unknown classes** for which neither samples (*e.g.*, images) nor their side semantic information is known during training, such as unknown objects encountered in self-driving contexts and chest X-ray images of unknown viral pneumonia. This is because they assume a closed-set learning setting where all classes in the test data are presented in the training data. As a result, ZSL methods would misclassify the unknown-class samples into one of the unseen classes, as demonstrated in Figure 1 (top). Open-Set Recognition (OSR) [7, 8] is dedicated to addressing the unknown class issue, but existing OSR methods are not designed to model the semantic information of the unseen classes. To tackle this problem, we consider the case of "*Zero-Shot Open-Set Recognition*" (ZS-OSR), where a model is trained under the ZSL setting but it is required to accurately classify samples from the unseen classes while being able to reject samples from the unknown classes during inference, as illustrated in Figure 1 (bottom). As shown in Table 1, ZS-OSR is a joint ZSL and OSR problem. A straightforward solution is thus to simply combine existing state-of-the-art (SOTA) ZSL and OSR models to build ZS-OSR models. As a contribution to Pattern Recognition research, we establish a set of such baselines and construct four ZS-OSR benchmark datasets adapted from widely-used ZSL datasets, *i.e.*, CUB [9], AWA2 [10], FLO [11] and SUN [12], to evaluate their performance. Our empirical results reveal that such combined solutions perform badly in differentiating the unseen-class and the unknown-class samples. These findings highlight the need for more effective solutions for ZS-OSR, which can contribute to the development of more robust and effective pattern recognition systems.

We further introduce a novel approach specifically designed for ZS-OSR, namely ASE, which learns to generate Adversarial Semantic Embeddings of the unknown classes to train an unknowns-informed open-set classifier. The key insight is to generate meaningful samples of both unseen and unknown classes to train the unknowns-informed open-set classifier. Existing generative ZSL models have demonstrated superior performance in generating the samples/features of the unseen classes based on their learned relations between the samples and the semantic embeddings of the seen classes. The key challenge lies in the generation of unknown-class samples. To address this challenge, we introduce the adversarial semantic embedding learning module that learns a set of

Set / Task	Training	Testing
ZSL	Seen classes & semantic information[1]	Unseen classes
OSR	Seen classes	Seen classes & unknown classes
ZS-OSR	Seen classes & semantic information[1]	Unseen classes & unknown classes

Table 1: ZS-OSR vs. ZSL and OSR. Here *semantic information*[1] is the semantic information of both seen and unseen classes.

adversarial semantic embeddings for the unknown classes so that they are tightly distributed around but separable from the unseen-class embeddings (see Figure 5 for a visualization of these generated samples). This is achieved by jointly minimizing a distance loss in the *semantic embedding space* that brings the unknown-class embeddings closer to the given unseen-class embeddings, and an adversarial loss in the *feature space* that pulls the corresponding unknown-class feature prototypes away from the unseen-class features generated by an off-the-shelf trained generative ZSL model. Given the learned semantic embeddings, samples of the unknown classes are generated using the trained generative ZSL model.

There have been OSR methods that generate adversarial samples to represent the unknown-class samples, *e.g.*, [13, 14]. However, unlike ASE that can work on both the semantic and feature spaces, they cannot model the rich semantics in the semantic embedding space as they were primarily designed to work in the single feature space, largely limiting their performance in the ZS-OSR setting.

In summary, this work makes the following contributions: 1) we explore the ZS-OSR problem and establish extensive performance benchmarks by building a set of baselines based on the combination of existing SOTA ZSL and OSR models on four adapted ZS-OSR widely-used datasets, 2) we propose a novel generative approach for ZS-OSR, which learns a set of adversarial semantic embeddings to represent the unknown classes to train an unknowns-informed open-set classifier, and 3) large-scale empirical results show that our method ASE substantially outperforms the baselines in detecting the unknown classes without degrading the classification accuracy on the unseen classes, and it also achieves similar superiority on generalized ZS-OSR and ZS-OOD (out-of-distribution) settings.

2 Related Work

Zero-shot learning [1] utilizes an additional class semantic embedding set to connect the seen and unseen classes. Current ZSL approaches are either to align the images and semantic embeddings [2, 3, 15–23], or to generate image features of unseen classes and train a closed-set classifier [24, 25, 4–6, 26]. Recently, big vision-language models like CLIP [27] have demonstrated significant potential in the realm of ZSL. Nevertheless, they are still focused on the standard ZSL setting. New designs are required to utilize such big models to handle unknown samples as neither samples nor side semantic information of the unknown classes is available during ZS-OSR training.

Open-set recognition [7, 8] targets the problem of learning a classifier to reject samples of classes that are unseen during training. A large number of OSR methods have been introduced. Some early studies focus on designing new network layers, such as the OpenMax layer [28], while most studies are dedicated to generating pseudo unknown-class samples to train open-set classifiers [14, 29–31]. The other studies explore new ways of representing the unknown classes, *e.g.*, through prototype mining [32].

Out-of-distribution (OOD) detection addresses a similar problem as OSR, but focuses on detecting data from a different distribution. For example, [33–35] tackle the problem by exploiting the prediction logits to define OOD scores and reject samples from different datasets, while [36, 37] focused on the class-agnostic information in feature space that is not recoverable from logits.

Zero-shot open-set recognition (ZS-OSR) has not been explored in previous studies, as far as we know. Some related but different explorations are done in [38–45]. Particularly, [38–40] treat

ZSL and OSR as two independent problems, meaning that ZSL-oriented methods handle seen and unseen classes only, while OSR-oriented methods handle seen and unknown classes only. Typically, these methods handle both unknown and unseen classes with the same branch in the network, so they cannot be used to solve ZS-OSR tasks. Consequently they do not have the ability to distinguish between unseen and unknown classes. [41, 42] attempt to utilize OOD detection models to address the problem of generalized ZSL, solving the GZSL problem while neglecting the potential presence of OOD scenarios. [43] explores a new task, compositional ZSL, different from the conventional ZSL. While conventional ZSL aims to recognize unseen classes, compositional ZSL is designed to explore and recognize unknown combinations of known patterns. Due to the absence of unknown classes in its assumption, it cannot handle ZS-OSR tasks. [44, 45] explore the use of a large CLIP model [27] pretrained with extensive auxiliary data for OOD detection without using any training samples, which differs from zero-shot learning as there are no seen classes involved. Additionally, the task of open-set recognition under a few-shot setting is explored in some recent studies [46, 47], which addresses a different task from ours as we focus on zero-shot settings.

3 Zero-Shot OSR and Its Challenges

3.1 Problem Statement

In ZS-OSR, there are three different types of classes, including seen classes $\mathcal{Y}^{seen}$, unseen classes $\mathcal{Y}^{unseen}$, and unknown classes $\mathcal{Y}^{unknown}$, where $\mathcal{Y}^{seen} \cap \mathcal{Y}^{unseen} \cap \mathcal{Y}^{unknown} = \emptyset$. Compared to the $\mathcal{Y}^{unknown}$ classes that the training data does not provide any prior information, the classes in $\mathcal{Y}^{seen}$ and $\mathcal{Y}^{unseen}$ are considered as known classes since the training data contains the image samples of the $\mathcal{Y}^{seen}$ classes and some side semantic information of both $\mathcal{Y}^{seen}$ and $\mathcal{Y}^{unseen}$ classes. ZS-OSR aims to accurately recognize the images of unseen classes while rejecting the images of unknown classes based on these training image samples and semantic information. Formally, given a training set consisting of $\mathcal{D}^{seen} = \{(\mathbf{x}, y_{\mathbf{x}}, \mathbf{a}_y) \mid \mathbf{x} \in \mathcal{X}^{seen}, y_{\mathbf{x}} \in \mathcal{Y}^{seen}, \mathbf{a}_y \in \mathcal{A}^{seen}\}$ and $\mathcal{D}^{unseen} = \{(\tilde{y}_i, \mathbf{a}_{\tilde{y}_i}) \mid \tilde{y}_i \in \mathcal{Y}^{unseen}, \mathbf{a}_{\tilde{y}_i} \in \mathcal{A}^{unseen}\}$, where $\mathbf{x}$ is an image from the seen class sample set $\mathcal{X}^{seen}$, $y_{\mathbf{x}}$ or $\tilde{y}_i$ is a class label, $\mathbf{a}_y, \mathbf{a}_{\tilde{y}_i} \in \mathbb{R}^M$ are M-dimensional class semantic embeddings (*i.e.*, class-level side information), $\mathcal{A}^{seen}$ and $\mathcal{A}^{unseen}$ contain the semantic embeddings of seen and unseen classes respectively. For test images consisting of images of both unseen and unknown classes, *i.e.*, $\mathcal{X}^{unseen} \cup \mathcal{X}^{unknown}$, ZS-OSR aims to learn a model ϕ^{open} that maps the images to the class label space $\mathcal{O} = \mathcal{Y}^{unseen} \cup \{'unknown'\}$, where $\mathcal{O}$ includes an '*unknown*' class label in addition to the labels of the unseen classes.

3.2 The Challenges

One of the key challenges in ZS-OSR is the difficulty in distinguishing between the unseen and unknown classes due to the lack of training data for both of them. In typical OSR problems, the most commonly used approach involves conducting OSR first, followed by classification. Nevertheless, in ZS-OSR, none of the currently existing OSR methods have been found to be effective due to the lack of images from the unseen class for training purposes. To address this challenge, a straightforward solution would be to use simple combinations of existing generative ZSL and OSR methods. Specifically, generative ZSL methods can be first applied to generate latent visual features of unseen classes. The ZS-OSR task is then converted to a general OSR task in the visual feature space that contains the features of both seen and unseen classes. OSR methods can then be directly used to learn an open-set classifier using the training set composed of these features to recognize known (unseen) classes while being capable of rejecting unknown classes. Figure 2 shows the results of such solutions that uses the widely-used TF-VAEGAN [6] and MSP [33], OpenMax [28], Placeholder [29], Energy [35], ODIN [34], LogitNorm [48], and MaxLogit [49], as the generative ZSL method and the OSR methods, respectively, where the open scores are the likelihood scores of being unknown class samples yielded by the open-set classifier, with larger open scores indicating higher likelihood (see Table 4 for detailed quantitative results of these methods).

The results show that the method performs poorly in distinguishing between unseen and unknown classes since the open scores of many unseen class samples are large, highly overlapping with that of the unknown class samples. The ineffective performance may be impacted by the fact that the generative ZSL model generates features for the known unseen classes based on the closed-set assumption, without being informed of the possible presence of unknown classes. As a result, these
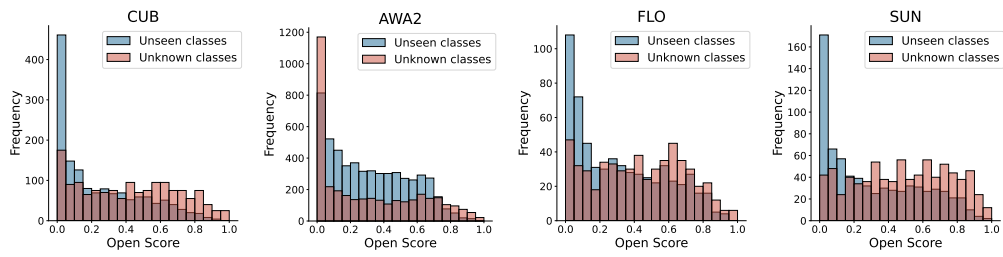
(a) MSP
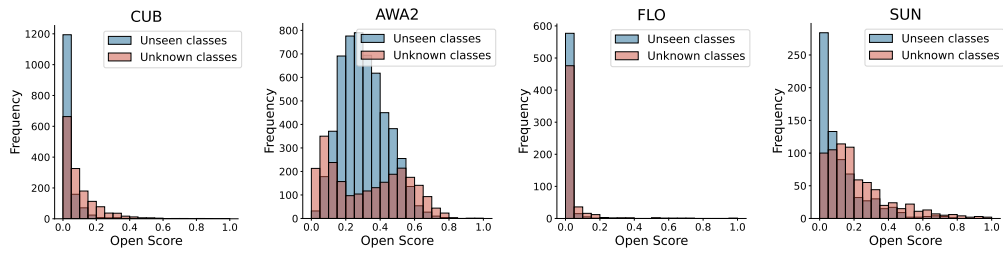
(b) OpenMax
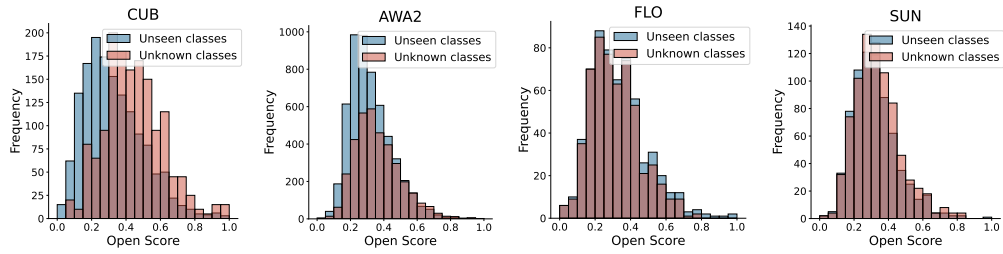
(c) Placeholder

(d) Energy

(e) ODIN

5

Figure 2: Distribution of the open scores yielded by combinations of TF-VAEGAN [6] and MSP [33], OpenMax [28], Placeholder [29], Energy [35], ODIN [34], LogitNorm [48], and MaxLogit [49], combination of APN [15] and MSP for the test images of unseen and unknown classes on CUB, AWA2, FLO and SUN datasets.

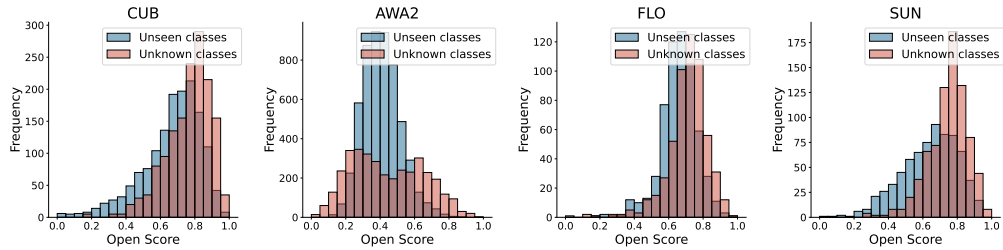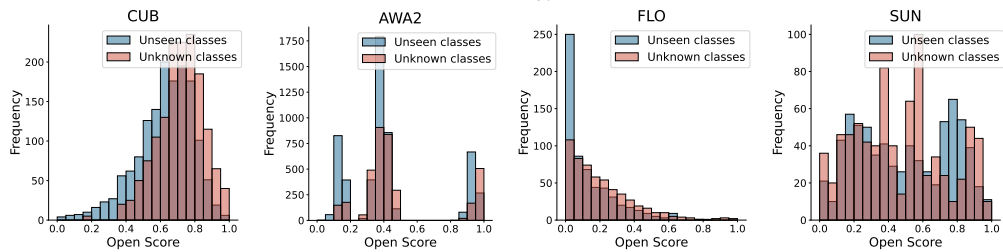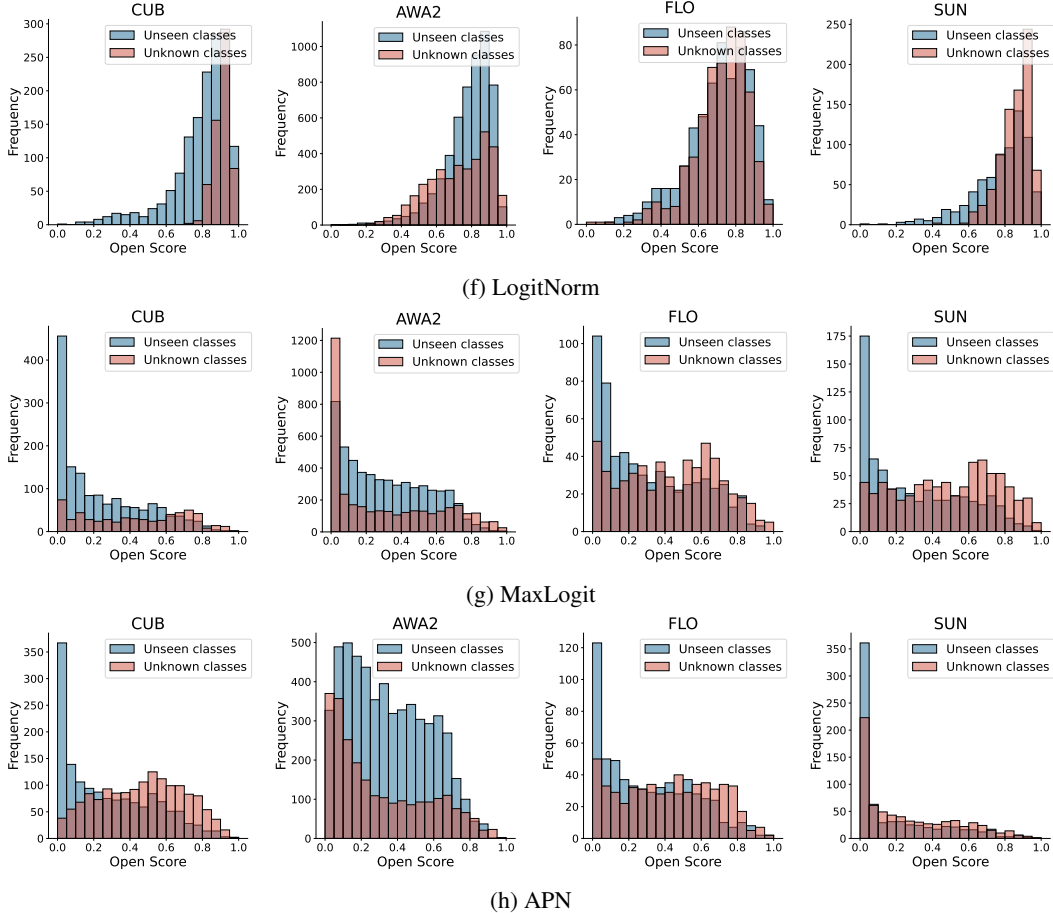generated unseen features can heavily overlap with that of the unknown classes in the feature space, rendering the subsequent OSR models ineffective. In the next section, we introduce the ASE approach that learns an open-set classifier for the zero-shot setting to address this issue.

4 The Proposed Approach

4.1 Overview of Our Approach

Our proposed approach, namely Adversarial Semantic Embeddings (ASE), is a generative framework specifically designed for the ZS-OSR problem. Since ZS-OSR does not provide training samples of unseen and unknown classes, the framework aims to directly generate these samples to train a classifier for this task. Unlike the simple solution in Sec. 3.2 that directly generates the features of unseen classes for the subsequent OSR task, ASE takes a step back and focuses on learning faithful semantic embeddings of unknown classes via an adversarial learning approach before training an unknowns-informed OSR model.

An overview of ASE is provided in Figure 3, which consists of three successive components, including *Using off-the-shelf generative ZSL models to generate unseen classes*, *learning adversarial semantic embeddings of unknown classes*, and *unifying these learned features and semantic embeddings to train a $K + 1$ open-set classifier* where $K = |\mathcal{Y}^{unseen}|$. The first component is to directly take an existing generative ZSL model to generate the visual features of the unseen classes based on
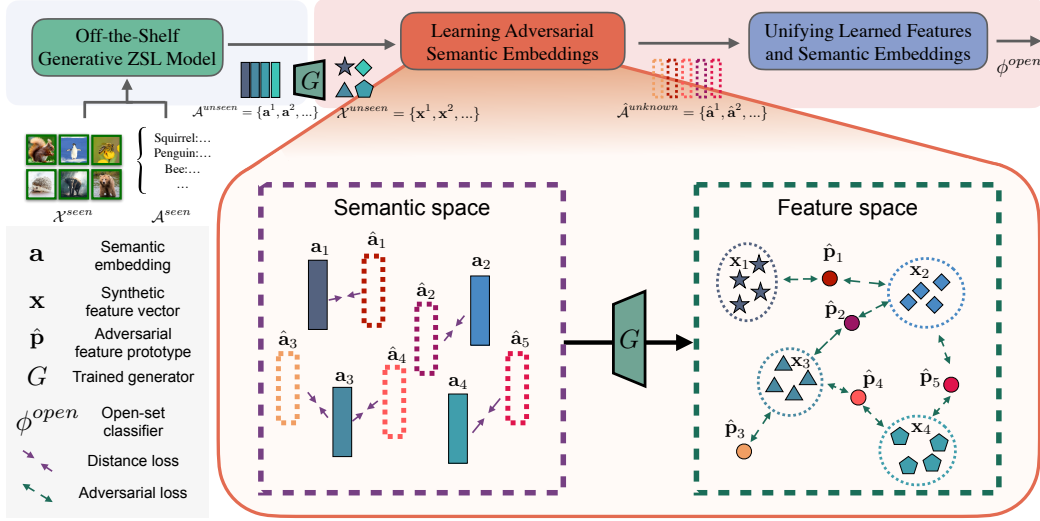
Figure 3: The proposed approach ASE. Given a pre-trained generative ZSL model, ASE first uses its generator G to generate the features of unseen classes. It then learns a set of adversarial semantic embeddings of unknown classes so that they are tightly distributed around but separable from the unseen-class embeddings. Lastly ASE uses the unknown-class embeddings to train an unknowns-informed open-set classifier ϕ^{open}.

their semantic embeddings. The second component is the key novelty of ASE, which is designed to take the trained generator and the closed-set classifier in the ZSL model as input to learn a set of adversarial semantic embeddings of unknown classes, such that the learned semantic embeddings are distributed around the boundary between the known and unknown classes in the semantic space. In the third component, these semantic embeddings are subsequently used to generate a set of adversarial feature vectors to represent the samples of the unknown classes, along with the previously generated unseen class samples, to train the $K + 1$ classifier for OSR. Below we introduce each component in detail.

4.2 Using Off-the-Shelf Generative ZSL Models to Generate Unseen-Class Features

Training a zero-shot open-set classifier in the visual feature space requires the features of unseen classes. Generative ZSL models have demonstrated superior performance in utilizing the relationship between semantic embeddings and image features of the seen classes to generate the unseen-class features. Therefore, existing off-the-shelf generative ZSL models are directly taken to generate these features. Briefly, generative ZSL methods learn a generator network G to generate sample $\tilde{\mathbf{x}} = G(\mathbf{a}, \epsilon)$ conditioned on a Gaussian noise $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and a class semantic embedding $\mathbf{a}$. Meanwhile, a discriminator network $D(\mathbf{x}, \mathbf{a})$ is learned by taking an input feature $\mathbf{x}$ and outputs a real value representing the probability that $\mathbf{x}$ is from the real data rather than from the generator network. G and D can be learned by optimizing the following adversarial objective:

$$\mathcal{L} = \mathbb{E}[D(\mathbf{x}, \mathbf{a})] - \mathbb{E}[D(\tilde{\mathbf{x}}, \mathbf{a})], \tag{1}$$

where $\mathbf{x} \in \mathcal{X}^{unseen}$ is an image from the seen classes and $\mathbf{a} \in \mathcal{A}^{seen}$ is its class semantic embedding. The trained generator generates synthetic features $\tilde{\mathcal{X}}^{unseen} = \{\tilde{\mathbf{x}}_l^{\tilde{y}_i} | \tilde{\mathbf{x}}_l^{\tilde{y}_i} = G(\mathbf{a}_{\tilde{y}_i}, \epsilon_l)\}$ for semantic embedding $\mathbf{a}_{\tilde{y}_i}$ of each class, where $\mathbf{a}_{\tilde{y}_i} \in \mathcal{A}^{unseen}$. Then we obtain $\tilde{\mathcal{D}}^{unseen} = \{(\tilde{\mathbf{x}}, \tilde{y}_{\tilde{\mathbf{x}}}) \mid \tilde{\mathbf{x}} \in \tilde{\mathcal{X}}^{unseen}, \tilde{y}_{\tilde{\mathbf{x}}} \in \mathcal{Y}^{unseen}\}$, and train a general ZSL (closed-set) classifier ϕ^{closed} to classify the test image samples from unseen classes.

Existing state-of-the-art (SOTA) generative ZSL models [24, 5, 6] can be directly used to implement this component. The generator G and the ϕ^{closed} classifier (or the generated unseen-class features) are taken as input to adversarially learn the semantic embeddings of unknown classes.

7

4.3 Learning Adversarial Semantic Embeddings of Unknown Classes

ASE is dedicated to learning an unknowns-informed open-set classifier. However, we are given neither semantic embeddings nor image samples of the unknown classes. ASE aims to learn adversarial representations of the unknown classes to train such a classifier. Given the generated features and the training data, the representations of the unknown classes can be adversarially learned in either the semantic embedding space or the visual feature space. However, as discussed in Sec. 3.2, the generated unseen-class features are not faithful enough as they are generated under the closed-set setting. Consequently, directly using these synthetic unseen-class features to generate the unknown-class features may accumulate and/or amplify the closed-set biases in the generator, leading to non-discriminative features for the unknown classes. Such unseen-class and unknown-class features are ineffective in training the open-set classifier (see Table 5).

Thus, ASE instead learns the unknown-class representations in the semantic space, while enforcing separable unseen-and-unknown representations in the visual feature space. This approach is more plausible since 1) it directly learns the unknown-class semantic embeddings based on the pre-defined embeddings of both seen and unseen classes, while the other approach above is an indirect way that heavily relies on the unstable quality of the generated unseen-class visual features; and 2) it seamlessly leverages the learned relation between the semantic space and the visual feature space in the ZSL models to learn the unknown-class semantic embeddings.

To this end, ASE introduces a class-wise adversarial semantic embedding learning approach to generate a set of semantic embeddings of the unknowns, $\mathcal{A}^{unknown}$. As highlighted in Figure 3, for each unseen class, ASE generates multiple adversarial semantic embeddings for the unknown classes that are tightly distributed around but separable from the unseen-class embeddings. To achieve this goal, it jointly minimizes a distance loss $\mathcal{L}_{\text{dis}}$ in the embedding space that brings the unknown-class embeddings closer to the unseen-class embeddings, and an adversarial loss $\mathcal{L}_{\text{adv}}$ in the feature space that pulls the corresponding prototypical unknown-class features away from the generated unseen-class features:

$$\mathcal{L}_{\text{ase}} = \mathcal{L}_{\text{adv}} + \beta \mathcal{L}_{\text{dis}}, \tag{2}$$

where β is a hyper-parameter, $\mathcal{L}_{\text{dis}}$ is defined as the Euclidean distance between the generated unknown-class embedding $\hat{\mathbf{a}} \in \mathcal{A}^{unknown}$ and the given unseen-class embedding $\tilde{\mathbf{a}} \in \mathcal{A}^{unseen}$:

$$\mathcal{L}_{\text{dis}} = \|\hat{\mathbf{a}} - \tilde{\mathbf{a}}\|_2, \tag{3}$$

and $\mathcal{L}_{\text{adv}}$ is defined as a *Helmholtz free energy*-based loss:

$$\mathcal{L}_{\text{adv}} = T \cdot \log \sum_{i}^{K} e^{\phi_i^{closed}(\hat{\mathbf{p}})/T}, \tag{4}$$

where ϕ^{closed} is the ZSL (closed-set) classifier obtained from the off-the-shelf ZSL model, and $\hat{\mathbf{p}} = G(\hat{\mathbf{a}}, \epsilon)$ is a generated prototypical unknown-class feature vector corresponding to the adversarial semantic embeddings $\hat{\mathbf{a}}$. Note that since ϕ^{closed} was trained using the unseen-class features and its weight parameters are fixed, the energy score that the unseen-class features receive are consistently low. Thus, Eq. (4) is designed to encourage high energy scores for unknown-class feature prototypes only. By minimizing $\mathcal{L}_{\text{ase}}$, the unknown-class and unseen-class class embeddings are close to each other, but they are discriminative from each other; this adversarial relation also applies to the corresponding unknown-class feature prototypes w.r.t. the unseen-class features.

4.4 Unknowns-Informed ZS-OSR

We then train an unknowns-informed ZS-OSR classifier with $K + 1$ classes, in which the extra (+1) class is for the '*unknown*' class and it is trained based on the learned unknown-class semantic embeddings.

Specifically, we first utilize the adversarial semantic embeddings $\hat{\mathcal{A}}^{unknown}$ and the trained generator G to obtain a set of unknown-class features $\hat{\mathcal{X}}^{unknown} = \{\hat{\mathbf{x}}_l^{\hat{y}_i} | \hat{\mathbf{x}}_l^{\hat{y}_i} = G(\mathbf{a}_{\hat{y}_i}, \epsilon_l), \mathbf{a}_{\hat{y}_i} \in \hat{\mathcal{A}}^{unknown}, \hat{y}_i \in \mathcal{Y}^{unknown}\}$, where $\mathcal{Y}^{unknown}$ is a set of unknown classes collectively labeled as '*unknown*', resulting in $\tilde{\mathcal{D}}^{unknown} = \{(\hat{\mathbf{x}}, 'unknown') | \tilde{\mathbf{x}} \in \hat{\mathcal{X}}^{unknown}\}$. The unknown-class features $\hat{\mathbf{x}}$ are expected to be centered around the unknown-class feature prototypes $\hat{\mathbf{p}}$. The unseen-class and unknown-class features are then combined to form the open-set training data, *i.e.*,

Dataset	#images	Dimension	$\lvert KS \rvert$	$\lvert KU \rvert$	$\lvert U \rvert$
CUB [9]	11788	312	150	25	25
AWA2 [10]	37322	85	40	5	5
FLO [11]	8189	1024	82	10	10
SUN [12]	14340	102	645	36	36

Table 2: Information on proposed ZS-OSR splits.

Dataset	Unseen	Unknown
AWA2	7, 23, 24, 31, 47	9, 30, 34, 41, 50
FLO	1, 3, 4, 5, 6, 8, 11, 16, 17, 20	2, 7, 9, 10, 12, 13, 14, 15, 18, 19
CUB	7, 19, 21, 34, 36, 56, 68, 79, 80, 88, 91, 98, 104, 108, 124, 142, 150, 152, 157, 166, 171, 179, 182, 187, 195	29, 50, 62, 69, 72, 87, 95, 100, 116, 120, 122, 125, 129, 139, 141, 159, 160, 167, 174, 176, 185, 189, 191, 192, 193
SUN	11, 25, 33, 39, 54, 73, 75, 76, 100, 146, 185, 217, 222, 238, 255, 263, 287, 316, 329, 337, 343, 359, 449, 483, 494, 510, 559, 561, 623, 632, 646, 651, 657, 659, 675, 712	4, 24, 58, 86, 96, 104, 113, 125, 131, 139, 153, 159, 197, 246, 247, 260, 299, 354, 380, 382, 421, 424, 426, 441, 472, 509, 518, 530, 581, 636, 680, 682, 696, 711, 713, 716

Table 3: Proposed splits on ZS-OSR datasets.

$\mathcal{D}^{open} = [\tilde{\mathcal{D}}^{unseen}, \tilde{\mathcal{D}}^{unknown}]$, which is used to train the open-set classifier ϕ^{open} by minimizing a standard cross-entropy loss:

$$\min_{\theta} \mathbb{E}_{(\mathbf{x}, y_\mathbf{x}) \sim \mathcal{D}^{open}} \left[-\log \phi_{y_\mathbf{x}}^{open}(\mathbf{x}) \right]. \tag{5}$$

During inference, given a test image $\mathbf{x}$, ϕ^{open} yields a softmax score in its $K+1$-th class prediction, which can be directly used as an open score. If the score exceeds a pre-defined threshold, $\mathbf{x}$ is predicted as '*unknown*', and otherwise $\mathbf{x}$ is predicted as the class with the highest logit among the K unseen classes. Alternatively, post-hoc OSR methods like MSP [33] and ODIN [34] can also be applied for obtaining the open score, but the ϕ^{open}-based open score is generally more effective (shown in Table 5), and is used by default.

5 Experiments

5.1 Experimental Setup

Datasets. To our knowledge, there are no publicly-available datasets designed for evaluating the performance of ZS-OSR, so we introduce four ZS-OSR datasets adapted from four existing widely-used ZSL datasets: Caltech-UCSD-Birds 200-2011 (CUB) [9], Animals with Attributes 2 (AWA2) [10], FLO [11] and SUN [12]. In particular, we first used the commonly-used seen/unseen class split as in [10] and [50]. However, this data split does not provide the unknown classes. In order to facilitate the ZS-OSR setup on different datasets without loss of generality, for the test data of each dataset, we further randomly take half of the unseen classes as the unknown classes. In other words, the test data is the same as ZSL datasets but it now contains both unseen and unknown classes; while part of semantic information in the original ZSL training data is removed to create the unknown classes in the test data. Detailed information of the datasets is presented in Tables 2 and 3.

6 Information of Datasets

Evaluation Metrics. Following [28], we evaluate both closed-set classification and open-set detection performance. Classification accuracy (Acc) is used to measure the performance of classifying the closed-set samples, *i.e.*, test samples from the unseen classes, while the FPR95 and Area Under ROC Curve (AUROC) are used to measure the performance of detecting open-set samples, *i.e.*, test samples from the unknown classes. All reported Acc, FPR95 and AUROC results are averaged over five independent runs.

Method	CUB			AWA2			FLO			SUN		
	Acc ↑	FPR95 ↓	AUC ↑	Acc ↑	FPR95 ↓	AUC ↑	Acc ↑	FPR95 ↓	AUC ↑	Acc ↑	FPR95 ↓	AUC ↑
MSP [33]	76.33	80.96	70.08	71.56	99.61	47.51	81.41	91.33	63.74	73.33	75.00	71.63
OpenMax [28]	76.41	79.54	74.98	70.33	99.43	49.86	81.34	97.16	52.80	72.94	74.03	69.75
Placeholder [29]	76.11	83.25	72.43	70.63	91.43	47.25	82.92	95.17	47.90	72.78	94.03	53.55
Energy [35]	76.19	82.04	71.52	71.01	99.96	67.87	81.37	92.33	68.14	73.33	70.45	71.06
ODIN [34]	76.25	84.94	69.39	70.94	89.51	62.86	81.15	92.00	63.53	72.97	73.75	71.15
LogitNorm [48],	75.00	72.92	73.41	69.04	99.29	45.79	78.76	92.50	67.65	66.11	76.25	65.21
MaxLogit [49]	76.74	82.92	71.87	71.64	99.58	46.70	80.16	91.00	62.11	72.92	74.86	69.70
APN [15]	77.41	69.82	71.82	71.41	97.66	42.71	76.53	87.62	63.31	67.50	85.00	62.58
ASE (Ours)	76.26	68.67	80.20	72.30	77.09	81.99	82.44	87.50	72.78	73.61	70.41	72.69

Table 4: Main results. Accuracy (%), FPR95 and AUROC of ASE and baselines on four ZS-OSR datasets. The best (second-best) results are boldfaced (underlined).

Implementation Details of ASE. Our ASE approach, outlined in Sec. 4, relies on an off-the-shelf generative ZSL method, TF-VAEGAN [6], to produce feature vectors for unseen classes based on their semantic embeddings. We use ResNet101 [51] to extract features for $\mathcal{X}$ and conduct a grid search on the validation set [10] to determine the optimal hyperparameter β for our model. We generate 50 unknown-class semantic embeddings around each unseen class and produce 1,000 adversarial samples for each unknown-class semantic embedding in the feature space, resulting in $|\tilde{\mathcal{D}}^{unknown}| = 1,000 \times |\mathcal{Y}^{unseen}|$ unknown-class samples per dataset. We train the unknowns-informed open-set classifier with a linear classifier featuring one fully connected layer and optimize it with the Adam optimizer using Eq. (5). We maintain the same hyperparameters as in TF-VAEGAN, which we hold fixed throughout the training process of the open-set classifier.

Comparison Baselines. Although there are no methods reported to deal with the ZS-OSR problem, SOTA ZSL and OSR methods can be combined to establish some good solutions. Similar to ASE, TF-VAEGAN is used as a SOTA ZSL method here and combined with eight diverse SOTA OSR methods, including **MSP** [33], **OpenMax** [28], **ODIN** [34], **Placeholder** [29], **Energy** [35], **LogitNorm** [48], and **MaxLogit** [49], Since all of these methods are for OSR or OOD detection and do not support ZS-OSR tasks, we adapt them to ZS-OSR using the following method: 1) We first generate the features of the unseen classes using the ZSL generative method TF-VAEGAN, and then 2) we treat the unseen classes as closed-set classes and apply one of these eight OSR methods to recognize the unseen classes while rejecting unknown-class samples. Additionally, we evaluate the performance of a popular non-generative ZSL method **APN** [15] in conjunction with **MSP**. All the hyperparameters of the baselines are tuned in the same way as ASE on each dataset to have a fair empirical comparison.

6.1 ZS-OSR Performance

The ZS-OSR results of ASE and the combined baselines on the four proposed datasets are shown in Table 4. ASE outperforms all eight baseline methods by a significant margin in detecting unknown-class samples and performs comparably well in terms of classification on the unseen-class samples on all four datasets. Details are discussed below.

Superior unknown-class detection. The baseline methods are inconsistent in detecting unknown-class samples on the four datasets, whereas ASE performs consistently well on all of them. ASE significantly improves the AUROC score compared to the best competing baseline on CUB, AWA2, and FLO by margins of 5.22%, 14.12%, and 4.64%, respectively. ASE is also the best performer on SUN, with a relatively marginal improvement. In terms of the FPR95 metric, ASE also demonstrates the best performance across all datasets. Notably, it reduces the FPR95 by 12.42% compared to the best baseline on AWA2.

Maintaining classification accuracy on unseen classes. ASE's superior ability to detect unknown-class samples does not affect its unseen-class classification ability. As seen in Table 4, ASE achieves the best overall accuracy performance across the four datasets, as competing methods such as Open-Max and Placeholder have large accuracy drops on certain datasets.

The reasons behind. As discussed in Sec. 3.2, the simply combined ZSL-and-OSR solutions fail to produce discriminative open scores for the unseen and unknown class samples, as illustrated in Figure 2. In contrast, ASE yields significantly more discriminative open scores for the samples of
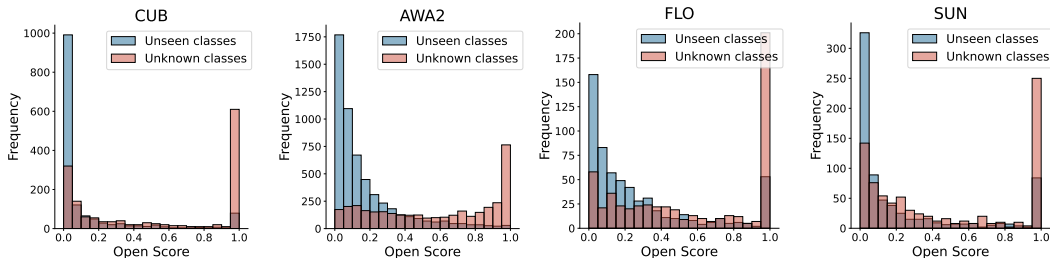
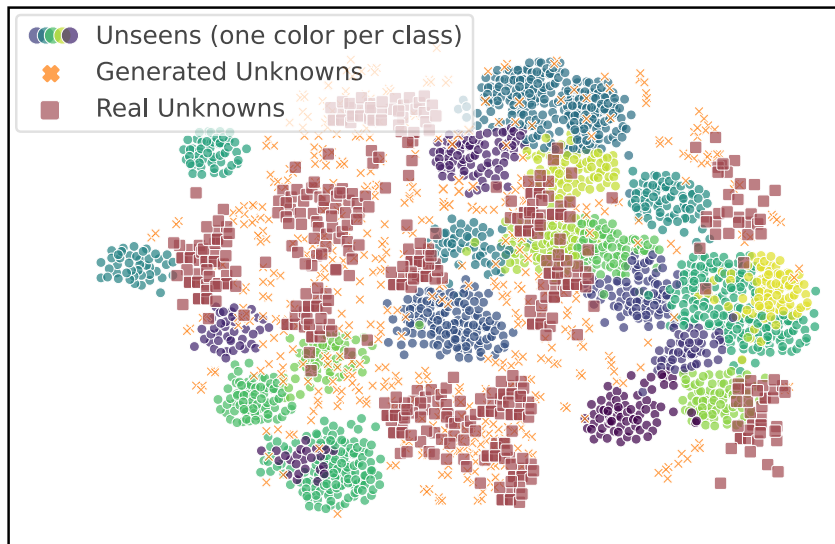Figure 4: Distribution of the open scores yielded by ASE.

Figure 5: T-SNE visualization of ASE features on CUB.

the unseen and unknown classes, as demonstrated in Figure 4. Furthermore, as shown in Figure 5, the unknown samples generated by ASE either lie between unseen and true unknowns or overlap with the real unknown samples, suggesting that ASE can effectively leverage the ZSL training data to generate unknown-class representations and distinguish them from the unseen-class samples through the adversarial unknown-class embedding learning in both the semantic and feature spaces.

6.2 Effectiveness on Data with Varying Openness

Following the OSR literature [52, 29, 53, 54], we conduct experiments on the CUB dataset to examine ZS-OSR performance under varying degrees of openness, defined as $1 - \sqrt{\frac{|\mathcal{Y}^{\text{unseen}}|}{|\mathcal{Y}^{\text{unseen}}| + |\mathcal{Y}^{\text{unknown}}|}}$. A larger openness indicates the presence of relatively more unknown classes in the test data. The experiment is focused on the CUB dataset. In particular, there are 50 unseen classes in CUB under the ZSL setting [10]. We create four ZS-OSR datasets based on CUB by retaining 10 classes as the unseen classes and the respectively remaining 10, 20, 30, and 40 classes as the unknown classes. This results in four datasets with respective openness of 29.3%, 42.3%, 50%, and 55.3%. The results of ASE and baselines on these four datasets are shown in Figure 6. It is clear that ASE maintains consistently superiority in unknown-class detection and comparably good unseen-class classification across different openness rates, demonstrating strong robustness and stability to the data openness.

6.3 Ablation Study

Table 5 shows the ablation study results of two main modules in ASE, including $\mathcal{L}_{ase}$ and ϕ^{open}. We discuss the results in detail below.
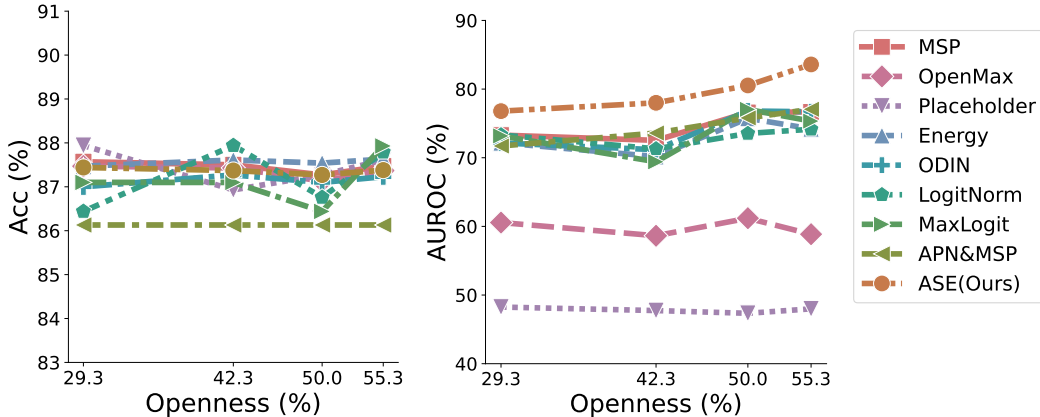
Figure 6: Performance w.r.t. openness rates on CUB.

Unknowns	**Score**	**AUROC**			
		CUB	AWA2	FLO	SUN
Mixup	ϕ^{open}	60.33	25.31	46.96	57.98
Uniform Noise	ϕ^{open}	63.58	46.11	50.42	57.29
Semantic + Noise	ϕ^{open}	63.74	69.02	63.22	54.77
Adversarial Features	ϕ^{open}	74.92	54.70	**73.19**	70.75
$\mathcal{L}_{ase}$	ϕ^{open}	**80.20**	**81.99**	72.78	72.69
$\mathcal{L}_{ase}$	MSP	71.69	75.07	69.63	**76.11**
$\mathcal{L}_{ase}$	ODIN	72.07	74.67	66.81	71.96

Table 5: Results of ASE (shaded) and its six variants

Adversarial semantic embedding learning via $\mathcal{L}_{ase}$. We compared the $\mathcal{L}_{ase}$-based adversarial embedding learning module with four alternative methods, including *Mixup* [55], *Uniform Noise*, *Semantic + Noise*, and *Adversarial Features* [13, 14], as described in Section 4.3. Table 5 shows that although some of the simpler methods, such as *Semantic + Noise*, can achieve fairly good performance compared to the baselines in Table 4, our full model ($\mathcal{L}_{ase} + \phi^{open}$) consistently outperforms them in AUROC on all four datasets. While *Adversarial Features* is more effective than the other baselines, it performs rather unstably. In summary, ASE is consistently the best performer.

Unknowns-informed open-set classifier via ϕ^{open}. As discussed in Section 4.4, Post-hoc OSR methods, such as ODIN and MSP, can also be applied to the final classifier trained by ASE. Table 5 shows that ASE-enabled ODIN and MSP methods can largely improve their unknown detection performance compared to the original ODIN and MSP methods (see Table 4), indicating that the final unknowns-informed classifier trained by ASE is more effective in discriminating the unknown samples than the classifier in the original generative ZSL model. However, ASE substantially outperforms both ASE-enabled ODIN and MSP methods in AUROC on CUB, AWA2, and FLO, with a maximal increase of about 9% on CUB and 7% on AWA2. Although ASE-enabled MSP obtains the best AUROC on SUN, it performs poorly on other datasets. Overall, since ϕ^{open} is end-to-end trained to detect unknown-class samples, it is much more effective than the heuristic methods, ASE-enabled ODIN and MSP.

6.4 Extending to Generalized ZS-OSR and ZS-OOD Settings

Our approach focuses on distinguishing unseen and unknown samples in ZS-OSR, but ASE can be extended to function under generalized ZS-OSR settings with a minor modification to Eq. (3). ASE generates tightly distributed adversarial semantic embeddings around each of the seen-class and unseen-class semantic embeddings, and can generate unknown-class features and train the ϕ^{open} classifier with seen-class, unseen-class, and unknown-class features. SOTA generalized ZS model **GCM-CF** [38] with MSP is included to extend our baselines under this setting. The results are

presented in Table 6, where ASE achieves the most effective detection of unknown samples across all four datasets, outperforming the best baseline per dataset by 0.1-4.9% in AUROC. In terms of FPR95, ASE also maintains a similarly leading position. Although APN surpasses ASE on two datasets in FPR95, ASE substantially outperforms it in both AUROC and FPR on the other datasets.

Method	CUB		AWA2		FLO		SUN	
	FPR95 ↓	AUC ↑	FPR95 ↓	AUC ↑	FPR95 ↓	AUC ↑	FPR95 ↓	AUC ↑
MSP	81.1	66.2	84.7	63.3	70.6	72.5	88.6	60.5
OpenMax	92.4	55.1	95.5	59.0	81.8	66.6	94.3	48.7
Placeholder	79.3	71.0	91.9	45.9	94.8	50.4	88.0	57.4
Energy	89.3	64.1	89.3	50.8	81.2	52.7	91.8	59.6
ODIN	85.1	67.8	78.5	68.2	80.5	75.1	89.9	60.7
LogitNorm	78.6	69.6	75.9	62.1	75.3	66.5	87.6	61.7
MaxLogit	84.2	64.8	83.7	62.6	70.2	65.4	89.1	59.5
APN	88.3	56.7	69.1	73.9	54.3	71.7	90.9	56.1
GCM-CF	84.3	64.5	74.0	67.0	67.3	75.7	90.2	60.6
ASE (Ours)	77.8	75.9	73.8	78.7	69.4	79.1	87.5	61.8

Table 6: Results under generalized ZS-OSR setting.

Method	CUB-AWA2		CUB-FLO		CUB-SUN	
	FPR95 ↓	AUC ↑	FPR95 ↓	AUC ↑	FPR95 ↓	AUC ↑
MSP	83.0	72.6	84.2	74.2	86.4	75.0
OpenMax	95.1	48.2	91.3	41.6	91.6	43.3
Placeholder	79.5	70.9	78.9	72.7	81.5	72.0
Energy	69.4	91.8	74.9	98.6	56.4	99.9
ODIN	91.7	69.2	78.9	62.4	83.9	70.5
LogitNorm	66.8	69.2	61.4	62.4	55.4	70.5
MaxLogit	86.7	79.4	83.1	83.7	81.4	89.0
APN	45.8	88.4	25.7	94.0	31.3	93.1
ASE (Ours)	4.0	97.2	5.3	99.0	0.3	99.9

Table 7: Results under generalized ZS-OOD setting.

While ZS-OSR focuses on the scenario where known classes and unknown classes belong to the same distribution, we believe that there exists scenarios of ZS-OOD (Out-of-Distribution), where we only know the semantic information of the known classes but need to detect images of unknown classes from a different distribution (*e.g.*, samples from a largely different dataset). To evaluate ASE's performance in the ZS-OOD setting, we randomly select 25 classes from AWA2, FLO, and SUN as unknown classes and combine them with 25 unseen classes of CUB to obtain three ZS-OOD test datasets, CUB-AWA2, CUB-FLO, and CUB-SUN. Table 7 shows that considering both AUROC and FPR95 metrics, ASE's performance surpasses all baselines by a significant margin, demonstrating the empirical effectiveness of ASE under the ZS-OOD setting.

7 Conclusions

This work introduces ZS-OSR, a problem setting which extends ZSL to open-set scenarios, and analyzes the challenge of distinguishing samples of unseen and unknown classes. To promote the development and evaluation of ZS-OSR methods, we build eight baselines that combine SOTA ZSL and OSR models, and establish performance benchmarks by applying them to four ZS-OSR datasets adapted from ZSL datasets. We further propose the ASE approach that learns adversarial semantic embeddings to accurately detect the unknown samples while maintaining preferable classification accuracy of the unseen-class samples. Empirical results show that ASE 1) outperforms the baselines on the four datasets in AUROC, 2) performs stably on datasets with varying openness, and 3) can be easily extended to detect the unknown samples under generalized ZS-OSR settings and ZS-OOD settings.

References

[1] C. H. Lampert, H. Nickisch, S. Harmeling, Learning to detect unseen object classes by between-class attribute transfer, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 951–958. 2, 3

[2] W. Xu, Y. Xian, J. Wang, B. Schiele, Z. Akata, Vgse: Visually-grounded semantic embeddings for zero-shot learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 9316–9325. 2, 3

[3] S. Chen, Z. Hong, Y. Liu, G.-S. Xie, B. Sun, H. Li, Q. Peng, K. Lu, X. You, Transzero: Attribute-guided transformer for zero-shot learning, Proceedings of the AAAI Conference on Artificial Intelligence 36 (1) (2022) 330–338. 2, 3

[4] V. K. Verma, G. Arora, A. Mishra, P. Rai, Generalized zero-shot learning via synthesized examples, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. 2, 3

[5] Y. Xian, S. Sharma, B. Schiele, Z. Akata, f-vaegan-d2: A feature generating framework for any-shot learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10275–10284. 7

[6] S. Narayan, A. Gupta, F. S. Khan, C. G. Snoek, L. Shao, Latent embedding feedback and discriminative features for zero-shot classification, in: European Conference on Computer Vision, Springer, 2020, pp. 479–495. 2, 3, 4, 6, 7, 10

[7] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, T. E. Boult, Toward open set recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (7) (2013) 1757–1772. 2, 3

[8] C. Geng, S.-j. Huang, S. Chen, Recent advances in open set recognition: A survey, IEEE transactions on pattern analysis and machine intelligence 43 (10) (2020) 3614–3631. 2, 3

[9] C. Wah, S. Branson, P. Welinder, P. Perona, S. Belongie, The caltech-ucsd birds-200-2011 dataset, Tech. Rep. CNS-TR-2011-001 (2011). 2, 9

[10] Y. Xian, C. H. Lampert, B. Schiele, Z. Akata, Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (9) (2019) 2251–2265. 2, 9, 10, 11

[11] M.-E. Nilsback, A. Zisserman, Automated flower classification over a large number of classes, in: 2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing, 2008, pp. 722–729. 2, 9

[12] G. Patterson, J. Hays, Sun attribute database: Discovering, annotating, and recognizing scene attributes, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2751–2758. 2, 9

[13] Y. Yu, W.-Y. Qu, N. Li, Z. Guo, Open category classification by adversarial sample generation, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 2017, pp. 3357–3363. 3, 12

[14] L. Neal, M. Olson, X. Fern, W.-K. Wong, F. Li, Open set learning with counterfactual images, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018. 3, 12

[15] W. Xu, Y. Xian, J. Wang, B. Schiele, Z. Akata, Attribute prototype network for zero-shot learning, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems, Vol. 33, Curran Associates, Inc., 2020, pp. 21969–21980. 3, 6, 10

[16] Y. Liu, L. Zhou, X. Bai, Y. Huang, L. Gu, J. Zhou, T. Harada, Goal-oriented gaze estimation for zero-shot learning, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 3793–3802.

[17] F. Zhang, G. Shi, Co-representation network for generalized zero-shot learning, in: K. Chaudhuri, R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, Vol. 97 of Proceedings of Machine Learning Research, PMLR, 2019, pp. 7434–7443.

[18] Y. Liu, X. Gao, J. Han, L. Liu, L. Shao, Zero-shot learning via a specific rank-controlled semantic autoencoder, Pattern Recognition 122 (2022) 108237.

[19] Z. Chen, Y. Gao, C. Lang, L. Wei, Y. Li, H. Liu, F. Liu, Integrating topology beyond descriptions for zero-shot learning, Pattern Recognition (2023) 109738.

[20] H. Kim, J. Lee, H. Byun, Discriminative deep attributes for generalized zero-shot learning, Pattern Recognition 124 (2022) 108435.

[21] H. Zhang, L. Liu, Y. Long, Z. Zhang, L. Shao, Deep transductive network for generalized zero shot learning, Pattern Recognition 105 (2020) 107370.

[22] J. Zhang, Q. Li, Y.-a. Geng, W. Wang, W. Sun, C. Shi, Z. Ding, A zero-shot learning framework via cluster-prototype matching, Pattern Recognition 124 (2022) 108469.

[23] H. Zhang, H. Bai, Y. Long, L. Liu, L. Shao, A plug-in attribute correction module for generalized zero-shot learning, Pattern Recognition 112 (2021) 107767. 3

[24] J. Li, M. Jing, K. Lu, Z. Ding, L. Zhu, Z. Huang, Leveraging the invariant side of generative zero-shot learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7402–7411. 3, 7

[25] N. Bendre, K. Desai, P. Najafirad, Generalized zero-shot learning using multimodal variational auto-encoder with semantic concepts, in: 2021 IEEE International Conference on Image Processing (ICIP), IEEE, 2021, pp. 1284–1288. 3

[26] T. Shermin, S. W. Teng, F. Sohel, M. Murshed, G. Lu, Integrated generalized zero-shot learning for fine-grained classification, Pattern Recognition 122 (2022) 108246. 3

[27] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, I. Sutskever, Learning transferable visual models from natural language supervision, in: M. Meila, T. Zhang (Eds.), Proceedings of the 38th International Conference on Machine Learning, Vol. 139 of Proceedings of Machine Learning Research, PMLR, 2021, pp. 8748–8763. 3, 4

[28] A. Bendale, T. E. Boult, Towards open set deep networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1563–1572. 3, 4, 6, 9, 10

[29] D.-W. Zhou, H.-J. Ye, D.-C. Zhan, Learning placeholders for open-set recognition, CoRR abs/2103.15086 (2021). 3, 4, 6, 10, 11

[30] S. Kong, D. Ramanan, Opengan: Open-set recognition via open data generation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 813–822.

[31] H. Cevikalp, B. Uzun, Y. Salk, H. Saribas, O. Köpüklü, From anomaly detection to open set recognition: Bridging the gap, Pattern Recognition 138 (2023) 109385. 3

[32] J. Lu, Y. Xu, H. Li, Z. Cheng, Y. Niu, Pmal: Open set recognition via robust prototype mining, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, 2022, pp. 1872–1880. 3

[33] D. Hendrycks, K. Gimpel, A baseline for detecting misclassified and out-of-distribution examples in neural networks, in: International Conference on Learning Representations, 2017. 3, 4, 6, 9, 10

[34] S. Liang, Y. Li, R. Srikant, Enhancing the reliability of out-of-distribution image detection in neural networks, in: International Conference on Learning Representations, 2018. 4, 6, 9, 10

[35] W. Liu, X. Wang, J. D. Owens, Y. Li, Energy-based out-of-distribution detection, in: Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS'20, Curran Associates Inc., Red Hook, NY, USA, 2020. 3, 4, 6, 10

[36] Y. Sun, C. Guo, Y. Li, React: Out-of-distribution detection with rectified activations, in: M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, J. W. Vaughan (Eds.), Advances in Neural Information Processing Systems, Vol. 34, Curran Associates, Inc., 2021, pp. 144–157. 3

[37] H. Wang, Z. Li, L. Feng, W. Zhang, Vim: Out-of-distribution with virtual-logit matching, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 4911–4920. 3

[38] Z. Yue, T. Wang, Q. Sun, X.-S. Hua, H. Zhang, Counterfactual zero-shot and open-set visual recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 15404–15414. 3, 12

[39] Y. Fu, X. Wang, H. Dong, Y.-G. Jiang, M. Wang, X. Xue, L. Sigal, Vocabulary-informed zero-shot and open-set learning, IEEE transactions on pattern analysis and machine intelligence 42 (12) (2019) 3136–3152.

[40] C. Geng, L. Tao, S. Chen, Guided cnn for generalized zero-shot and open-set recognition using visual and semantic prototypes, Pattern Recognition 102 (2020) 107263. 3

[41] O. Gune, A. More, B. Banerjee, S. Chaudhuri, Generalized zero-shot learning using open set recognition., in: BMVC, 2019, p. 213. 4

[42] X. Chen, X. Lan, F. Sun, N. Zheng, A boundary based out-of-distribution classifier for generalized zero-shot learning, in: European Conference on Computer Vision, Springer, 2020, pp. 572–588. 4

[43] M. Mancini, M. F. Naeem, Y. Xian, Z. Akata, Open world compositional zero-shot learning, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 5222–5230. 4

[44] N. Liao, Y. Liu, L. Xiaobo, C. Lei, G. Wang, X.-S. Hua, J. Yan, Cohoz: Contrastive multimodal prompt tuning for hierarchical open-set zero-shot recognition, in: Proceedings of the 30th ACM International Conference on Multimedia, MM '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 3262–3271. 4

[45] S. Esmaeilpour, B. Liu, E. Robertson, L. Shu, Zero-shot out-of-distribution detection based on the pre-trained model clip, in: AAAI, 2022. 3, 4

[46] B. Liu, H. Kang, H. Li, G. Hua, N. Vasconcelos, Few-shot open-set recognition using meta-learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 8798–8807. 4

[47] H. Wang, G. Pang, P. Wang, L. Zhang, W. Wei, Y. Zhang, Glocal energy-based learning for few-shot open-set recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7507–7516. 4

[48] H. Wei, R. Xie, H. Cheng, L. Feng, B. An, Y. Li, Mitigating neural network overconfidence with logit normalization, in: K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvári, G. Niu, S. Sabato (Eds.), International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA, Vol. 162 of Proceedings of Machine Learning Research, PMLR, 2022, pp. 23631–23644. 4, 6, 10

[49] D. Hendrycks, S. Basart, M. Mazeika, A. Zou, J. Kwon, M. Mostajabi, J. Steinhardt, D. Song, Scaling out-of-distribution detection for real-world settings, in: K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvári, G. Niu, S. Sabato (Eds.), International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA, Vol. 162 of Proceedings of Machine Learning Research, PMLR, 2022, pp. 8759–8773. 4, 6, 10

[50] S. Reed, Z. Akata, H. Lee, B. Schiele, Learning deep representations of fine-grained visual descriptions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 49–58. 9

[51] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778. 10

[52] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, T. E. Boult, Toward open set recognition, IEEE transactions on pattern analysis and machine intelligence 35 (7) (2012) 1757–1772. 11

[53] X. Sun, Z. Yang, C. Zhang, K.-V. Ling, G. Peng, Conditional gaussian distribution learning for open set recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 13480–13489. 11

[54] P. Perera, V. I. Morariu, R. Jain, V. Manjunatha, C. Wigington, V. Ordonez, V. M. Patel, Generative-discriminative feature representations for open-set recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11814–11823. 11

[55] H. Zhang, M. Cisse, Y. N. Dauphin, D. Lopez-Paz, mixup: Beyond empirical risk minimization, arXiv preprint arXiv:1710.09412 (2017). 12