Class-Specific Semantic Generation and Reconstruction Learning for Open Set Recognition

Haoyang Liu[1,2], **Yaojin Lin**[3,4], **Peipei Li**[1,2], **Jun Hu**[5,*] and **Xuegang Hu**[1,2,6]
[1]Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education
[2]School of Computer Science and Information Engineering, Hefei University of Technology
[3]School of Computer Science, Minnan Normal University
[4]Key Laboratory of Data Science and Intelligence Application, Minnan Normal University
[5]School of Computing, National University of Singapore
[6]Anhui Province Key Laboratory of Industry Safety and Emergency Technology
haoyangliu.cn@gmail.com, zzlinyaojin@163.com, peipeili@hfut.edu.cn, jun.hu@nus.edu.sg, jsjxhuxg@hfut.edu.cn

Abstract

Open set recognition is a crucial research theme for open-environment machine learning. For this problem, a common solution is to learn compact representations of known classes and identify unknown samples by measuring deviations from these known classes. However, the aforementioned methods (1) lack open training consideration, which is detrimental to the fitting of known classes, and (2) recognize unknown classes on an inadequate basis, which limits the accuracy of recognition. In this study, we propose an open reconstruction learning framework that learns a union boundary region of known classes to characterize unknown space. This facilitates the isolation of known space from unknown space to represent known classes compactly and provides a more reliable recognition basis from the perspective of both known and unknown space. Specifically, an adversarial constraint is used to generate class-specific boundary samples. Then, the known classes and approximate unknown space are fitted with manifolds represented by class-specific auto-encoders. Finally, the auto-encoders output the reconstruction error in terms of known and unknown spaces to recognize samples. Extensive experimental results show that the proposed method outperforms existing advanced methods and achieves new state-of-the-art performance. The code is available at https://github.com/Ashowman98/CSGRL.

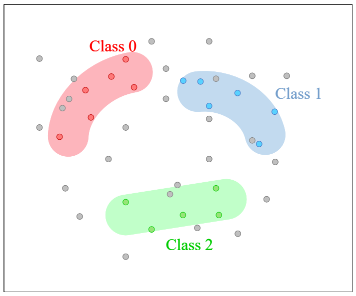
1 Introduction

In recent years, artificial intelligence models have achieved human-like performance in numerous settings, most of which are based on a closed assumption. This assumption implies that the model must be provided with knowledge of all classes during the training phase. However, the real world is full of unknowns, and the testing phase may present unknown classes [Yoshihashi *et al.*, 2019]. This is a key issue that has prevented artificial intelligence models from reaching human-level performance in areas such as autonomous driving and medical diagnostics. This issue has drawn significant attention from researchers and engineers in diverse fields such as data mining [Lv *et al.*, 2022; Wu *et al.*, 2020; Zhu *et al.*, 2023; Liu *et al.*, 2023a], computer vision [Hu *et al.*, 2024; Gao *et al.*, 2023], and multimedia processing [Zhang *et al.*, 2020; Qian *et al.*, 2023; Cai *et al.*, 2023].

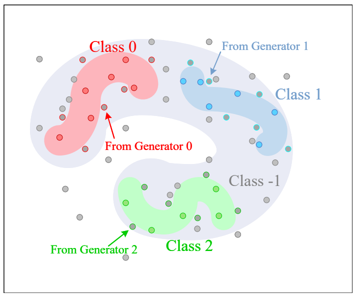
To address classification and recognition tasks with openness, Open Set Recognition (OSR) defines a realistic task objective, which is to accurately classify samples from known classes and reject samples from unknown classes [Geng *et al.*, 2020]. The main challenge of OSR is that the model has not received any information about unknown classes, which makes it impossible to directly divide the feature space of the unknown classes [Scheirer *et al.*, 2013]. Therefore, many existing works have been proposed to learn compact representations of known classes, which can isolate the unknown space and thus reduce the open space risk.

Prototype learning and reconstruction learning show great power in the representation of known classes. Prototype learning represents each class by learning class-specific prototype points based on the Gaussian distribution assumption [Yang *et al.*, 2020], which encourages known samples to approach prototype points, thus excluding unknown samples. Unfortunately, the assumption can easily be violated in practice, which limits the fitting ability of the method. Compared to the prototype representation, reconstruction learning can learn a more compact feature representation based on the Auto-Encoder (AE). These methods reconstruct the input through a compression-recovery process to learn latent representations of the original features, which can also be viewed as learning a low-dimensional manifold to fit the distribution of known samples [Huang *et al.*, 2023]. Therefore, when the input comes from an unknown sample outside the manifold, the AE outputs a higher reconstruction error, which can be

*Corresponding author.



(a) Closed training paradigms fail to characterize the unknown space.



(b) Learn the union boundary region of the known classes to approximately characterize the unknown space.

Figure 1: Comparison of reconstruction learning methods. Method (a) have a limited ability to fit and compact representations, and for unknown samples on the margins of the manifolds, the low reconstruction error of known classes is not sufficient to reject them. Method (b) can obtain a compact representation. In addition, it can identify confusable samples that are still located at the margins of the manifolds based on known and unknown reconstruction errors.

used as a basis for recognition [Oza and Patel, 2019a]. However, representing all known classes with a single continuous manifold is detrimental to the learning of class-specific features and potentially devours inter-class regions, thus introducing the open space risk.

Combining the respective advantages of prototype learning and AE is an effective solution for OSR, i.e., modeling each known class by class-specific manifolds. This method incorporates the class-specificity of prototype learning and the compactness of AE. However, the model is based on a closed paradigm during the training phase, so it is only allowed to fit the known classes while completely ignoring the unknown feature space. As illustrated in Figure 1 (a), this gives rise to two problems: (1) without the characterization of the unknown classes, the manifolds of the known classes may still not be completely compact, and (2) it is risky to rely exclusively on the distinction from the known classes to recognize the unknown classes.

In light of the above observations, we explore (1) a way to characterize the unknown feature space and (2) a more reliable method for distinguishing between known and unknown classes. As any samples from the unknown classes are invisible during the training phase, characterizing the global unknown space is unrealistic. As depicted in Figure 1(b), we propose that an alternative solution is to learn the boundaries of known classes to approximately characterize the unknown space. Once the model learns the union boundary region of the manifolds of the known classes, similar to the reconstruction error of the known classes, it can also be used as a basis for identifying the unknown classes. It is easier to find a threshold to distinguish the known space from the unknown space.

In this paper, we propose a Class-specific Semantic Generation and Reconstruction Learning (CSGRL) method that involves a reconstruction module and a generation module. In CSGRL, in order to characterize the unknown space, the generation module is adversarially constrained to generate samples around the boundary of the manifolds of known classes. Then, the reconstruction module fits the known classes with class-specific AEs and fits the unknown space with an additional AE. The joint optimization of all AEs keeps the unknown regions out of known samples and known manifolds out of unknown space, resulting in more compact representations of known classes. Finally, integrating the reconstruction errors of the AEs representing the known classes and the AE representing the unknown classes, CSGRL makes a decision of classification or rejection. Our contributions are the following:

(1) We propose a training framework with openness, which learns the union boundary region of known classes by an independent AE to approximately characterize the unknown space. This reduces the open space risk and facilitates the fitting of known classes.

(2) We define class-specific generators that can generate slack or tight boundary samples as required, which serve as inputs to augment the reconstruction module.

(3) We recognize the testing samples from two perspectives, the reconstruction error of known classes and unknown classes, which effectively improves the recognition accuracy.

(4) Experimental results under different protocols demonstrate CSGRL outperforms baseline methods and achieves state-of-the-art performance on several public datasets.

2 Related Work

We propose CSGRL to address the OSR problem, which is naturally related to Out-Of-Distribution (OOD) detection, and we will discuss their existing work in this section.

2.1 Open Set Recognition

Initially, solutions for OSR were based on traditional machine learning methods [Scheirer *et al.*, 2014; Mendes Júnior *et al.*, 2017; Zhu *et al.*, 2018]. However, the performance of these methods was limited by the technology available at the time. With the development of deep neural networks, which have achieved excellent performance in a closed setting, more attention has been paid to their application in open scenarios. Deploying deep neural networks to solve OSR problems, prototype-like and reconstruction learning methods are the two feasible research paths.

Prototype-like learning methods. [Yang *et al.*, 2020] presented the first convolutional prototype network for the open-world, which aimed to represent each specific class abstractly. As a variant of the prototype network, reciprocal point learning (RPL) [Chen *et al.*, 2020] utilizes the differences between samples and the reciprocal points rather than their commonalities for classification and recognition. In order to enhance the recognition of unknown classes, RPL further integrated the adversarial training strategy, named ARPL [Chen *et al.*, 2021]. [Liu *et al.*, 2023b] derived a multiple Gaussian prototype framework from Bayesian inference, thus preserving some space for unknown classes. Constrained by Gaussian distribution assumptions, these methods fail to represent real-world complexities.

Reconstruction learning methods. [Yoshihashi *et al.*, 2019] proposed the first Classification Reconstruction learning for Open Set Recognition (CROSR) based on deep representation, which inspired subsequent such methods. C2AE [Oza and Patel, 2019a] was constructed as a two-stage method, where a closed-set classifier is first trained, and then an open set recognition model was trained by reconstructing conditional on class identity. Based on the unknown sample detection capability of the variational AE, [Sun *et al.*, 2020] proposed a conditional Gaussian distribution learning for open set recognition (CGDL), which classifies known samples by forcing latent features to approximate different Gaussian models. To reduce the class activation value of unknown samples and maximize their difference from known samples, [Perera *et al.*, 2020] extended the optimal closed set features and learned both the original and reconstructed images. Further, [Huang *et al.*, 2023] proposed the CSSR model that replaces prototype points with manifolds represented by class-specific AEs, which enhances the representation of the individual classes and releases the inter-class space. However, the above work neglects the characterization of unknown classes, which may impair the fitting ability of the model. In addition, reconstructions available only for known classes are not conducive to finding distinguishable thresholds.

2.2 Out-of-Distribution Detection

A subtask of OSR is to reject unknown classes, which shares the same target as some problems. For reasons of AI system safety, [Hendrycks and Gimpel, 2017] advocated reliable methods for detecting samples in the test set that come from outside the distribution of the training set. Without changing the pre-trained neural network, [Liang *et al.*, 2018] used temperature scaling and added small perturbations to improve the separability of in- and OOD images. [Lee *et al.*, 2018] minimized the Kullback-Leibler divergence from the predictive distribution of out-of-distribution samples to the uniform distribution. To this end, they introduced an OOD dataset for training and augmented it with boundary samples generated from low-density regions of known classes. Outlier exposure [Hendrycks *et al.*, 2019] generalizes to unseen anomalies by feeding the anomaly detectors with auxiliary data. Similarly, outlier detection [Zhou and Paffenroth, 2017] and novelty detection [Perera *et al.*, 2019] have been extensively studied to improve the robustness of models. Most of these works focus on rejecting OOD (unknown) samples without considering the ability of the model to classify in-distribution (known) samples. In addition, some methods introduce OOD data, which violates the OSR protocol. Obviously, an OSR setup without any unknown data is more general and realistic in practice.

3 Preliminaries

In this section, we define the research problem and briefly describe the relevant research components.

Problem definition. Given a training set of n labeled samples $\mathcal{X} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n)\}$, where $y_i \in \{1, 2, ..., k\}$ is the label of $\mathbf{x}_i$. In the testing phase, any one sample $\mathbf{x}_t$ corresponds to the label $y_t \in \{1, 2, ..., k, k+1\}$, where $k+1$ is a uniform label of all unknown classes. The goal of OSR is to learn a model f from $\mathcal{X}$ that establishes a one-to-one mapping of $\mathbf{x}_t$ to y_t.

A good OSR model should be able to correctly classify known classes, i.e., minimize the empirical classification risk $\mathcal{R}_{\mathcal{E}}$ of labeled known data. In addition, unlike traditional classification models, OSR models also need to have the ability to reject unknown classes. As defined in [Scheirer *et al.*, 2013], open space risk measures the uncertainty that f recognizes sample $\mathbf{x}$ as known or unknown classes, which is calculated as a nonzero integral function

$$\mathcal{R}_{\mathcal{O}}(g(f)) = \frac{\int_{\mathcal{O}} g(f(\mathbf{x}))}{\int_{\mathcal{X} \bigcup \mathcal{O}} g(f(\mathbf{x}))} \tag{1}$$

where $\mathcal{O}$ is the open (unknown) space and g is a binary function whose output is

$$g(f(\mathbf{x})) = \begin{cases} 1 & \text{if } f(x) \in [1, k] \\ 0 & \text{otherwise} \end{cases} . \tag{2}$$

Therefore, the OSR problem can be transformed into an optimization objective as follows:

$$\arg\min_{f} \{\mathcal{R}_{\mathcal{O}}(g(f(\mathbf{x}))) + \lambda \mathcal{R}_{\mathcal{E}}(f(\mathbf{x}))\}, \tag{3}$$

where λ is a positive regularization parameter.

Class-specific semantic reconstruction learning. Class-specific Semantic Reconstruction (CSSR) learning utilizes a set of AEs to learn class-specific low-dimensional manifolds to fit the distribution of the data, where each AE is a bottleneck structure consisting of an encoder $\mathcal{E}$ and decoder $\mathcal{D}$. The encoder $\mathcal{E}$ learns a mapping of the original data from the low-dimensional embedding space, and then the decoder $\mathcal{D}$ recovers the low-dimensional vectors to the original feature space. In this process, the loss of compression and the difference in recovery need to be reduced as much as possible, i.e., minimizing the reconstruction error. CSSR considers the reconstruction error as the distance between the sample and a low-dimensional manifold, and it assigns the test sample to the nearest manifold. For OSR, samples far from all manifolds are regarded as unknown samples. These concepts can be formalized as follows:

$$\begin{aligned} P(y = i|\mathbf{x}, \mathcal{A}, 1 \leq i \leq k) &\propto (-\left\|\mathbf{f} - \mathcal{D}_i(\mathcal{E}_i(\mathbf{f}))\right\|_1), \\ P(y = k+1|\mathbf{x}, \mathcal{A}) &\propto \min_{1 \leq i \leq k} \left\|\mathbf{f} - \mathcal{D}_i(\mathcal{E}_i(\mathbf{f}))\right\|_1, \end{aligned} \tag{4}$$

where $\mathbf{f} = \mathcal{B}(\mathbf{x})$ is the embedding feature of input $\mathbf{x}$ extracted by the backbone network $\mathcal{B}$. AE corresponding to class c is denoted as $\mathcal{A}_c$. During training, this method ignores modeling the unknown class and only maximizes the probability of the ground-truth class.

4 The Proposed Method

In order to address the reported problems, we propose a Class-specific Semantic Generation and Reconstruction Learning (CSGRL) method. Figure 2 illustrates the overall architecture of the proposed method. The key theme is to characterize the unknown space. We first present the overall architecture of the fitting class. Then, a method for characterizing unknown space is introduced. Finally, we describe the classification of known classes and the recognition of unknown classes.

4.1 Learning Class-Specific Manifolds

To make decisions about sample $(\mathbf{x}, c) \in \mathcal{X}$, we define a reconstruction module R. The module consists of a set of AEs $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_{k+1}\}$. Different from the existing methods, an additional AE $\mathcal{A}_{k+1}$ is used to represent the unknown classes. Sample $\mathbf{x}$ is output by the backbone network as $\mathbf{f}$, the reconstruction error specific to class i can be obtained by $\mathcal{A}$ as $d(\mathbf{f}, \mathcal{A}_i) = \|\mathbf{f} - \mathcal{A}_i(f)\|_1$. Similar to Eq. (4), the classification probability of sample $\mathbf{x}$ is related to the distance of the manifolds. Applying the SoftMax activation function, the reconstruction error can be normalized to probability as follows:

$$P(R(\mathbf{f}) = i|\mathbf{x}, \mathcal{A}) = \frac{\exp(-\gamma d(\mathbf{f}, \mathcal{A}_i))}{\sum_1^{k+1} \exp(-\gamma d(\mathbf{f}, \mathcal{A}_j))}, \tag{5}$$

where γ is a hyperparameter used to control the hardness of the error-probability transformation, and it is set to 0.1.

We hope that the probability that sample $\mathbf{x}$ belongs to the ground truth class c should be infinitely close to 1, while the probability that it belongs to the other classes is approximately equal to 0. Therefore, the model optimization is guided by a loss function as follows:

$$\mathcal{L}_{kn} = -\sum_{c=1}^{m} \log P(R(\mathbf{f}) = c|\mathbf{x}, \mathcal{A}). \tag{6}$$

To minimize the loss, the reconstruction module should guarantee a minimum distance mapping from the features to the manifolds of the true class, based on which the distance to other manifolds is maximized.

4.2 Characterizing Unknown Space

Due to there being no unknown samples available for training, the model can only fit known classes. The latent unknown sample space can be infinite and thus impossible to cover. We propose to use $\mathcal{A}_{k+1}$ to fit the union boundary region. This intuitively approximately characterizes the unknown space, because most unknown samples are always closer to their boundary region than the manifolds of known classes. For this purpose, we need to build a generation module $G = \{G_1, G_2, ..., G_k\}$ to generate boundary samples for each known class. As an effective generative model,

Algorithm 1 The training procedure of CSGRL

Input:
$\mathcal{X}$: Training data; ε: Hyperparameter.

Output:
$\mathcal{B}$: Backbone network; R: Reconstruction module;
G: Generation module.

```
Initialize B, R and G.
/** Step-I: Update proposed R **/
repeat
    Update B, R by descending their stochastic gradient
    according to Eq. (6)
until convergence;
/** Step-II: Alternate update proposed G and R **/
repeat
    Update G by descending its stochastic gradient according to Eq. (9)
    Update B, R by descending their stochastic gradient
    according to Eq. (10)
until convergence;
return B, R and G;
```

Generative Adversarial Networks (GAN) [Goodfellow *et al.*, 2014] consists of a generator $G^{'}$ and a binary discriminator $D^{'}$. The generator synthesizes pseudo-samples to augment the discriminator, while the discriminator guides the generator to obtain confusable samples. GAN performs a minimax optimization over $D^{'}$ and $G^{'}$ as

$$\min_{G^{'}} \max_{D^{'}} [\log D^{'}(\mathbf{x}) + \log D^{'}(G^{'}(\tilde{\mathbf{x}}))], \tag{7}$$

where the input of $G^{'}$ is a Gaussian noise $\tilde{\mathbf{x}} \sim \mathrm{N}(0, 1)$.

However, we need boundary samples to approximate the unknown space, instead of samples within manifolds of known classes. Obviously, the boundary samples should not be too far away from the known samples, but they should be located near the extreme points. The current maximum point for class c is collected in each mini-batch $\{(\mathbf{x}_1, c), (\mathbf{x}_2, c), ..., (\mathbf{x}_m, c)\}$ as

$$d_c = \max(d_c, \max_{1 \le i \le m} d(\mathbf{f}_i, \mathcal{A}_c)), \tag{8}$$

where d_c is set to 0 at the beginning of each epoch. Generating high-dimensional realistic images is challenging [Pidhorskyi *et al.*, 2018; Zenati *et al.*, 2018], and existing work [Kong and Ramanan, 2021] has demonstrated that generating samples at the feature-level works better than at the pixel-level. Therefore, G directly generates feature-level samples, and we impose constraints on G to obtain class-specific boundary samples

$$\begin{aligned}\mathcal{L}_G = -\sum_{c=1}^{m} [&\log P(R(\tilde{\mathbf{f}}) = c|\mathbf{x}, \mathcal{A}) \\ &+ \log \mathrm{Sigmoid}(d(G_c(\tilde{\mathbf{x}}), \mathcal{A}_c) - d_c - \varepsilon)],\end{aligned} \tag{9}$$

where ε is a slackness, the larger its value, the looser the resulting boundary. Conversely, the tighter the resulting boundary. To minimize the second term of $\mathcal{L}_G$, the generator G_c is

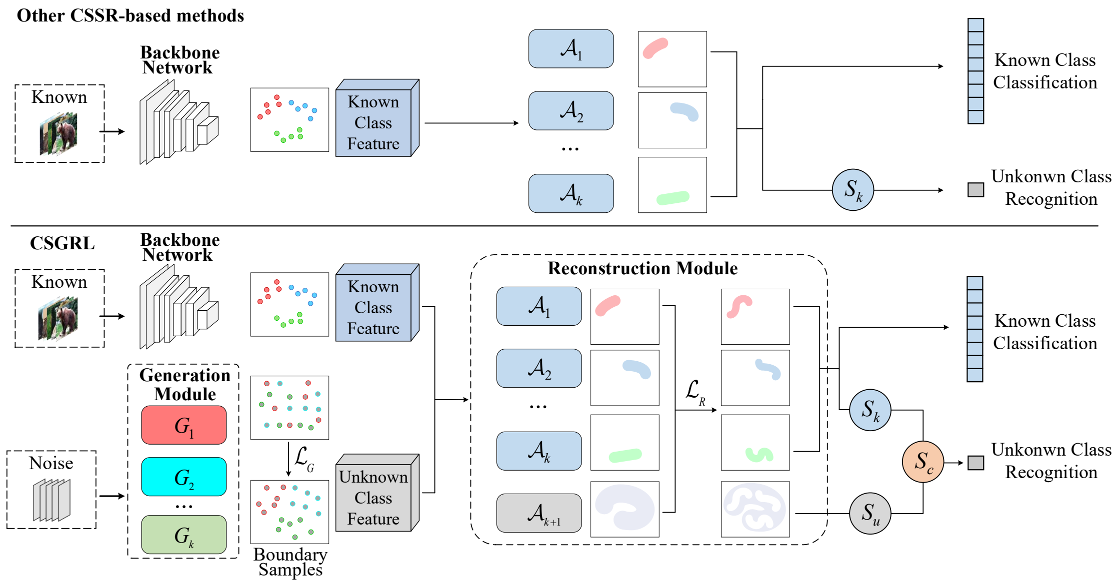


Figure 2: The architecture of CSGRL compared to previous CSSR-based methods. Our network architecture consists of a backbone network, a reconstruction module, and a generation module. The backbone network is used to extract features $\mathbf{f}$ from the input $\mathbf{x}$. The reconstruction module reconstructs the features $\mathbf{f}$ (Section 4.1). The generation module generates boundary samples from Gaussian noise to augment the reconstruction module (Section 4.2). Based on the reconstruction errors, the model rejects sample $\mathbf{x}$, or outputs the known class with the highest probability as the class of sample $\mathbf{x}$ (Section 4.3).

forced to generate samples with reconstruction errors higher than $d_c - \varepsilon$. However, if these samples are far from $\mathcal{A}_c$, the first term of $\mathcal{L}_G$ should be high. Thus, one can expect that $\mathcal{L}_G$ can encourage G to generate valid boundary samples for each class.

Finally, input real and generated samples to update the reconstruction module, and Eq. (6) is expanded as

$$\mathcal{L}_R = -\sum_{c=1}^{m}[\log P(R(\mathbf{f}) = c|\mathbf{x}, \mathcal{A}) + \log P(R(G_c(\tilde{\mathbf{x}})) = m+1|\mathbf{x}, \mathcal{A})]. \tag{10}$$

To minimize the loss, $\mathcal{L}_R$ encourages the separation of known and unknown space.

Due to $\mathcal{L}_G$ involves the computation of the maximum value of the reconstruction errors for known classes, R is preferentially trained to obtain a stable representation of the known classes. Then, G and D are alternately optimized. The complete training procedure for the proposed method is summarized in Algorithm 1.

4.3 Recognition and Classification

In order to obtain a reliable score to reject unknown classes, CSGRL integrates the reconstruction errors relative to known and unknown classes. OSR methods typically quantify the difference (or similarity) between a test sample and the known samples as a score S, and classify the sample into an unknown class if S exceeds (or is below) than a threshold. Similarly, referring to Eq. 4, we can utilize the reconstruction errors of the known classes to recognize the unknown samples. Known classes should have low relative reconstruction errors, which can be formulated as

$$S_k = -\min_{1\le i\le k} d(\mathbf{f}, \mathcal{A}_i). \tag{11}$$

For unknown classes that are vastly different from all known classes, e.g., far OOD datasets [Chen *et al.*, 2021], their samples can lead to significant reconstruction errors. Therefore, they can be easily recognized based on S_k. However, there are also unknown samples in the unknown space with small differences from the known classes, e.g., near OOD datasets [Chen *et al.*, 2021]. These data may be located at the margins of the manifolds of known classes, resulting in lower reconstruction errors. In CSGRL, because the $\mathcal{A}_{k+1}$ learns the union boundary region of known classes, a natural idea is to make inferences with the help of reconstruction errors from $\mathcal{A}_{k+1}$. This is given by

$$S_u = d(\mathbf{f}, \mathcal{A}_{k+1}). \tag{12}$$

For a sample located at the margin of known manifolds, although it has a low score $|S_k|$, it also has a low score $|S_u|$. Therefore, it is probably from an unknown class. A normalization is applied to S_k and S_u to unify the scales:

$$\tilde{S}_k = \frac{S_k - E(s_k)}{Std(S_k)}, \quad \tilde{S}_u = \frac{S_u - E(s_u)}{Std(S_u)}. \tag{13}$$

where $E(*)$ and $Std(*)$ denote the mean and standard deviation of the scores, respectively. Finally, they are integrated into a final score:

$$S_c = \alpha\, \tilde{S}_k + (1-\alpha)\tilde{S}_u, \tag{14}$$

where α is a hyperparameter that controls the weight of the two reconstruction errors on the recognition result. The transformation from "Is it close to known classes?" to "Is it closer to known classes or to their boundaries?" may be more favorable for selecting a threshold in practical scenarios.

A test sample is recognized as an unknown class when S_c is below a threshold δ, otherwise it is assigned to the known class with the highest probability. The process can be formulated as

$$y = \begin{cases} \underset{1\leq i\leq k}{\arg\min}\, d(\mathbf{f}, \mathcal{A}_i) & \text{if } S_c \geq \delta \\ k+1 & \text{otherwise} \end{cases}. \tag{15}$$

5 Experiment

5.1 Implementation Details

The proposed method CGSRL does not involve feature extraction, so it can be plugged into any backbone network. To be fair, we use the same backbone network and data augmentation technique with [Huang *et al.*, 2023]. In the training phase, the loss is minimized using the stochastic gradient descent optimizer with momentum = 0.9. We train the network for 250 epochs with a mini-batch size of 128, including 200 epochs for Step-I and 50 epochs for Step-II in Algorithm 1. The learning rate was initially set to 0.4

5.2 Experiments for Unknown Detection

The evaluation protocol defined in [Neal *et al.*, 2018] is widely used for unknown detection.

Datasets. We construct experiments on three image datasets, including Cifar10 [Krizhevsky, 2009], SVHN [Netzer *et al.*, 2011], and Tinyimagenet [Le and Yang, 2015]. For Cifar10 and SVHN, 6 classes are randomly sampled as known classes and the remaining 4 classes are set as unknown classes. For Tinyimagenet with 200 classes, the ratio of known to unknown classes is 20:180. To avoid chance, the splits of known and unknown classes are randomized five times, and then their average results are reported. In this experiment, we directly use the same data split with [Huang *et al.*, 2023].

Evaluation metrics. For most OSR methods, selecting the appropriate threshold is crucial. Therefore, a threshold-independent metric, the Area Under the Receiver Operating Characteristic (AUROC) curve is often used for evaluation. It reflects the true positive rate against the false positive rate at different thresholds. In our research problem, AUROC can be interpreted as the probability that a known sample is assigned a higher recognition score than an unknown sample.

Baselines. CSGRL was compared to advanced related methods, including prototype-like learning methods [Yang *et al.*, 2020], [Chen *et al.*, 2021], [Liu *et al.*, 2023b] and reconstruction learning methods [Yoshihashi *et al.*, 2019], [Oza and Patel, 2019a], [Oza and Patel, 2019b], [Sun *et al.*, 2020], [Perera *et al.*, 2020], [Huang *et al.*, 2023]. Among these methods, [Chen *et al.*, 2021] uses GAN for OSR.

Method	Cifar10	SVHN	Tinyimagenet
GCPL [Yang *et al.*, 2020]	82.8	92.6	-
ARPL [Chen *et al.*, 2021]	91.0	96.7	78.2
MGPL [Liu *et al.*, 2023b]	86.9	94.1	64.3
CROSR [Yoshihashi *et al.*, 2019]	88.3	89.9	58.9
C2AE [Oza and Patel, 2019a]	89.5	92.2	74.8
MLOSR [Oza and Patel, 2019b]	84.5	95.5	71.8
CGDL [Sun *et al.*, 2020]	90.3	93.5	76.2
GFROSR [Perera *et al.*, 2020]	83.1	93.5	64.7
CSSR [Huang *et al.*, 2023]	91.3	97.9	**82.3**
RCSSR [Huang *et al.*, 2023]	91.5	97.8	81.9
CSGRL (Ours)	**94.0**	**98.3**	**82.3**

Table 1: Performance of comparison methods under various validation setups.

Method	IN-C	IN-R	LS-C	LS-R
PROSER[Zhou *et al.*, 2021]	84.9	82.4	86.7	85.6
ODL [Liu *et al.*, 2022]	86.1	84.2	87.1	85.6
ConOSR [Xu *et al.*, 2023]	89.1	84.3	91.2	88.1
MGPL [Liu *et al.*, 2023b]	86.2	86.2	86.9	86.8
CROSR [Yoshihashi *et al.*, 2019]	72.1	73.5	72.0	74.9
C2AE [Oza and Patel, 2019a]	83.7	82.6	78.3	80.1
CGDL [Sun *et al.*, 2020]	84.0	83.2	80.6	81.2
GFROSR [Perera *et al.*, 2020]	75.7	79.2	75.1	80.5
CSSR [Huang *et al.*, 2023]	92.9	90.9	94.1	93.5
RCSSR [Huang *et al.*, 2023]	93.3	91.5	94.0	94.0
MEDAF [Wang *et al.*, 2024]	91.5	90.0	92.2	92.6
CSGRL (Ours)	**94.0**	**93.3**	**94.8**	**95.1**

Table 2: Performance of comparison methods under various OSR setups.

Result comparison. Table 1 shows the AUROC results for the thirteen methods on the three benchmark datasets. In all the tables presenting the experimental results, the results are presented as percentages and the best results are in bold. The results show that CSGRL outperforms other methods on all three datasets. This is attributed to the reasonable characterization of the unknown space, thus enabling the model to separate the known and unknown samples more efficiently.

5.3 Experiments for Open Set Recognition

An effective OSR method not only rejects unknown samples, but also categorizes known samples. [Yoshihashi *et al.*, 2019] provided a setup for OSR experiments, and we apply it in this section.

Experimental setup. Cifar10 is used as the training set, and samples from other datasets are collected into the test set. We collect 10,000 unknown samples from ImageNet and LSUN respectively, which is the same number of samples as the Cifar10 test set. Then, we crop or resize them to ensure that their image sizes are consistent with the known samples. In this experiment, we use the version of the four datasets released by Liang *et al.* [Liang *et al.*, 2018], i.e., ImageNet-Crop (IN-C), ImageNet-Resize (IN-R), LSUN-Crop (LS-C), and LSUN-Resize (LS-R). Since AUROC can only evaluate the recognition of unknown classes and not the classification performance of the method on known classes, we use the macro-averaged F1-scores to evaluate the overall performance of the method on 10 known classes and 1 unknown class. Furthermore, except for prototype-like and reconstruc-

Method	In:CIFAR10/Out:CIFAR100				In:CIFAR10/Out:SVHN			
	DTACC	AUROC	AUIN	AUOUT	DTACC	AUROC	AUIN	AUOUT
ODL [Liu *et al.*, 2022]	-	-	-	-	90.2	93.5	91.7	95.3
GCPL [Yang *et al.*, 2020]	80.2	86.4	86.6	84.1	86.1	91.3	86.6	94.8
ARPL [Chen *et al.*, 2021]	83.4	90.3	91.1	88.4	91.6	96.6	94.8	98.0
CSI [Tack *et al.*, 2020]	-	92.2	-	-	-	97.9	-	-
OpenGAN [Kong and Ramanan, 2021]	84.2	89.7	87.7	89.6	92.1	95.9	93.4	97.1
CSSR [Huang *et al.*, 2023]	83.8	92.1	89.4	89.3	95.7	99.1	98.2	99.6
RCSSR [Huang *et al.*, 2023]	85.3	92.3	92.9	91.0	95.7	99.1	98.3	99.6
MEDAF [Wang *et al.*, 2024]	85.4	92.5	93.2	91.1	95.3	99.1	98.0	99.6
CSGRL (Ours)	**87.7**	**94.1**	**94.5**	**93.2**	**98.0**	**99.8**	**99.5**	**99.9**

Table 3: Performance of comparison methods under different OOD setups.

tion learning methods, other types of OSR methods [Zhou *et al.*, 2021], [Liu *et al.*, 2022], [Xu *et al.*, 2023], [Wang *et al.*, 2024] are introduced for comparison.

Results comparison. As shown in Table 2, CSGRL outperforms all methods on four datasets. This indicates that the proposed method is generally more in line with the goal of OSR, i.e., CSGRL possesses higher classification accuracy for known samples and stronger capability for unknown detection. In addition, CSSR and its reciprocal version RCSSR have mutual advantages in different setups. In order to reject unknown classes, instead of considering whether to represent the known classes by prototype or reciprocal points, CSGRL characterizes both known and unknown classes to separate them directly.

5.4 Experiments for Out-of-Distribution Detection

OOD detection aims to detect samples that come from out-distribution, which may be similar to, or very different from, in-distribution samples. The setup of the OOD experiment in [Chen *et al.*, 2021] is adopted in this section.

Experimental setup. We conduct two sets of experiments where the models are trained on Cifar10 as the in-distribution dataset. Then, they are tested on Cifar100 or SVHN as the near OOD dataset and far OOD dataset, respectively. In addition to the previously introduced methods, we compare CSI [Tack *et al.*, 2020] and OpenGAN [Kong and Ramanan, 2021], which enhance the representation of known classes from different perspectives. Following [Chen *et al.*, 2021], we use DTACC, AUROC, AUIN, and AUOUT as evaluation metrics. DTACC calculates the maximum value of classification accuracy at different thresholds. AUPR measures the area of graph plotting precision against recall at different thresholds. When in-distribution (or out-distribution) samples are specified as positive, AUPR is further represented by AUIN (or AUOUT).

Results comparison. Based on the results in Table 3, we can summarize the following conclusions:

(1) Under different OOD settings, CSGRL outperformed OpenGAN on four metrics, which indicates that the conventional GAN has limited improvement in unknown detection.

(2) It is easier to recognize far OOD data than near OOD data, so its performance is close to saturation on most methods. Despite this, CSGRL improves performance over other methods.

(3) For near OOD data, CSGRL is ranked first. This indicates that CSGRL is better able to distinguish unknown samples that are similar to known samples.

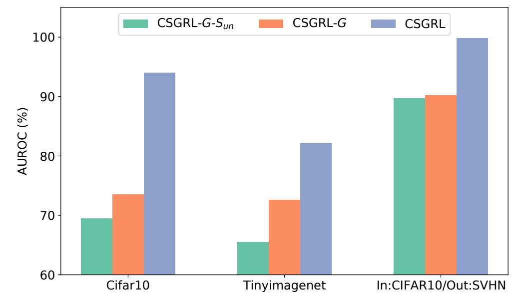


Figure 3: Ablation study of various components.

5.5 Ablation Study

In this section, we construct ablation experiments to investigate the effectiveness of two key designs for CSGRL. As in the dataset setup in Section 5.2 and Section 5.4, the experiment includes three datasets commonly used in OSR: Cifar10, SVHN, and Tinyimagenet.

Ablation terms. (1) Generation Module G: the model learns the union boundary region to approximately characterize the unknown space. (2) Reconstruction Score of Unknown Classes S_{un}: the model introduces the reconstruction error of the unknown classes as the recognition score.

Result comparison. Figure 3 shows the AUROC results of the ablation experiments. From these results, we can obtain the following three observations:

(1) Learning a union boundary region benefits more compact representations of known classes.

(2) The reconstruction score of unknown classes can effectively improve the recognition accuracy of unknown classes.

(3) As more CSGRL components are integrated into the model, the performance of the model gradually improves. This indicates the necessity of the components for OSR.

6 Conclusion

In this paper, we proposed CSGRL, a class-specific semantic generation and reconstruction learning. In contrast to previous reconstruction learning for OSR, this method follows an open paradigm in the training phase, where it learns a union boundary region to approximately characterize the unknown space. In addition, the reconstruction error of the unknown classes is introduced as a decision score, which further separates the confusing unknown samples from the known space. Experimental results under multiple protocols demonstrated that this method outperforms other state-of-the-art methods.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants No. 62376085, 62076085, 62120106008, 62076116 and 62106262, and the Natural Science Foundation of Fujian Province through Grant No. 2021J02049.

References

[Cai *et al.*, 2023] Desheng Cai, Shengsheng Qian, Quan Fang, Jun Hu, and Changsheng Xu. User cold-start recommendation via inductive heterogeneous graph neural network. *ACM Transactions on Information Systems*, 41(3):1–27, 2023.

[Chen *et al.*, 2020] 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*, pages 507–522. Springer, 2020.

[Chen *et al.*, 2021] 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*, 44(11):8065–8081, 2021.

[Gao *et al.*, 2023] Junyu Gao, Mengyuan Chen, and Changsheng Xu. Vectorized evidential learning for weakly-supervised temporal action localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12):15949 – 15963, 2023.

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

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[Hendrycks and Gimpel, 2017] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

[Hendrycks *et al.*, 2019] Dan Hendrycks, Mantas Mazeika, and Thomas G. Dietterich. Deep anomaly detection with outlier exposure. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019.

[Hu *et al.*, 2024] Yufan Hu, Junyu Gao, Jianfeng Dong, Bin Fan, and Hongmin Liu. Exploring rich semantics for open-set action recognition. *IEEE Transactions on Multimedia*, 26:5410 – 5421, 2024.

[Huang *et al.*, 2023] Hongzhi Huang, Yu Wang, Qinghua Hu, and Ming-Ming Cheng. Class-specific semantic reconstruction for open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):4214–4228, 2023.

[Kong and Ramanan, 2021] Shu Kong and Deva Ramanan. Opengan: Open-set recognition via open data generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 813–822, 2021.

[Krizhevsky, 2009] Alex Krizhevsky. Learning multiple layers of features from tiny images. *Tech Report*, 2009.

[Le and Yang, 2015] Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015.

[Lee *et al.*, 2018] Kimin Lee, Honglak Lee, Kibok Lee, and Jinwoo Shin. Training confidence-calibrated classifiers for detecting out-of-distribution samples. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

[Liang *et al.*, 2018] Shiyu Liang, Yixuan Li, and R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

[Liu *et al.*, 2022] Zhun-ga Liu, Yi-min Fu, Quan Pan, and Zuo-wei Zhang. Orientational distribution learning with hierarchical spatial attention for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[Liu *et al.*, 2023a] Haoyang Liu, Yaojin Lin, Chenxi Wang, Lei Guo, and Jinkun Chen. Semantic-gap-oriented feature selection in hierarchical classification learning. *Information Sciences*, 642:119241, 2023.

[Liu *et al.*, 2023b] Jiaming Liu, Jun Tian, Wei Han, Zhili Qin, Yulu Fan, and Junming Shao. Learning multiple gaussian prototypes for open-set recognition. *Information Sciences*, 626:738–753, 2023.

[Lv *et al.*, 2022] Junwei Lv, Ying He, Xuegang Hu, Desheng Cai, Yuqi Chu, and Jun Hu. Dual confidence learning network for open-world time series classification. In *International Conference on Database Systems for Advanced Applications*, pages 574–589. Springer, 2022.

[Mendes Júnior *et al.*, 2017] Pedro R Mendes Júnior, Roberto M De Souza, Rafael de O Werneck, Bernardo V Stein, Daniel V Pazinato, Waldir R de Almeida, Otávio AB Penatti, Ricardo da S Torres, and Anderson Rocha. Nearest neighbors distance ratio open-set classifier. *Machine Learning*, 106(3):359–386, 2017.

[Neal *et al.*, 2018] Lawrence Neal, Matthew Olson, Xiaoli Fern, Weng-Keen Wong, and Fuxin Li. Open set learning with counterfactual images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 613–628, 2018.

[Netzer *et al.*, 2011] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. In *Neural Information Processing Systems*, pages 1–9, 2011.

[Oza and Patel, 2019a] Poojan Oza and Vishal M Patel. C2ae: Class conditioned auto-encoder for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2307–2316, 2019.

[Oza and Patel, 2019b] Poojan Oza and Vishal M Patel. Deep cnn-based multi-task learning for open-set recognition. *arXiv preprint arXiv:1903.03161*, 2019.

[Perera *et al.*, 2019] Pramuditha Perera, Ramesh Nallapati, and Bing Xiang. Ocgan: One-class novelty detection using gans with constrained latent representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2898–2906, 2019.

[Perera *et al.*, 2020] Pramuditha Perera, Vlad I Morariu, Rajiv Jain, Varun Manjunatha, Curtis Wigington, Vicente Ordonez, and Vishal M Patel. Generative-discriminative feature representations for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11814–11823, 2020.

[Pidhorskyi *et al.*, 2018] Stanislav Pidhorskyi, Ranya Almohsen, and Gianfranco Doretto. Generative probabilistic novelty detection with adversarial autoencoders. *Advances in neural information processing systems*, 31, 2018.

[Qian *et al.*, 2023] Shengsheng Qian, Hong Chen, Dizhan Xue, Quan Fang, and Changsheng Xu. Open-world social event classification. In *Proceedings of the ACM Web Conference 2023*, pages 1562–1571, 2023.

[Scheirer *et al.*, 2013] Walter J. Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E. Boult. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1757–1772, 2013.

[Scheirer *et al.*, 2014] Walter J Scheirer, Lalit P Jain, and Terrance E Boult. Probability models for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2317–2324, 2014.

[Sun *et al.*, 2020] Xin Sun, Zhenning Yang, Chi Zhang, Keck-Voon Ling, and Guohao Peng. Conditional gaussian distribution learning for open set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13480–13489, 2020.

[Tack *et al.*, 2020] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. *Advances in neural information processing systems*, 33:11839–11852, 2020.

[Wang *et al.*, 2024] Yu Wang, Junxian Mu, Pengfei Zhu, and Qinghua Hu. Exploring diverse representations for open set recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

[Wu *et al.*, 2020] Man Wu, Shirui Pan, and Xingquan Zhu. Openwgl: Open-world graph learning. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 681–690. IEEE, 2020.

[Xu *et al.*, 2023] Baile Xu, Furao Shen, and Jian Zhao. Contrastive open set recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 10546–10556, 2023.

[Yang *et al.*, 2020] Hong-Ming Yang, Xu-Yao Zhang, Fei Yin, Qing Yang, and Cheng-Lin Liu. Convolutional prototype network for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2358–2370, 2020.

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

[Zenati *et al.*, 2018] Houssam Zenati, Manon Romain, Chuan-Sheng Foo, Bruno Lecouat, and Vijay Chandrasekhar. Adversarially learned anomaly detection. In *2018 IEEE International conference on data mining (ICDM)*, pages 727–736. IEEE, 2018.

[Zhang *et al.*, 2020] Huaiwen Zhang, Shengsheng Qian, Quan Fang, and Changsheng Xu. Multimodal disentangled domain adaption for social media event rumor detection. *IEEE Transactions on Multimedia*, 23:4441–4454, 2020.

[Zhou and Paffenroth, 2017] Chong Zhou and Randy C Paffenroth. Anomaly detection with robust deep autoencoders. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 665–674, 2017.

[Zhou *et al.*, 2021] Da-Wei Zhou, Han-Jia Ye, and De-Chuan Zhan. Learning placeholders for open-set recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2021.

[Zhu *et al.*, 2018] Yue Zhu, Kai Ming Ting, and Zhi-Hua Zhou. Multi-label learning with emerging new labels. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1901–1914, 2018.

[Zhu *et al.*, 2023] Yifan Zhu, Fangpeng Cong, Dan Zhang, Wenwen Gong, Qika Lin, Wenzheng Feng, Yuxiao Dong, and Jie Tang. Wingnn: dynamic graph neural networks with random gradient aggregation window. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3650–3662, 2023.