Computer Vision and Image Understanding 229 (2023) 103651

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Spatial location constraint prototype loss for open set recognition

Ziheng Xia, Penghui Wang *, Ganggang Dong, Hongwei Liu *

National Laboratory of Radar Signal Processing, Xidian University, No. 2 South Taibai Road, Xi'an, Shaanxi 710071, PR China

ARTICLE INFO

Communicated by Nikos Paragios

Keywords:
Open set recognition
Spatial location constrain prototype loss
Matching theory
Open space risk
Empirical risk

ABSTRACT

One of the challenges in pattern recognition is open set recognition. Compared with closed set recognition, open set recognition needs to reduce not only the empirical risk, but also the open space risk, and how to reduce the open space risk is the key of open set recognition. Compared with the previous work, this paper analyzes the distribution rules of the known and unknown features, which is highly related to the open space risk. Then, this paper proposes the matching theory to explain the origin of the open space risk. On this basis, the spatial location constraint prototype loss function is proposed to reduce both risks simultaneously. Extensive experiments on multiple benchmark datasets and many visualization results verify the validity of the proposed matching theory and the effectiveness of the proposed method.

1. Introduction

With the development of deep learning technology, pattern recognition based on deep neural network has been greatly developed, such as image recognition (Zhang et al., 2020) and speech recognition (Granell et al., 2019). However, there are still many challenges and difficulties that need to be addressed in real-world applications, and one of them is that the test set may contain some classes that are not present in the training phase. Unfortunately, the traditional closed set recognition (CSR) is incapable of effectively dealing with this situation, whose test set and training set contain the same data categories. Therefore, open set recognition (OSR) was proposed to solve this kind of problem that the model needs to not only correctly classify the known classes, but also identify the unknown classes (Scheirer et al., 2013). Generally, classifying the known classes correctly needs to reduce the empirical risk on the training set. Apart from reducing the empirical risk, OSR also needs to reduce the open space risk (Scheirer et al., 2013), which corresponds to identifying the unknown classes effectively (see Fig. 1).

Generally, it has reached a consensus that deep neural network can perform well in CSR. As shown in Fig. 2(a), LeNet++ trained with SoftMax can classify MNIST effectively. However, Fig. 2(d) shows that there is a considerable overlap between the known and unknown features. Under this circumstance, the network is incapable of distinguishing the known and unknown samples, which creates the open space risk. According to Scheirer et al. (2013), the size of the open space occupied by the known features determines the open space risk. In other words, making the known features more compact can reduce the open space risk. To accomplish this purpose, Yang et al. (2018) proposed the generalized convolutional prototype learning (GCPL) for robust classification and OSR (Yang et al., 2018, 2022). Fig. 2(b) shows

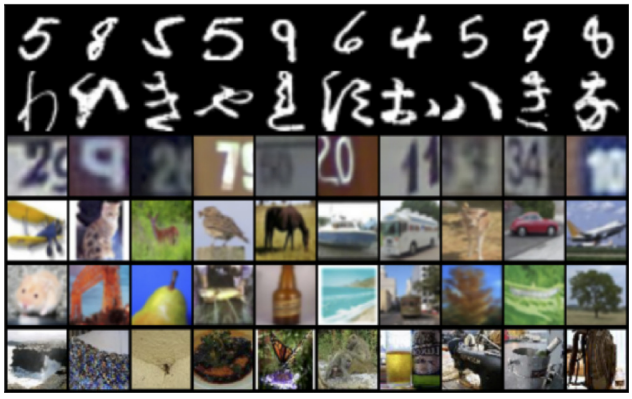
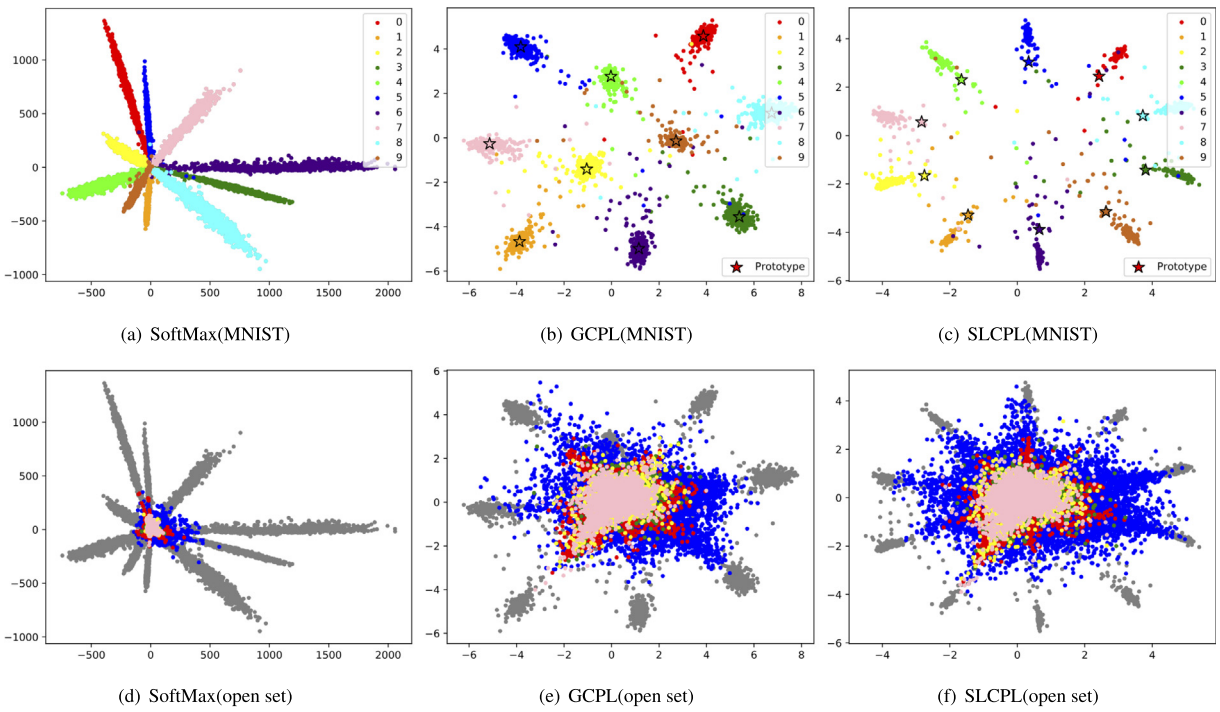


Fig. 1. This figure shows some samples in Fig. 2 for open set evaluation. From top to bottom, they are from MNIST (Lecun and Bottou, 1998), KMNIST (Clanuwat et al., 2018), SVHN (Netzer et al., 2011), CIFAR10 (Krizhevsky, 2012), CIFAR100 (Krizhevsky, 2012) and TinyImageNet (Russakovsky et al., 2015), respectively. It is obvious that their complexity gradually increases.

that the known features extracted with GCPL are much more compact than those extracted with SoftMax, and LeNet++ trained with GCPL also classify MNIST effectively. However, Fig. 2(e) shows there are still three clusters of the known features that overlap with the unknown features. Obviously, this overlap of features will make it difficult for the model to identify the unknown classes, which leads to the open space risk. Therefore, it is the features overlap that causes the open space risk.

* Corresponding authors.
E-mail addresses: wangpenghui@mail.xidian.edu.cn (P. Wang), hwliu@xidian.edu.cn (H. Liu).

https://doi.org/10.1016/j.cviu.2023.103651
Received 10 May 2022; Received in revised form 8 January 2023; Accepted 10 February 2023
Available online 16 February 2023



(a) SoftMax(MNIST) (b) GCPL(MNIST) (c) SLCPL(MNIST)

(d) SoftMax(open set) (e) GCPL(open set) (f) SLCPL(open set)

Fig. 2. The feature visualization results of LeNet++ on known and unknown classes, and MNIST is used for training the network (Dhamija et al., 2018). In the 1st row of the figure, dots in different colors represent different classes in the MNIST test set. In the 2nd row of the figure, MNIST (gray), KMNIST (blue), SVHN (red), CIFAR10 (green), CIFAR100 (yellow) and TinyImageNet (pink) are used for open set evaluation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

According to the above analysis of the open space risk, if we can figure out why they overlap, it is possible to significantly reduce, or even eliminate the open space risk. Therefore, this paper analyzes the distribution of the known and unknown features from Figs. 2(d) and 2(f), and it seems that the distribution of features satisfy these rules:

- There are two obvious different distribution patterns of the known features.
- The unknown features tend to be distributed in the center region of the feature space.
- As the complexity of the unknown classes increases, the distribution range of the unknown features decreases gradually.

These issues are closely related to the origin of the open space risk, which is explained by the proposed matching theory (**MT**). On the basis of the proposed theory about the open space risk, this paper proposes the spatial location constraint prototype loss (SLCPL) for OSR, which adds a constraint term to control the spatial location of the prototypes in the feature space. As shown in Fig. 2(c), apart from effectively reducing the empirical risk, SLCPL controls the known features to the edge region of the feature space, which can reduce the open space risk more effectively than SoftMax and GCPL.

Our contributions mainly focus on the following:

- The distribution of the known and unknown features is analyzed in detail.
- A matching theory is proposed to explain the origin of the open space risk when a deep neural network is used for OSR.
- A novel loss function, SLCPL, is proposed to address the OSR problem.
- Many experiments and analyses are performed to verify the proposed theory about the origin of the open space risk and the effectiveness of the proposed loss function.

2. Related work

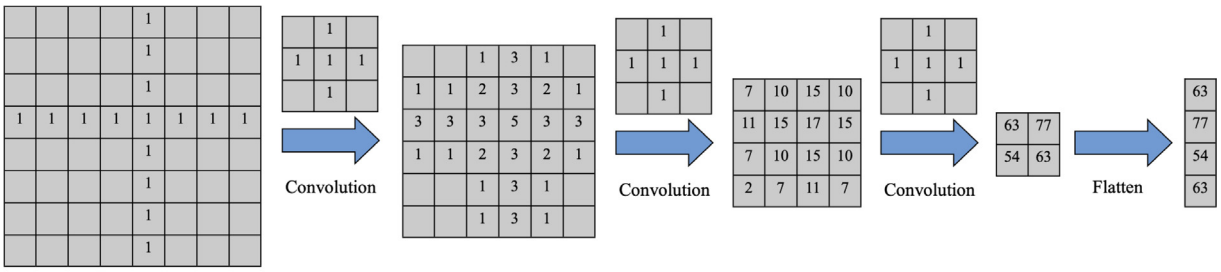
2.1. Open set recognition

Scheirer et al. (2013) first defined the OSR issue in 2013, and most of the current methods were based on support vector machine (SVM), such as 1-vs-set (Scheirer et al., 2013), W-SVM (Walter et al., 2014) and P_I-SVM (Jain et al., 2014). This kind of method mainly uses kernel function of specific form to extract features, which limits its feature extraction ability, and then limits the improvement of its OSR performance. Subsequently, the traditional method is gradually replaced by the deep neural network. Apart from using LeNet++ to assess the open space risk, Dhamija et al. (2018) proposed a novel loss function 'Objectosphere' for OSR. Bendale and Boult (2016) proposed the OpenMax model, which replaces the SoftMax layer by the OpenMax layer to predict the probability of the unknown classes. Ge et al. (2017) combined the characteristics of generative adversarial network (Goodfellow et al., 2014) and OpenMax, and proposed the G-OpenMax model, which trains deep neural network with generated unknown classes. Neal et al. (2018) proposed the OSRCI model for OSR, which also adds generated samples into the training phase. Some works based on the reconstruction idea also made important contributions to OSR, such as classification-reconstruction open set recognition (CROSR, Yoshihashi et al. (2019)), class conditioned auto-encoder (C2AE, Oza and Patel (2019)) and conditional Gaussian distribution learning (CGDL, Sun et al. (2020)). These methods took the reconstruction error of the test sample as the basis to judge the sample category.

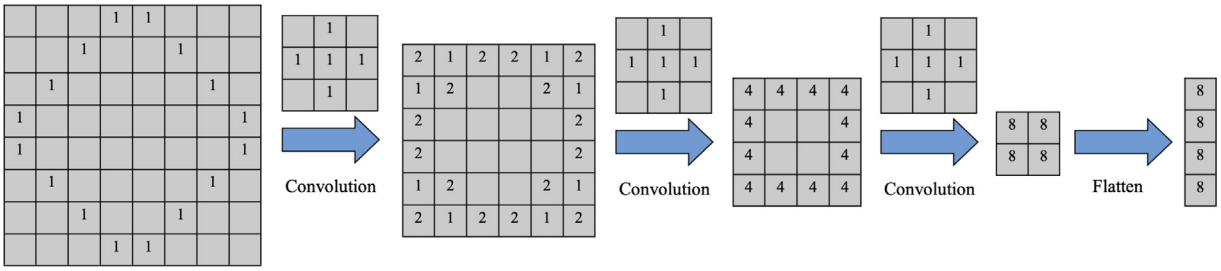
Although various algorithms or theories have been put forward to address the OSR problem, few works analyze the origin of the open space risk, and this paper will discuss this problem in Section 3.

2.2. Prototype learning

The prototype is often regarded as one or more points in the feature space to represent the clustering of a specific category. Wen et al.



(a) The case of convolution kernel matching sample.



(b) The case of convolution kernel mismatching sample.

Fig. 3. This figure shows the convolution processes of matching and mismatching. In fact, the actual numerical operation process of the network is much more complicated than that of this figure, while this paper just wants to illustrate the effect of the proposed MT by using these two examples.

(2016) proposed a novel center loss to improve the face recognition accuracy, and it can learn more discriminative features by making the known features more compact. Similarly, Fig. 2(b) shows that GCPL also has this capability. Moreover, Chen et al. (2020) proposed a novel reciprocal prototype loss (RPL) function for OSR, which encourages the known features to be far away from the corresponding prototypes, and it also achieved good performance.

Although GCPL can effectively improve the compactness of the feature clustering, simply increasing intra-class compactness still creates the open space risk slightly. Therefore, this paper propose a novel loss function that adds spatial location constraints to the prototype to further reduce the open space risk.

3. The nature of open set recognition

According to Fig. 2, it is believed that studying the distribution of the known and unknown features is the key to further understanding the origin of the open space risk, which is very significant to address the OSR problem. Therefore, This section figures out why the known and unknown features may overlap in the feature space.

3.1. Known features distribution

According to the distribution characteristics of the known features, this paper divides the known feature distribution into two categories: the half open space distribution, and the prototype subspace distribution.

The half-open space distribution. When the neural network is trained with SoftMax, the feature space is divided into several half-open spaces by several hyper-planes, and each known class occupies the corresponding half-open space. As shown in Fig. 2(a), the 10 known classes are distributed into 10 wedge-shaped clusters from the spatial origin, and entire 2D feature space is divided into several sector area

The prototype subspace distribution. When the neural network is trained with the prototype learning, the prototype subspace distribution is created. In this kind of distribution, the whole feature space is divided into several known subspace and residual space (also called the open space) by the prototypes, and each known subspace is represented as a hyper-sphere. As shown in Fig. 2(b), each known feature is distributed in the circle determined by the each prototype. Because GCPL is incapable of limiting the spatial location of the prototypes, some known features are possible to be distributed in the center region of the feature space. Moreover, the feature distribution of the proposed SLCPL also falls into this category, but the proposed model constrains the spatial location of the prototypes to be distributed in the edge region of the feature space, as shown in Fig. 2(c).

3.2. Unknown features distribution

From Figs. 2(d)–2(f), it can be seen that whatever the unknown class is, its features usually tend to be distributed in the center region of the entire feature space. Next, this section explains the inevitability of this phenomenon by the proposed MT:

> *The trained network and the training data are highly matched. If there is a difference in the distribution of test data and training data, the mismatch between test data and the trained network will weaken the outputs of the network.*

As is known, the neural network realizes recognition function by a series of complex numerical operations, such as convolution computing and fully connection computing. When a network is trained well, the training data is highly matched with these complex numerical operations. Therefore, when test data and training data are independently and identically distributed, the network will show good performance. However, when the distribution of test data and training data is different, this matching relationship between input data and complex numerical operations will be severely damaged. In other words, this mismatch will cause the internal friction of the network system and weaken the final output. Fig. 3 tries to illustrate the above idea with two simple examples of convolution. When the convolution kernel matches the sample well, the convolution operation can get higher feature value. These flattened features determine the position of the sample in the feature space, and higher feature values correspond to higher coordinate values in the feature space. On the contrary, as shown in Fig. 3(b), the mismatch between the sample and the convolution kernel will result in small coordinate values of the samples in the feature space, which will make the unmatched sample distributed in the center region of the feature space as shown in Figs. 2(d)–2(f).

3.3. Summary

Finally, this section summarizes the origin of the open space risk: First, a network is trained with the known classes, and most current training methods (such as SoftMax, GCPL) rarely avoid the known features to be distributed in the center region of the feature space. Then, because of the mismatch between the unknown classes and the trained network, the unknown classes usually tend to be distributed in the center region of the feature space. Therefore, the known and unknown features will overlap in the center region, which causes the open space risk. Section 5 will conduct lots of experiments to prove the rationality of the proposed MT.

4. Proposed method

According to the origin of the open space risk we explored in Section 3, if we can control the known features to be distributed in the edge region of the feature space, it will reduce the open space risk effectively. Therefore, this paper proposes SLCPL to accomplish this purpose. Since SLCPL is an improvement on GCPL, GCPL should be introduced firstly.

GCPL. Given training set $D = \{(x_1, y_1), (x_2, y_2), \dots\}$ with N known classes, the label of data x_i is $y_i \in \{1, \dots, N\}$. A neural network is used as a classifier, whose parameters and embedding function are denoted as θ and Θ, respectively. The prototypes $O = \{O^i, i = 1, 2, \dots, N\}$ are initialized randomly, and O^i corresponds to the ith known class. For a training sample (x, y), the GCPL loss function can be expressed as

$$l_G(x, y; \theta, O) = l(x, y; \theta, O) + \lambda pl(x; \theta, O), \tag{1}$$

where the optimization of $l(x, y; \theta, O)$ is used to classify the different known classes, λ is a hyper-parameter, and the constraint term $pl(x; \theta, O)$ is used to make the known features more compact. Specifically, $l(x, y; \theta, O)$ can be expressed as

$$\begin{aligned} l(x, y; \theta, O) &= -\log p(y = k|x, \Theta, O) \\ &= -\log \frac{e^{-d(\Theta(x), O^k)}}{\sum_{i=1}^{N} e^{-d(\Theta(x), O^i)}}, \end{aligned} \tag{2}$$

where $d(\Theta(x), O^i)$ is Euclidean distance between $\Theta(x)$ and O^i. And the constraint term $pl(x; \theta, O)$ can be expressed as

$$pl(x; \theta, O) = \sum_{k=1}^{N} d(\Theta(x^k), O^k), \tag{3}$$

where x^k is the training samples of kth class. Under the optimization of Eq. (1), the known features extracted from the network will present the prototype subspace distribution, as shown in Fig. 2(b).

SLCPL. According to the proposed MT, as long as the known feature clusters can be constrained in the edge region of the feature space, the open space risk can be effectively reduced. Therefore, this paper proposes SLCPL based on GCPL:

$$l_S(x, y; \theta, O) = l_G(x, y; \theta, O) + slc(O). \tag{4}$$

In Eq. (4), the spatial location constraint term $slc(O)$ can be expressed as

$$slc(O) = \frac{1}{N-1} \sum_{i=1}^{N} (r_i - \frac{1}{N} \sum_{j=1}^{N} r_j)^2, \tag{5}$$

where $r_i = d(O^i, O_c)$ and $O_c = \frac{1}{N} \sum_{i=1}^{N} O^i$.

In other words, the term $slc(O)$ is the variance of the distances r_i. Here, choosing O_c instead of coordinate origin as the center is more beneficial to the optimization of the training process. By controlling the variance of these distances, the known feature distribution in Fig. 2(b) can be manipulated into the distribution in Fig. 2(c).

How to detect unknown classes? In the proposed method, the distance between the feature and the prototype can be used to measure the probability of which category it belongs to. Specifically, the probability that x in the test set belongs to a known class can be determined by the following formula:

$$p(\hat{y} = k|x) \propto \exp\big(-\min_{k \in \{1, \dots, N\}} d(\Theta(x), O^k)\big). \tag{6}$$

Based on the distance distribution from the known features to the corresponding prototypes, a threshold value τ can be determined. When the distance between the sample feature and the prototype is greater than the threshold τ, the sample will be identified as unknown class; Otherwise, the category is further determined according to Eq. (6).

5. Experiments and discussion

5.1. Experimental settings

Evaluation Metrics. This paper choose the accuracy (ACC) and Area Under the Receiver Operating Characteristic (AUROC) curves to evaluate the performance of classifying known classes and identifying unknown classes respectively. Moreover, Macro Average F1-score and open set classification rate (OSCR, Dhamija et al. (2018)) are used to comprehensively evaluate the performance of OSR. OSCR is an indicator similar to AUROC, which evaluates the model by calculating the area under the corresponding curve. Let ξ be a score threshold. The correct classification rate (CCR) is the fraction of the samples where the correct class k has maximum probability and has a probability greater than ξ:

$$CCR(\xi) = \frac{|\{x|x \in (D_{kt}) \wedge \arg\max_k P(k|x) = \hat{k} \wedge P(\hat{k}|x) \geq \xi\}|}{|D_{kt}|}, \tag{7}$$

where D_{kt} represents the known test data. The false positive rate (FPR) is the fraction of samples from the unknown test data D_{ut} that are classified as any known class k with a probability greater than ξ:

$$FPR(\xi) = \frac{|\{x|x \in D_{ut} \wedge \max_k P(k|x) \geq \xi\}|}{|D_{ut}|}. \tag{8}$$

Similar to AUROC, the higher the value of OSCR, the better the performance of the model.

Network Architecture. To make a fair comparison with the evaluation results of Neal et al. (2018), this paper adopts the same encoder network structure (VGG32) for experiments. Moreover, to further prove the applicability of the method proposed in this paper, the large-scale network ResNet (He et al., 2016) is also used for experiments in Section 5.6.

Other Settings. In our experiments, the hyper-parameter λ is set to 0.1. The momentum stochastic gradient descent (SGD-M, Ning (1999)) optimizer is used to optimize the classifier. The initial learning rate of the network is set to 0.1, dropping to one-tenth of the original rate every 30 epochs, and the network is trained for 100 epochs. Additionally, the batch size is set to 128.

5.2. Ablation study

An important factor affecting OSR performance is the *openness*, which is defined by Scheirer et al. (2013), and it can be expressed as

$$\mathbb{O} = 1 - \sqrt{\frac{2 \times N_{train}}{N_{test} + N_{target}}}, \tag{9}$$

where N_{train} is the number of the known classes, N_{test} is the number of test classes that will be observed during testing, and N_{target} is the number of target classes that needs to be correctly recognized during test. We can see that with the increase of $\mathbb{O}$, the difficulty of OSR is also increasing.

Experiment 1. This experiment analyzes the proposed method on CIFAR100, which consists of 100 classes. We randomly sample 15 classes out of 100 classes as known and varying the number of unknown classes from 15 to 85, which means $\mathbb{O}$ is varied from 18% to 49%. The performance is evaluated by the Macro Average F1-score in 16

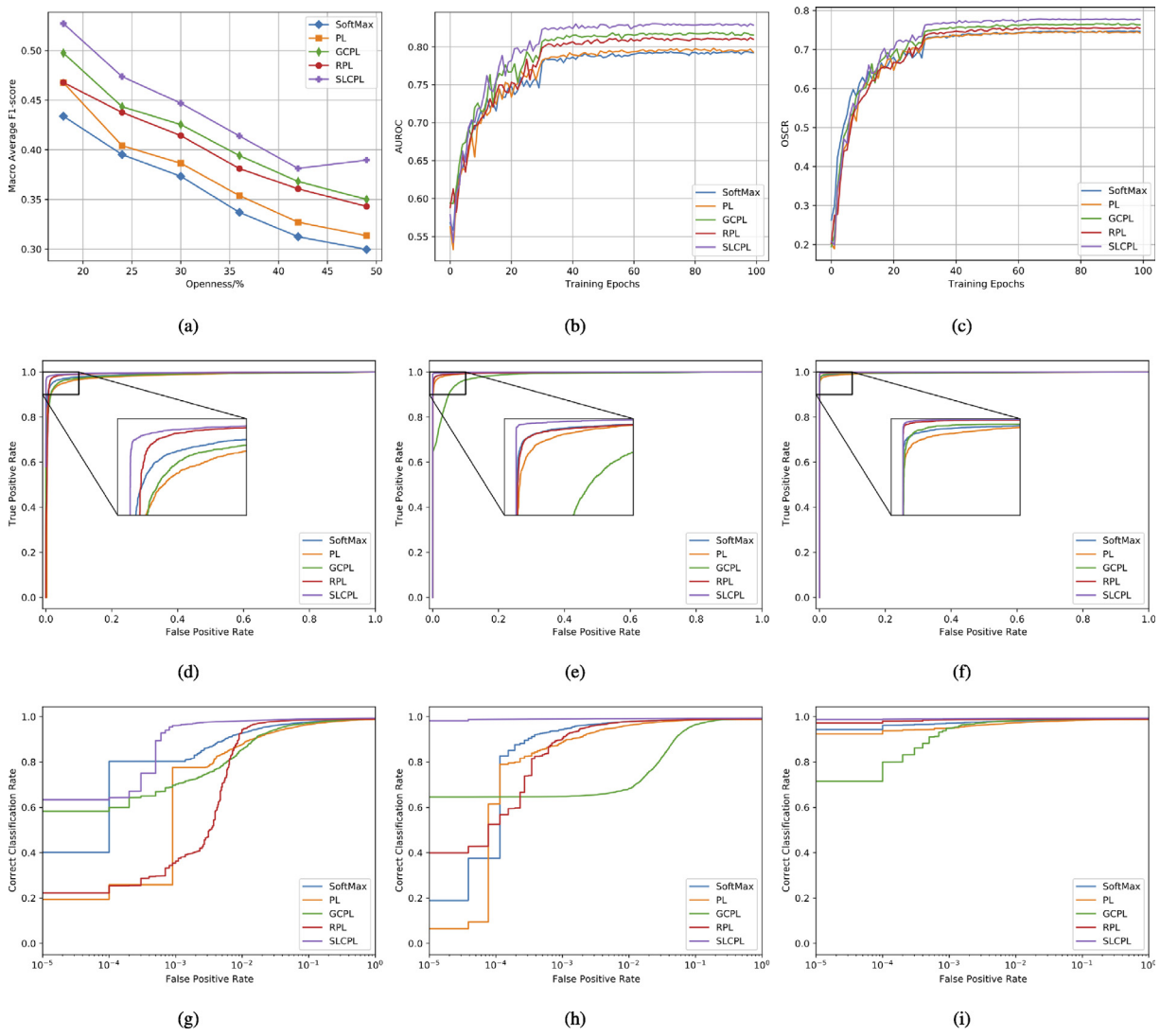


Fig. 4. Experiment 1, Macro Average F1-score against varying openness in Fig. 4(a); Experiment 2, AUROC/OSCR with training epochs in Figs. 4(b)/4(c); Experiment 3, 3 columns (2 rows) correspond to 3 kinds of data settings (2 kinds of metrics) in Figs. 4(d)∼4(i).

classes (15 known classes and unknown), and the threshold is set to 0.5 like (Yoshihashi et al., 2019; Sun et al., 2020). For ablation study, the network only trained with $l(x, y; \theta, O)$ is denoted as prototype loss (**PL**). The experimental results are shown in Fig. 4(a). The performance of PL, GCPL, and SLCPL increases in successively, which proves the effectiveness of the proposed method. Moreover, the proposed method is also better than another prototype method, RPL.

Experiment 2. This experiment conducts another ablation experiment to analyze the proposed method. In this experiment, the network is trained with CIFAR10, and TinyImageNet is used as the unknown classes for open set evaluation. The experimental results are shown in Figs. 4(b) and 4(c). Compared with other methods, the proposed method achieves the best results on AUROC and OSCR.

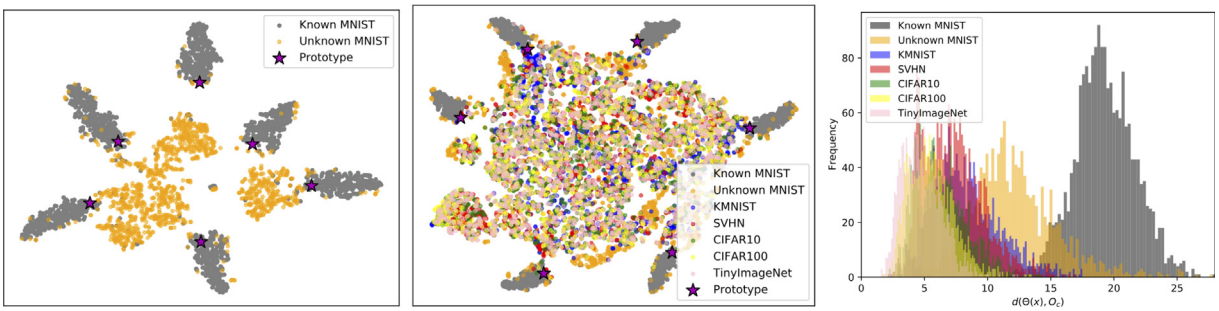
Experiment 3. In this experiment, the network is trained with MNIST, and KMNIST, SVHN, and CIFAR10 are used as the unknown classes for open set evaluation. The experimental results are shown in Figs. 4(d)∼4(i). As the difference between the unknown and the known classes increases, the OSR performance of the method increases gradually, which also explains the validity of the proposed MT. In addition, the proposed method achieves the best results in these 3 out-of-distribution detection experiments.

5.3. Fundamental experiments

This subsection provides a simple summary of these protocols for each data set. Except MNIST, some data augmentations (RandomCrop, RandomHorizontalFlip) are used in our experiments.

- MNIST, SVHN, CIFAR10. Each of these three datasets contains 10 categories. 6 categories are randomly selected as the known classes, and the remaining 4 categories are the unknown classes.
- CIFAR+10, CIFAR+50. For these two datasets, 4 classes are randomly sampled from CIFAR10 as the known classes for training and test. 10 and 50 classes are randomly sampled from CIFAR100, respectively, which are used as the unknown classes.
- TinyImageNet. 20 known classes and 180 unknown classes are randomly sampled for evaluation.

Table 1 shows the performance ACC of the different methods for CSR tasks, whose test set excludes the unknown classes. Moreover, ACC in this table is calculated by Eq. (6) without thresholds. As shown in Table 1, ACC of the proposed method has an advantage in both mean and variance, which shows that the proposed method can well reduce the empirical risk. Compared with GCPL, the constraint term $slc(O)$ proposed in this paper does not reduce the ACC of CSR, but improves it.



(a) t-SNE(MNIST) (b) t-SNE(open set) (c) Distance distribution

Fig. 5. The feature space dimension of network used in Section 5.3 is 128, so the t-SNE (Maaten and Hinton, 2012) technology is used to visualize its feature distribution. Fig. 5(c) shows the distance distribution results of the high dimensional space in Fig. 5(b).

Table 1
The ACC (%) test on various methods and datasets. Every value is averaged among five randomized trials. This table reports the results from Yang et al. (2022), Neal et al. (2018) and Yoshihashi et al. (2019), except reproduce the RPL results from Chen et al. (2020).

Method	MNIST	SVHN	CIFAR10
SoftMax (Neal et al., 2018)	99.5 ± 0.2	94.7 ± 0.6	80.1 ± 3.2
OpenMax (Bendale and Boult, 2016)	99.5 ± 0.2	94.7 ± 0.6	80.1 ± 3.2
G-OpenMax (Ge et al., 2017)	99.6 ± 0.1	94.8 ± 0.8	81.6 ± 3.5
OSRCI (Neal et al., 2018)	99.6 ± 0.1	95.1 ± 0.6	82.1 ± 2.9
CROSR (Yoshihashi et al., 2019)	99.2 ± 0.1	94.5 ± 0.5	93.0 ± 2.5
RPL (Chen et al., 2020)	**99.8** ± 0.1	96.9 ± 0.4	94.5 ± 1.7
GCPL (Yang et al., 2022)	99.7 ± 0.1	96.7 ± 0.4	92.9 ± 1.2
SLCPL	**99.8** ± 0.1	**97.1** ± 0.3	**94.6** ± 1.2

Similar to Table 1, Table 2 shows that the performance difference between different methods becomes larger with the increase of the difficulty of OSR. In this experiment, the proposed method has the strongest ability to identify the unknown classes on all datasets. In particular, the comparison with GCPL further proves the validity of the constraint term $slc(O)$ proposed in this paper. Compared with GCPL, only adding a constraint term to the loss function can greatly improve the OSR performance, which is derived from the analysis of the origin of the open space risk.

5.4. Visualization analysis

To further verify the effectiveness of the proposed method, this subsection presents the experiment results on SLCPL through visualization and distance analysis.

Fig. 5(a) visualizes the MNIST experimental results in Section 5.3. To compare with Fig. 2(f), this experiment also gives the corresponding visualization results as shown in Fig. 5(b). It can be seen from these 2 figures that the proposed method can distinguish the known and unknown classes well, except that a few unknown features overlap with the known features. As for the distribution of the unknown features, Fig. 5(c) further demonstrates that the unknown features tend to be distributed in the center region of the feature space. In addition, Fig. 5(c) also shows that the distribution rule of the unknown features is similar to Fig. 2: with the increase of the difference between the known and unknown classes, the distribution of the unknown features is closer to the center region. In general, the above results verify the proposed MT about the origin of the open space risk and the effectiveness of the proposed method.

5.5. Further experiments

Experiment I. All samples from the 10 classes in CIFAR10 are considered as the known classes, and samples from ImageNet and LSUN (Yu et al., 2016) are selected as the unknown classes. This experiment resizes or crops the unknown samples to make them have the same size with the known samples. The Settings for datasets are the same as in Yoshihashi et al. (2019), and the other experimental settings are the same as in Section 5.1. For convenience in performance comparison, threshold is also set to 0.5 like (Yoshihashi et al., 2019; Sun et al., 2020). The results are shown in Table 3.

Experiment II. All samples from the 10 classes in MNIST are considered as the known classes, and samples from Omniglot (Accent, 1968), MNIST-noise and Noise are selected as the unknown classes. Omniglot is a data set containing various alphabet characters. Noise is a synthesized data set by setting each pixel value independently from a uniform distribution on $[0, 1]$. MNIST-Noise is also a synthesized data set by adding noise on MNIST testing samples. The Settings for datasets are the same as in Yoshihashi et al. (2019), and the other experimental settings are the same as in Section 5.1. For convenience in performance comparison, threshold is also set to 0.5 like (Yoshihashi et al., 2019; Sun et al., 2020). The results are shown in Table 4.

In above 2 experiments, it can seen from the results that on all given datasets, the proposed method is more effective than previous methods and achieves a new state-of-the-art performance. Especially, compared with GCPL, these results once again prove the validity of the constraint term $slc(O)$ proposed in this paper.

5.6. Adversarial experiment

To better compare the proposed method with SoftMax and GCPL, this subsection conducts an adversarial experiment on the TinyImageNet and ImageNet-O (Hendrycks et al., 2021). ImageNet-O is an adversarial data set with 200 classes based on ImageNet, whose samples are very similar to that of ImageNet. This experiment is conducted on 43 classes shared by TinyImageNet and ImageNet-O: ResNet152 (He et al., 2016) is trained on 43 classes of TinyImageNet and the corresponding classes of ImageNet-O are used as the unknown classes for open set evaluation. Fig. 6 shows some samples of 5 classes in this experiment, and it can seen that the samples from ImageNet-O are very similar to the samples from TinyImageNet. Therefore, it is difficult for the models to identify them.

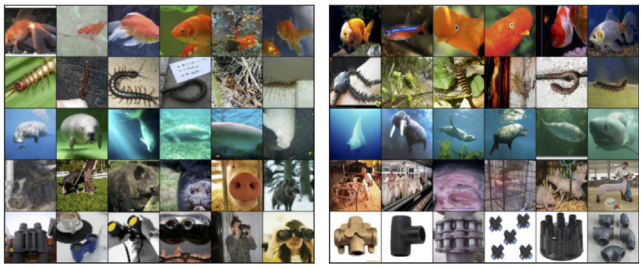
The experimental results are shown in Table 5. As we predicted, all of methods cannot detect the unknown ImageNet-O samples well, but the proposed method still achieves the best results. Apart from ImageNet-O, this experiment also uses KMNIST, SVHN, and CIFAR10 as the unknown classes to evaluate the OSR performance of these methods. To further verify the proposed MT, this experiment also provides the distance distribution results in the feature space. As shown in Fig. 7, because ImageNet-O is so similar to TinyImageNet, their spatial distance distribution is almost overlap. In terms of other unknown classes, due to the difference between the unknown and the known, the spatial distribution of unknown classes is almost completely separate with known classes, and they are closer to the center O_c than the known classes. Above all, Figs. 5(c) and 7 show that the MT proposed in this paper is valid regardless of the complexity of the known classes and the dimension of the feature space.

Table 2
The AUROC (%) of the OSR test on various methods and datasets. Every value is averaged among five randomized trials. This table reports the corresponding results from Yang et al. (2022), Neal et al. (2018), Yoshihashi et al. (2019) and Chen et al. (2020). Standard deviation values for state of the art are not available for CIFAR+10, CIFAR+50 and TinyImageNet.

Method	MNIST	SVHN	CIFAR10	CIFAR+10	CIFAR+50	TinyImageNet
SoftMax (Neal et al., 2018)	97.8 ± 0.6	88.6 ± 1.4	67.7 ± 3.8	81.6	80.5	57.7
OpenMax (Bendale and Boult, 2016)	98.1 ± 0.5	89.4 ± 1.3	69.5 ± 4.4	81.7	79.6	57.6
G-OpenMax (Ge et al., 2017)	98.4 ± 0.5	89.6 ± 1.7	67.5 ± 4.4	82.7	81.9	58.0
OSRCI (Neal et al., 2018)	98.8 ± 0.4	91.0 ± 1.0	69.9 ± 3.8	83.8	82.7	58.6
CROSR (Yoshihashi et al., 2019)	99.1 ± 0.4	89.9 ± 1.8	–	–	–	58.9
RPL (Chen et al., 2020)	99.3	95.1	**86.1**	85.6	85.0	70.2
GCPL (Yang et al., 2022)	99.0 ± 0.2	92.6 ± 0.6	82.8 ± 2.1	88.1	87.9	63.9
SLCPL	**99.4** ± 0.1	**95.2** ± 0.8	**86.1** ± 1.4	**91.6** ± 1.7	**88.8** ± 0.7	**74.9** ± 1.4

Table 3
OSR results on CIFAR10 with various outliers added to the test set as unknowns. The performance is evaluated by Macro Average F1-scores (%) in 11 classes (10 known classes and 1 unknown class). This table reports the experiment results from Yoshihashi et al. (2019) and Sun et al. (2020), and the results of GCPL and RPL are reproduced from Yang et al. (2022) and Chen et al. (2020).

Method	ImageNet-crop	ImageNet-resize	LSUN-crop	LSUN-resize
SoftMax (Neal et al., 2018)	63.9	65.3	64.2	64.7
OpenMax (Bendale and Boult, 2016)	66.0	68.4	65.7	66.8
Ladder+SoftMax (Yoshihashi et al., 2019)	64.0	64.6	64.4	64.7
Ladder+OpenMax (Yoshihashi et al., 2019)	65.3	67.0	65.2	65.9
DHR+SoftMax (Yoshihashi et al., 2019)	64.5	64.9	65.0	64.9
DHR+OpenMax (Yoshihashi et al., 2019)	65.5	67.5	65.6	66.4
OSRCI (Neal et al., 2018)	63.6	63.5	65.0	64.8
CROSR (Yoshihashi et al., 2019)	72.1	73.5	72.0	74.9
C2AE (Sun et al., 2020)	83.7	82.6	78.3	80.1
CGDL (Sun et al., 2020)	84.0	83.2	80.6	81.2
RPL (Chen et al., 2020)	84.6	83.5	85.1	87.4
GCPL (Yang et al., 2022)	85.0	83.5	85.3	88.4
SLCPL	**86.7**	**85.9**	**86.5**	**89.2**



(a) TinyImageNet (b) ImageNet-O

Fig. 6. From top to bottom, the classes are "goldfish", "centipede", "dugong", "pig", and "binoculars", respectively. The samples from ImageNet-O are very similar to that of TinyImageNet, but they are not from the same class, actually.

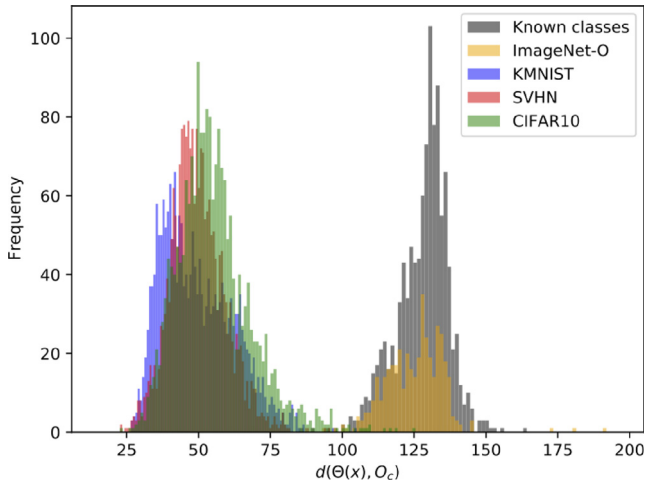


Fig. 7. In this figure, 'Known classes' represents TinyImageNet.

Table 4
OSR results on MNIST with various outliers added to the test set as unknowns. The rest of the information is similar to Table 3.

Method	Omniglot	MNIST-noise	Noise
SoftMax (Neal et al., 2018)	59.5	80.1	82.9
OpenMax (Bendale and Boult, 2016)	78.0	81.6	82.6
CROSR (Yoshihashi et al., 2019)	79.3	82.7	82.6
CGDL (Sun et al., 2020)	85.0	88.7	85.9
RPL (Chen et al., 2020)	96.4	92.6	99.1
GCPL (Yang et al., 2022)	97.1	93.5	99.5
SLCPL	**97.9**	**93.7**	**99.6**

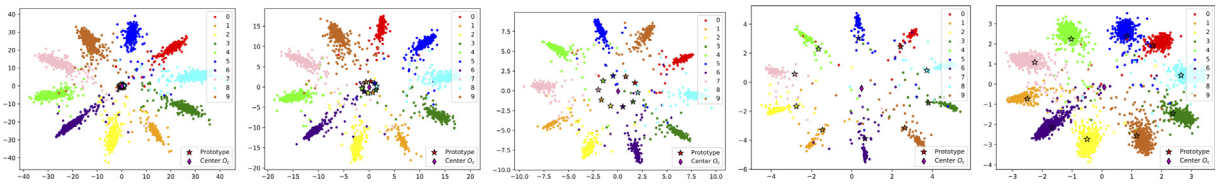
Table 5
Adversarial experiment results and Out-of-distribution detection results.

Method	ImageNet-O		KMNIST		SVHN		CIFAR10	
	AUROC	OSCR	AUROC	OSCR	AUROC	OSCR	AUROC	OSCR
SoftMax	54.3	50.0	84.0	64.3	76.0	61.0	75.8	60.7
PL	59.2	47.0	95.2	59.5	96.4	59.6	94.8	59.2
GCPL	57.6	51.9	98.9	68.1	99.1	68.1	98.9	68.0
RPL	55.5	42.7	99.4	58.3	98.5	58.1	92.3	56.3
SLCPL	**62.2**	**54.0**	**99.9**	**70.0**	**99.8**	**69.9**	**99.7**	**69.9**

5.7. Further analysis

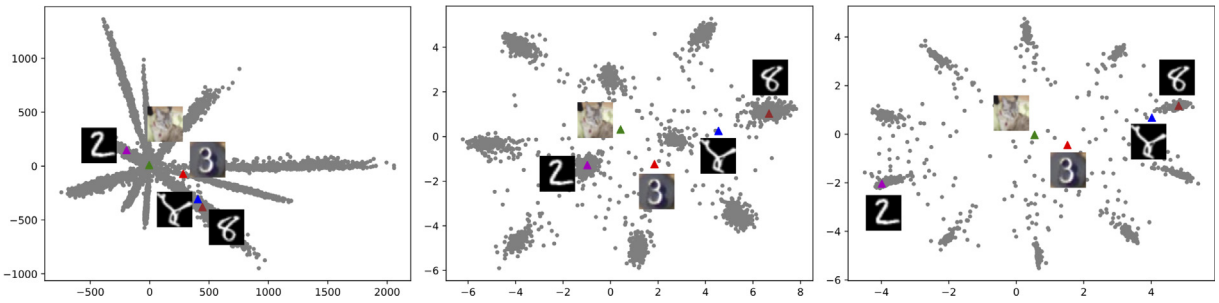
5.7.1. About the hyper-parameter

This subsection discusses the effect of the hyper-parameter λ on the proposed method. Fig. 8 shows the MNIST visualization results when λ is set to 0.0001, 0.001, 0.01, 0.1 and 1.0, respectively. According to Eq. (1), λ controls the distance between training features and the corresponding prototype, which is presented in Fig. 8. At the same time, the reduction of the distance between sample features and the corresponding prototypes also increases the relative dispersion of the prototypes. Considering the distribution of the unknown features, the value of $\lambda = 0.1$ is recommended for real applications.



(a) $\lambda = 0.0001$ (b) $\lambda = 0.001$ (c) $\lambda = 0.01$ (d) $\lambda = 0.1$ (e) $\lambda = 1.0$

Fig. 8. The visualization results with different value of hyper-parameter λ.



(a) SoftMax (b) GCPL (c) SLCPL

Fig. 9. The triangles represent the spatial location of the five samples ('2' and '8' from MNIST, a character from KMNIST, '3' from SVHN, 'cat' from CIFAR10).

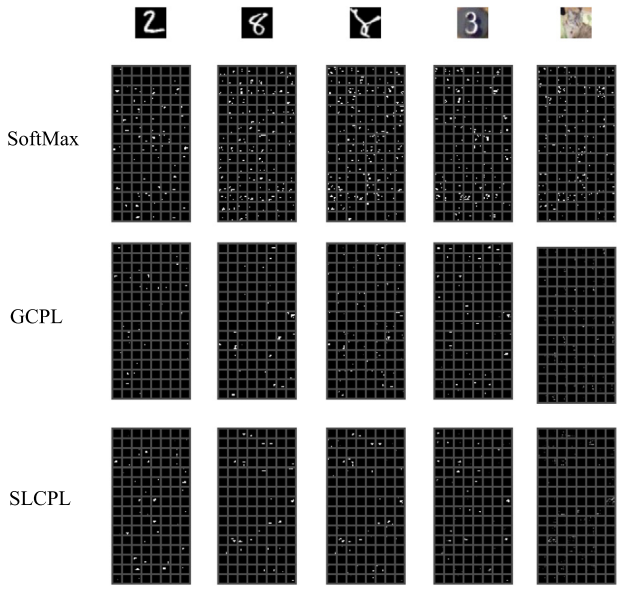


Fig. 10. The visualization results of the last convolution feature map in LeNet++. Different columns (rows) correspond to different methods (samples). With the increase of the difference between the unknown samples and the known samples, the brightness of the feature map decreases gradually.

5.7.2. About the unknown features distribution

To further illustrate the proposed MT about the open space risk, five specific samples are randomly selected as shown in Fig. 9 for visualization analysis. In this figure, 2 samples ('2' and '8' in MNIST) are the known classes and other 3 samples are the unknown classes. As the 3 unknown samples become more and more different from the known MNIST, they get closer and closer to the center of the feature space.

LeNet++ has 6 convolution layers and 2 fully connection layers, and we visualized the feature map after the last convolution operation as shown in Fig. 10 (see Appendix for more details). In this figure, the brighter the area, the higher the value. And the following issues can be find in this figure:

- In 3 rows, from left to right the brightness of the feature map decreases gradually.
- The brightness of '2' is not as bright as that of '8' in SoftMax and GCPL.
- The feature maps in SoftMax are generally brighter than other methods.

The 1st item indicates that with the increase of the mismatch between the unknown samples and the trained network, the outputs of the network decrease gradually. Especially the sample 'cat', its feature map brightness of GCPL and SLCPL are very low. The 1st and 2nd items explain the relationship between the brightness of the feature map and the position in the feature space. Specifically, the darker the feature map is, the closer it is to the center region of the feature space. The 3rd item indicates that the brightness of feature map is also related to the absolute scale of the spatial location coordinates.

On the whole, it is the mismatch between the unknown samples and the trained network that makes the unknown features usually tend to be distributed the center of the feature space.

5.7.3. Advantages and disadvantages

This subsection provides a brief summary of the advantages and disadvantages of the proposed method.

The advantages of the proposed method are as follows:

- Easy to understand; Fig. 2 clearly shows the technical ways in which the proposed approach improves performance.
- Training cost is small; The constraints term $slc(O)$ in the proposed SLCPL only involve the optimization of N prototype positions, and its computational amount is almost negligible compared with GCPL.
- Robust performance; In this paper, a large number of experiments are used to verify the performance advantages of the proposed method under different network structures, different scale datasets and different evaluation metrics.

The disadvantages of the proposed method are as follows:

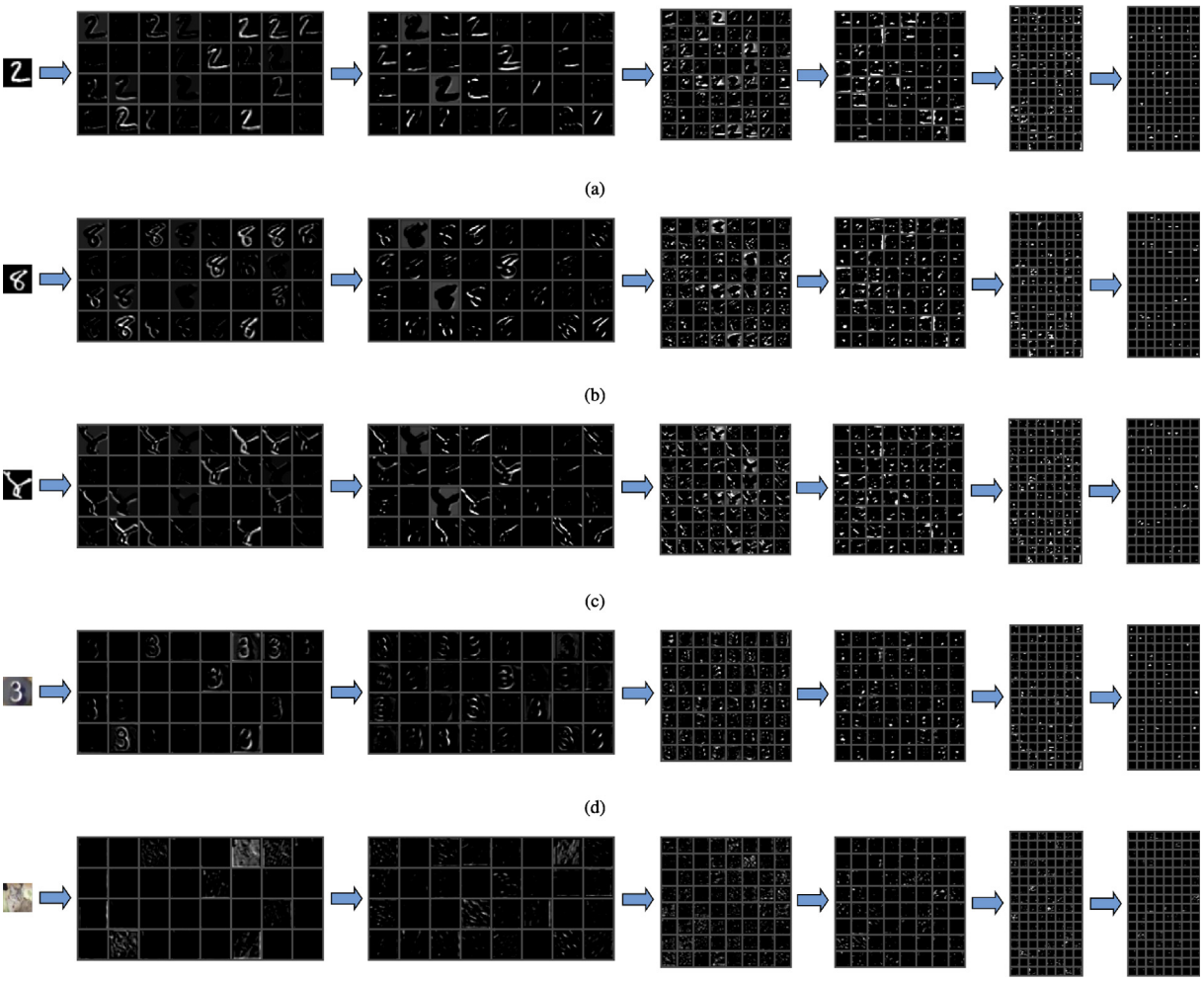


Fig. A.11. In this figure, these 5 convolution output results correspond to the 3rd column of Fig. 10.

- Hyper-parameter; Although this article recommends 0.1 for λ, in practice it usually needs to take several trials to determine the value that achieves the best performance.
- Network structure; Simple datasets can achieve good performance by choosing simple network, but complex datasets still choose simple network will affect the performance of the method. Although other methods also have this problem, we found that the proposed method seems to be more sensitive to this problem in real practice. It is believed that this issue is related to the excessive dependence of the proposed method on the feature distribution in the feature space.
- Threshold; Like many other works (Scheirer et al., 2013; Yang et al., 2018, 2022; Walter et al., 2014; Yoshihashi et al., 2019; Oza and Patel, 2019; Sun et al., 2020; Chen et al., 2020), the proposed method also rarely discuss threshold selection in detail. This is because different unknown classes often need to set different thresholds to obtain better performance. However, it is difficult to obtain prior information of unknown classes in practical applications.

6. Conclusion

As mentioned in this paper, the difficulty of OSR lies in how to reduce the open space risk. However, previous OSR works rarely discuss the open space risk in detail. As discussed in Section 3, the distribution of the known and unknown features determines the open space risk. Because of the mismatch between the unknown classes and the trained network, the unknown features usually tend to be distributed in the center region of the feature space. Therefore, it is natural for us to think that if the known features can be distributed in the edge region of the feature space by adjusting the loss function, then the open space risk will be effectively reduced. Based on the above analysis and discussion, this paper proposes SLCPL for OSR. Finally, a large number of experiments and analysis have proved the proposed MT about the open space risk and the effectiveness of the proposed method. In the future, more advanced loss functions may be proposed to further improve OSR performance based on the proposed MT.

CRediT authorship contribution statement

Ziheng Xia: Conceptualization, Methodology, Software, Validation, Writing – original draft. **Penghui Wang:** Supervision, Writing – review & editing, Funding acquisition. **Ganggang Dong:** Writing – review & editing. **Hongwei Liu:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code has been shared on GitHub.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61701379 and Grant 62192714, in part by the Stabilization Support of National Radar Signal Processing Laboratory under Grant KGJ202204, in part by the Fundamental Research Funds for the Central Universities under Grant QTZX22160, and in part by the Shanghai Aerospace Science and Technology Innovation Fund under Grant SAST2021-011.

Appendix

More experimental details in Section 5.7.2 are presented here. As shown in Fig. A.11, with the increase of the mismatch between the unknown samples and the trained network, the outputs of each convolution layer decrease gradually. Therefore, the greater the difference between the unknown sample and the known sample, the more the unknown features tend to be distributed in the center region of the feature space. In general, it is believed that the proposed MT can explain the origin of the open space risk well.

References

Accent, A., 1968. Omniglot writing systems and languages of the world. Z. Psychol. 175 (1), 64–91.

Bendale, A., Boult, T.E., 2016. Towards open set deep networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 1563–1572.

Chen, Guangyao, Qiao, Limeng, Shi, Yemin, Peng, Peixi, Li, Jia, Huang, Tiejun, Pu, Shiliang, Tian, Yonghong, 2020. Learning open set network with discriminative reciprocal points. In: Computer Vision – ECCV 2020. pp. 507–522.

Clanuwat, Tarin, Bober-Irizar, Mikel, Kitamoto, Asanobu, Lamb, Alex, Yamamoto, Kazuaki, Ha, David, 2018. Deep learning for classical Japanese literature. CoRR, abs/1812.01718.

Dhamija, Akshay Raj, Günther, Manuel, Boult, Terrance, 2018. Reducing network agnostophobia. In: Advances in Neural Information Processing Systems, Vol. 31.

Ge, ZongYuan, Demyanov, Sergey, Chen, Zetao, Garnavi, Rahil, 2017. Generative OpenMax for multi-class open set classification. In: 2017 British Machine Vision Conference Proceedings.

Goodfellow, Ian, Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron, Bengio, Yoshua, 2014. Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems, Vol. 27. Curran Associates, Inc..

Granell, Emilio, Romero, Verónica, Martínez-Hinarejos, Carlos-D., 2019. Image–speech combination for interactive computer assisted transcription of handwritten documents. Comput. Vis. Image Underst. 180, 74–83.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hendrycks, Dan, Zhao, Kevin, Basart, Steven, Steinhardt, Jacob, Song, Dawn, 2021. Natural adversarial examples.

Jain, L.P., Scheirer, W.J., Boult, T.E., 2014. Multi-class open set recognition using probability of inclusion. In: 2014 European Conference on Computer Vision. pp. 393–409.

Krizhevsky, Alex, 2012. Learning Multiple Layers of Features from Tiny Images. University of Toronto.

Lecun, Y., Bottou, L., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.

Maaten, Lvd, Hinton, G., 2012. Visualizing non-metric similarities in multiple maps. Mach. Learn. 87 (1), 33–55.

Neal, Lawrence, Olson, Matthew, Fern, Xiaoli, Wong, Weng-Keen, Li, Fuxin, 2018. Open set learning with counterfactual images. In: Ferrari, Vittorio, Hebert, Martial, Sminchisescu, Cristian, Weiss, Yair (Eds.), Computer Vision – ECCV 2018. pp. 620–635.

Netzer, Yuval, Wang, Tao, Coates, Adam, Bissacco, Alessandro, Wu, Bo, Ng, Andrew, 2011. Reading digits in natural images with unsupervised feature learning. In: NIPS.

Ning, Q., 1999. On the momentum term in gradient descent learning algorithms. Neural Netw. 12 (1), 145–151.

Oza, Poojan, Patel, Vishal M., 2019. C2AE: Class conditioned auto-encoder for open-set recognition. In: 2019 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2302–2311.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115 (3), 211–252.

Scheirer, Walter, Rocha, Anderson, Sapkota, Archana, Boult, Terrance, 2013. Toward open set recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35, 1757–1772.

Sun, X., Yang, Z., Zhang, C., Peng, G., Ling, K. V., 2020. Conditional gaussian distribution learning for open set recognition. In: 2020 IEEE Conference on Computer Vision and Pattern Recognition. pp. 13480–13489.

Walter, J., Scheirer, Lalit, P., Jain, Terrance, E., Boult, 2014. Probability models for open set recognition. IEEE Trans. Pattern Anal. Mach. Intell. 36 (11), 2317–2324.

Wen, Yandong, Zhang, Kaipeng, Li, Zhifeng, Qiao, Yu, 2016. A discriminative feature learning approach for deep face recognition. In: Computer Vision – ECCV 2016. pp. 499–515.

Yang, H.M., Zhang, X.Y., Yin, F., Liu, C.L., 2018. Robust classification with convolutional prototype learning. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition.

Yang, H.M., Zhang, X.Y., Yin, F., Yang, Q., Liu, C.L., 2022. Convolutional prototype network for open set recognition. IEEE Trans. Pattern Anal. Mach. Intell. 44 (5), 2358–2370.

Yoshihashi, Ryota, Shao, Wen, Kawakami, Rei, You, Shaodi, Iida, Makoto, Naemura, Takeshi, 2019. Classification-reconstruction learning for open-set recognition. In: 2019 IEEE Conference on Computer Vision and Pattern Recognition. pp. 4011–4020.

Yu, Fisher, Seff, Ari, Zhang, Yinda, Song, Shuran, Funkhouser, Thomas, Xiao, Jianxiong, 2016. LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop.

Zhang, Bowen, Tondi, Benedetta, Barni, Mauro, 2020. Adversarial examples for replay attacks against CNN-based face recognition with anti-spoofing capability. Comput. Vis. Image Underst. 197–198, 102988.