Open Set Recognition Using the Feature Space of Deep Neural Networks

doi: https://doi.org/10.1109/ISPACS51563.2021.9650985

Open Set Recognition Using the Feature Space of Deep Neural Networks

1st Daiju Kanaoka
Graduate School of Life Science and Systems Engineering
Kyushu Institute of Technology
Fukuoka, Japan
kanaoka.daiju327@mail.kyutech.jp

2nd Yuichiro Tanaka
Research Center for Neuromorphic AI Hardware
Kyushu Institute of Technology
Fukuoka, Japan
tanaka-yuichiro@brain.kyutech.ac.jp

3rd Hakaru Tamukoh
Graduate School of Life Science and Systems Engineering
Research Center for Neuromorphic AI Hardware
Kyushu Institute of Technology
Fukuoka, Japan
tamukoh@brain.kyutech.ac.jp

***Abstract*—Image classification assumes that all classes used in testing are known. Therefore, when an unknown class data is input, it cannot be recognized correctly. A method that enables unknown classes to be identified is called open set recognition. In this paper, we propose a method of open set recognition focusing on the feature space of the classifier and a Mahalanobis-based threshold. The experimental results show that the proposed method surpasses state-of-the-art methods on some datasets, demonstrating the potential of a method focusing on the feature space.**

***Index Terms*—open set recognition, deep learning, neural networks, image classification, unknown class**

I. Introduction

Recently, most image classification methods have become capable of classifying learned images, but when an untrained image is input, it is classified as one of the classes of the learned images. Such image classification is referred to as closed set recognition. In contrast, when an untrained image is input, the image classification that enables it to be classified as unknown is called open set recognition [1]. Fig. 1 shows the difference between closed set recognition and open set recognition. Although open set recognition has been studied, its accuracy is still low, and further research is required.

An anomaly detection method using the feature space of a recognition model pre-trained by ImageNet has been proposed [2]. This research confirmed that normal and abnormal images appear at different locations in feature space. This method achieved state-of-the-art performance, and it is expected that methods focusing on feature space will be studied in future.

In this study, we propose an open set recognition method that focuses on the feature space of deep neural networks. In the feature space of a trained classifier, a cluster is formed for each class. When an untrained image is an input, it should appear in a position in the feature space that differs from the cluster locations of each class.

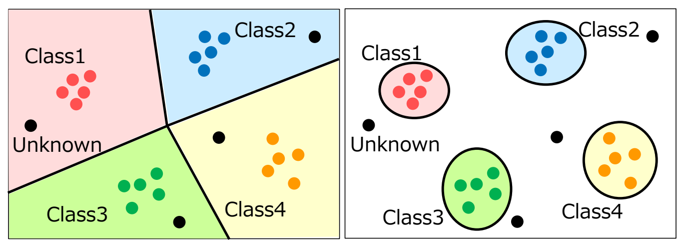

Fig. 1. Comparison between closed set recognition and open set recognition

II. Previous Research

Deep neural network-based open set recognition can be divided into two categories: discriminative models and reconstructive models.

A. Discriminative Model (DM) Based Methods

If an output of an image classifier exceeds a certain threshold, it is classified as unknown. This is a simple method that can be easily extended from existing closed set recognition models. It is also used in traditional machine learning methods, and many variants exist [3], [4].

B. Reconstructive Model (RM) Based Methods

This method combines a reconstruction network with an image classifier to determine an unknown by comparing an input image with a reconstructed image. This method is currently the mainstream method, and many research methods have been reported [5]–[7] that perform better than discriminative model-based methods.

III. Proposed Method

In this study, we propose an open set recognition method that uses the feature space of deep neural networks. We input training data to a trained classifier and apply multivariate Gaussian fitting for each class of clusters appearing in the

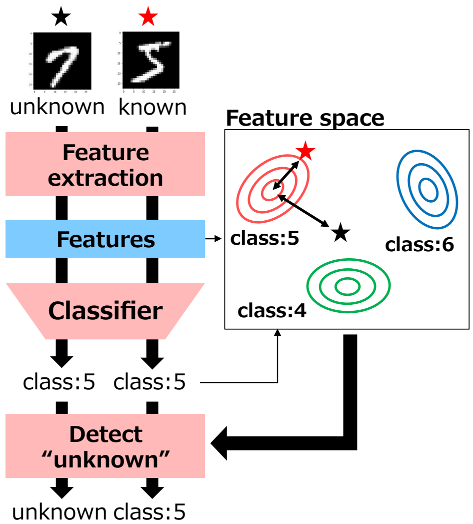

Fig. 2. Block diagram of the proposed method

feature space. Its operation is presented in Fig. 2. First, the test data are input and the labels are estimated. Next, we calculate the Mahalanobis distance between the multivariate Gaussian distribution of the estimated labels and the test data. A distance with a confidence interval of 95% of the multivariate Gaussian distribution is set as the threshold. If the calculated Mahalanobis distance is higher than this threshold, the data are classified as unknown.

IV. Experiment and Result

We compared the performance of the proposed method with those of previous methods on three datasets: MNIST, SVHN, and CIFAR10. For open set recognition, we trained the networks on only some classes of the dataset. During testing the untrained classes were treated as unknown classes to evaluate performance. In this experiment, we trained on six of the ten classes in each dataset and used the remaining four classes as unknown classes.

The experimental results are listed in Table I. The scores of the comparison methods were taken from [7]. We trained the classification model using ResNet [8] and evaluated its performance.

TABLE I
Macro-F1 scores

	Method	MNIST	SVHN	CIFAR10
DM	Softmax [3]	0.768	0.725	0.600
	Openmax [4]	0.798	0.737	0.623
RM	CROSR [6]	0.803	0.753	0.668
	GDFR [5]	0.821	0.716	0.700
	CGDL [7]	0.837	0.776	0.655
	Proposed	0.843	0.741	0.610

On MNIST, we achieved a score that exceeded those of the state-of-the-art methods. On SVHN, the proposed method achieved a score close to that of a reconstructive model-based method. However, the score of the proposed method on CIFAR10 was the next lowest score after [3].

V. Discussion and Future Work

The proposed method achieved a high score on MNIST and SVHN, but it was inferior to the other methods on CIFAR10. MNIST and SVHN are datasets of digit images, and there is not much difference in the shape of the recognition target within the same class. However, CIFAR10 is a dataset of object images, so there is substantial variation in the shape of the recognition target within the same class. Therefore, it is thought that the variance in the feature space of deep neural networks is high, resulting in a low score. In conclusion, we expect that our method to be effective when the recognition target is simple or a huge dataset can be used for constructing the feature space of deep neural networks.

In future, we aim to improve the performance of the method by setting constraints to minimize the within-class variance and maximize the between-class variance in a feature space used for face recognition.

VI. Conclusion

In this study, we proposed and verified an open set recognition method using the Mahalanobis distance on the feature space of neural networks. The experimental results show that our method achieves better scores than the state-of-the-art methods in MNIST, which demonstrates the potential of a method focusing on the feature space. In the future, we aim to improve the performance by setting constraints on the feature space, such as a feature space for face recognition.

Acknowledgment

This paper is based on results obtained from a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

[1] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 35, no. 7, pp. 1757–1772, 2013.

[2] O. Rippel, P. Mertens, and D. Merhof, "Modeling the distribution of normal data in pre-trained deep features for anomaly detection," in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 6726–6733, 2020.

[3] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *5th International Conference on Learning Representations (ICLR)*, 2017.

[4] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 1563–1572, 2016.

[5] P. Perera, V. I. Morariu, R. Jain, V. Manjunatha, C. Wigington, V. Ordonez, and V. M. Patel, "Generative-discriminative feature representations for open-set recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[6] R. Yoshihashi, W. Shao, R. Kawakami, S. You, M. Iida, and T. Naemura, "Classification-reconstruction learning for open-set recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4016–4025, 2019.

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

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 770–778, 2016.