POC: Periodical Orthogonal Center Loss For Open Set Classification

Yusheng Pu
College of Intelligence and Computing
Tianjin University
Tianjin, China
pys@tju.edu.cn

Ruonan Liu
College of Intelligence and Computing
Tianjin University
Tianjin, China
ruonan.liu@tju.edu.cn

Qiang Chen
LEM3-UMR 7239 CNRS, Arts et
Me´tiers ParisTech Metz, 4
Rue Augustin Fresnel, France.
qiang.chen@ensam.eu

Dongyue Chen
College of Intelligence and Computing
Tianjin University
Tianjin, China
dyue_chen@163.com

Wenlong Yu
College of Intelligence and Computing
Tianjin University
Tianjin, China
751909848@qq.com

Di Cao
College of Intelligence and Computing
Tianjin University
Tianjin, China
cd1147583042@126.com

Abstract*—When designing classification models, people usually do not assume that there will be unknown classes in the test set, which never appeared in the training set. However, this tricky situation is very common in practical applications. Such test conditions are called Open Set environments. Now, how to make models have the ability to identify unknown classes in the open environment has become a topic of great concern to researchers. In this paper, we follow up on previous research, which focusses on using orthogonal class centers to detect the unknown. We explain the reasons for the poor performance of the previous class center update strategy and propose using the orthogonal loss applied to the class centers to restrict the update direction. In addition, we use the multi-head attention layer for centers' calculation to find suitable projection space adaptively. Experiments show that our method improves the performance of preceding orthogonal center methods.

***Keywords*—Open-set; classification; metric learning**

I. Introduction

With the continuous development of deep learning technology, applications combining artificial intelligence with real-life are constantly emerging, such as Pedestrian Detection and Person Re-Identification [9], [10], and other target detection [11], [12], [13], which requires more reliability. At the same time, more practical issues have been widely discussed, among which the shortcomings of the deep network classification model in an open environment have been widely concerned. The traditional machine learning model framework uses a limited training set to train the model, and tests the classification performance with the same limited data. That is, it assumes that the class of the test set is a subset of the training set [14]. But the reality is that there are likely to be categories in the test set that doesn't appear in the training set. In the real world, it is almost impossible to cover all possible classes, and closed set classifiers are error-prone for samples of unknown classes, which limit their availability [16].

As is shown in Fig. 1-(a), owl represents the unknown class. In traditional classification settings, the existence of unknown classes is not taken into consideration. Any image in the test set is forced into a known class. At the same time, in the field of face recognition, there is a natural problem of open environment recognition, that is, the recognition of unknown faces, especially when dealing with tasks of fine-grained classification.

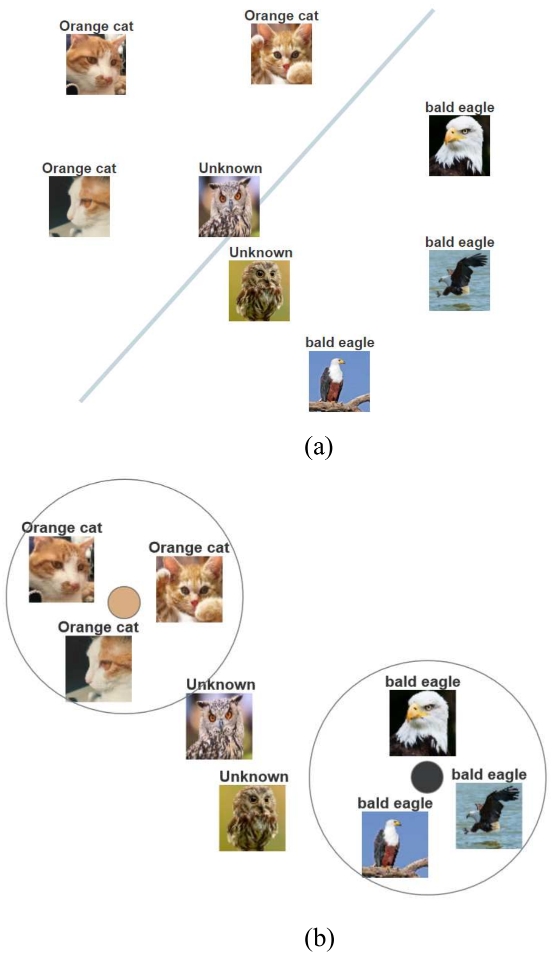


Fig. 1. Open set risk schematic diagram

Predecessors have made fruitful explorations in the field of face to improve the detection performance of unknown faces, such as [17], [18], among which measurement learning has a significant effect. Metric learning can improve the characteristics of the same class proximity and different classes separation [19]. Therefore, it is a good choice to use metric learning to improve the unknown class recognition performance of the classifier.

In general, the cross-entropy used by people only encourages the model to project samples of different classes to different regions, instead of pursuing the separation and compactness between these regions as far as possible. As a result, the entire feature space is likely to be filled with known classes, that is, there is not enough room to reject unknown classes.

On this point, there are many predecessors who have done related research. In [4], the author proposed using an orthogonal fixed class center to guide training, and improved the recognition ability of unknown classes, with good performance. This is shown in Figure 1-b. However, the author does not point out the fundamental reason that why the orthogonality fixed class center is effective in improving the open set recognition performance. Moreover, in order to maintain this coercion effect, the author was forced to set the class center not to be updated with the training. In fact, this leads to uncertainty of setting the length of the center of the class. After our verification, due to the limitation of the feature extraction ability of backbones, the network may not be able to map the features of the corresponding category of the specified fixed point nearby. This leads to the result that the final fixed class center is not the actual class center, and the projection of the corresponding class in the eigenspace does not guarantee the best compactness. In fact, due to the influence of feature extraction ability of backbones, the true boundary is not round, but elliptical. During classification, the distance from the sample to the center of the corresponding class is sent to Softmax to calculate the probability, so in fact, the long-tail sample with the farthest distance will be taken as the circular boundary, and the radius will be the farthest distance from the sample within the class to the center of the class. As a result, there will be a lot of blank space on the other end. This class boundary is not optimal, and it will accommodate more unknown classes into the decision boundary, which makes it impossible to reject the unknown classes accurately.

Based on this, we propose an additional orthogonal loss to restrict the centers' update. It allows updating iteration of class center to find the optimal length of class center adaptively and ensure the compactness. Specifically, we divide it into three stages to set the class center, and use multi-head attention layer to get fruitful features. In the first stage, the initial orthogonal class center is placed at guiding the initial direction of model feature extraction. In the second stage, we start to update the class center to avoid the feature deviating from the class center, and ensure the orthogonality of the update with the loss of the orthogonality class center. In addition, we used the multi-head attention mechanism in the hope of extracting richer information. In the third stage, it is fixed again and enters the stage of searching for the optimal solution of a small step length. We find that fixing the class center again at this time helps to speed up the search for the optimal solution. And then we repeat the second phase again, and then we go to the third phase until we reach the maximum epoch.

Contributions of this paper are as follows:

1. Reveals the fundamental reason why orthonormal class centers are effective in improving the performance of open sets, as well as the shortcomings of existing researches
2. A semantic orthogonality loss for class centers is proposed, which gives fixed orthogonality semantic centers the ability to update under orthogonality constraints.
3. The use of stage alternating center and multi-head attentional layer provides both stable and flexible.

II. RELATED WORKS

A. Open set Recognition

As deep learning models become ever more closely integrated with practical applications, a new problem is discovered, namely the risk problem of the open world. Much research has been done on how to better identify unknown classes. Bnadle[1] analyzed in detail the risks of traditional deep learning models in the open world, and used EVT (Extreme Value Theory) to construct the OpenMax layer and replace the Softmax layer, so as to improve the ability of the model to identify unknown classes. In addition to using EVT to improve authentication performance, Kong also uses a GAN network as a discriminator for unknown classes. He put the long mantissa data of the class into training to generate a dichotomous network to identify the unknown class[2]. And then, Sun proposed an open-set recognition method based on conditional Gaussian distribution in [3]. He applied Gaussian approximation to the features of different layers and used probability ladder structure to extract deep information. Miller tries to improve the open set recognition performance from the point of view of feature space. In [4], he proposed using fixed orthogonal class center to guide feature extraction and achieved good performance. In [5], Oza innovatively uses the autoencoder to train the known classes and uses the reconstruction error to conduct EVT modeling.

B. Metric Learning

Metric learning is widely used in human faces. Wen Y et al. proposed Center Loss[6]. They proved that the distribution of feature space was not compact after using only cross-entropy Loss. Therefore, the distance between samples and the Center of each class was taken as additional Loss to improve the performance of the model. In addition to guiding feature extraction through class centers, F. Schroff goes a step further and uses triples to calculate similarity measurement losses. In [7], he applied Triplet Loss to face recognition, reducing the distance between samples of the same category and increasing the distance between samples of different categories. In addition, there are studies that focus on the inter class margin. CY Wu studied the importance of margin and its adaptive calculation method in [8], which also achieved a good performance improvement.

C. Open set Recognition with Metric Learning

In terms of the combination of metric learning and open set recognition, we affirm the relevant research of Miller in [4]. He starts from the feature mapping level and applies metric learning to improve open set performance (CAC). The orthogonal one-hot class center is set up to guide the feature extraction of different categories. Although this has played a positive effect, there are still many deficiencies and room for improvement. Forcibly fixed class centers can not reflect the real class center, which will lead to an inaccurate division of boundary discrimination. Our work is to carry out further research on the basis of this work.

III. PROPOSED METHODS

We focus on: (1) the meaning of the orthogonality of centers and the difference between fixed and non-fixed centers, (2) the use of the multi-attentional mechanism to obtain multiple differentiated mappings for calculating distances to the center of the class. (3) the reason for performance degradation of the previous non-fixed mode is revealed. Orthogonal loss is introduced to the update of the class center to restrict the update direction. A periodic orthogonal renewal scheme is proposed, which takes into account both stability and flexibility. And Fig. 2 is a schematic diagram of the training stage of our proposed method.

A. Center Loss and Semantic Orthogonal Loss of Centers

Let x be a sample of a known class, and f be a feature extraction network. $f(x)$ represents the dense features of x extracted after backbone, and the dimension number of these features is equal to the number of known classes. We have class centers $C = (c_1, c_2, \dots, c_N)$. The N denotes the number of known classes. Firstly, we initialize the class centers as,

$$Initialize(c_i) = alpha \times e_i, \tag{1.}$$

where e_i is the i-th dimensional unit vector.

As is shown in Fig. 3, we have three losses in total, namely, classification loss, central convergence loss and central orthogonal loss.

We use the cross entropy loss as the classification loss, which is given as follows,

$$\mathcal{L}_{CE} = -\sum_{j=1}^{m}\sum_{i=1}^{n} y_j log \frac{e^{f(x_j)} + b}{\sum_{i=1}^{n} e^{f(x_{ij})+b}}, \tag{2.}$$

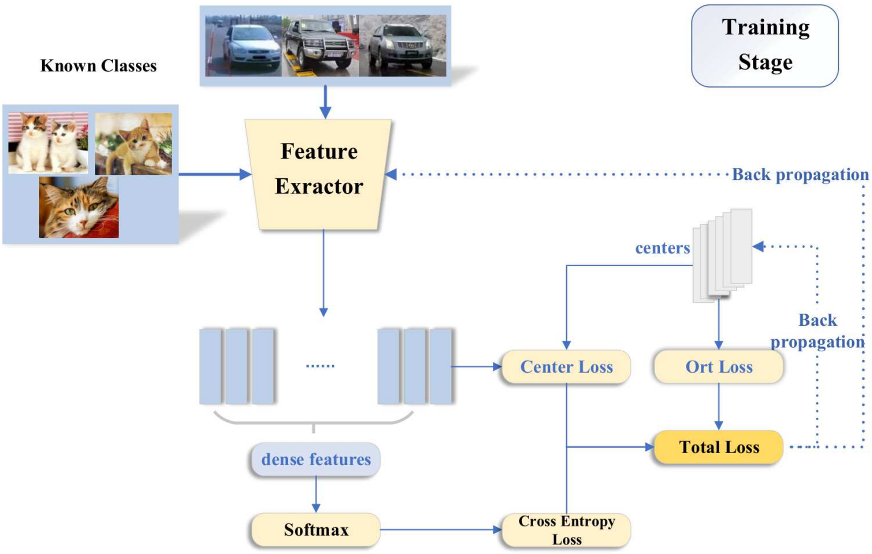


Fig. 2. The model framework of the proposed method

This is cross entropy loss in the form of softmax. In the formula 2, the size of one batch is m and the number of known classes is n. Parameter i denotes the class sample x belonging to, and j means the j-th sample in one batch. Besides, $f(x)$ is the features of sample x, and y is the class one-hot vector, which is subjected to whether x belongs to class i. In addition, b is the offset.

At the same time, we also need to use the class center convergence loss to bring the sample closer to the class center, let $ptc(x_j, i)$ be the flag that whether x belongs to corresponding class. It can be formulated as,

$$ptc(x_j, i) = \begin{cases} True, & x_j\ belongs\ to\ class\ i, \\ False, & x_j\ doesn't\ belong\ to\ class\ i. \end{cases} \tag{3.}$$

Sequentially, the center loss can be calculated as,

$$\mathcal{L}_C = \sum_{j=1}^{m}\sum_{i=1}^{n} \frac{\|x_j - c_i\|_2^2}{2} \times ptc(x_j, i), \tag{4.}$$

Here, the measure of distance is Euclidean.

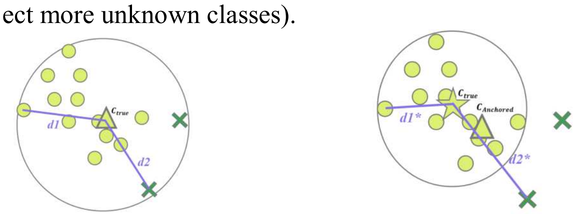
B. The Orthogonal Constraint for updating the Center

The use of semantic center loss is actually to promote the orthogonal effect of the total samples of different categories by making all kinds of samples close to the orthogonal center of each class. The central semantic orthogonal loss can be calculated as follows,

$$pto = \begin{cases} True, & if\ center\ updated, \\ False, & if\ center\ anchored, \end{cases} \tag{5.}$$

$$\mathcal{L}_O = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{c_i \cdot c_j}{|c_i| \times |c_j|} \times pto \times (i! = j). \tag{6.}$$

As we can see from Fig. 3, the center convergence loss causes all samples in the same class to converge near center $\boldsymbol{C_{Anchored}}$, that is, minimize d. Finally, the longest edge of all d (d_1 in the figure) is taken as the radius boundary (the actual discriminant boundary may be smaller than d_1 to reject more unknown classes).



(a) (b)

Fig. 3. .Feature projection diagram of two methods

However, when the class center is fixed, the data distribution will be offset. The fixed class center is not the real class center. As is shown in Fig. 3-(a), the distance, which is between the unknown class sample and the class center is smaller, that is, $d_2 < d_1$ .This can lead to performance degradation.

In Fig. 3-(b), this problem can be avoided by using an update class center strategy with semantic orthogonal loss constraints. As is shown in the figure, the distance from the center of the actual class $\boldsymbol{C_{true}}$ to the boundary point is d1*<d2<d2*, the unknown class falls outside the discriminant boundary.

To sum up, the total loss is,

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{CE} + \alpha_c \times \mathcal{L}_C + \alpha_o \times \mathcal{L}_O \\ &= -\sum_{j=1}^{m}\sum_{i=1}^{n} y_i log \frac{e^{f(x_j)} + b}{\sum_{i=1}^{n} e^{f(x_{ij})+b}} \\ &+\alpha_c \times \sum_{j=1}^{m}\sum_{i=1}^{n} \frac{\left\|x_j - c_i\right\|_2^2}{2} \times ptc(x_j, i) \\ &+\alpha_o \times \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{c_i \cdot c_j}{|c_i| \times |c_j|} \times pt \times (i! = j)\ . \end{aligned} \quad (7.)$$

C. The use of multi-head attention

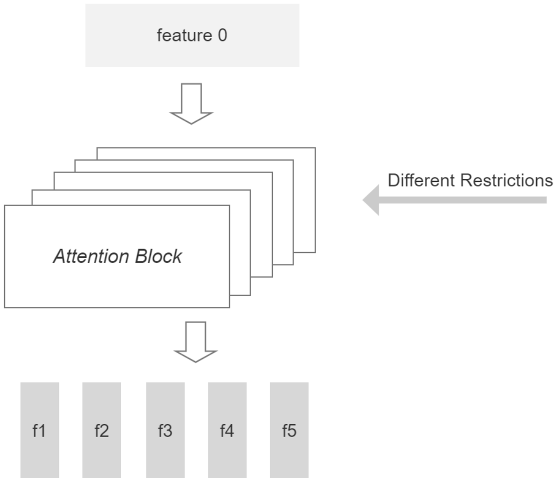


Fig. 4. Get richer representation of features

Our core approach is to bring a given class close to its center in some feature space. And the class is divided by the distance from the class center to the sample. Sometimes, however, this feature space is not always ideal. For example, the feature space obtained by the model may be the best space for dividing certain known classes, or it may be only a selection of deep features. The distinguishing features of the picture are not necessarily the same size. Some categories may be more important in terms of their overall features, while others may be more important in terms of their local features. Therefore, we believe that the feature mapping method can be adaptively selected according to the image features through the self-attention mechanism. And by limiting the differences in the parameters of multiple attention, the model is allowed to learn relevant information in different representation subspaces[27]. Finally, each feature calculates the distance to the centers.

IV. EXPERIMENTS

We test on common image datasets, on the one hand compared with some current open-set recognition algorithms, on the other hand compared with fixed class center, unconstrained center update, and orthogonal constrained center update performance.

A. Metrics

ACC (Accuracy): It describes the total recognition Accuracy of the model for positive and negative classes. This index is mainly used to evaluate the classification performance of the model in the closed set case. ACC has a process to the reject threshold in cases where unknown classes have to be identified. Therefore, it is not sufficient to use it alone to evaluate open set performance.

AUROC (Area Under the Receiver Operating Characteristic): We use the area under the ROC curve (AUROC)[21]. AUROC is a very suitable metric for quantifying dichotomy performance. It simplifies the comparison between methods, eliminates the adjustment of the threshold value, and can reflect the recognition rate of known class and unknown class comprehensively.

B. Dataset Description

MNIST: It is a very classic and simple data set, used to recognize handwritten numbers. Among them, an image is 28 * 28 pixels, the color is a single channel, and there are a total of 60,000 training samples and 10,000 test samples [22]. Six known classes and four unknown classes were randomly selected.

SVHN: This data set is derived from Google Street View door number. The numbers are cropped to 32x32 size. The training set contained 73,257 numbers and the test set 26,032 numbers. Six known classes and four unknown classes were randomly selected[23].

CIFAR10: A small data set for identifying universal objects. There are 10 categories of RGB color images. The size of the image is 32×32, with a total of 50,000 training samples and 10,000 test samples[24]. Six known classes and four unknown classes were randomly selected.

CIFAR+10/+50: Set 4 non-animal classes of CIFAR10 as known, and 10 or 50 animal classes randomly selected from CIFAR100[24].

C. Methods of Comparison

Softmax: Only use cross-entropy as a classification loss. For specific formulas, refer to Formula 2.

OpenMax: The Weibull distribution is used to fit the long mantissa data, which locates on the class boundary, to calculate the probability belonging to the unknown class. We used the experimental data in the original paper[1].

G-OpenMax: Based on the decision boundary established by the method of OpenMax, this method uses generative adversarial networks (GANs) to compound novel category images[20].

OSRCI: A counterfactual image enhancement method is used to generate samples that do not belong to any known class using GAN network[26].

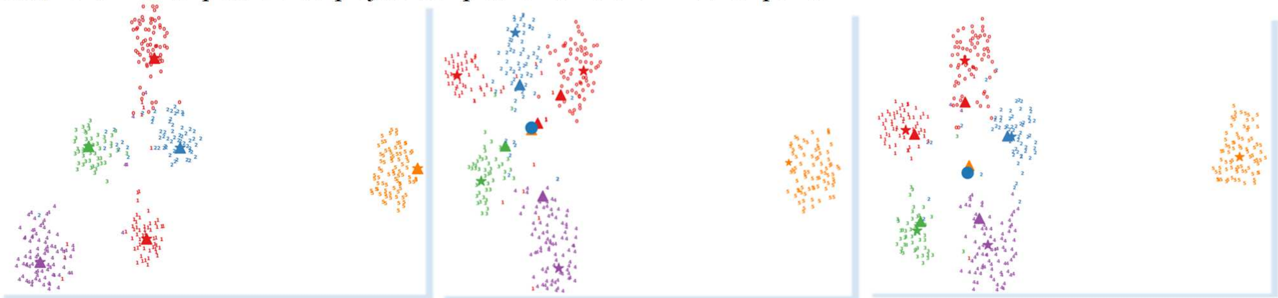
CAC: The method uses a fixed orthogonality class center to guide known classes to be projected into a compact feature space, which is apart from each other[4].

Besides, we use a modified ResNet50[25] to be our backbone. We used Adam as the optimizer, and the learning rate was set to decrease linearly from 3e-4.

D. Experiment Analysis

As we can see in TABLE I. and TABLE II. , our approach performs better in most cases than the fixed class center strategy (CAC). It should be pointed out that the experimental data are extracted from the corresponding papers [1][4]. Although the index improvement is limited, and the actual performance certainly cannot exceed the current SOTA algorithms such as C2AE (not posted in the table), the significance of our work is to prove that by limiting the direction of class center updates, we can not only not lose performance, but also improve performance appropriately. Also, as a loss term, our methods can be used in other methods to improve their performance. We believe that this will provide feasible ideas for the researchers in future research.

As can be seen from the figure, if a fixed class center is used, it is indeed possible to project samples of different categories into a relatively compact cluster area, but as is shown in Fig. 5. As can be seen from the samples of class 0 and class 5 cannot well surround their fixed class centers in the feature space. This is because the extraction ability of backbones and the difficulty of feature extraction of all categories in the whole data set will jointly affect the projection position of the category in the feature space. However, the fixed semantic orthogonality of the class center will lead to the failure of the sample to reach the expected center point.



(a) Anchored (b) Updated without constraint (c) Updated with constraint

Fig. 5. The data distribution of three strategies using t-SNE[15]

TABLE I. AUROC OF METHODS

Methods	**MNIST**	**SVHN**	**CIFAR 10**	**CIFAR +10/+50**
Softmax	0.978	0.886	0.677	0.816/0.805
OSRCI	**0.988**	0.910	0.699	0.838/0.827
G-OpenMax	0.984	0.896	0.675	0.827/0.819
CAC	0.987	0.942	0.803	0.863/0.872
PSOC(ours)	**0.988**	**0.946**	**0.804**	**0.866/0.879**

TABLE II. ACC OF DIFFERENT CENTERS UPDATE STRATEGY

Datasets	**Softmax**	**Anchored**	**Updated without constraint**	**Updated with constraint**
MNIST	0.997	**0.998**	0.997	**0.998**
SVHN	0.97	0.972	0.969	**0.974**
CIFAR10	0.939	0.943	0.936	**0.945**
CIFAR+10	0.948	0.951	0.946	**0.954**
CIFAR+50	0.949	0.953	0.952	**0.958**

Such an effect may not greatly affect the classification accuracy, but it will be detrimental to the rejection of unknown classes. As is shown in 0, the unconstrained update of the center and the fixed center strategy are not as good as the strategy we used, although the use of central orthogonal initialization gives the model better performance than softmax cross entropy alone.

The open set recognition method using the distance measurement criterion actually calculates the classification probability by using the distance from the sample to the respective class center in the corresponding feature space. However, if the class center used for calculation does not reflect the real class center, it will inevitably lead to the distance between the party farther from the class center and the center (As is shown in Fig. 3), greater than the distance from the unknown class to the class center that is closer to the class center. This also confuses known classes with unknown classes. In [4], we have discussed the effect of updating the class center on performance and concluded that fixing the class center (△) is more effective. However, this is because the author does not restrict updates to the class center. During the iteration, the class center may move in any direction in order to minimize center loss 4. As is shown in Fig. 5-(b), although the class center (☆) at this time can reach the true center position of the class, the feature extractor does not project samples of different classes to the orthogonal region in the end. The orthogonality class center works because of the artificial addition of orthogonality to the center guidance, but if the class center is updated without the addition of directional constraints, the orthogonality effect will be lost, and the result will not be so good. But in practice, if the class center is updated with a directional constraint, you can reach the true class center and still maintain the orthogonality effect.

V. CONCLUTION

Based on the fixed orthogonality class center, we illustrate the practical meaning of central orthogonality and demonstrate the necessity of updating the class center. Moreover, we use a strategy of periodic updates, alternating between updates and fixed tightening. In the update phase, orthogonal loss of the class center is used to constrain the updating direction of the class center. We demonstrate the excellent performance of the proposed method for experiments on multiple data sets. We are also noted that the transformation of the class center to measure the loss is sufficient to change the distribution of the entire feature space. In practice, this means that the traditional triple loss can be replaced by an improved central loss, which can result in significant savings in computational costs. In addition, in the model of feature pyramid structure, the feature orthogonalization of different layers may also have superb results. We will continue to study this direction in the future.

ACKNOWLEDGMENT

I would like to express my gratitude to the teachers and students in the lab for discussing relevant knowledge with me, and to the Department of Intelligent Computing of Tianjin University for providing experimental equipment for the researchers. Wish we continue to devote ourselves to research and contribute to solving practical problems.

REFERENCES

[1] Bendale and T. E. Boult, "Towards Open Set Deep Networks," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1563-1572, doi: 10.1109/CVPR.2016.173.

[2] Kong S , Ramanan D . OpenGAN: Open-Set Recognition via Open Data Generation[J]. 2021.

[3] X. Sun, Z. Yang, C. Zhang, K. -V. Ling and G. Peng, "Conditional Gaussian Distribution Learning for Open Set Recognition," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 13477-13486, doi: 10.1109/CVPR42600.2020.01349.

[4] D. Miller, N. Sünderhauf, M. Milford and F. Dayoub, "Class Anchor Clustering: A Loss for Distance-based Open Set Recognition," 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 2021, pp. 3569-3577, doi: 10.1109/WACV48630.2021.00361.

[5] P. Oza and V. M. Patel, "C2AE: Class Conditioned Auto-Encoder for Open-Set Recognition," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 2302-2311, doi: 10.1109/CVPR.2019.00241.

[6] Wen Y , Zhang K , Li Z , et al. A Discriminative Feature Learning Approach for Deep Face Recognition[C]// European Conference on Computer Vision. Springer, Cham, 2016.

[7] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.

[8] Wu C Y , Manmatha R , Smola A J , et al. Sampling Matters in Deep Embedding Learning[C]// 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017.

[9] Cheng Chi, Shifeng Zhang, Junliang Xing, Zhen Lei, Stan Z Li, and Xudong Zou. Pedhunter: Occlusion robust pedestrian detector in crowded scenes. arXiv preprint arXiv:1909.06826, 2019..

[10] Songtao Liu, Di Huang, and Yunhong Wang. Adaptive nms: Refining pedestrian detection in a crowd. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019..

[11] S. Beery, G. Wu, V. Rathod, R. Votel and J. Huang, "Context R-CNN: Long Term Temporal Context for Per-Camera Object Detection," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 13072-13082, doi: 10.1109/CVPR42600.2020.01309.

[12] Y. Liu and X. Tang, "BFBox: Searching Face-Appropriate Backbone and Feature Pyramid Network for Face Detector," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 13565-13574, doi: 10.1109/CVPR42600.2020.01358.

[13] P. Hu, J. Ziglar, D. Held and D. Ramanan, "What You See is What You Get: Exploiting Visibility for 3D Object Detection," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10998-11006, doi: 10.1109/ CVPR42600.2020.01101.

[14] C. Geng, S. -J. Huang and S. Chen, "Recent Advances in Open Set Recognition: A Survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2020.2981604.

[15] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," Journal of machine learning research, vol. 9, no. Nov, pp. 2579–2605, 2008.

[16] R. Yoshihashi, W. Shao, R. Kawakami, S. You, M. Iida and T. Naemura, "Classification-Reconstruction Learning for Open-Set Recognition," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4011-4020, doi: 10.1109/CVPR.2019.00414.

[17] H. Yu et al., "Unknown Identity Rejection Loss: Utilizing Unlabeled Data for Face Recognition," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 2662-2669, doi: 10.1109/ICCVW.2019.00325.

[18] R. H. Vareto and W. R. Schwartz, "Unconstrained Face Identification using Ensembles trained on Clustered Data," 2020 IEEE International Joint Conference on Biometrics (IJCB), 2020, pp. 1-8, doi: 10.1109/IJCB48548.2020.9304882.

[19] B. Chen and W. Deng, "Hybrid-Attention Based Decoupled Metric Learning for Zero-Shot Image Retrieval," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 2745-2754, doi: 10.1109/ CVPR.2019.00286.

[20] Sergey Demyanov Zongyuan Ge and Rahil Garnavi. Generative openmax for multi-class open set classification. In Proceedings of the British Machine Vision Conference (BMVC), pages 42.1–42.12, 2017.

[21] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In ICML, 2006. 4

[22] Yann LeCun, Corinna Cortes, and CJ Burges. Mnist handwritten digit database. 2010.

[23] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. In NIPS workshop on deep learning and unsupervised feature learning， volume 2011, page 5, 2011.

[24] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778

[26] 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.

[27] Vaswani A , Shazeer N , Parmar N , et al. Attention Is All You Need[J]. arXiv, 2017.