

Article

Multi-Class Hypersphere Anomaly Detection Based on Edge Outlier Exposure Set and Margin

Min Gao¹, Xuan Liu¹, Di Xu¹ and Guowei Yang^{2,*}

¹ School of Computer Science (School of Intelligent Auditing), Nanjing Audit University, Nanjing 211815, China; ma2209104@stu.nau.edu.cn (M.G.); mp2209145@stu.nau.edu.cn (X.L.); mp2209136@stu.nau.edu.cn (D.X.)

² School of Electronic Information, Qingdao University, Qingdao 266071, China

* Correspondence: 270178@nau.edu.cn

Abstract: Currently, the decision boundary of the multi-class anomaly detection algorithm based on deep learning does not sufficiently capture the positive class region, posing a risk of abnormal sample features falling into the domain of normal sample features and potentially leading to misleading outcomes in practical applications. In response to the above problems, this paper proposes a new method called multi-class hypersphere anomaly detection (MMHAD) based on the edge outlier exposure set and margin. The method aims to utilize convolutional neural networks for joint training of all normal object classes, identifying a shared set of outlier exposures, learning compact identification features, and setting appropriate edge parameters to guide the model in mapping outliers outside the hypersphere. This approach enables more comprehensive detection of various types of exceptions. The experiments demonstrate that the algorithm is superior to the most advanced baseline method, with an improvement of 26.0%, 8.2%, and 20.1% on CIFAR-10 and 14.8%, 12.0%, and 20.1% on FMNIST in the cases of (2/8), (5/5), and (9,1), respectively. Furthermore, we investigate the challenging (2/18) case on CIFAR-100, where our method achieves approximately 17.4% AUROC gain. Lastly, for a recycling waste dataset with the (4/1) case, our MMHAD yields a notable 22% enhancement in performance. Experimental results show the effectiveness of the proposed model in multi-classification anomaly detection.

Keywords: deep learning; multi-class anomaly detection; outlier exposure; convolutional neural network

MSC: 68T10



Citation: Gao, M.; Liu, X.; Xu, D.; Yang, G. Multi-Class Hypersphere Anomaly Detection Based on Edge Outlier Exposure Set and Margin.

Mathematics **2024**, *12*, 2340.

<https://doi.org/10.3390/math12152340>

Received: 19 June 2024

Revised: 19 July 2024

Accepted: 25 July 2024

Published: 26 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Anomaly detection (AD), i.e., the task of identifying irregular or unexpected patterns in data, is a complex challenge in many applications, including autonomous driving [1], safety monitoring [2], and industrial detection [3]. In an open-world environment, new class distributions may emerge during testing, necessitating the development of classifiers capable of detecting new class sample examples while maintaining high classification accuracy for known class distributions. Given the wide range of anomalies, a common approach involves modeling the distribution of normal samples and subsequently identifying abnormal samples through outlier detection [4,5]. However, to date, there is no practical AD framework directly applicable to real-world image data [6–8]. Real-world datasets often exhibit high dimensionality, noise, and heterogeneity, making it challenging to accurately identify anomalies across multiple classes. Multi-class AD remains one of the most intricate challenges within this domain. In multi-class AD, the distribution of data points across classes is often highly imbalanced, with normal instances far outnumbering anomalous ones and some anomalous classes being significantly rarer than others. Current multi-class exception detection techniques often rely on individual models tailored to distinct object classes [9,10]. Nevertheless, as the diversity of classes expands, this approach becomes

increasingly resource-demanding and may significantly hinder computational efficiency. The primary reasons for this limitation stem from the inherent flaws in the design of these detectors. Firstly, the decision surface fails to align closely with the boundary of the positive class region, leading to suboptimal detection performance, especially on datasets with extensive variance. Secondly, the absence of an effective mechanism to discern samples belonging to different classes often results in false-positive predictions [11–13], indicating their inability to focus on the specific normal features that differentiate each class from the rest.

Obtaining labeled data for anomalous classes can be expensive and time-consuming, especially when dealing with rare or emerging anomalies. Unsupervised or semi-supervised learning techniques are often employed, but they struggle to accurately differentiate between multiple anomalous classes. Hence, there lies a significant risk of abnormal sample features being mistakenly classified as normal, particularly when the initial spatial configurations of abnormal samples overlap with those of normal ones. Addressing this urgent issue, a plethora of solutions have been devised and will be thoroughly examined in the upcoming chapter. Notably, research by Hendricks et al. [14] underscores the substantial enhancement in AD performance that can be achieved through the utilization of outlier exposure (OE). To mitigate this hazard, it is indispensable to consider incorporating an OE set, which effectively constrains outliers and modulates the boundary of the positive feature space.

To address these issues, this paper proposes a novel deep multi-class hypersphere AD algorithm (MMHAD) based on an edge OE set. It constructs a compact sphere that encompasses the multi-class feature region and a parallel sphere that extends the outlier exposure region outwards, maximizing the distance between the two parallel spheres. The algorithm learns compact and discriminative features that encompass multiple normal object categories embedded across multiple datasets. It minimizes the closed hypersphere that encompasses multiple positive sample feature regions and controls the optimized outlier exposure set features that lie outside this closed hypersphere. The framework diagram of the algorithm is illustrated in Figure 1. The main contributions of this paper are as follows:

- We design an innovative multi-class AD method, the multi-class hypersphere algorithm grounded in edge OE sets and margin optimization. In the proposed model, we use neural networks to learn the distribution changes of normal class and edge OE sets, alleviate the problem of inter-class boundary overlap, and greatly improve the performance of the classifier on typical databases.
- In addition, we use data enhancement to constrain outliers or control the boundaries of positive feature regions, and we summarize a set of innovative steps to construct and optimize OE sets in open set identification or AD, reduce the risk of existing AD methods mapping abnormal sample features into normal sample feature domains, and improve the generalization ability of anomaly detectors.
- By setting margin parameters to create a clear boundary between representations of different categories, the model is motivated to map outliers or potential data points that do not meet expectations far enough from the center of the category to improve classification performance.
- We propose new extreme values for decision making. The decision threshold of most other AD methods is artificially set according to experience.

The remaining sections of this paper are structured as follows: Section 2 outlines the integration of common AD methods with data enhancement algorithms. In Section 3, a comprehensive explanation of the proposed model, MMHAD, is provided. To validate the efficiency and stability of the proposed algorithm, simulation tests on various datasets are conducted in Section 4. In Section 5, the limitations of the algorithm are analyzed. Finally, Section 6 presents the conclusions drawn from this study.

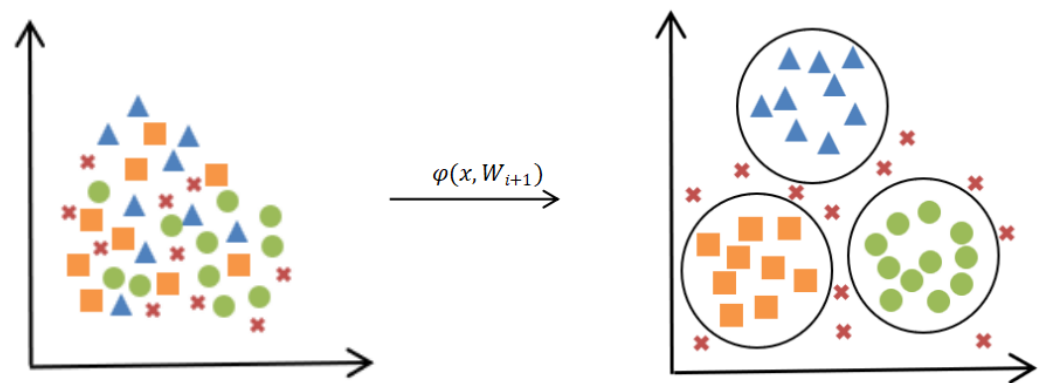


Figure 1. The block diagram depicts a multi-class AD frame diagram, where green, blue, and orange respectively represent normal samples of different categories, while red represents abnormal samples. The proposed model learns a neural network transformation $\phi(\cdot; W)$ from input space \mathcal{X} to output space \mathcal{F} , aiming to minimize the closed hypersphere containing the feature region of the positive sample while controlling the optimized anomalous exposure set feature outside the closed hypersphere.

2. Related Work

2.1. AD

Previous methods for data description primarily relied on kernel-based approaches, transforming data into high-dimensional feature spaces to capture intricate patterns and nonlinear relationships [15–17]. However, when confronted with data stemming from multiple distributions, each occupying a distinct niche in the feature space, a solitary hyperplane or hypersphere proves inadequate for comprehensively describing all distributions. The representation learning ability of deep neural networks (DNNs) surpasses that of kernel functions, making them a preferred choice for addressing the challenges posed by large-scale and complex data distributions. Ruff et al. [18,19] extended the classic support vector data description (SVDD) paradigm to deep SVDD, harnessing DNNs instead of kernel methods to efficiently encapsulate the normality of high-dimensional data and learn a latent representation that maps the normal class into a minimum-volume hypersphere. Nevertheless, this advancement is not without its limitations [20,21]. It fails to ascertain whether the derived features retain the crucial information of the data's inherent structure, thereby rendering it less suitable for intricate and dynamic datasets.

To address the challenges of processing large-scale datasets, Ghafoori et al. [22] introduced deep multi-sphere SVDD (DMSVDD). Similarly, Goyal et al. [23] proposed deep robust one-class classification (DROCC), leveraging a projection ascent technique centered on the normal class to identify the most illustrative outliers and bolster the model's robustness. Moreover, the essence of most AD techniques [24–28] lies in training individual models tailored to distinct classes of objects. You et al. [29] introduced the unified model for multi-class AD (UniAD), which leverages multiple one-class anomaly detectors jointly to tackle the challenge of multi-class AD. However, as the number of classes proliferates, relying on individual models for each object class becomes prohibitively resource-intensive, placing a significant strain on computational resources and impeding direct applicability to intricate real-world image scenarios. To alleviate this issue, Singh et al. [30] introduced deep multi-class AD (Deep MAD), an innovative method that learns concise and discriminative features across multiple normal object classes, thereby augmenting the efficacy of multi-class AD while mitigating computational overheads.

The above algorithm alternately trains the neural network model and the traditional classification model so that the parameters of the neural network are constantly updated in the direction of improving the classification performance of the traditional classification model, which gives full play to the representation learning ability of DNNs. Nonetheless, these models, by solely focusing on training normal sample models, risk inadvertently

admitting abnormal sample features into the normal sample feature space, which can adversely affect classification accuracy and model performance. Our approach distinguishes itself through its adept utilization of neural networks for intricate feature extraction, its inventive strategy for assessing and refining model limitations via class-specific acceptance regions, and its versatility in addressing a wide array of open-image datasets.

2.2. Integrated with Data Augmentation

The cornerstone of generative AD technology revolves around crafting abnormal samples from target data and integrating them into classifier training to bolster the precision of detecting anomalous class samples. However, the validity of the generated abnormal samples remains a significant concern. A commonly adopted approach for learning the distribution of normal class relies on image (or feature) reconstruction [31–35]. This method assumes that a well-trained model consistently produces normal samples, disregarding any imperfections in the input, which frequently results in a significant reconstruction error for abnormal samples, thus posing challenges in distinguishing them from normal ones.

OE performs admirably when employed as an auxiliary abnormal dataset for training anomaly detectors with large-scale photos, garnering significant attention from both academic and industrial researchers due to its broad applicability. Currently, numerous scholars have delved into this realm of research. Liznerski et al. [36] introduced the fully convolutional data description (FCDD) network, leveraging random samples from an auxiliary out-of-distribution (OOD) dataset to detect abnormal images and extract additional semantic insights. However, the performance of FCDD is intimately tied to the diversity and quality of its training data, with biases or shortcomings in the data potentially undermining the model's generalization capabilities. Expanding on this, Ruff et al. [37] examined the influence of OE data diversity on detection outcomes, emphasizing that OE techniques' exceptional performance often hinges on their ability to access a vast and comprehensive collection of auxiliary anomalous data. Kirchheim et al. [38] proposed multi-class hypersphere AD (MCHAD), extending existing hypersphere learning methodologies to accommodate the inclusion of example anomalies in the training process. Yet, MCHAD's intricate structure necessitates meticulous parameter tuning. Cevikalp et al. [39] introduced the deep compact hypersphere (DCHS), utilizing random OE dataset samples as anomalous data to demonstrate the enhanced robustness of their algorithm. Nonetheless, DCHS's training and inference processes demand substantial computational resources and time. The selection of an exposure set presents a formidable challenge, as the appropriateness of the exposure can significantly impact the accuracy of the AD algorithm. If the OE is not representative, the AD algorithm may produce misleading results in practical applications. If the distribution of the selected samples is similar to that of the normal class, it is easy for the network to fail to train, or if the distribution characteristics of the normal samples are not considered, it is easy to overfit, and the detection effect cannot be guaranteed.

3. Proposed Model

This section presents a comprehensive introduction to the proposed MMHAD model, which aims to accurately and effectively perform multi-class AD on hyperspheres. First, we provide a concise overview of DMSVDD, followed by an extensive description of the proposed detectors and their theoretical properties.

3.1. DMSVDD

Deep-SVDD is not able to classify input samples into multiple classes given in training data because it only focuses on detecting novel samples (or outliers) and does not utilize class labels at all, so it cannot be directly applied to real, serious image scenes. Similar to DSVDD, DMSVDD generates useful and discriminative features by embedding the normal class with a multi-modal distribution into multiple data-enclosing hyperspheres with minimum volume. Hence, the feature representations of the normal class are distributed inside the hypersphere, while those of the anomalous data are distributed outside. Different

from DSVDD, which uses a single hypersphere, DMSVDD allows multiple hyperspheres to adapt to complex data distributions.

Let $\phi(\cdot; \mathcal{W})$ be a DNN with the set of network parameters $W = \{W_1, \dots, W_L\}$ and the spherical boundary for class K be specified by its center $C = \{c_1, \dots, c_K\}$ and radius $R = \{R_1, \dots, R_K\}$. Given N training samples from K different classes, the loss function expression is as follows:

$$\min_{W,R,c} \frac{1}{K} \sum_{k=1}^K R_k^2 + \frac{1}{vn} \sum_{i=1}^n \max(0, \|\phi(x_i; \mathcal{W}) - c_j\|^2 - R_j^2) + \frac{\lambda}{2} \sum_{l=1}^L \|W^l\|_F^2 \quad (1)$$

where c_j is the cluster center assigned to $\phi(x_j; W)$ using NPR, $R_j > 0$ is its corresponding radius, $\|\cdot\|_F^2$ represents the Frobenius norm, $v \in (0, 1]$ represents the weight parameter that controls the proportion of outliers, and λ represents the weight parameter of the regularization term. The first term minimizes the volume of hyperspheres, while the second term penalizes for mapping points outside of the hyperspheres and is controlled by parameter v . This parameter allows a fraction of points to be mapped outside the hyperspheres to compensate for noisy instances or unknown anomalies in the training data X .

3.2. Our Model

The increase in the number of classes has led to the use of separate models for different object classes to solve multi-class AD tasks, which has become resource-intensive and significantly impacts computing power. Therefore, our proposed method aims to identify the common OE set of N normal classes, recognize shared anomaly characteristics among different normal classes, and enhance computing efficiency and resource utilization. The model simplified structure of MMHAD is illustrated in Figure 2.

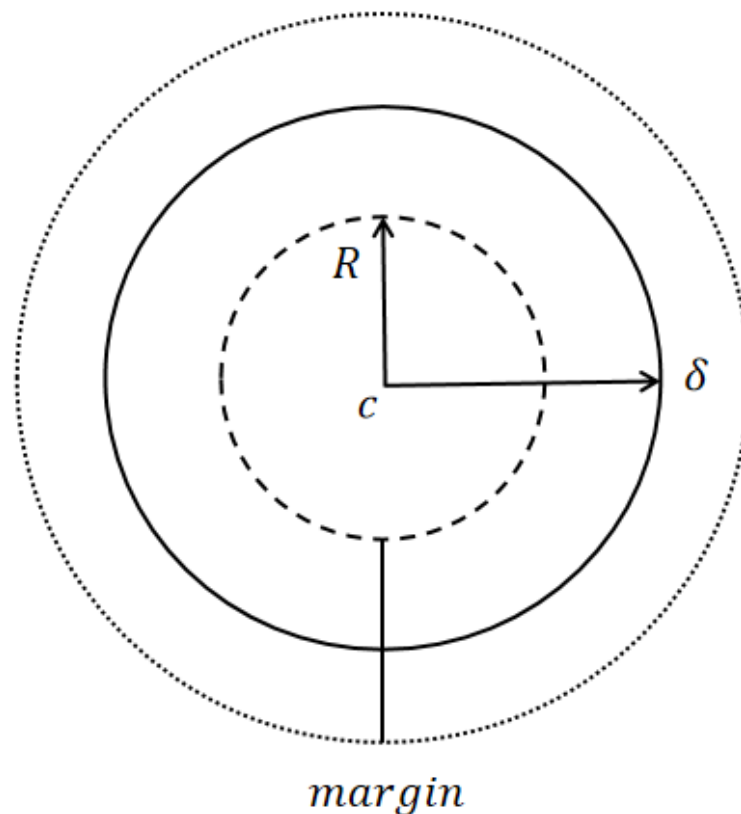


Figure 2. Single hypersphere simplified structure of class y .

Given training data $X = \{X_1 \cup X_2 \cup \dots \cup X_k\}$, we first find the appropriate common exception exposure set $D_{out}^{OE} = \{x_q \mid q = 1, \dots, N_{OE}\}$ for K class, such that $U = \{1, 2, \dots, k, \wedge\}$ represents the normal classes K (1 to k) and a set of all exceptions labeled as \wedge . A general classifier assigns a probability to the probability $P(u|x)$ that an image x belongs to a set of u classes, where $U \in 2^u$.

$$P(\wedge|x) = \frac{1}{N} \sum_{\neg\{1\} \cap \neg\{2\} \cap \dots \cap \neg\{k\} = \wedge} \prod_{i=1}^k P_i(\neg i|x) \tag{2}$$

which simplifies to

$$P(\wedge|x) = \frac{1}{N} \prod_{i=1}^k P_i(\neg i|x) \tag{3}$$

$$K = 1 - \sum_{u_1 \cap u_2 \cap \dots \cap u_k} \prod_{i=1}^k P_i(u_i|x) \tag{4}$$

The training data of all k classes are combined and treated as a class, and our objective function is as follows:

$$\begin{aligned} \min_{W,r,c} & \frac{1}{K} \sum_{k=1}^K R_j^2 + \frac{1}{N} \sum_{k=1}^N \left(\frac{1}{N-1} \sum_{k \neq j, 1 \leq k \leq N} \left(\frac{1}{\text{vol}_k} \max\{0, \|\phi(x_i^k; \mathcal{W}) - c_j\|^2 - R_j^2\} \right) \right) \\ & + \frac{1}{N} \sum_{j=1}^N \left(\frac{1}{\mu N_{OE}} \sum_{q=1}^{N_{OE}} \max\{0, (R_j + m_j)^2 - \|\phi(x_q; \mathcal{W}) - c_j\|^2\} \right) + \frac{\lambda}{2} \sum_{l=1}^L \|W_l\|_F^2 \\ \text{s.t.} & \|\phi(x_i^k; \mathcal{W}) - c_j\|^2 \leq R_j^2 \\ \text{s.t.} & \|\phi(x_q; \mathcal{W}) - c_j\|^2 \geq R_j + m_j^2 \end{aligned} \tag{5}$$

where $N = |X_1| + |X_2| + \dots + |X_k|$ is the total number of normal class samples and is the number of normal class i samples. $\frac{1}{N} R_j^2$ is all j hypersphere volumes, the third term tries to find a common set of anomalous exposures for N classes, the second term expects to maximize the distance between normal classes, the exposure set radius is $R_j + m_j$, and in the inference process for class j , the point of distance is rejected.

Furthermore, the algorithm uses the nearest neighbor algorithm, where $j = \arg \min \|\phi(x; \mathcal{W}) - c_j\|^2$ to assign the anomaly scores. Note that our proposed model finds the optimal decision threshold if it is greater than the normal class, otherwise it is the abnormal class. Any sample that falls outside the boundary hypersphere of this estimate is considered an anomaly.

$$s(x) = R_j^2 - \|\phi(x_i^k; \mathcal{W}) - c_j\|^2 \tag{6}$$

For the backpropagation algorithm used in objective function optimization, the deep network architecture can be set as a feed-forward neural network. Given a set of normal samples (x_i, y_+) for $1 \leq i \leq N$ and a set of negative samples (x_j, y_-) for $1 \leq j \leq N_{OE}$ as M , the output target feed-forward neural network uses $\phi(x; W)$. The optimization formula can be expressed as follows:

$$J(W, R, c) = J_1(W, R, c) + J_2(W, R, c) + J_3(W, R, c) \tag{7}$$

We define J_1, J_2, J_3 as the follow formula:

$$J_1(W, R, c) = \frac{1}{N} \sum_{j=1}^N R_j^2 + \frac{\lambda}{2} \sum_{l=1}^L \|W^l\|_F^2 \tag{8}$$

$$J_2(W, R, c) = \frac{1}{N} \sum_{k=1}^N \frac{1}{N-1} \sum_{k \neq j, 1 \leq k \leq N} \frac{1}{vn_k} \max \left\{ 0, \left\| \phi(x_i^k; \mathcal{W}) - c_j \right\|^2 - R_j^2 \right\} \quad (9)$$

$$J_3(W, R, c) = \frac{1}{N} \sum_{j=1}^N \frac{1}{\mu N_{OE}} \sum_{q=1}^{N_{OE}} \max \left\{ 0, (R_j + m_j)^2 - \left\| \phi(x_q; \mathcal{W}) - c_j \right\|^2 \right\} \quad (10)$$

If we set $L_{in}(W, R, c) = \left\| \phi(x_i; w) \right\|^2 - R^2$, we get

$$J_2(W, R, c) = \frac{1}{vn_j} \sum_{i=1}^N L_{in}(W, R, c, x_{n_i}, y_+) \quad (11)$$

$$J_3(W, R, c) = -\frac{1}{\mu m_q} \sum_{i=1}^{N_{OE}} L_{in}(W, R, c, x_{m_q}, y_-) \quad (12)$$

where n_1, n_2, \dots, n_j stands for N positive sample indices for which the function $L_{in}(W, R, c, x, y_+) > 0$, and m_1, m_2, \dots, m'_q is the index of N'_{OE} negative samples for which the function $L_{out}(W, R, c, x, y_-) < 0$. To minimize the objective function, we employ the gradient descent method and update specific parameters as follows:

$$W^l = W^l - \alpha \frac{\partial J(W, R, c)}{\partial W^l} = W^l - \alpha \left(\frac{1}{vn_k} \sum_{i=1}^{n_k} \frac{\partial L_{in}(W, R, c, x_{n_i}, y_+)}{\partial W^l} + \frac{1}{\mu N_{OE}} \sum_{q=1}^{N_{OE}} \frac{\partial L_{out}(W, R, c, x_{m_q}, y_-)}{\partial W^l} \right). \quad (13)$$

$$R^l = R^l - \alpha \frac{\partial J(W, R, c)}{\partial R^l} = R^l - \alpha \left(2R + \frac{1}{vn_k} \sum_{i=1}^{n_k} \frac{\partial L_{in}(W, R, c, x_{n_i}, y_+)}{\partial R^l} + \frac{1}{\mu N_{OE}} \sum_{q=1}^{N_{OE}} \frac{\partial L_{out}(W, R, c, x_{m_q}, y_-)}{\partial R^l} \right). \quad (14)$$

$$c^l = c^l - \alpha \frac{\partial J(W, R, c)}{\partial c^l} = c^l - \alpha \left(\frac{1}{vn_k} \sum_{i=1}^{n_k} \frac{\partial L_{in}(W, R, c, x_{n_i}, y_+)}{\partial c^l} - \frac{1}{\mu N_{OE}} \sum_{q=1}^{N_{OE}} \frac{\partial L_{out}(W, R, c, x_{m_q}, y_-)}{\partial c^l} \right). \quad (15)$$

During the model training procedure, we used the validation and repeated iteration approach. We observed that setting $m_q = \sqrt{d}$ produces favorable outcomes in our experiments, where d represents the dimensionality of the output space. Specifically, lower margin values result in a more compact hypersphere and a more sensitive model to anomalous samples, but they may also increase the false-positive rates. Conversely, a higher margin value may decrease the false-positive rate but at the expense of some reduced sensitivity to anomalous samples. Consequently, introducing a margin of m_j around each hypersphere with radius R_j serves to encourage the model to map anomalies to a distance $\geq \sqrt{d}$ from all class centers; details of the experiment are given in the next section.

Theorem 1 (Influence theorem for common OE set). *We assume that both the AD model based on DSVDD and the AD model based on MMHAD have been successfully trained and achieved an ideal state. Given that the positive class is uniformly distributed with training samples within*

a constrained region, the outlier exposure set is uniformly sampled from training samples in the outlier distribution and the overlap region. Theoretically, the AD error rate of DSVDD is more than twice that of MMHAD.

Proof. Assuming there are N positive test samples, N_{OE} anomalous test samples, and $2n_j$ test samples in the overlap region, DSVDD may misclassify half or more of these samples as anomalous (since they are uniformly distributed and the hypersphere of DSVDD may not perfectly encompass the normal class). Therefore, the AD error rate of DSVDD is at least $\frac{2n_j}{N+N_{OE}}$. In contrast, the AD error rate of MMHAD will be solely determined by the number of anomalous samples it incorrectly classifies as normal. Thus, the error rate of MMHAD does not exceed $\frac{n_j}{N+N_{OE}}$. \square

In summary, under ideal conditions, the error rate of the AD modeling method with abnormal exposure set participation is lower than that of the AD modeling method without abnormal exposure set participation. Furthermore, we summarize an innovative set of steps to optimize OE sets in open set identification or AD as follows.

Step 1. First, we add 1/3 noise samples to the positive sample set to achieve positive sample set enhancement.

Step 2. Second, we increase the image with 2/3 noise of each positive sample image to enhance the negative class sample set. The negative class sample set can be enhanced by increasing the upper, lower, left, and right 2/5 images of each positive sample image (corresponding to the lower, upper, right, and left 3/5 images, respectively).

Step 3. In addition, owing to the complex image background environment of the CIFAR-10 dataset, not all samples are suitable for this operation. For instance, the sample of birds is relatively small.

Step 4. Finally, we increase the tightly packed point set of the positive class sample set (using the tightly packed point set to learn the generation algorithm and alternate training algorithm) to enhance the negative class samples.

4. Experiment

4.1. Datasets

We performed experiments utilizing benchmark datasets, namely, FMNIST [40], CIFAR-10 [41], CIFAR-100 [41], and RECYCLE [42]. The characteristics of each dataset are outlined as follows: (1) The FMNIST dataset comprises 70,000 grayscale images, each with dimensions of 28×28 pixels. It encompasses 10 distinct commodity categories, with 60,000 images allocated to the training set and 10,000 images reserved for testing. (2) The CIFAR-10 dataset features 10 distinct object classes, totaling 60,000 color images, each with dimensions of 32×32 pixels. Of these, 50,000 images belong to the training set, while the remaining 10,000 images constitute the test set. (3) Analogous to CIFAR-10, CIFAR-100 also contains 60,000 color images, each measuring 32×32 pixels, but across 100 different categories. Notably, each image in CIFAR-100 is associated with both a coarse-grained label and a fine-grained label, with the coarse-grained tag encompassing 20 superclasses. (4) The RECYCLE dataset comprises 11,500 color images of 32×32 pixels across five categories, with 2300 images in each category. Among them, 10,000 images are used for training and 1500 images are reserved for testing.

4.2. Evaluation Metrics

Most ADs only use the AUROC (area under receiver operating characteristic curve) as an evaluation index, which cannot objectively evaluate the difference between known instances and outliers. To comprehensively evaluate the overall classification of the model, this section considers multiple indicators FPR95, AUPR, and ACCURACY to reveal the accuracy of the model in different aspects.

4.3. Benchmark Model

We conducted a comparative analysis to assess the performance of five algorithms: DROCC, DROCC (m), DeepSVDD, DeepSVDD (m), and DeepMAD. For each dataset, images corresponding to classes 2, 5, and 9 in CIFAR-10 and FMNIST were designated as normal classes, whereas images belonging to the remaining classes were categorized as abnormal classes. For all DROCC experiments, we utilized the following parameters: radius $r = 8$, $\mu = 1$, learning rate $lr = 0.001$, ascent step $t = 0.001$, and the Adam optimizer. The parameter settings for DeepSVDD included $\eta = 1$, learning rate $lr = 0.0001$, and also employed the Adam optimizer. Similar to previous experiments, MMHAD used the Adam optimizer to train the model for 100 epochs while dynamically adjusting the learning rate at each milestone using a learning rate scheduler. At each milestone, the learning rate was multiplied by a factor of $\gamma = 0.1$ to gradually decrease it and facilitate more stable convergence in later stages of training. Specifically, in the case of (2/8) scenarios, there exist 45 combinations; for (5/5) scenarios, there are 252 combinations; and in the case of (9/1) scenarios, there are 10 combinations. Regarding CIFAR-100, our focus was on studying scenarios encompassing 20 superclasses with a ratio of (2/18). Section 4.5 presents a comprehensive evaluation of the AUROC performance of various algorithms across diverse datasets, emphasizing the results achieved by the top-performing algorithms. It is worth mentioning that the benchmark dataset results have been sourced from [Deep MAD]. During the training phase, only the normal class from the original training set and the edge OE set were utilized to train various AD methods. Subsequently, during the testing phase, test sets derived from each original dataset were employed to assess the performance of the trained AD methods.

4.4. Environment and Configuration

On a server with two Intel Xeon E5-2630 v3 2.40 GHz CPUs, 188 GB memory, GTX1660 Ti with 6 GB GPU memory, and CentOS Linux 7 (Core) operating systems, Python 3.8.13, Pytorch 1.11.0, NVIA-SMI 460.84, Driver Version: 460.84, CUDA Version 11.2, etc., were used for category quantitative analysis and ablation experiments.

4.5. Results and Analysis

Tables 1–3 present observations on the AUROC values achieved by five algorithms across CIFAR-10, FMNIST, CIFAR-100, and RECYCLE datasets. (1) Evidently, the AUROC value of the combined training across all categories typically surpasses that of training on a single category alone, indicating that multi-category training enhances the model's generalization capability and recognition accuracy. (2) A close examination of Tables 1 and 2 reveals that among all tested algorithms, the MMHAD algorithm emerges as the top performer in various scenarios. Specifically, in instances such as (2/8), (5/5), and (9/1), the algorithm boasts improvements of approximately 26.0%, 8.2%, and 20.1% on the CIFAR-10 dataset and similar gains of 14.8%, 12.0%, and 20.1% on FMNIST, respectively. These remarkable enhancements not only substantiate the efficacy of the MMHAD algorithm in handling unbalanced datasets but also showcase its potent feature extraction and classification capabilities. (3) From Table 3, it is evident that our algorithm achieves a remarkable improvement of approximately 17.4% in AUC value on the CIFAR-100 dataset compared to other algorithms. This accomplishment stands out particularly as CIFAR-100, a complex image recognition benchmark comprising 100 categories, poses a significantly higher classification challenge than the commonly used CIFAR-10 dataset. This demonstrates the robust discriminative capability of MMHAD when dealing with images featuring a large number of categories and highly similar characteristics. Furthermore, the staggering 22% increase in AUC value on the RECYCLE dataset underscores the superiority of our algorithm in tackling tasks within specific domains, such as recycling item recognition. The RECYCLE dataset, focused on recycling items, presents a unique challenge with substantial variations in appearance, shape, and color among these items, yet it also harbors numerous subtle similarities that complicate the recognition process. Our algorithm's ability to accurately

discern these minute differences underscores its excellence in handling the intricacies of specialized domains.

Table 1. AUROC range for CIFAR-10 (10 repetitions for each).

Methods	Two in, Eight out	Five in, Five out	Nine in, One out
DROCC	$0.4728 \pm 0.0119 \leftrightarrow 0.7252 \pm 0.0081$	$0.4316 \pm 0.0257 \leftrightarrow 0.7219 \pm 0.0039$	$0.4107 \pm 0.0454 \leftrightarrow 0.7146 \pm 0.0079$
Mean	0.5990	0.5753	0.5627
DROCC (m)	$0.4216 \pm 0.0424 \leftrightarrow 0.6912 \pm 0.0188$	$0.3806 \pm 0.0047 \leftrightarrow 0.7023 \pm 0.0648$	$0.3439 \pm 0.1034 \leftrightarrow 0.6896 \pm 0.0453$
Mean	0.5564	0.5415	0.5168
DSVDD	$0.4216 \pm 0.0424 \leftrightarrow 0.6912 \pm 0.0188$	$0.3806 \pm 0.0047 \leftrightarrow 0.7023 \pm 0.0648$	$0.3439 \pm 0.0134 \leftrightarrow 0.6896 \pm 0.0453$
Mean	0.5990	0.5753	0.5627
DSVDD (m)	$0.4147 \pm 0.0129 \leftrightarrow 0.7516 \pm 0.0093$	$0.3482 \pm 0.0123 \leftrightarrow 0.6909 \pm 0.0133$	$0.3580 \pm 0.0166 \leftrightarrow 0.5864 \pm 0.0167$
Mean	0.5832	0.5196	0.4722
DMAD	$0.5396 \pm 0.0031 \leftrightarrow 0.7647 \pm 0.0014$	$0.4929 \pm 0.0046 \leftrightarrow 0.7738 \pm 0.0022$	$0.5437 \pm 0.0028 \leftrightarrow 0.7230 \pm 0.0084$
Mean	0.6522	0.6334	0.6359
Ours	$0.8384 \pm 0.0139 \leftrightarrow 0.9863 \pm 0.0104$	$0.5020 \pm 0.0081 \leftrightarrow 0.9291 \pm 0.0194$	$0.8049 \pm 0.0128 \leftrightarrow 0.8687 \pm 0.0075$
Mean	0.9124	0.7156	0.8368

Table 2. AUROC range for F-MNIST (10 repetitions for each).

Methods	Two in, Eight out	Five in, Five out	Nine in, One out
DROCC	$0.6873 \pm 0.0937 \leftrightarrow 0.9774 \pm 0.0049$	$0.5738 \pm 0.0397 \leftrightarrow 0.9260 \pm 0.0307$	$0.5408 \pm 0.0961 \leftrightarrow 0.8247 \pm 0.0507$
Mean	0.8161	0.7448	0.6992
DSVDD	$0.6622 \pm 0.0502 \leftrightarrow 0.9871 \pm 0.0033$	$0.5438 \pm 0.0274 \leftrightarrow 0.9279 \pm 0.0325$	$0.4551 \pm 0.0285 \leftrightarrow 0.8825 \pm 0.0137$
Mean	0.8538	0.7269	0.6523
DMAD	$0.6434 \pm 0.0640 \leftrightarrow 0.9714 \pm 0.0011$	$0.5732 \pm 0.0485 \leftrightarrow 0.8832 \pm 0.0137$	$0.4860 \pm 0.0267 \leftrightarrow 0.9395 \pm 0.3466$
Mean	0.8329	0.7739	0.7613
Ours	$0.9651 \pm 0.0030 \leftrightarrow 0.9993 \pm 0.0004$	$0.7955 \pm 0.0179 \leftrightarrow 0.9914 \pm 0.0009$	$0.9510 \pm 0.0175 \leftrightarrow 0.9896 \pm 0.0071$
Mean	0.9807	0.8935	0.9703

Table 3. AUROC range for CIFAR-100 and RECYCLE (10 repetitions for each).

	CIFAR-100	RECYCLE
Method	Two in, Eighteen out	Four in, One out
DROCC	$0.3548 \pm 0.0006 \leftrightarrow 0.7325 \pm 0.0971$	$0.4447 \pm 0.0176 \leftrightarrow 0.7997 \pm 0.0719$
Mean	0.5638	0.6128
DSVDD	$0.4192 \pm 0.0077 \leftrightarrow 0.7185 \pm 0.0180$	$0.3703 \pm 0.0207 \leftrightarrow 0.8728 \pm 0.0079$
Mean	0.5559	0.5791
DMAD	$0.5384 \pm 0.0018 \leftrightarrow 0.8213 \pm 0.0012$	$0.5906 \pm 0.0035 \leftrightarrow 0.8283 \pm 0.0073$
Mean	0.6580	0.6966
MMHAD	$0.7366 \pm 0.0125 \leftrightarrow 0.9279 \pm 0.0106$	$0.8897 \pm 0.0163 \leftrightarrow 0.9441 \pm 0.0036$
Mean	0.8323	0.9169

The AUROC objectively assesses the difference between known and abnormal instances but ignores the accuracy of correctly identifying known categories. The same experimental settings were used to evaluate the detection performance, including AUPR, FPR95, and ACCURACY indexes as shown in Table 4. Three other performance indicators for different combinations of normal classes are presented. MMHAD shows similar overall performance in CIFAR-10 and FMNIST. When there are only 2 normal classes, the classifier can more easily distinguish between them, leading to optimal overall classification performance. When there are only two normal classes, the classifier can distinguish them more easily, resulting in the best overall classification performance. However, when there are five normal classes, the complexity of the model makes it difficult for the classifier to accurately distinguish between them, leading to a decrease in accuracy. With nine normal classes, although there is a slight decrease in performance indicators such as AUROC, AUPR, and

FPR95 due to increased model complexity, the accuracy remains high as most samples can be correctly classified. Since CIFAR-100 contains more classes, the model’s performance, especially FPR95, declines compared to CIFAR-10, but it still maintains a high AUPR and accuracy. The performance on the RECYCLE dataset is between CIFAR-10 and CIFAR-100, with fluctuations in AUPR, FPR95, and accuracy, but the overall performance is good.

Table 4. AUPR, FPR95, and accuracy range for CIFAR-10, FMNIST, CIFAR-100 and RECYCLE (10 repetitions for each).

Datasets	Case	AUPR	FPR95	ACCURACY
CIFAR-10	2/8	0.9508 ↔ 0.9962	0.5642 ↔ 0.6410	0.8348 ↔ 0.9610
	5/5	0.5047 ↔ 0.9069	0.9358 ↔ 0.4392	0.5002 ↔ 0.8705
	9/1	0.5518 ↔ 0.6243	0.4100 ↔ 0.3440	0.8997 ↔ 0.9313
F-MNIST	2/8	0.9883 ↔ 0.9998	0.1941 ↔ 0.2600	0.9327 ↔ 0.9914
	5/5	0.7832 ↔ 0.9869	0.6912 ↔ 0.0286	0.4999 ↔ 0.9620
	9/1	0.7002 ↔ 0.9440	0.3040 ↔ 0.6300	0.9227 ↔ 0.9773
CIFAR-100	2/18	0.9538 ↔ 0.9884	0.7793 ↔ 0.4130	0.9046 ↔ 0.9459
RECYCLE	4/1	0.8666 ↔ 0.6887	0.5167 ↔ 0.2433	0.8360 ↔ 0.9187

Table 5 presents the results of the Friedman test, comparing and analyzing the performance of four different methods in this study in the case of (2/8) on CIFAR-10. Specifically, the median response rates for DROCC and DSVDD were relatively low at 0.595 and 0.573, respectively, suggesting similar and relatively weak performance under test conditions. In contrast, the median response rate for DMAD increased to 0.669, indicating some improvement in performance but with limited impact. The MMHAD method exhibited the highest median response rate at 0.897 with a small standard deviation (0.039), signifying significantly better performance compared to the other three methods with stable results. Cohen’s F-number, serving as a measure of effect size, quantitatively indicates differences between methods. In this study, it was calculated as 1.902; this number is considered significant when exceeding 0.4 according to general belief.

Table 5. Analysis results of Friedman test in 2-in, 8-out case on CIFAR-10.

Methods	Total	Median	Standard	Statistic	p	Cohen’s f
DROCC	45	0.595	0.080	122.947	0.000 ***	1.902
DSVDD	45	0.573	0.093			
DMAD	45	0.669	0.050			
MMHAD	45	0.897	0.039			

Note: *** represent significance levels of 1%.

4.6. Category Quantity Analysis

When addressing class imbalance, it is crucial to ensure that the classifier maintains high performance even when there is a significant variation in the number of samples from different classes. Imbalanced classification tends to result in the classifier favoring the prediction of the class with a larger sample size. This section provides a detailed discussion of the ability of the MMHAD algorithm to address class imbalance.

Figure 3 presents a scatter plot depicting all 45 combinations of five algorithms under normal category conditions. The observed combination of different categories can have a substantial impact on classifier performance when sorted from smallest to largest. Particularly when one category has significantly more samples than others, there may be a bias toward predicting the category with more samples. Compared with the benchmark algorithm, MMHAD can reach or exceed the AUROC value of 0.8 under most category combinations, which indicates that it has higher accuracy and reliability in classification tasks. MMHAD notably diminishes the influence of normal class combinations on AUROC

values. Figure 4 illustrates AUROC values for both the top and bottom 10 cases under normal conditions for all five categories. It was observed that our algorithm showed significant advantages. In particular, the AUROC values of our algorithm exhibit extremely low variance, more effectively suppressing the effects of noise and outliers. It can maintain stable performance under different category combinations and can adapt to different data distributions and category combinations. Our algorithm indicates its capability to learn robust recognition features, map the inner points to the low-dimensional output space, and reduce the computational complexity.

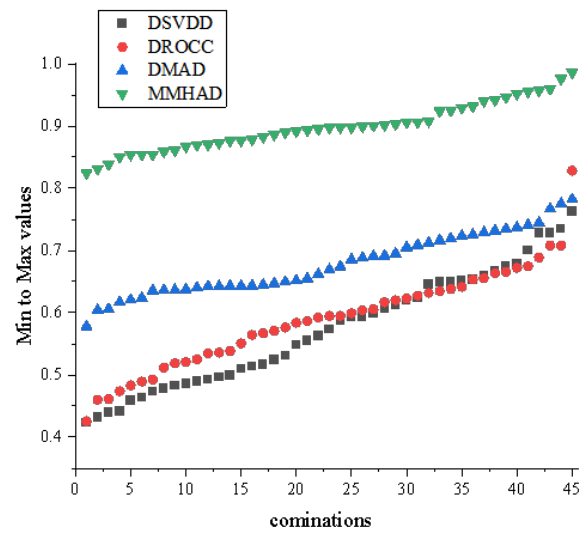


Figure 3. Two-in and eight-out case on CIFAR-10.

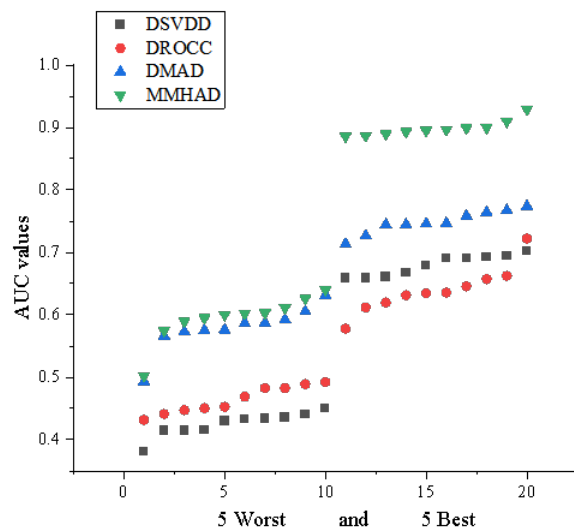


Figure 4. Five-in and five-out case on CIFAR-10.

Due to considerable variations in the time costs of different methods, we take logarithms of the training results to ensure consistent visualization in a single plot. Figure 5 illustrates that, from a temporal perspective, the proposed method demonstrates higher computational efficiency when dealing with a small number of categories. However, as the number of categories increases, there is a significant rise in training time due to the model’s need to acquire more internal representations and boundaries for handling numerous target categories, thereby increasing the computational load. In all three scenarios, test times were relatively short (ranging from 0.556 to 1.578 s) and did not exhibit substantial growth

with an increase in classes. This can be attributed to the fact that testing primarily involves forward propagation, which is significantly less computationally intensive compared to backpropagation and parameter updating during training, thus affirming that the model can swiftly classify new samples or detect anomalies post-training.

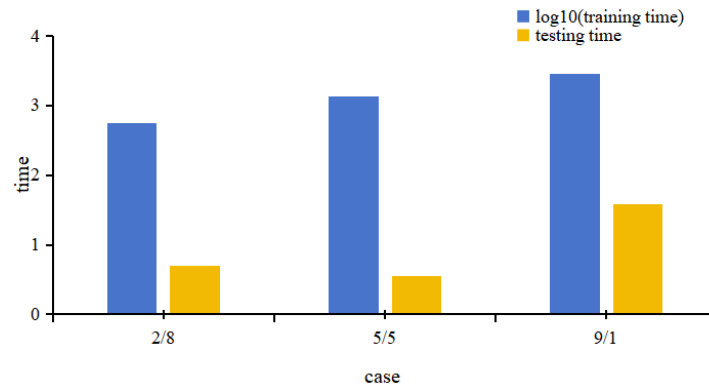


Figure 5. The training and testing times of different cases on CIFAR-10.

4.7. Ablation Experiment

To demonstrate the influence of hyperparameters on our AD approach, we conducted ablation studies on the CIFAR-100 dataset, specifically focusing on the validation of the edge OE set and margin parameters. The first two rows of Table 6 show the significant enhancement of the detection performance metrics after the inclusion of the edge OE set. AUROC increased by 10.42%, AUPR increased by 5.12%, FPR95 decreased by 32.33%, and ACCURACY increased by 1.89%. Notably, FPR95 experienced a substantial decline, indicating a reduction in false positives during m AD tasks. Furthermore, the adjustment of the margin parameters is aimed at optimizing the decision boundary of the model to better distinguish between normal and abnormal samples. It can be observed that AUROC increased by 4.82%, AUPR increased by 0.94%, FPR95 increased by 1.01%, and ACCURACY increased by 0.58%. FPR95 is an indicator measured at a specific TPR (typically 95%) that can be very sensitive to small changes in the model’s decision boundaries. Therefore, even if the adjustment of margin parameters brings about an improvement in the overall performance of the model, FPR95 may show a different trend from other indicators because of this sensitivity.

Table 6. Ablation was performed on CIFAR-100 (10 repetitions for each).

μ	m	AUROC	AUPR	FPR95	ACCURACY
×	×	0.5384 ↔ 0.8213	0.8805 ↔ 0.9405	0.9411 ↔ 0.8777	0.9000 ↔ 0.9008
✓	×	0.6613 ↔ 0.9068	0.9397 ↔ 0.9837	0.6913 ↔ 0.4808	0.9001 ↔ 0.9388
✓	✓	0.7366 ↔ 0.9279	0.9538 ↔ 0.9884	0.7793 ↔ 0.4130	0.9046 ↔ 0.9459

Note: × indicates that the method was not used in the experiment, ✓ indicates that the method was used in the experiment.

4.8. Hyperparameter Analysis

Figure 6 demonstrates the influence of different values of the trade-off parameters μ and m on the MMHAD performance on the CIFAR-10 dataset. Note that the algorithm selects the mean of the 9/1 case, with each experiment repeated 10 times. To comprehensively investigate the impact of parameters μ and m while keeping other parameters constant, we systematically varied each variable within a predetermined range, often based on experience or cross validation.

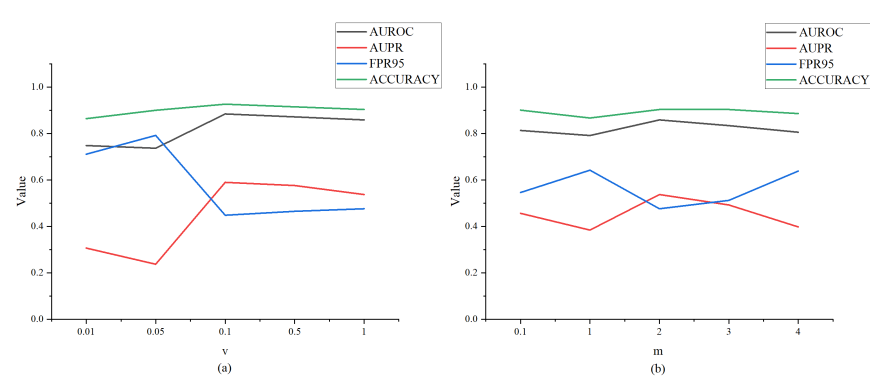


Figure 6. Sensitivity analysis with μ, m in terms of AUROC, AUPR, FPR95, and ACCURACY on CIFAR-10. (a) The effect of different values of μ ; (b) The effect of different values of m .

μ controls the loss contribution of anomalous (or outlier) samples. The shape of the enclosing sphere tightens as μ decreases. As observed in Figure 6a, choosing an excessively small value for μ can lead to overfitting, where some normal samples are mistakenly classified as anomalous, increasing the error rate. When μ is set to an optimal value, MMHAD achieves the best test results across almost all evaluation metrics. The impact of different values of the trade-off parameter m on MMHAD's performance is shown in Figure 6b. m defines the safety margin between the hypersphere for normal samples and the potential anomaly region. The margin m is a key parameter balancing sensitivity to anomalies and the risk of false positives. When m is set to its optimal value, all metrics achieve the best performance. A smaller margin value makes the hypersphere more compact, making the model more sensitive to outlier samples but potentially increasing the false-positive rate. Conversely, a larger margin value can reduce the false-positive rate but may sacrifice some sensitivity to anomalous samples. By systematically tuning μ and m , we can gain insight into how these parameters influence the model's performance and find optimal values that balance these competing objectives.

5. Discussion

The algorithm demonstrates various advantages, yet it also encounters certain limitations. Firstly, employing convolutional neural networks to train all normal classes collectively enables the learning of compact distinguishing features. However, this approach may have high computational complexity, particularly when dealing with large-scale datasets or multi-class tasks. Secondly, while MMHAD focuses on abnormal samples at the class boundary, it may not effectively detect such samples when encountering new and unknown classes. In practical applications, significant differences among normal samples from different classes could pose challenges for accurately capturing class boundaries. The current network architecture may not be optimal for handling specific data types or exceptions. Future work should contemplate the introduction of a dynamic margin, which can flexibly adjust the boundary position by altering data distribution, and facilitate capturing the boundary between normal and abnormal samples more precisely, thereby enhancing the detection accuracy. However, how to maintain the stability of the dynamic margin and prevent its excessive fluctuation or premature convergence constitutes a key issue. Unstable margins might result in fluctuations in model performance and impact the final detection outcome, thereby increasing the complexity and computational cost of the model. All in all, future studies can ameliorate and optimize these limitations to enhance the model's performance, generalization, and practicability.

6. Conclusions

The primary contribution of this paper lies in proposing a novel deep multi-class hypersphere-based AD method (MMHAD) that incorporates an edge OE set and margin. This method effectively addresses the issue of traditional multi-class AD algorithms where

the decision boundary does not closely align with the boundary of the normal sample region. By studying the embedding of multiple normal object classes into various datasets under the constraint of an OE set that encompasses class boundaries, MMHAD learns a discriminative latent space to distinguish between normal and abnormal samples. Normal samples are tightly clustered within their respective hyperspheres, while abnormal samples are pushed outside these hyperspheres. Additionally, by introducing a margin parameter, MMHAD further enhances the model's ability to reject anomalous values, ensuring that abnormal samples are effectively mapped outside the hyperspheres, thereby improving the accuracy and robustness of AD. Experimental results demonstrate that MMHAD exhibits significant performance improvements across multiple benchmark datasets and under different class settings, validating its effectiveness and practical value in multi-class AD tasks.

Author Contributions: Conceptualization, G.Y.; methodology, M.G. and G.Y.; validation, M.G.; writing—original draft preparation, M.G.; writing—review and editing, X.L. and D.X.; funding acquisition, M.G. and G.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Postgraduate Research Practice Innovation Program of Jiangsu Province KYCX23_2347, the National Natural Science Foundation of China under Grant 62172229, and the Natural Science Foundation of Jiangsu Province BK20211295.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Baccari, S.; Haddad, M.; Ghazzai, H.; Touati, H.; Elhadeif, M. Anomaly Detection in Connected and Autonomous Vehicles: A Survey, Analysis, and Research Challenges. *IEEE Access* **2024**, *12*, 19250–19276. [[CrossRef](#)]
- Fernando, T.; Gammulle, H.; Denman, S.; Sridharan, S.; Fookes, C. Deep Learning for Medical Anomaly Detection—A Survey. *ACM Comput. Surv.* **2021**, *54*, 1–37. [[CrossRef](#)]
- Tao, X.; Gong, X.; Zhang, X.; Yan, S.; Adak, C. Deep Learning for Unsupervised Anomaly Localization in Industrial Images: A Survey. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–21. [[CrossRef](#)]
- Pang, G.; Shen, C.; Cao, L.; Hengel, A.V.D. Deep Learning for Anomaly Detection: A Review. *ACM Comput. Surv.* **2022**, *54*, 1–38. [[CrossRef](#)]
- Munir, M.; Chattha, M.A.; Dengel, A.; Ahmed, S. A Comparative Analysis of Traditional and Deep Learning-Based Anomaly Detection Methods for Streaming Data. In Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 561–566.
- Ma, X.; Wu, J.; Xue, S.; Yang, J.; Zhou, C.; Sheng, Q.Z.; Xiong, H.; Akoglu, L. A Comprehensive Survey on Graph Anomaly Detection with Deep Learning. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 12012–12038. [[CrossRef](#)]
- Hojjati, H.; Ho, T.K.K.; Armanfard, N. Self-supervised anomaly detection in computer vision and beyond: A survey and outlook. *Neural Netw.* **2024**, *170*, 1–23. [[CrossRef](#)] [[PubMed](#)]
- Subbiah, U.; Kumar, D.K.; Thangavel, S.K.; Parameswaran, L. An Extensive Study and Comparison of the Various Approaches to Object Detection using Deep Learning. In Proceedings of the 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 20–22 August 2020; pp. 1018–1030.
- Ruff, L.; Kauffmann, J.R.; Vandermeulen, R.A.; Montavon, G.; Samek, W.; Kloft, M.; Dietterich, T.G.; Müller, K. A Unifying Review of Deep and Shallow Anomaly Detection. *Proc. IEEE* **2021**, *109*, 756–795. [[CrossRef](#)]
- Defard, T.; Setkov, A.; Loesch, A.; Audigier, R. Padim: A patch distribution modeling framework for anomaly detection and localization. In Proceedings of the International Conference on Pattern Recognition (ICPR 2021), Milan, Italy, 10–15 January 2021; pp. 475–489.
- Yang, G.; Zhou, S.; Wan, M. Open-Set Recognition Model Based on Negative-Class Sample Feature Enhancement Learning Algorithm. *Mathematics* **2022**, *10*, 4725. [[CrossRef](#)]
- Yann, Y.; Yoshua, B.; Geoffrey, H. Deep learning. *Nature* **2015**, *521*, 436–444.
- Amarasinghe, K.; Kenney, K.; Manic, M. Toward Explainable Deep Neural Network Based Anomaly Detection. In Proceedings of the 2018 11th International Conference on Human System Interaction (HSI), Gdansk, Poland, 4–6 July 2018; pp. 311–317.
- Hendricks, D.; Mazeika, M.; Dietterich, T. Deep anomaly detection with outlier exposure. *arXiv* **2018**, arXiv:1812.04606.
- Schölkopf, B.; Platt, J.C.; Shawe-Taylor, J.; Smola, A.J.; Williamson, R.C. Estimating the Support of a High-Dimensional Distribution. *Neural Comput.* **2001**, *13*, 1443–1471. [[CrossRef](#)]

16. Cheng, G.; Tong, X. Fuzzy Clustering Multiple Kernel Support Vector Machine. In Proceedings of the 2018 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), Chengdu, China, 15–18 July 2018; pp. 7–12.
17. Tax, D.M.; Duin, R.P. Support Vector Data Description. *Mach. Learn.* **2004**, *54*, 45–66. [[CrossRef](#)]
18. Ruff, L.; Vandermeulen, R.; Goernitz, N.; Deecke, L.; Siddiqui, S.A.; Binder, A.; Müller, E.; Kloft, M. Deep one-class classification. In Proceedings of the International Conference on Machine Learning (PMLR), Stockholm, Sweden, 10–15 July 2018; pp. 4393–4402.
19. Ruff, L.; Vandermeulen, R.A.; Grnitz, N. Deep Semi-Supervised Anomaly Detection, International Conference on Learning Representations. *arXiv* **2019**, arXiv:1906.02694.
20. Park, J.; Moon, J.; Ahn, N.; Sohn, K. What is Wrong with One-Class Anomaly Detection? *arXiv* **2021**, arXiv:2104.09793.
21. Chong, P.; Ruff, L.; Kloft, M.; Binder, A. Simple and Effective Prevention of Mode Collapse in Deep One-Class Classification. In Proceedings of the 2020 International Joint Conference on Neural Networks(IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–9.
22. Ghafoori, Z.; Leckie, C. Deep multi-sphere support vector data description. In Proceedings of the 2020 SIAM International Conference on Data Mining (SDM 2020), Cincinnati, OH, USA, 7–9 May 2020; pp. 109–117.
23. Goyal, S.; Raghunathan, A.; Jain, M.; Simhadri, H.V.; Jain, P. Drocc: Deep robust one-class classification. In Proceedings of the 37th International Conference on Machine Learning(ICML'20), Virtual, 13–18 July 2020; pp. 3711–3721.
24. Hojjati, H.; Armanfard, N. DASVDD: Deep autoencoding support vector data descriptor for anomaly detection. *arXiv* **2021**, arXiv:2106.05410.
25. Hu, W.; Wang, M.; Qin, Q.; Ma, J.; Liu, B. HRN: A Holistic Approach to One Class Learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 19111–19124.
26. Zhang, Z.; Deng, X. Anomaly detection using improved deep SVDD model with data structure preservation. *Pattern Recognit. Lett.* **2021**, *148*, 1–6. [[CrossRef](#)]
27. Zhou, Y.; Liang, X.; Zhang, W.; Zhang, L.; Song, X. VAE-based Deep SVDD for anomaly detection. *Neurocomputing* **2021**, *453*, 131–140. [[CrossRef](#)]
28. You, Z.; Yang, K.; Luo, W.; Cui, L.; Zheng, Y.; Le, X. ADTR: Anomaly detection transformer with feature reconstruction. In Proceedings of the Neural Information Processing: 29th International Conference, Berlin/Heidelberg, Germany, 22–26 November 2023; pp. 298–310.
29. You, Z.; Cui, L.; Shen, Y.; Yang, K.; Lu, X.; Zheng, Y.; Le, X. A Unified Model for Multi-class Anomaly Detection. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 4571–4584.
30. Singh, S.; Luo, M.; Li, Y. Multi-Class Anomaly Detection. In Proceedings of the Neural Information Processing: 29th International Conference, Berlin/Heidelberg, Germany, 22–26 November 2023; pp. 359–371.
31. Shi, Y.; Yang, J.; Qi, Z. Unsupervised anomaly segmentation via deep feature reconstruction. *Neurocomputing* **2021**, *424*, 9–22. [[CrossRef](#)]
32. Li, C.-L.; Sohn, K.; Yoon, J.; Pfister, T. CutPaste: Self-Supervised Learning for Anomaly Detection and Localization. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 9659–9669.
33. Zaheer, M.Z.; Lee, J.-H.; Astrid, M.; Lee, S.-I. Old Is Gold: Redefining the Adversarially Learned One-Class Classifier Training Paradigm. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 14171–14181.
34. Abati, D.; Porrello, A.; Calderara, S.; Cucchiara, R. Latent Space Autoregression for Novelty Detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 481–490.
35. Yi, J.; Yoon, S. Patch SVDD: Patch-level SVDD for anomaly detection and segmentation. In Proceedings of the Computer Vision—ACCV 2020: 15th Asian Conference on Computer Vision, Kyoto, Japan, 30 November–4 December 2020; pp. 375–390.
36. Liznerski, P.; Ruff, L.; Vandermeulen, R.A.; Franks, B.J.; Kloft, M.; Müller, K.-R. Explainable deep one-class classification. *arXiv* **2020**, arXiv:2007.01760.
37. Ruff, L.; Vandermeulen, R.A.; Franks, B.J.; Müller, K.; Kloft, M. Rethinking Assumptions in Deep Anomaly Detection. *arXiv* **2020**, arXiv:2006.00339.
38. Kirchheim, K.; Filax, M.; Ortmeier, F. Multi-class hypersphere anomaly detection. In Proceedings of the 26th International Conference on Pattern Recognition(ICPV), Montreal, QC, Canada, 21–25 August 2022; pp. 2636–2642.
39. Cevikalp, H.; Uzun, B.; Salk, Y.; Saribas, H.; Köpüklü, O. From anomaly detection to open set recognition: Bridging the gap. *Pattern Recognit.* **2023**, *138*, 109385. [[CrossRef](#)]
40. Xiao, H.; Rasul, K.; Vollgraf, R. Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms. *arXiv* **2017**, arXiv:1708.07747.
41. Krizhevsky, A.; Hinton, G. Learning Multiple Layers of Features from Tiny Images. Master's Thesis, Department of Computer Science, University of Toronto, Toronto, ON, USA, 2009.
42. Suresh, S.; Minwei L.; Yu, L. Generalized Anomaly Detection. *arXiv* **2021**, arXiv:2110.15108.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.