Journal of Image and Signal Processing 图像与信号处理, 2026, 15(1), 49-63
Published Online January 2026 in Hans. https://www.hanspub.org/journal/jisp
https://doi.org/10.12677/jisp.2026.151005

Hans汉斯

对抗紧密包裹与超球面约束的异常检测模型

付纯博[1]，杨国为[2*]，徐俊华[1]，汤文杰[1]

[1]南京审计大学计算机学院，江苏 南京
[2]南通理工学院信息工程学院，江苏 南通

收稿日期：2025年12月9日；录用日期：2025年12月30日；发布日期：2026年1月14日

摘 要

无监督异常检测模型常面临检测边界模糊和泛化能力弱的难题。现有的深度异常检测模型，例如DeepSVDD，虽然通过超球面约束使正常模式特征分布更加紧密，但难以适配不规则的特征分布。此外，传统检测方法在对“紧贴正常边界且具有高度迷惑性的异常分布”进行精准建模方面还存在不足。为解决这些问题，本文提出了一种基于对抗紧密包裹与超球面约束的异常检测模型。该模型融合了紧密包裹学习、对抗学习以及深度超球面约束的思想，其博弈损失函数与现有方法的博弈损失函数完全不同。具体而言，该损失函数综合了正常样本的紧密包裹损失、异常样本的排斥损失以及对抗样本的生成损失，目标是缩小已知类别特征分布区域与检测模型所确定的类别特征分布区域之间的差异。其中，紧密包裹学习有助于使模型的检测边界更加清晰和准确，而对抗学习则使模型能够学习到更加鲁棒的特征表示，深度超球面约束使异常样本与正常样本分隔更大，从而优化模式特征分布建模并增强模型的泛化能力。实验结果表明，在多个数据集上，与多种现有的异常检测模型相比，本文提出的模型表现更为出色。

关键词

异常检测，单类分类，对抗学习，超球面学习，紧密包裹

Adversarial Compact Wrapping with Hyperspherical Constraint for Anomaly Detection

Chunbo Fu[1], Guowei Yang[2*], Junhua Xu[1], Wenjie Tang[1]

[1]School of Computer Science, Nanjing Audit University, Nanjing Jiangsu
[2]School of Information Engineering, Nantong Institute of Technology, Nantong Jiangsu

Received: December 9, 2025; accepted: December 30, 2025; published: January 14, 2026

*通讯作者。

文章引用: 付纯博, 杨国为, 徐俊华, 汤文杰. 对抗紧密包裹与超球面约束的异常检测模型[J]. 图像与信号处理, 2026, 15(1): 49-63. DOI: 10.12677/jisp.2026.151005

Abstract

Unsupervised anomaly detection models often encounter the challenges of ambiguous detection boundaries and weak generalization ability. Existing deep anomaly detection models, such as Deep SVDD, although they make the feature distribution of normal patterns more compact through hypersphere constraints, are difficult to adapt to irregular feature distributions. Moreover, traditional detection methods still have deficiencies in precisely modeling "anomalous distributions that closely adhere to the normal boundary and are highly deceptive". To address these issues, this paper proposes an anomaly detection model based on adversarial tight wrapping and hypersphere constraints. This model integrates the ideas of tight wrapping learning, adversarial learning, and deep hypersphere constraints, and its game loss function is completely different from that of existing methods. Specifically, this loss function combines the tight wrapping loss of normal samples, the repulsion loss of anomalous samples, and the generation loss of adversarial samples, aiming to minimize the difference between the feature distribution region of known categories and the category feature distribution region determined by the detection model. Among them, tight wrapping learning helps to make the detection boundary of the model clearer and more accurate, while adversarial learning enables the model to learn more robust feature representations. Deep hypersphere constraints increase the separation between anomalous and normal samples, thereby optimizing the modeling of pattern feature distributions and enhancing the generalization ability of the model. Experimental results show that on multiple datasets, the proposed model outperforms many existing anomaly detection models.

Keywords

Anomaly Detection, One-Class Classification, Adversarial Learning, Hypersphere Learning, Tight Wrapping

1. 引言

异常检测(Anomaly Detection)作为机器学习与数据挖掘的重要分支，旨在从大规模数据集中识别显著偏离“正常”行为模式的异常样本，广泛应用于网络安全入侵检测、工业设备故障预警、金融欺诈识别和医学影像分析等领域。传统方法如一类支持向量机(OCSVM) [1]、孤立森林(Isolation Forest) [2]、核密度估计(KDE) [3]和高斯混合模型(GMM) [4]在低维结构化数据上表现良好，但在高维复杂数据上因“维度灾难”和非线性分布建模能力有限而性能受限[5]。

随着深度神经网络的发展，深度异常检测方法通过神经网络自动学习数据的低维紧密表示，并在该表示空间中构建判别边界。代表性工作包括基于自编码器的重构误差法[6]和基于 GAN 的 AnoGAN [7]。然而，这些方法本质是间接学习正常数据分布，缺乏直接面向异常检测任务的优化目标。Ruff 等人[8]提出的 Deep SVDD 首次将单类分类目标嵌入端到端深度学习框架，通过最小化网络输出特征到球心的欧氏距离之和实现对正常样本的紧凑包裹。Kim 等人[9]则提出通过最大化类内角度一致性与类间角度间隔来优化单类表示。

“紧密包裹”(Compact Wrapping)是近年来在开集识别与单类分类中兴起的核心思想，强调在几何空间中对正常类样本形成紧密边界约束，确保类内特征高度聚集和类间边界清晰分离。Zhang 等人[10]设计

了“对抗紧密包裹学习”框架，通过构造“负类增强样本”填充正常类边界外围潜在异常区域，实现更精确的边界拟合。

对抗学习源于 Goodfellow 等人[11]提出的 GAN，通过生成器与判别器的博弈实现共同优化。将对抗学习引入异常检测具有重大意义。Hendrycks 等人[12]提出的“Outlier Exposure”方法利用大规模外部异常数据集进行联合训练，Sabokrou 等人[13]则在训练中动态生成对抗样本并加入训练集。

尽管当前异常检测方法取得进展，但仍面临三大挑战：高维稀疏性与特征漂移问题；模型可解释性与边界可控性不足；对抗鲁棒性缺失。为应对这些挑战，研究者提出了多种解决方案，但往往未能同时兼顾特征紧密性、边界明确性与对抗鲁棒性。

针对上述问题，本文提出基于对抗紧密包裹与超球面约束的异常检测模型(ACHC)。该模型融合紧密包裹学习、对抗学习与深度超球面约束，通过联合优化正常样本的紧密包裹损失、异常样本的排斥损失以及对抗样本的生成损失，使模型的检测边界更加清晰准确，同时提升对未知异常的鲁棒判别能力。

本文的主要贡献为：

1) 提出融合紧密包裹与超球面约束的新型深度异常检测框架，通过联合优化网络参数与超球中心，实现对正常类特征空间的最小体积包裹，提升特征表示的紧密性与判别力。

2) 将对抗学习机制系统性融入基于紧密包裹的异常检测，设计新的对抗训练策略，利用生成器构造决策边界附近的“虚拟异常样本”作为困难负样本参与训练，形成对抗紧密包裹机制，显著提升模型对细微异常、边缘案例及对抗攻击的敏感度与拒识能力。

3) 验证新方法的优越性能。通过实验验证了所提方法在多个真实场景下的优越性能与鲁棒性。实验结果表明，本文方法在 CIFAR-10、CIFAR-100 等关键数据集上均取得领先表现。

2. 相关工作

2.1. 异常检测

异常检测作为机器学习与数据挖掘的重要分支，旨在从大量正常样本中识别偏离常规模式的异常实例。由于异常事件稀少、多样且难以标注，无监督或单类学习成为主流范式。近年来，随着深度神经网络在表示学习方面的突破，异常检测技术经历了从传统统计模型到深度架构的深刻变革。

单类分类是异常检测的核心理论基础之一。最具代表性的单类分类算法包括支持向量数据描述(Support Vector Data Description, SVDD)和一类支持向量机(One-Class Support Vector Machine, OC-SVM)。SVDD 由 Tax 提出，其核心思想是寻找一个最小体积的超球体，使得所有正常样本都被包含在该球体内，同时最小化球体的半径[14]。该方法通过核技巧将数据映射到高维特征空间，从而实现对非线性分布的建模。OC-SVM 本质上也是如此。尽管两者在低维数据上表现良好，但其计算复杂度随样本数量增长而急剧上升，且在高维、大规模数据集上易受维度影响，导致性能下降。

随着深度学习技术的兴起，深度神经网络(DNN)因其强大的非线性映射和自动特征提取能力，被广泛应用于异常检测任务。目前，主流的深度异常检测方法可分为三类：基于重构、基于生成和基于嵌入的方法。基于重构的方法以自编码器(AE)及其变体为代表，包括去噪自编码器(Denoising AE)、稀疏自编码器(Sparse AE)和变分自编码器(VAE)。例如，Xu *et al.* [15]利用卷积自编码器进行异常检测，取得了优于传统方法的效果。然而，重构误差并不总是与异常程度成正比，某些结构复杂的正常样本也可能产生高重构误差，导致误报。基于生成的方法主要依赖生成对抗网络(GAN)或流模型(Flow-Based Models)。AnoGAN (Schlegl *et al.*, 2019)是最早将 GAN 用于异常检测的工作之一，其通过在潜在空间中搜索最能重构测试样本的编码向量，并以搜索误差作为异常评分。后续工作如 f-AnoGAN 进一步优化了搜索过程，提升了效率[16]。

为了融合传统异常检测方法的理论优势与深度学习的强大表示能力，研究者提出了多种混合模型。其中最具代表性的是深度支持向量数据描述(Deep SVDD)，由 Ruff 等人在 *Deep One-Class Classification* 中首次提出。实验表明，Deep SVDD 在 MNIST、CIFAR-10 等图像数据集上显著优于一些传统模型，但它存在超球塌陷问题。为了更深入地理解这个问题，让我们详细查看一下 DSVDD 的目标函数：

$$\min_{R,W} R^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\max\left\{0, \left\|\phi\left(x_i;W\right)-c\right\|^2 - R^2\right\} + \frac{\lambda}{2}\sum_{l=1}^{L}\left\|W^l\right\|_F^2 \tag{1}$$

其中 c 是超球面的中心，表示神经网络的输出，W 表示网络权重，是控制正则化项贡献的超参数，n 为样本总数。不难看出，上述方程具有 $W=0$ 和 $c=0$ 的平凡解。会引发超球崩溃的问题。这个限制条件在很多应用场景中制约了深度支持向量数据描述的性能表现。

为进一步提升性能，研究者提出了多种改进版本。例如，Cevikalp 等人将球心 c 也作为可优化参数参与训练，增强了模型灵活性[17]；另一些工作引入极值理论(EVT)对尾部分布建模，以更准确地设置决策阈值[18]。此外，也有研究尝试将 OC-SVM 与 DNN 结合，如将 DNN 作为特征提取器，其输出送入 OC-SVM 进行最终分类。这类方法虽能提升特征质量，但仍受限于 OC-SVM 本身的计算瓶颈，难以实现完全端到端优化。

近年来，研究者基于 Deep SVDD 等框架提出多项改进以应对复杂的检测需求。例如，异常暴露方法通过联合训练外部异常数据集，提高了异常识别能力。同时，结合记忆模块的模型通过存储正常模式增强了对局部异常的敏感性，另有研究探索对比学习在异常检测中的应用，通过拉大正负样本对的距离提高判别性能。

尽管现有异常检测方法取得了显著进展，但仍面临诸多挑战与局限性。首先，高维稀疏性与特征漂移问题依然严峻。在真实应用场景中，正常数据内部可能存在较大变异(如不同光照、姿态下的同一物体)，而异常样本可能仅在局部区域呈现微弱偏差，传统基于距离或密度的方法在高维空间中易失效。其次，模型可解释性与边界可控性不足。多数深度方法缺乏明确的几何解释，难以精确控制接受域范围。第三，对抗鲁棒性缺失。现有方法大多假设测试环境为理想状态，忽视了对抗性扰动的存在。

2.2. 紧密包裹学习

紧密包裹学习(Tight Enclosure Learning)是一种旨在通过构造最小化、几何上紧密的决策边界来建模正常类别的特征分布，从而实现高效异常拒识的机器学习范式。其核心思想是：在特征空间中寻找一个体积尽可能小但能完全覆盖所有正常样本的闭合区域，任何位于该区域之外的样本均被视为异常。

Cevikalp 等人[19]提出的 DCHC 模型的核心思想是将正常类别的特征映射到一个紧凑的超球面区域内，并通过联合优化深度网络参数与球体几何属性，实现对正常类别的高效包裹。后续研究进一步拓展了该思想，Snell 等人[20]提出了原型网络(Prototypical Networks)，通过学习多个原型向量来表征每个类的中心，实现了多类场景下的紧凑包裹。

杨等人[21]在 2021 年提出一套行之有效的紧密包裹学习算法进行分类器优化，文中提出了一套系统性的同类特征区域紧密包裹面构造理论。该工作首次将“包裹面”作为一个可显式构造与求解的数学对象进行研究。其核心贡献在于提出“同类特征集合的紧密包裹集”概念，并构建了三阶段算法。杨等人的工作为紧密包裹学习提供了从理论定义到算法实现的完整链条，其“致密性参数”与“两阶段构造”机制为后续研究如何在保证覆盖完整性的前提下实现边界紧密性提供了清晰的解决路径。

对抗性紧凑包裹分类器学习用于开放集识别中提出的对抗紧密包裹分类学习(ACWCL)框架通过引入“紧密包裹点”(CWP)与分类器的对抗机制，实现包裹边界的动态优化。首先，深度超球冠分类器(DCHCC)在单位超球面上构建紧凑的分类区域；随后，CWP 生成器生成虚拟负样本点以紧密环绕正类特

征，与分类器进行对抗博弈——分类器力求正确分类并拒绝 CWP，而生成器则不断调整 CWP 位置以逼近决策边界，最终通过交替优化使边界紧密贴合真实分布，实现对凹形区域的适应与开放空间风险的显式控制。Zhang 等人的工作将生成对抗网络(GAN)的思想引入紧密包裹学习，开创了“用对抗方式优化包裹边界”的新范式，为构建更精确、更鲁棒的异常检测模型提供了全新的思路。

2.3. 对抗学习

对抗学习(Adversarial Learning)作为深度学习领域的重要分支，在最初的生成对抗网络中，其核心思想是通过两个或多个模型之间的动态博弈过程，实现对数据分布的更精细建模与决策边界的持续优化。近年来，该机制被广泛引入异常检测任务中，旨在提升模型对细微异常、分布外样本以及恶意攻击的识别能力。

对抗学习的理论根基建立在博弈论与优化理论之上，其核心是通过极小极大(minimax)优化框架实现两个网络之间的对抗性训练。最具代表性的模型是 Goodfellow 等人于 2014 年提出的生成对抗网络(GAN)，其目标函数为：

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}}\left[\log D(x)\right] + E_{z\sim p_z}\left[\log\left(1-D\left(G(z)\right)\right)\right] \tag{2}$$

该损失函数被称为原始 GAN 损失，其含义是：判别器 D 试图最大化正确区分真实样本与生成样本的概率，而生成器 G 试图最小化被识别为“假”的概率。这一框架首次实现了无需显式概率密度建模的高质量数据生成。然而，原始 GAN 存在训练不稳定、梯度消失等问题，尤其在生成器性能较差时，判别器可能过早饱和，导致生成器无法获得有效梯度。

为解决这一问题，Arjovsky 等人(2017)提出了 Wasserstein GAN (WGAN) [22]，采用 Wasserstein 距离(又称 Earth-Mover 距离)作为衡量生成分布与真实分布之间差异的指标，WGAN 通过缓解梯度消失问题，提升了梯度传播效率，使得生成器能够获得更稳定的性能。其损失函数为：

$$\min_G \max_{D\in\mathcal{F}} \mathbb{E}_{x\sim p_{data}}\left[D(x)\right] - \mathbb{E}_{z\sim p_z}\left[D\left(G(z)\right)\right] \tag{3}$$

WGAN 的关键贡献在于，其损失值与生成样本质量具有更强的相关性，且在训练过程中更加稳定，避免了模式崩溃和梯度消失问题。此外，Mao 等人[23]提出的最小二乘 GAN (LSGAN)采用最小二乘损失替代对数损失，其判别器损失为：

$$\mathcal{L}_D = \frac{1}{2}\mathbb{E}_{x\sim p_{data}}\left[\left(D(x)-1\right)^2\right] + \frac{1}{2}\left[\mathbb{E}_{z\sim p_z}\left(D\left(G(z)\right)-0\right)^2\right] \tag{4}$$

生成器损失为：

$$\mathcal{L}_G = \frac{1}{2}\mathbb{E}_{z\sim p_z}\left[\left(D\left(G(z)\right)-1\right)^2\right] \tag{5}$$

模型通过使用最小二乘损失替代传统 GAN 的 sigmoid 交叉熵损失，缓解了饱和区梯度消失问题，增强了模型稳定性。

对抗学习在异常检测中的应用主要体现在三个方面：利用对抗样本增强训练数据、通过对抗训练提升模型鲁棒性，以及借助对抗机制优化决策边界。或者将对抗样本生成引入异常检测框架，在训练中动态生成对抗扰动样本作为“合成异常”，迫使模型关注决策边界附近的敏感区域，从而在无真实异常标签的情况下实现主动数据增强。Madry 等人[24]提出的 PGD 对抗训练进一步将对抗学习系统化为一种正则化方法，通过多步投影梯度下降生成最坏情况下的扰动，提高模型对异常的判别能力。

另一条重要路径是利用 GAN 框架进行异常检测。Schlegl 等人提出的 AnoGAN 通过 GAN 模型重构正常数据，并以重构误差与判别器反馈误差作为异常评分，实现无监督分布外检测。为提升效率，f-AnoGAN 引入一个编码器网络直接学习从输入到潜在空间的映射，避免了测试时的迭代搜索。Zenati 等人[25]提出的 EfficientGAN 则通过联合训练生成器、判别器和编码器，实现端到端的异常检测。Deecke 等人[26]提出 AnoVAE，将变分自编码器与对抗学习结合，利用 KL 散度与对抗损失共同约束潜在空间结构，进一步提升了检测性能。

随着研究的深入，对抗学习与紧密包裹学习的融合成为趋势。传统方法(如 Deep SVDD)能几何包裹正常样本，但缺乏主动探索边界的能力。Chen 等人[27]提出对抗超球面聚类框架，在隐空间生成挑战样本以增强模型的边界鲁棒性。Zhang 等人提出对抗紧密包裹学习框架，通过主分类器与 CWP 生成器的动态博弈，使模型在生成的“边界样本”约束下逐步收缩接受域，提升拒绝未知类别的能力。Ruff 等人[28]的 Deep SAD 在半监督条件下利用少量异常样本辅助边界学习；Hendrycks 等人的 Outlier Exposure 方法则利用外部异常数据集扩展决策边界。这些研究共同表明，引入对抗或外部挑战信号能显著增强模型的开放集识别与鲁棒性。

3. 基于对抗紧密包裹与超球面约束的无监督异常检测模型(ACHC)

在本节中，我们提出了基于对抗紧密包裹与超球面约束的无监督异常检测模型(ACHC)，旨在解决现有方法在高维数据下边界模糊、鲁棒性不足及对未知异常泛化能力弱的问题。为有效建模正常类分布并增强对异常样本的拒绝能力，我们构建了一个端到端的学习框架。该框架以深度神经网络作为特征提取器，用于提取正常样本的特征表示。在此基础上，我们引入紧密包裹点集与对抗生成机制，通过优化正常样本到球心的欧氏距离，确保正常样本在特征空间中的聚集，从而实现几何包裹。此外，对抗学习通过生成边界附近的虚拟异常样本，增加训练的挑战性，推动决策边界向内收缩。分类器和生成器交替优化，使模型在学习过程中不断增强对细微异常的检测能力和对抗攻击的鲁棒性，提升模型的准确性与稳定性。最终，我们对所提出的 ACHC 及其理论基础进行全面阐述。

3.1. 紧密包裹点集(由紧密包裹点构造算法生成紧密包裹点集)

在无监督异常检测中，缺乏完整的异常样本集使得模型难以学习开放空间的边界特性，限制了其对未知异常的拒识能力。采集足够多样化的异常样本成本高且难以实现。通过分析现有异常检测框架并借鉴开集识别中负类增强的相关研究[29]，发现“边界异常”样本对优化决策边界更为有效，它们能更好地引导模型压缩接受区域，提高边界刻画精度。因此，紧密包裹学习算法为异常检测中构建紧密且泛化的正常类接受区域提供了新的解决思路。它通过在同类样本点集 C 的外围，沿各坐标轴方向生成距离为 ε 的候选探测点，并筛选出那些位于 C 的 $\varepsilon/\sqrt{rN}$ 致密单连通区域之外的点，从而形成一个“紧贴”样本分布轮廓的点集。这一集合有效避免了传统分类器(如超球 SVM)决策面过度扩张而侵占其他类别特征空间的问题，确保了决策边界既能完整包裹同类特征，又最大限度地减少了对未知类或邻近类的误判风险，为构建高正确识别率和具备明确拒识区域的分类器提供了精确的几何基础。紧密包裹点集示例如图 1 所示，其中图 1(a)为添加紧密包裹点之前的 ACHC 图 1(b)添加紧密包裹点图 1(c)添加紧密包裹点的 ACHC。

本文引入紧密包裹学习策略，应用于神经网络特征提取层，通过对抗训练生成边界性增强样本，使正常类特征聚拢，优化决策面，同时通过降低嵌入空间维度，提升算法效率与稳定性。致密连通的定义以及致密凸集上的紧密包裹学习定义如下：

致密 $S-\beta$ 连通定义：设 C 是 N 维特征空间 $\mathbb{R}^N$ 的一点集，ε 是大于零的小常数，r 是大于 1 的常数。设 S 是 C 中的一个某固定点。若以 S 和 C 中任意另外点为端点的连线都在 $C_{\varepsilon/\sqrt{rN}}$ 中，这些不同连线上任

意两点 A、B，以及以 S 为锥顶，线段 AS、BS 为锥侧线的圆锥面 Z_S 在 $C_{\varepsilon/\sqrt{rN}}$ 中，且存在由 Z_S 和的 $C_{\varepsilon/\sqrt{rN}}$ 面所包围的面积 $\tilde{Z}_S$ 中点 D，使线段线 AD、BD 为锥侧线，以 D 为锥顶的圆锥面，即以 $\angle ADB > \beta$ 为锥角的锥面(β 往往大于 $\pi/2$)，与 $C_{\varepsilon/\sqrt{rN}}$ 的面所围区域在 $C_{\varepsilon/\sqrt{rN}}$ 中，则称 C 是 $\varepsilon/\sqrt{rN}$ 致密 $S-\beta$ 凸集。

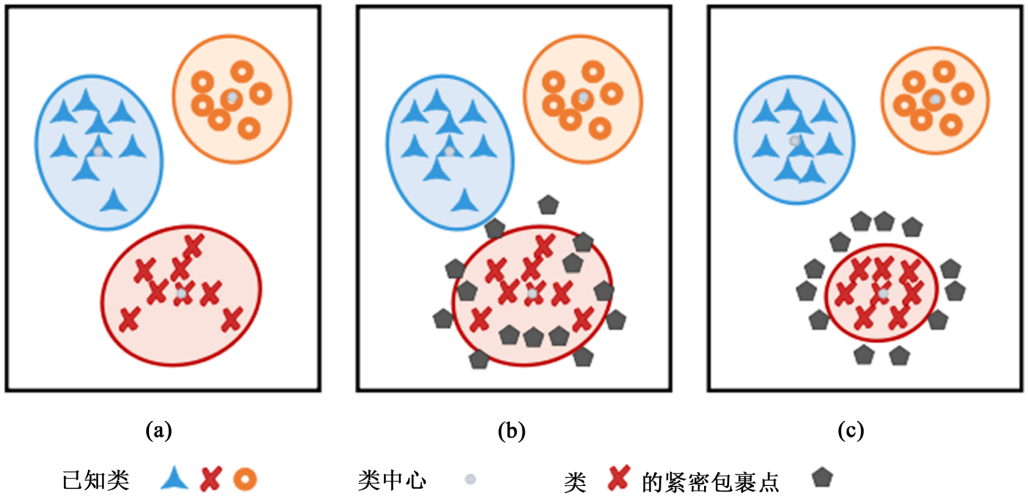

Figure 1. Example diagram of a compact packing point set
图 1. 紧密包裹点集示例图

实际的紧密包裹点数据增强如算法 3.1 所示。首先，根据致密连通定理，确定致密性参数的次优估计 ε。通过同类特征点集以及该点集的致密性参数 ε 构造同类特征训练点集的近似覆盖区域 $C_{\varepsilon/\sqrt{rN}}$。其次，依托紧密包裹集存在性条件，利用同类特征点集的 ε 派生点构造近似覆盖区域的紧密包裹点集。初步构造方法为使用致密性参数 ε 对同类特征点集进行数据增强，保留不在近似覆盖区域 $C_{\varepsilon/\sqrt{rN}}$ 的增强样本 $Z=\{z_i \mid i=1,2,\cdots,\kappa\}, z_i \in \mathbb{R}^d$。

算法 3.1：紧密包裹点构造方法

输入： 训练集 $\{(x_i,y_i) \mid i=1,\cdots,n\}, x_i \in \mathbb{R}^d, y_i \in \mathcal{Y}$。

1. 为类 $j \in \mathcal{Y}$ 计算点集 $\{x_i \mid y_i = j\}$ 的致密性参数 ε_j 的次优估计以及构造相应的类近似覆盖区域 Ω_j

2. 每个训练样本点 $x_i=(x_{i,1},x_{i,2},\cdots,x_{i,d})$ 派生出以下 $2d$ 点：

   $(x_{i,1},\cdots,x_{i,k} \pm \epsilon_{y_i},\cdots,x_{i,d}), 1 \le k \le d, k \in \mathbb{N}$

3. 按类别检测是否所有派生点 $(x_{i,1},\cdots,x_{i,k} \pm \epsilon_{y_i},\cdots,x_{i,d}), 1 \le i \le n, 1 \le k \le d$ 是否在近似覆盖区域 Ω_{y_i} 内。不在近似覆盖区域 Ω_{y_i} 内的派生点构成紧密包裹集：

   $Z=\{z_i \mid i=1,2,\cdots,\kappa\}, z_i \in \mathbb{R}^d$

3.2. 基于紧密包裹策略的超球面异常检测模型(CWD-SVDD)

我们提出了一种融合超球异常检测与紧密包裹学习的异常检测建模方法，旨在克服现有深度异常检测方法在特征空间中决策边界模糊、异常样本易与正常样本特征重叠的问题。本文在 DSVDD 框架基础

上创新性地引入紧密包裹学习机制，通过构建和优化边缘异常暴露集，显式最大化正类样本与异常样本之间的球面距离，从而将异常特征有效推离正常特征域。该策略不仅增强了模型对异常样本的排斥能力，也显著降低了特征重叠的概率。我们的模型采用 ResNet 作为主干网络进行高维特征提取，并在隐空间中构造一个同心单超球体，用于包裹正常样本并排斥异常样本。整个框架实现了从“被动包裹”到“主动排斥”的范式转变，提升了异常检测的准确性与鲁棒性。

我们提出的模型核心思想是将异常检测建模为一个联合优化问题，同时优化正常样本的紧密性、异常样本的排斥性以及网络参数的稳定性。对 CWD-SVDD 模型，给定输入空间 $x\in\mathbb{R}^d$ 的正常训练数据集 $D_{in}=\{x_1,\cdots,x_m\}$，和异常数据集 $D_{out}=\{x_1,\cdots,x_n\}$，其中 m 和 n 分别表示正常样本和异常样本的数量，输出空间 $F\in\mathbb{R}^p$，网络的连接权重为 $w=\{w_1,w_2,\cdots,w_l\}$，表示通过网络 $\Phi(x;W)$ 映射得到的参数数据。我们方法的目标损失函数可以表述如下：

$$\begin{aligned}&R^2+\frac{1}{\upsilon n}\sum_{i=1}^{n}\max\left\{0,\left\|\phi(x_i;W)-c\right\|^2-R^2\right\}+\frac{\lambda}{2}\sum_{l=1}^{L}\left\|W^l\right\|_F^2\\&+\frac{1}{\mu m_1}\sum_{k_0\text{为紧密包裹点集类},i=1}^{m_1}\max\left\{0,R^2-\left\|\phi\left(x_i^{k_0};W\right)-c\right\|^2\right\}\\&+\frac{1}{\mu m_2}\sum_{k_1\text{为其他未知类},j=1}^{m_2}\max\left\{0,R^2-\left\|\phi\left(x_j^{k_1};W\right)-c\right\|^2\right\}\end{aligned}\tag{6}$$

该损失函数由五个关键项构成，每一项都对应特定的优化目标，共同引导模型学习鲁棒且判别性强的特征表示。第一项 R^2 是超球体积最小化项，旨在驱动模型寻找一个体积最小的超球来包裹所有正常样本。第二项用于处理训练数据中可能存在的噪声或边界正常样本。参数 ν 为正则化系数。第三项旨在防止模型过拟合，提升泛化能力。第四项是紧密包裹点集的排斥项，其核心作用是主动引导决策边界向内收缩。这些点集样本通过对正常样本施加轻微扰动生成，模拟那些在特征空间中接近正常类但又不属于该类的“最难区分”异常。当这些样本落入超球内部时，该项激活并产生梯度，推动其远离中心，从而迫使超球边界压缩，提升模型的判别精度。第五项则利用来自其他未知类别的真实样本作为负例，进一步增强模型对语义合理但分布外实例的拒绝能力。避免将语义相近但非目标类别的样本误判为正常。

3.3. 结合对抗生成网络的紧密超球面异常检测模型(ACHC)

在构建基于紧密包裹策略的超球面异常检测模型的过程中，对抗生成网络(GAN)的引入为模型提供了进一步增强异常检测能力的重要机制。与传统 GAN 不同，本文不追求生成样本的视觉真实性，而是利用生成器 G 作为“边界探索者”，主动生成逼近当前正常类决策边界的紧密包裹点，以实现对特征空间的主动压缩与紧密建模。

具体而言，生成器 G 以随机噪声 $z\sim P_z(z)$ 为输入，生成伪样本 x_g，其目标并非欺骗判别器，而是尽可能逼近由判别器 D 所学习的超球边界 $\left\|\phi(x)-c\right\|^2=R^2$。判别器(即主网络)在训练中一方面将正常样本 x_n 的特征映射紧密化至超球内部，最小化 $\left\|\phi(x_n)-c\right\|^2$；另一方面，对生成样本 x_g 施加排斥项，最大化 $\left\|\phi(x_g)-c\right\|^2$，防止其进入正常区域。二者在对抗过程中形成博弈：生成器不断调整策略，生成更接近边界的挑战样本；判别器则持续收紧超球半径 R 并优化中心 c，迫使特征空间向更紧凑、更具判别性的方向演化。

为了实现这一目标，生成器通过深度卷积神经网络将低维噪声映射到高维数据空间，生成具有真实数据特性的伪样本。在训练中，生成器专注于探索判别器学习的超球边界附近的脆弱区域，动态调整生成样本，以便更难被判别器识别。通过这种对抗性交互，生成器与判别器协同进化，提升边界探索能力和建模精度，最终增强模型对未知异常模式的泛化和拒识能力。我们方法的目标损失函数可以表述如下：

$$\omega_1\sum_{i=1}^{n}\max\left\{0,R^2-\phi\left(G\left(z_i\right);W\right)-c^2\right\}-\omega_2\sum_{i=1}^{n}\phi\left(G\left(z_i\right);W\right)-\phi\left(x_i;W\right)^2 \tag{7}$$

第一项为生成样本的排斥项。该损失项要求生成的样本 $G\left(z_i\right)$ 在特征空间中应位于超球体之外。若落入超球内则产生惩罚，促使模型调整以排斥生成样本。第二项是生成样本与正常样本的相似度项。该损失项鼓励生成样本在特征空间中尽可能接近正常样本但不重合。通过计算特征差异并作为惩罚项，帮助模型在特征空间中形成有效的“边界模糊区”，识别潜在异常模式。

在实际训练过程中，对抗生成网络的训练是一个动态博弈的过程。生成器 G 和异常检测模型交替进行优化。具体而言，生成器的目标是生成能够欺骗异常检测模型的样本，而异常检测模型的目标则是准确区分正常样本和生成样本，同时保持对真实异常样本的识别能力。在每一轮迭代中，首先固定异常检测模型的参数，优化生成器 G，使其生成的样本能够最大程度地挑战当前决策边界。然后固定生成器的参数，优化异常检测模型，使其能够准确区分正常样本和生成样本，并保持对真实异常样本的识别能力。通过这种方式，生成器和异常检测模型在不断的博弈中逐渐提升各自的性能，最终达到一个相对稳定的平衡状态。最终，我们的整体损失函数定义如下：

$$\begin{aligned}
L_{ACHC}=\min_{R,c,W}\max_{G}\Bigg(&R^2+\frac{1}{\upsilon n}\sum_{i=1}^{n}\max\left\{0,\left\|\phi\left(x_i;W\right)-c\right\|^2-R^2\right\}+\frac{\lambda}{2}\sum_{l=1}^{L}\left\|W^l\right\|_F^2\\
&+\frac{1}{\mu m_1}\sum_{k_0\text{为紧密包裹点集类},i=1}^{m_1}\max\left\{0,R^2-\left\|\phi\left(x_i^{k_0};W\right)-c\right\|^2\right\}\\
&+\frac{1}{\mu m_2}\sum_{k_1\text{为其他未知类},j=1}^{m_2}\max\left\{0,R^2-\left\|\phi\left(x_j^{k_1};W\right)-c\right\|^2\right\}\\
&+\omega_1\sum_{i=1}^{n}\max\left\{0,R^2-\left\|\phi\left(G\left(z_i\right);W\right)-c\right\|^2\right\}\\
&-\omega_2\sum_{i=1}^{n}\left\|\phi\left(G\left(z_i\right);W\right)-\phi\left(x_i;W\right)\right\|^2\Bigg)
\end{aligned} \tag{8}$$

$$\begin{aligned}
&\text{S.T.}\quad \left\|\phi\left(x_i;W\right)-c\right\|^2-R^2\le 0,1\le i\le n,\\
&\left\|\phi\left(x_i^{k_0};W\right)-c\right\|^2-R^2\ge 0,\left\|\phi\left(x_i^{k_1};W\right)-c\right\|^2-R^2\ge 0,\\
&k_0\text{ 为紧密包裹点集类};\left\|\phi\left(G\left(z_i\right);W\right)-c\right\|^2-R^2\ge 0,\\
&k_1\text{ 为其他未知点类}\left(\text{异常点暴露生成}\right),\\
&G\left(z_i\right)\text{为 GAN 网络对应生成点}.
\end{aligned}$$

该模型结合超球异常检测、紧密包裹学习和对抗生成机制，提升了判别能力。通过紧密性约束和显式排斥异常样本，确保正常数据高度聚集，排除潜在异常。引入对抗生成网络生成挑战性样本，增强对细微异常和未知模式的敏感性，避免边界松弛。正则化项保障训练稳定性，防止过拟合，整体实现联合优化，在高维数据中表现出更强的鲁棒性和泛化能力，优于传统单类分类方法。

3.4. 异常标准

为了对样本进行正常或异常分类，我们的模型异常评判标准建立在优化后的特征空间结构之上，其核心依据为样本到正常类中心的欧氏距离是否超出预设的超球半径阈值：

$$\boldsymbol{s}\left(x\right)=\left\|\phi\left(x;\boldsymbol{W}^*\right)-\boldsymbol{c}^*\right\|^2 \tag{9}$$

若 $S\left(x\right)>R^{*2}$，则判定该样本为异常。该评判机制利用紧密包裹集与对抗生成共同塑造的边界特性，

确保正常类边缘样本被充分考虑，避免传统方法边界松弛。生成器提供逼近边界的伪异常样本，迫使判别器压缩接受区域，提升判别精度与稳定性。损失函数中的正向惩罚项增强了模型对未知分布的敏感性，确保其在开放世界场景下有效识别未见异常

算法 3.2 详细描述了基于对抗紧密包裹与超球面约束的异常检测模型的训练流程。该算法采用交替优化策略，首先更新生成器生成更具挑战性的潜在异常样本，然后联合优化特征网络和超球参数，确保正常样本紧密包裹并排斥异常样本。训练过程中动态调整决策边界，测试阶段通过计算样本与超球中心的距离实现高效异常判别。

算法 3.2：基于紧密包裹与对抗生成的异常检测训练算法

输入：训练集正常样本集 $\mathcal{X}_n=\{x_i\}_{i=1}^{n}$

- 紧密包裹点集 $\mathcal{X}_{cwp}=\{x_i^{k_0}\}_{i=1}^{m_1}$
- 未知类*样本*集 $\mathcal{X}_{unknown}=\{x_j^{k_1}\}_{j=1}^{m_2}$
- 初始网络参数 W、超球中心 c、生成器 $G(z)$、半径 R
- 超*参数*：$\nu,\mu,\lambda,\omega_1,\omega_2$，学习率 η
- 最大训练轮数 T

输出：

- 优化后的网络参数 W^*，超球中心 C^* 与半径 R^*
- 异常评分函数

初始化：

1.
- 使用预训练 ResNet 网络初始化特征提取器 $\phi(\cdot;W)$
- 在正常样本集上计算初始中心。
- 初始化生成器 $G(z)$ 为随机噪声映射网络
- 设置迭代计数器 $t=0$

2. **迭代训练($t=1$ 到 T)：**

1) 固定生成器 G 及其参数，训练判别器 D：
- 随机采样一批正常样本 $\mathcal{X}_i\in\mathcal{X}_n$，紧密包裹点 $x_i^{k_0}\in\mathcal{X}_{cwp}$，其他未知类样本 $x_j^{k_1}\in\mathcal{X}_{unknown}$
- 对每个样本前向传播得到特征表示后，计算判别器损失。
- 执行随机梯度下降更新：

$W\leftarrow W-\eta\nabla_W\mathcal{L}_D \quad c\leftarrow c-\eta\nabla_c\mathcal{L}_D \quad R\leftarrow R-\eta\nabla_R\mathcal{L}_D$

2) 固定判别器 D 及其参数，训练生成器 G：
- 从噪声先验 $P_z(z)$ 中采样一批噪声 $z_i\sim\mathcal{N}(0,I)$
- 生成伪样本 $x_g=G(z_i)$。
- 提取特征后计算生成器损失。
- 最大化生成器目标，执行梯度上升：

$G\leftarrow G+\eta\nabla_G\mathcal{L}_G$

3) 执行(1)和(2)，开始交替循环：
- 在每一轮迭代中，先执行(1)更新判别网络，再执行(2)更新生成网络。
- 重复此过程，直到 $W^{(t)}-W^{(t-1)}<\varepsilon, c^{(t)}-c^{(t-1)}<\varepsilon, R^{(t)}-R^{(t-1)}<\varepsilon$，或达到最大迭代次数 T。

3. **输出最优模型参数：W^*,c^*,R^***

续表

4.	**异常评分** 对测试样本 x，计算其与中心距离： $s(x)=\left\|\phi\left(x;W^{*}\right)-c^{*}\right\|^{2}$ 若 $S(x)>R^{*2}$，则判为异常。

4. 实验

在这一部分中，在多个常用的异常检测基准数据集上对我们的方法进行了评估。我们在四个公开可用的数据集——CIFAR-10、CIFAR-100、MVTec AD [30]和 DIOR [31]上进行了实验，报告了我们的方法在标准基准数据集上的单类分类(OCC)结果。为了评估检测性能，我们采用 ROC-AUC 指标作为性能评分[32]。并与之前几种成熟的算法 DeepSVDD、Autoencoder、OCSVM、CSI [33]和 PANDA [34]，以及最近的算法 MSC [35]进行了比较，以验证所提出的 AD 方法的有效性。此外，我们进一步分析了所提出的目标函数，并进行了消融实验，以考察各个组件的贡献。实验结果表明该模型在较复杂的场景下均获得了良好的性能和鲁棒性。实验结果的细节如下所示。

4.1. 实验结果

如表 1 所示，本文提出的基于对抗紧密包裹与超球面约束的异常检测模型(OURS)在所有测试数据集上均取得了最优性能，显著优于 DeepSVDD 及 MSC 等主流方法，充分体现了其在复杂场景下的鲁棒性与泛化能力。模型性能的提升得益于所引入的对抗学习与紧密包裹学习机制，通过生成器与判别器的博弈训练，模型能够更精确地学习正常样本的边界分布，避免常见的模式崩溃问题。此外，超球面约束有效引导特征空间向紧凑、可分的方向演化，使高维多类背景下的正常样本仍能被紧密聚集，提升对未知类别的敏感度。

Table 1. The AUROC (%) performance of different algorithms on different datasets
表 1. 不同算法在不同数据集上的 AUROC (%)性能

Method	CIFAR10	CIFAR100	MVTec AD	DIOR
DeepSVDD	64.8%	67.0%	62.4%	68.5%
Autoencoder	68.1%	69.3%	58.7%	65.4%
OCSVM	66.5%	64.2%	55.1%	62.8%
CSI	94.3%	89.6%	63.6%	78.5%
PANDA	96.2%	94.1%	86.5%	94.3%
MSC (ResNet152)	97.2%	96.4%	87.2%	97.7%
OURS	**97.7%**	**98.2%**	**88.5%**	**98.4%**

如表 2 所示，在小样本数据集下，我们的模型也达到了较好的效果，表明其对有限数据具有高效利用能力。

如表 3 所示，在 ResNet [36]、EfficientNet [37]、DenseNet [38]和 ViT 四种骨干网络上，我们的模型

均取得最高或接近最高性能，验证了方法的结构兼容性与泛化能力。综合实验表明，对抗生成机制与紧密包裹学习约束的协同作用，有效增强了模型在多样场景下的异常判别能力。

Table 2. Anomaly detection accuracy (average ROC-AUC%) on small datasets
表 2. 在小数据集上的异常检测精度(平均 ROC-AUC%)，最好的用粗体

Dataset	CSI	PANDA	DeepSVDD	OCSVM	Autoencoder (AE)	MSC (ResNet18)	OURS
CIFAR-10 (100samples)	86.5	95.8	81.2	75.3	78.9	89.5	**98.5**
CIFAR-10 (200 samples)	90.1	97.1	84.5	79.8	82.3	92.0	**98.1**
CIFAR-10 (500 samples)	81.3	95.4	88.2	83.5	86.1	93.1	**96.9**
CIFAR-100 (200 samples)	84.2	93.1	82.7	77.6	80.5	90.8	**97.2**
CIFAR-100 (500samples)	88.1	95.6	85.6	80.9	83.2	93.8	**97.8**

Table 3. Performance improvement under different network architectures (CIFAR-10, average ROC-AUC%)
表 3. 不同网络体系结构下的性能提高(CIFAR-10，平均 ROC-AUC%)

Method	ResNet	EfficientNet	DenseNet	ViT
DN2	92.5	89.3	85.6	95.7
PANDA	96.2	95.3	82.4	95.8
MSC	97.2	97.0	95.7	98.6
DeepSVDD	84.5	82.1	80.3	88.2
Autoencoder (AE)	82.3	80.7	78.9	84.1
OURS	**98.1**	**98.7**	**97.6**	**98.7**

4.2. 超参数分析

在本研究中，所提出的模型基于一个复杂的多目标损失函数，旨在同时优化异常检测的判别能力、生成一致性与类间边界约束。该损失函数包含四个关键超参数：α_1, α_2, v 和 μ，分别控制不同正则化项的权重。为深入理解这些超参数对模型性能的影响，本文开展了系统的超参数敏感性分析实验。实验结果如图 2 所示，展示了各超参数变化对模型异常检测性能的影响趋势。从图 2 中可以看出，所有四个超参数均表现出明显的非线性响应特性，且存在最优取值点，表明模型性能对超参数的选择具有一定的依赖性。

5. 总结

本研究提出了一种基于对抗紧密包裹与超球面约束的新型异常检测模型，通过将超球学习、对抗生成网络与紧密包裹机制有机结合，构建了一个统一的端到端学习框架。该模型在特征空间中引导正常样本紧密聚集于低体积超球体内，同时利用对抗学习增强边界判别能力，有效提升了对未知异常的敏感性。

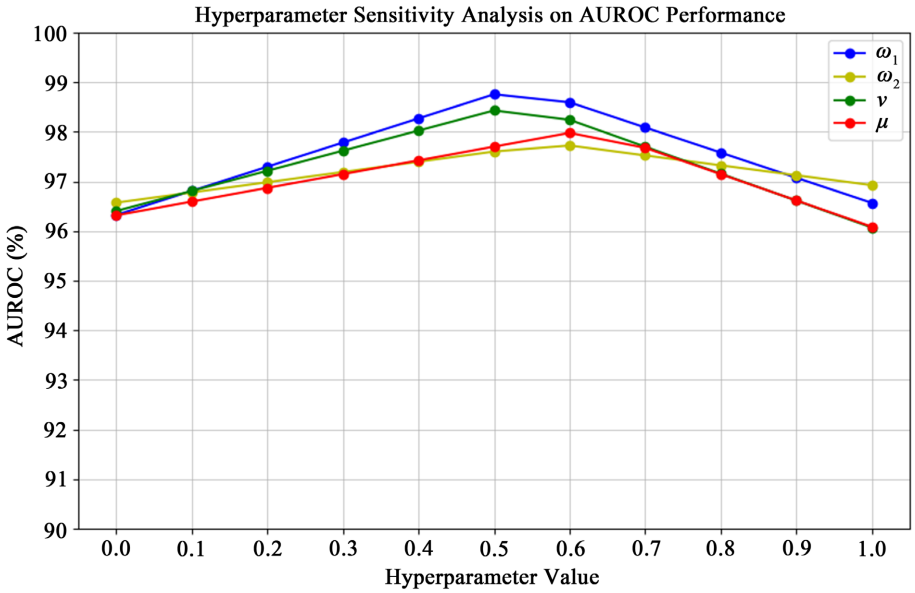

Figure 2. The influence of hyperparameters on model anomaly detection
图 2. 超参数对模型异常检测的影响

在 CIFAR-10、CIFAR-100、MVTec AD 和 DIOR 等多个公开数据集上的实验结果表明，本方法在常规、小样本及不同骨干网络设置下均取得最优性能。

参考文献

[1] Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C. (2001) Estimating the Support of a High-Dimensional Distribution. *Neural Computation*, **13**, 1443-1471. https://doi.org/10.1162/089976601750264965

[2] Liu, F.T., Ting, K.M. and Zhou, Z. (2008). Isolation Forest. 2008 8*th IEEE International Conference on Data Mining*, Pisa, 15-19 December 2008, 413-422. https://doi.org/10.1109/icdm.2008.17

[3] Parzen, E. (1962) On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, **33**, 1065-1076. https://doi.org/10.1214/aoms/1177704472

[4] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the *em* Algorithm. *Journal of the Royal Statistical Society Series B*: *Statistical Methodology*, **39**, 1-22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

[5] Duda, R.O., Hart, P.E. and Stork, D.G. (2001) Pattern Classification. 2nd Edition, Wiley-Interscience.

[6] Vincent, P., Larochelle, H., Bengio, Y. and Manzagol, P. (2008) Extracting and Composing Robust Features with Denoising Autoencoders. *Proceedings of the* 25*th International Conference on Machine Learning*, Helsinki, 5-9 July 2008, 1096-1103. https://doi.org/10.1145/1390156.1390294

[7] Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U. and Langs, G. (2017) Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. In: Niethammer, M., *et al*., Eds., *Information Processing in Medical Imaging*, Springer International Publishing, 146-157. https://doi.org/10.1007/978-3-319-59050-9_12

[8] Ruff, L., Vandermeulen, R.A., Görnitz, N., Binder, A., Müller, E., Müller, K. and Kloft, M. (2018) Deep One-Class Classification. Proc. *Proceedings of the* 35*th International Conference on Machine Learning*, Stockholm, July 2018, Vol. 80, 4393-4402.

[9] Kim, Y., Lee, S. and Hwang, S.J. (2023) Hypersphere Embedding with Angular Margin for One-Class Classification. in *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, June 2023, 14567-14576.

[10] Zhang, Z., Chen, X., Li, J. and Liu, Y. (2024) Adversarial Compact Wrapping Classifier Learning for Open Set Recognition. *The IEEE Transactions on Neural Networks and Learning Systems*, **35**, 345-358.

[11] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014) Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, Montreal, Vol. 27, 2672-2680.

[12] Hendrycks, D., Mazeika, M. and Dietterich, T.G. (2019) Deep Anomaly Detection with Outlier Exposure. https://arxiv.org/abs/1812.04606

[13] Sabokrou, M., Khalooei, M., Fathy, M. and Adeli, E. (2018) Adversarially Learned One-Class Classifier for Novelty Detection. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Las Vegas, 18-22 June 2018, 3379-3388. https://doi.org/10.1109/cvpr.2018.00356

[14] Tax, D.M.J. and Duin, R.P.W. (2004) Support Vector Data Description. *Machine Learning*, **54**, 45-66. https://doi.org/10.1023/b:mach.0000008084.60811.49

[15] Xu, B., Wu, X., Wang, P. and Chen, L. (2015) Convolutional Autoencoder for Video Anomaly Detection. *IEEE Signal Processing Letters*, **22**, 1791-1795.

[16] Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G. and Schmidt-Erfurth, U. (2019) f-AnoGAN: Fast Unsupervised Anomaly Detection with Generative Adversarial Networks. *Medical Image Analysis*, **54**, 30-44. https://doi.org/10.1016/j.media.2019.01.010

[17] Cevikalp, H., Trutschel, M. and Trancón y Widemann, B. (2020) Deep Support Vector Data Description for Anomaly Detection. 15*th International Conference on Computer Vision Theory and Applications* (*VISAPP*), Valletta, 27-29 February 2020, 532-539.

[18] Russell, S.J., Moreira, L.C., Pimentel, M.A., Bentley, P.J. and Constantinides, G.A. (2016) Anomaly Detection in Wearable Medical Sensors Using Extreme Value Theory. *IEEE Engineering Medicine and Biology Society* (*EMBC*), Orlando, 16-20 August 2016, 2122-2125.

[19] Cevikalp, H. and Trancón y Widemann, B. (2017) Deep Compact Hypersphere Classification. 16*th International Conference on Machine Learning and Applications* (*ICMLA*), Cancun, 18-21 December 2017, 387-392.

[20] Snell, J., Swersky, K. and Zemel, R. (2017) Prototypical Networks for Few-Shot Learning. 31*st International Conference on Neural Information Processing Systems* (*NeurIPS*), Long Beach, 4-9 December 2017, 4077-4087.

[21] 杨国为, 万鸣华, 赖志辉, 等. 具有合适拒识机制的高正确识别率分类器设计[J]. 电子学报, 2021, 49(8): 1569.

[22] Arjovsky, M., Chintala, S. and Bottou, L. (2017) Wasserstein Generative Adversarial Networks. *Proceedings of the* 34*th International Conference on Machine Learning*, Vol. 70, 214-223.

[23] Mao, X., Li, Q., Xie, H., Lau, R.Y.K., Wang, Z. and Smolley, S.P. (2017) Least Squares Generative Adversarial Networks. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, October 2017, 2794-2802. https://doi.org/10.1109/iccv.2017.304

[24] Madry, A., Makelov, A., Schmidt, L., Tsipras, D. and Vladu, A. (2018) Towards Deep Learning Models Resistant to Adversarial Attacks. *International Conference on Learning Representations* (*ICLR*), Vancouver, 30 April-3 May 2018. https://arxiv.org/abs/1706.06083

[25] Zenati, M., Foo, C.-S., Lecouat, B., Manek, G. and Lee, V.R.P. (2018) Adversarially Regularized Autoencoders for Generating Realistic Images. *International Joint Conference on Artificial Intelligence*, Stockholm, 13-19 July 2018, 5338-5344.

[26] Deecke, L., Vandermeulen, R.A., Ruff, L., Mandt, S. and Kloft, M. (2018) Anomaly Detection with Robust Deep Variational Autoencoders. *The* 24*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Londo, 19-23 August 2018, 1038-1047.

[27] Chen, L., Wang, Z., Zhu, F. and Yuen, P.C. (2024) Adversarial Hypersphere Clustering for Unsupervised Anomaly Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **46**, 1892-1905.

[28] Ruff, L., Vandermeulen, R.A., Görnitz, N., Binder, A., Müller, E., Müller, K.-R. and Kloft, M. (2019) Deep Semi-Supervised Anomaly Detection. in Proc. 7*th International Conference on Learning Representations*, New Orleans, 6-9 May 2019. https://arxiv.org/abs/1906.02694

[29] Cevikalp, H., Uzun, B., Salk, Y., Saribas, H. and Köpüklü, O. (2023) From Anomaly Detection to Open Set Recognition: Bridging the Gap. *Pattern Recognition*, **138**, Article ID: 109385. https://doi.org/10.1016/j.patcog.2023.109385

[30] Bergmann, P., Fauser, M., Sattlegger, D. and Steger, C. (2019) MVTec AD—A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 9584-9592. https://doi.org/10.1109/cvpr.2019.00982

[31] Li, G., Liu, W., Zhou, L. and Li, H. (2019) DIOR: A New Dataset for Object Detection in Optical Remote Sensing Images. *ISPRS Journal of Photogrammetry and Remote Sensing*, **151**, 119-132.

[32] Fawcett, T. (2006) An Introduction to ROC Analysis. *Pattern Recognition Letters*, **27**, 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

[33] Tack, J., Mo, S., Jeong, J. and Shin, J. (2020) CSI: Novelty Detection via Contrastive Learning on Distributionally Shifted Instances. *Advances in Neural Information Processing Systems* (NeurIPS), Vol. 33, 11839-11852.

[34] Reiss, T., Cohen, N., Bergman, L. and Hoshen, Y. (2021) PANDA: Adapting Pretrained Features for Anomaly Detection and Segmentation. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 19-25

June 2021, 2806-2814. https://doi.org/10.1109/cvpr46437.2021.00283

[35] Reiss, T. and Hoshen, Y. (2023) Mean-Shifted Contrastive Loss for Anomaly Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, **37**, 2155-2162. https://doi.org/10.1609/aaai.v37i2.25309

[36] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[37] Tan, M. and Le, Q.V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the* 36*th International Conference on Machine Learning*, Long Beach, 9-15 June 2019, 6105-6114.

[38] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 4700-4708. https://doi.org/10.1109/cvpr.2017.243