基于协同对抗增强生成模型的智能无人机网络异常检测方法

隋嚞[1]，马春燕[1]，龙岭春[1]，顾兆军[2]，刘佳佳[3]，丁磊[4]
（1.中国民航大学航空工程学院，天津 300300；2.中国民航大学信息安全测评中心，天津 300300；
3.中北大学机械工程学院，山西 太原 030051；4.广州大学网络空间安全学院，广东 广州 510006）

摘　要：针对智能无人机网络异常检测中数据多类别不平衡的问题，提出了一种基于协同对抗增强生成模型的异常检测方法。生成器中采用动态类标签概率向量，逐渐增加少数异常样本增扩概率，并通过权值共享与“微调”机制，提高多生成器训练的稳定性和学习效率。判别器和分类器中，引入具有特征缩聚和激励模块的编码器，重新校准特征权重，显著提高小样本场景下关键特征提取能力。训练策略上，提出分类器和判别器共同引导异常样本生成的协同对抗机制，纠正生成数据与真实样本分布偏差。在4个开源数据集上与5种基线方法对比实验的结果表明，所提方法可将F_1分数提高3%，AUC值提升5%，G-mean值提升10%，Friedman测试和Nemenyi后续检验结果也证明了所提方法具有显著正向差异性。

关键词：无人机网络；异常检测；生成对抗模型；多类别不平衡
中图分类号：TP391.0
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2025218

Anomaly detection method for intelligent UAV networks based on collaborative adversarial enhanced generative model

SUI He[1], MA Chunyan[1], LONG Lingchun[1], GU Zhaojun[2], LIU Jiajia[3], DING Lei[4]
1. College of Aeronautical Engineering, Civil Aviation University of China, Tianjin 300300, China
2. Information Security Evaluation Center, Civil Aviation University of China, Tianjin 300300, China
3. School of Mechanical Engineering, North University of China, Taiyuan 030051, China
4. School of Cyber Science and Technology, Guangzhou University, Guangzhou 510006, China

Abstract: To address the problem of multi-class imbalance in the anomaly detection of intelligent UAV networks, a collaborative adversarial enhanced generative model-based anomaly detection method was proposed. For generators, dynamic class label probability vectors were adopted to gradually increase the probability of minority anormal sample. Moreover, through weight sharing and “fine-tuning” mechanism, the stability and learning efficiency of the multiple generators training were improved. For discriminators and the classifier, encoders with feature aggregation and excitation module were designed. The feature weights were recalibrated, significantly enhancing the model’s ability to extract key features in the few-shot learning scenario. For training strategy, a collaborative adversarial mechanism was proposed, in which the classifier and discriminators jointly guide the generation of samples. The distribution bias between the generated samples and the real ones were effectively corrected. A series of comparative experiments were conducted on four open-source datasets, against five baseline methods. Results show that the proposed method increases the F_1, AUC and G-mean by 3%, 5% and 10%, respectively. The results of the Friedman tests and Nemenyi post-hoc tests also demonstrate that the proposed method exhibits a significant positive difference.

Keywords: UAV network, anomaly detection, generative adversarial model, multi-class imbalance

收稿日期：2025–08–19；**修回日期：**2025–11–05
通信作者：隋嚞，hsui@cauc.edu.cn
基金项目：国家自然科学基金资助项目（No.U2333201）
Foundation Item：The National Natural Science Foundation of China (No.U2333201)

0　引言

无人机（UAV, unmanned aerial vehicle）、卫星、地面基站以及边缘服务器在人工智能（AI, artificial intelligence）的推动形成了全新的智能无人机网络[1]，已在航拍图像检测[2]、农业病虫害防护[3]、环境保护监测[4]等多个领域广泛应用，引起了业界的广泛关注。

传统的无人机网络多面向特定任务和封闭场景。与之不同，智能无人机网络是一类具有自主感知和决策能力的智能无人系统[5]。非结构化、未知、动态和开放的任务环境要求智能无人机网络具有自主学习能力，可以在与外界交互过程中提取有效信息，实时调整和优化自身行为策略[6]。作为典型的智能信息物理系统，智能无人机网络除了面临的无人机功能故障、通信线路故障、信号干扰等传统风险外，更容易受到复杂多变的安全威胁[7]，如图 1 所示。这些威胁不仅会影响智能无人机网络的通信，破坏其任务执行，甚至可能导致无人机坠毁等航空安全事故，已经成为低空安全面临的严峻挑战[8]。异常检测是解决这一问题的关键技术，为监控智能无人机网络异常行为和自动识别潜在高危风险提供了有效技术手段。然而，智能无人机网络异常检测中样本更稀缺、类型更多样、异常特征更隐蔽，如何在不充足的历史样本中学习到高隐蔽的异常模式，并精准定位多样化的异常类别，是智能无人机网络异常检测面临的新难题。

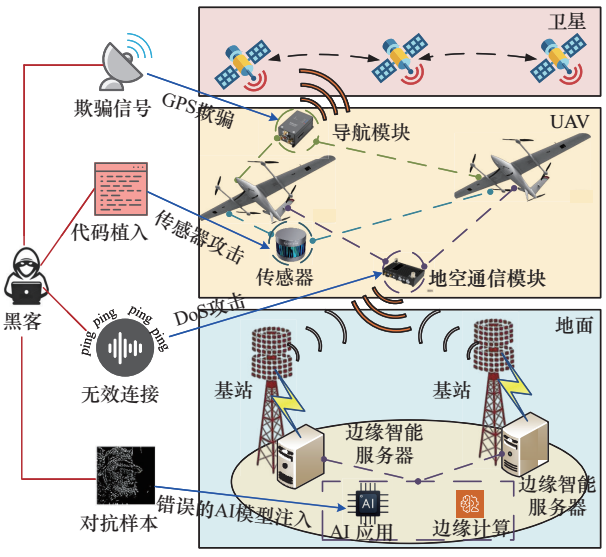

图 1　智能无人机网络面临的安全威胁

近年来，研究人员针对无人机网络提出了多种异常检测方法，主要可概括为基于知识的方法[9]、基于模型的方法[10]和数据驱动的方法[11]。然而，智能无人机网络异常检测中面临的多类别不平衡情况，是由多分类、小样本、不平衡、分布偏差等问题耦合形成的复杂难题。首先，已有异常检测方法多依赖于对大量历史数据蕴含异常模式的充分学习。但在智能无人机网络中，很难获取足够的异常样本，而且样本标注成本很高，导致小样本学习问题[12]。其次，在智能无人机网络中，异常样本的数量远远少于正常样本，检测模型容易忽略异常样本以保持总体性能，即类不平衡问题[13]。再次，通过传统采样方法生成的新数据可能会由于异常样本的罕见性而呈现出与真实样本分布不一致的现象，从而使检测模型产生错误的结果，即分布偏差问题[14]。此外，智能无人机网络异常检测是典型的多分类任务场景，既要考虑故障导致的异常（如无人机本体故障、通信网络故障等），也要考虑网络攻击导致的异常（如传感器攻击、拒绝服务攻击（DoS, denial of service）攻击等）。正是上述智能无人机网络异常检测方面的特殊性，导致传统异常检测方法未能很好匹配实际任务需求。

生成对抗网络（GAN, generative adversarial network）[15]具有强大的推理和学习能力，可以从有限的输入提示中生成高质量的数据，在各个领域已显示出显著的应用前景。然而，现有基于 GAN 的异常检测研究主要集中在二分类任务场景[16]。在多分类任务场景中，不同类别对之间的不平衡比（IR, imbalance ratio）存在显著差异，某一异常类别数据分布上的小偏差也会严重影响模型总体检测精度。

为了解决上述问题，本文在条件生成对抗网络（CGAN, conditional GAN）[17]基础上，提出了一种用于智能无人机网络多类别不平衡数据异常检测的协同对抗增强生成模型（CAE-GAN, collaborative adversarial enhanced GAN）。该模型从异常数据增扩、关键特征提取和分布偏差纠正多个层面协同强化异常检测能力。本文的主要工作如下。

1) 在数据生成方面，设计了动态类标签概率向量，在对抗训练中逐渐提高数量更少的异常类样本的生成概率，以实现对数据集类别平衡性的动态调控。同时，在不同生成器之间采用权重共享和

“微调”机制，提升多生成器训练的学习效率和稳定性，解决多类型异常数据增扩质量参差不齐的问题。

2) 在模型结构方面，在判别器和分类器中引入特殊设计的编码器（En, encoder），其中的特征缩聚和激励模块通过放大有效特征权重并减少无效或低效特征权重，为判别器判别真伪和分类器检测分类提供更为关键的深度特征，解决小样本学习特征提取不充分的问题。

3) 在学习策略方面，提出了一种分类器和判别器共同引导样本生成的协同对抗训练（CAT, collaborative adversarial training）策略，并设计一种新的损失函数来支持这种协同对抗训练策略的实施，实现分布偏差的协同感知，解决生成数据与真实样本分布偏差的问题。

1　相关工作

面向无人机网络的异常检测方法可概括为3类：知识驱动的方法[18]、模型驱动的方法[19]和数据驱动的方法[20]。知识驱动的方法依赖先验知识，而模型驱动的方法在建立准确物理模型上面临挑战。同时，由于复杂的气象条件、多样的任务需求和无人机系统的高度非线性，这两类方法在处理复杂无人机数据时灵活性和适应性相对受限。随着大数据和深度学习技术的发展，数据驱动的方法，尤其是深度学习模型，在最近的研究中受到广泛关注。Deng等[21]提出了一种基于混合多模态神经网络的检测方法，在不同时频域特征上训练2个相对独立的深度神经网络，进而通过混合软投票机制来整合结果，实现了对关键异常类型的检测。Brulin等[22]提出了一种多旋翼无人机推进链模型，通过向系统注入各种故障来生成故障行为数据集，为检测器提供了更为丰富和高质量的训练样本。唐立等[23]定义了合作型和非合作型无人机网络的异常行为，并对两类无人机网络异常行为的运行参数进行分析，明确了各种异常行为运行特征及其判定参数的提取方法。在此基础上，提出了基于Sobel Operator-CNN算法的异常类型判定方法。顾兆军等[24]通过最大信息系数、图注意力网络和Transformer实现了无人机异常特征选择优化、时空特征融合和异常判据自定义。

针对类不平衡问题，目前主要有数据级[25]和算法级方法[26]两类解决方案。数据级方法包括过采样、欠采样和混合采样。过采样通过插入少数类样本来平衡数据集，但容易导致类重叠加剧，使异常检测模型难以捕捉准确的分类边界。欠采样则通过删除多数类样本来平衡数据集。基于聚类、矩阵分解、优化和动态加权的过采样方法均取得了良好的应用效果[27]。但确定欠采样的最佳样本数量仍然具有挑战性。混合采样结合过采样和欠采样各自优势，综合性能更好，但其时空复杂度高也使其实际应用受到极大限制。算法级方法直接改进训练过程以减轻分类器对少数类样本的偏见[28]，最常用的策略是根据任务设计特定的损失函数，以增加少数样本被误分类的损失，如代价敏感机制[29]。Abdelmonem等[30]提出了一种具有类不平衡鲁棒性的自适应高斯过程检测器，平衡了后验近似均值，并采用非对称条件预测，使模型在训练过程中更关注少数类。陆克中等[31]提出了一种具有自适应遗忘因子的加权在线顺序极限学习机集成算法，融合加权和遗忘机制，初步构建在线顺序极限学习机集成算法。进一步设计包含自适应遗忘因子和概念漂移检测器的在线集成策略，实现了更稳定、更平衡和更准确的检测效果。

生成式人工方法已被证明在一些极端条件下的异常检测中具有良好效果。通常，用于异常检测的生成式人工方法包括变分自编码器（VAE, variational autoencoder）、深度信念网络（DBN, deep belief net）、扩散模型（DM, diffusion model）、Transformer模型和GAN。例如，卷积变分自编码器（CVAE, convolution VAE）成功用于无人机传感器的零样本异常检测[32]。作为一种概率生成模型，DBN允许神经网络以最大概率生成训练数据，已用于具有分布偏差的数据异常检测[33]。扩散模型不需要后验分布校准，这使它在无人机图像异常检测中受到青睐[34]。Transformer模型则几乎成为众多安全领域大语言模型的标准模型[35]。自GAN被提出以来，在不同的场景下发展出了大量改进模型，已成为一种最为成熟的生成式人工方法。Zheng等[36]进一步引入惩罚系数增强了GAN的性能，通过梯度范围限制提高了数据生成质量和模型稳定性。Kullback-Leibler散度[37]和Wasserstein距离[38]也被引入GAN模型，用于引导学习过程向少数类倾斜，并克服梯度消失

问题。此外，Luo等[39]提出了利用动态聚类解决GAN中的模式崩溃问题。Kong等[40]将集成学习和注意力机制与GAN相结合，增强了时间序列依赖性建模在重建和生成损失中的作用，提高了模型的泛化能力。Chen等[41]则采用集成GAN策略，将多个鉴别器和分类器依据集成学习策略进行训练，并引入了一种主动学习算法，降低对现实世界数据进行标注的成本。王坤等[42]提出了一种基于深度学习的软件定义网络（SDN, software defined network）异常流量分布式检测方法，该方法将部署在云端服务器的判别器与若干部署在SDN控制器的生成器构造为“一对多”的分布式生成对抗网络，实现了大规模SDN环境下各控制器子网中异常流量的分布式检测。

然而，上述研究成果大多关注二分类异常检测的不平衡问题，多类不平衡问题的特殊性尚未得到充分重视。一些基于GAN的异常检测方法面向的任务场景虽然与智能无人机网络具有一定的相似性，但也尚未能全面考虑多分类、小样本、不平衡、分布偏差等问题耦合所造成的复杂性。针对这些问题，第2节提出了一种全新的GAN框架，用以解决智能无人机网络异常检测难题。

2 CAE-GAN模型

2.1 问题描述

智能无人机网络异常检测本质上是一个多分类问题，CAE-GAN模型期望能够找到有效将正常样本和不同类型异常样本区分开的分类边界。具体来说，给定一个数据集$X=\{x_1,x_2,\cdots,x_N\},x_i\in R^D$包含$N$个样本和$K$个类别（其中包含1个正常类，$K-1$个异常类），且每个样本有$D$维特征。样本类标签表示为$Y=\{y_1,y_2,\cdots,y_N\},y_i\in R^K$。由于$X$是一个多类别数据集，所以样本$x_i$对应的类别标签$y_i$是一个$K$维向量$y_i=[y_{i1},y_{i2},\cdots,y_{iK}],y_{ij}\in\{0,1\}$，其中$y_{ij}=1$表示样本$x_i$属于$j$类；否则样本$x_i$不属于该类。考虑到类别的唯一性，$y_i$中只能有一个元素为1，其他所有元素为0。不失一般性，设y_{i1}表示正常类，$y_{i2},y_{i3},\cdots,y_{iK}$表示不同的异常类。考虑到多类别不平衡问题，将各类别按包含样本数量进行降序排列，其中正常样本最多。假设每个类别的样本数量为$N(j)$，则$N(j)$可表示为

$$\sum_{j=1}^{K}N(j)=N$$
$$\text{s.t.}\quad N(1)>N(2)>\cdots>N(K) \tag{1}$$

多分类任务本质上是训练一个可以为样本x_i分配类标签的多类分类器C，最大化其预测准确率$C(x_i)$。$C(x_i)$是一个概率向量，表示为

$$C(x_i)=[C(x_{i1}),C(x_{i2}),\cdots,C(x_{iK})] \tag{2}$$

其中，$C(x_{ij})$是多类分类器C预测样本x_i属于j类的概率值。一般来说，训练分类器的目的是最小化其全局分类错误。基于交叉熵理论，多类分类器的损失函数为

$$L=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}\ln C(x_{ij}) \tag{3}$$

考虑到数据不平衡性，样本较少的类别更容易被误分类，应该得到更多关注。因此，将类别权重引入式(3)，其定义为

$$\psi(j)=\frac{N}{N(j)} \tag{4}$$

则多类分类器的损失函数可以优化为

$$L=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}\psi(j)\ln C(x_{ij}) \tag{5}$$

给定一个样本充足和分布均衡的训练集，最小化式(5)可以使预测概率收敛。然而，智能无人机网络中数据多类别不平衡情况下异常样本稀少，导致由该损失函数引导的模型面临学习不充分、泛化能力不足。为此从异常数据增扩、关键特征提取和分布偏差纠正多个维度设计了一个基于GAN的检测模型，并通过协同对抗训练强化模型在多类别不平衡场景下的异常检测能力。

2.2 基本框架

生成对抗网络由2个对抗性深度网络组成：一个是捕获数据分布的生成器G，另一个是估计样本来自真实数据而非生成器G的判别器D。G和D同时进行训练：调整G的参数以最小化其损失，同时调整D的参数以最小化其损失，这是一个极小极大的零和博弈过程，如式(6)所示。

$$L_{\text{GAN}}=\min_G\max_D V(D,G)=E_{x\sim p_{\text{data}}(x)}[\ln D(x)]+E_{x\sim p_z(z)}[\ln(1-D(G(z)))] \tag{6}$$

其中，p_{data}是数据x的先验分布，p_z是随机噪声的先验分布，$D(x)$是判别器判断样本x为真实样本的概率，$D(G(z))$是判别器判断生成样本$G(z)$为真实样本

的概率，z是随机噪声。如果生成器和判别器都以某些额外信息为辅助条件来支撑这一对抗训练过程，则GAN可以扩展为CGAN模型。在本文中以类别标签作为辅助信息，则CGAN模型的目标函数为

$$L_{\mathrm{CGAN}} = \min_{G}\max_{D} V(D,G) = E_{x\sim p_{\mathrm{data}}(x)}[\ln D(x|y)] + E_{x\sim p_z(z)}[\ln(1-D(G(z|y)))] \tag{7}$$

其中，$V(D, G)$表示生成器G和判别器D的对抗博弈。进一步，在CGAN模型基础上，CAE-GAN在数据生成、模型结构和学习策略3个方面进行了综合改进，如图2所示。具体如下。

1) 采用多个生成器$G_1,G_2,\cdots,G_K$进行不同类别异常数据生成，并使用类标签概率向量来调整不同类别异常样本的生成数量，逐渐将学习重点向数量更少的异常类倾斜。此外，在串行对抗训练过程中，每个后续生成器通过权值共享策略继承前一个生成器的参数并进行微调，以提高学习效率，并保持模型稳定性。

2) 设计了一个带有特征缩聚和激励模块的编码器，用于从不充足的异常样本中进行深度特征挖掘，提取异常模式关键特征，提升隐空间特征重构能力，以提高模型在小样本场景下的特征提取能力。

3) 为感知并纠正生成数据和真实数据之间的分布偏差，提出了多类分类器和判别器共同引导样本生成的机制，建立了一个协同对抗训练方案，并为其设计匹配了一种新的损失函数。

2.3　面向多类别不平衡的数据增扩

CGAN通过附带条件的生成器G和判别器D之间的对抗训练生成带标签的数据。然而，标准CGAN并未针对多类别不平衡问题进行设计，异常类样本的不充足可能导致模型失效。受大语言模型的预训练和微调模式启发，本节在数据增扩方面设计了一个动态类标签概率向量，以实现对数据集类别平衡性的动态调控。同时，在不同生成器之间采用权重共享和“微调”机制，提升多个生成器训练的学习效率和稳定性。

具体地，提出了一种新的具有多个生成器G_1, $G_2,\cdots,G_K$的GAN框架，在每一轮对抗训练中，优先考虑更少数异常类样本的生成。然而，在多分类任务中，多数类和少数类是相对的。例如，一个异常类相对于正常类是少数类，但相对于另一个数量更少的异常类则可能是多数类。为了更好地指导异常样本的生成，为每个生成器配备一个类标签概率向量$P_1(y^g), P_2(y^g), \cdots, P_K(y^g)$。需要注意的是，这个类标签概率向量也通过对抗训练进行微调，以逐渐倾向于样本更少的少数类。

对于生成器G_j，其类标签概率向量为$P_j(y^g)$ =

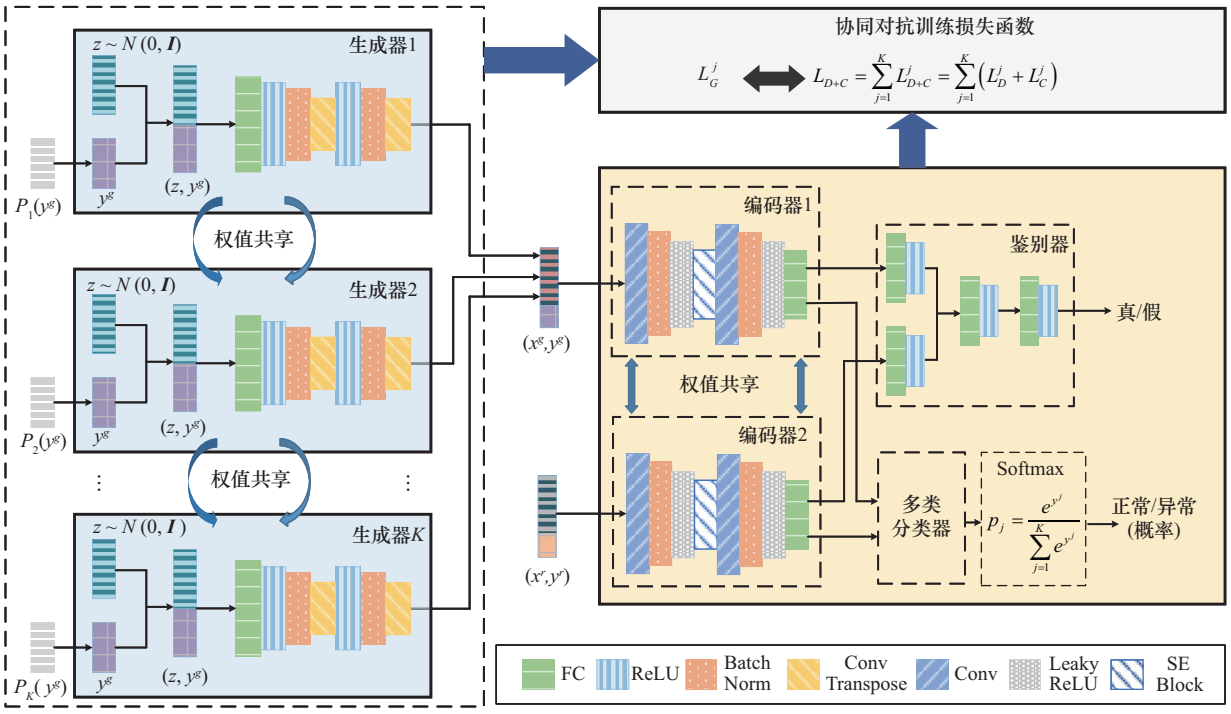

图2　CAE-GAN模型框架

$[P_{j,1}(y^s), P_{j2}(y^s), \cdots, P_{jK}(y^s)]$，其中每个类的标签概率为

$$p_{j,u}=\begin{cases}\dfrac{\sum_{u=j}^{K}N_u}{N}\dfrac{N_u}{\sum_{u=1}^{j-1}N_u}, & u<j\\ \dfrac{\sum_{u=1}^{j-1}N_u}{N}\dfrac{N_u}{\sum_{u=j}^{K}N_u}, & u\geqslant j\end{cases} \tag{8}$$

式(8)是一个包含全局和局部不平衡系数的分段函数。从全局角度，训练样本可以分为两大类：多数类和少数类。前者是标签为$1, 2, \cdots, j-1$的类，后者是标签为$j+1, \cdots, K$的类。在前$j-1$轮对抗训练中，少数类被认为是学习不足的类。因此，应该通过全局不平衡系数增加少数类的生成概率，而减少多数类的生成概率。具体地，将少数类的比例$\sum_{u=j}^{K}\frac{N_u}{N}$设置为多数类的全局不平衡系数。类似地，将$\sum_{u=1}^{j-1}\frac{N_u}{N}$设置为少数类的全局不平衡系数。

此外，在多数类和少数类内部还存在多种类别的数据，各类别之间也存在不平衡，即为局部不平衡。因此，引入局部不平衡系数$\frac{N_u}{\sum_{u=1}^{j-1}N_u}$和$\frac{N_u}{\sum_{u=j}^{K}N_u}$用于防止过度关注少数类而导致过拟合。通过上述设计，在与不同生成器进行对抗训练的过程中，该生成器生成不同类别样本的数量是存在差异的，其更多地生成原本数量较少的类别样本，以实现对数据集的再平衡。动态类标签概率向量通过微调机制逐渐将重点转向更少数的异常类。这种机制有效地解决了多类别数据异常检测的全局和局部不平衡共存问题。

2.4 小样本学习的深度特征提取

特征提取在判别器D和分类器C中起着关键作用。尽管生成器可以使用特定类别的概率向量来增强样本，但相对少数类中的某些类样本仍然严重不足。从不充足的样本中提取关键特征是十分困难的，是导致模型性能下降的重要原因。本节在模型结构层面为判别器和分类器设计了一种具有特殊结构的编码器，并在编码器中引入特征缩聚和激励模块，通过放大有效特征权重并减少无效或低效的特征权重，实现异常行为关键特征的深度提取。

编码器的具体结构是一个8层深度网络，如图2所示。其中，使用卷积层对应生成器中的反卷积层，转置层将低维随机噪声映射到高维空间以生成样本。利用卷积层的下采样操作提高对深度特征的提取，并通过权重共享和稀疏连接确保单层卷积的参数计算最小化。编码器采用Leaky ReLU作为激活函数。与标准ReLU不同，Leaky ReLU可以处理小于0的输入。基于此，在反向传播过程中，最大限度地避免了模型梯度方向的锯齿模式问题。特征缩聚和激励模块使用全局平均池化层压缩卷积特征作为聚合操作，并通过2个全连接层建立特征之间的相关性。最后，通过Sigmoid函数对权重进行归一化以完成激励操作。

特征缩聚是对输入特征进行全局池化，即将输入数据x_0的$H\times W\times C$特征压缩为$1\times 1\times C$特征，如式(9)所示。

$$z=F_{sq}(x_0)=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}x_0(i.j) \tag{9}$$

激励模块则通过2个全连接层对压缩后的$1\times 1\times C$特征进行跨通道融合，如式(10)所示。

$$s=F_{ex}(z,\boldsymbol{W})=\sigma\left(\boldsymbol{W}_2\delta\boldsymbol{W}_1(z)\right) \tag{10}$$

其中，δ是Leaky ReLU激活函数，$\boldsymbol{W}_1\in R^{\frac{C}{r}\times C}$，$\boldsymbol{W}_2\in R^{C\times\frac{C}{r}}$，$r$是超参数。激励后，将式(10)的结果与式(9)的输入相乘，得到最终输出为

$$x=x_0\cdot s \tag{11}$$

需要注意的是，为判别器D和分类器C同时配备了编码器E。通过权重共享，增强它们在对抗训练期间的协作，并减轻了编码损失对判别和分类的影响。

2.5 协同对抗训练

大多数数据增扩技术在协同对抗训练过程中会引入分布偏差，导致模型无法收敛或收敛非常缓慢。为此，引入一个与判别器互补的多类分类器C与生成器进行对抗训练。一方面，多类分类器C通过共享编码器权重与判别器D进行协同对抗训练，不断提高模型框架的生成和判别能力。另一方面，在反向传播过程中，提取的误差梯度从多类分类器C传播到生成器，通过协同对抗训练逐渐纠正生成数据分布。训练完成后，多类分类器C即可被作为智能

无人机网络中多类异常检测任务的有效检测模型。在学习策略方面，通过分类器和判别器共同引导样本生成的协同对抗训练，实现分布偏差的协同感知与自适应纠正，并设计了一种新的损失函数来支持这种协同对抗训练策略的实施。下面具体讨论次损失函数。

首先考虑生成器损失。随着多类分类器C的引入，生成器G不仅与判别器D对抗以生成假数据来欺骗D，还与多类分类器C对抗以使其进行误分类。在实际训练中，数据以批处理方式输入。考虑生成器G_j与判别器D和多类分类器C进行对抗训练的N^r个真实样本$\{(x_i^r,y_i^r)\}_{i=1}^{N^r}$和$N^g$个生成样本。其中，$y_i^r=[y_{i1}^r,y_{i2}^r,\cdots,y_{iK}^r]$和$y_i^g=[y_{i1}^g,y_{i2}^g,\cdots,y_{iK}^g]$都是$K$维向量。因此，生成器$G_j$的损失函数包括两部分：生成样本判别对抗损失$L_{G,D}^j$和生成样本预测对抗损失$L_{G,C}^j$，分别为

$$L_{G,D}^j=-\frac{1}{N^g}\sum_{i=1}^{N^g}\ln D(G(z_i,y_i^g)) \tag{12}$$

$$L_{G,C}^j=-\frac{1}{N^g}\sum_{i=1}^{N^g}\sum_{j=1}^{K}\psi(j)\ln C(y_{ij}=y_{ij}^g|G(z_i,y_i^g)) \tag{13}$$

在对抗训练的分类器中，还应考虑样本个体的误分类影响。在某轮训练中应优先考虑先前训练中被错误分类的样本。因此，通过结合先前的分类结果，引入误分类系数对$L_{G,D}^j$进行加权处理。

$$L_{G,C}^j=-\frac{1}{N^g}\sum_{i=1}^{N^g}(1-\overline{p_i^{j-1,g}})\sum_{j=1}^{K}\psi(j)\ln C(y_{ij}=y_{ij}^g|G(z_i,y_i^g)) \tag{14}$$

其中，$1-\overline{p_i^{j-1,g}}$为误分类系数，$\overline{p_i^{j-1,g}}=\overline{p_i^{j-1,g}(y_i=y_i^g|G(z_i,y_i^g))}$为前$j$–1轮对抗训练中样本的平均正确分类概率。综上，生成器G_j的总损失为

$$L_G^j=L_{G,D}^j+L_{G,C}^j=-\frac{1}{N^g}\sum_{i=1}^{N^g}\ln D(G(z_i,y_i^g))-\frac{1}{N^g}\left\{\sum_{i=1}^{N^g}(1-\overline{p_i^{j-1,g}(y_i=y_i^g|G(z_i,y_i^g))})\cdot\sum_{j=1}^{K}c(j)\ln C(y_{ij}=y_{ij}^g|G(z_i,y_i^g))\right\} \tag{15}$$

在与生成器G_j的对抗训练中，判别器D必须识别样本是真实样本还是生成样本。真实样本的判别损失为

$$L_D^{j,r}=-\frac{1}{N_r}\sum_{i=1}^{N_r}\ln D(x_i^r) \tag{16}$$

生成样本的判别损失为

$$L_D^{j,g}=-\frac{1}{N_g}\sum_{i=1}^{N_g}\ln\left[1-D(G(z_i,y_i^g))\right] \tag{17}$$

则判别器D在与生成器G_j的第j轮对抗训练中的判别损失为

$$L_D^j=L_D^{j,r}+L_D^{j,g}=-\frac{1}{N_r}\sum_{i=1}^{N_r}\ln D(x_i^r)-\frac{1}{N_g}\sum_{i=1}^{N_g}\ln\left[1-D(G(z_i,y_i^g))\right] \tag{18}$$

类似地，多类分类器C必须同时对真实数据和生成数据进行分类和预测。因此，多类分类器C的损失也包括预测真实数据的损失和预测生成数据的损失，分别为

$$L_C^{j,r}=-\frac{1}{N_r}\sum_{i=1}^{N_r}\sum_{j=1}^{K}\psi(j)\ln C(y_{ij}=y_{ij}^r|x_i^r) \tag{19}$$

$$L_C^{j,g}=-\frac{1}{N_g}\sum_{i=1}^{N_g}\sum_{j=1}^{K}\psi(j)\ln C(y_{ij}=y_{ij}^g|G(z_i,y_i^g)) \tag{20}$$

类似于式(14)，也使用误分类系数对误分类样本的损失进行加权，可表示为

$$L_C^{j,r}=-\frac{1}{N_r}\left\{\sum_{i=1}^{N_r}(1-\overline{p_i^{j-1,r}(y_i=y_i^r|x_i^r)})\sum_{j=1}^{K}\psi(j)\ln C(y_{ij}=y_{ij}^r|x_i^r)\right\} \tag{21}$$

$$L_C^{j,g}=-\frac{1}{N_g}\left\{\sum_{i=1}^{N_g}(1-\overline{p_i^{j-1,g}(y_i=y_i^g|G(z_i,y_i^g))})\sum_{j=1}^{K}\psi(j)\ln\left[1-C(y_{ij}=y_{ij}^g|G(z_i,y_i^g))\right]\right\} \tag{22}$$

因此，多类分类器C在与生成器G_j的第j轮对抗训练中的损失为

$$L_C^j=L_C^{j,r}+L_C^{j,g}=-\frac{1}{N_r}\left\{\sum_{i=1}^{N_r}(1-\overline{p_i^{j-1,r}(y_i=y_i^r|x_i^r)})\sum_{j=1}^{K}\psi(j)\ln C(y_{ij}=y_{ij}^r|x_i^r)\right\}-\frac{1}{N_g}\left\{\sum_{i=1}^{N_g}(1-\overline{p_i^{j-1,g}(y_i=y_i^g|G(z_i,y_i^g))})\sum_{j=1}^{K}\psi(j)\ln\left[1-C(y_{ij}=y_{ij}^g|G(z_i,y_i^g))\right]\right\} \tag{23}$$

综上，在与生成器 G_j 的对抗训练中，判别器 D 和多类分类器 C 的总损失为

$$L^j_{D+C} = L^j_D + L^j_C \tag{24}$$

考虑到判别器 D 和多类分类器 C 与所有生成器 $G_1, G_2, \cdots, G_K$ 进行多轮的串行对抗训练，则判别器和多类分类器的总损失为

$$L_{D+C} = \sum_{j=1}^{K} L^j_{D+C} = \sum_{j=1}^{K} (L^j_D + L^j_C) \tag{25}$$

2.6 基于 CAE-GAN 的异常检测

训练到 CAE-GAN 模型收敛，多类分类器 C 就成了一个高效的多类别异常检测器，可以有效地对智能无人机网络中的多类型异常和正常数据进行分类和检测。需要注意的是，在确定实例 x 是否为异常时，CAE-GAN 将样本 x 输入多类分类器 C 中，然后获得一个 K 维得分向量，表示为

$$\boldsymbol{S} = \left[S_1, S_2, \cdots, S_j, \cdots, S_K\right], \quad 0 < S_j < 1 \tag{26}$$

其中，每个元素 S_j 表示样本 x 属于特定类别的概率。通常，S_1 表示正常类，而其他元素对应不同类型的异常类。为了最大限度地提高智能无人机网络的安全性，正常类检测阈值预设为 0.5。如果 S_1<0.5，应考虑智能无人机网络出现异常，尽管 S_1 可能仍然是向量 $\boldsymbol{S}$ 中的最大值。CAE-GAN 异常检测算法如算法 1 和算法 2 所示。

算法 1 CAE-GAN 异常检测（训练）

输入 有标签的真实样本 (x^r, y^r)，随机噪声向量 $z \sim N(0, \boldsymbol{I})$，批量 N_b，样本数量 N，类别数量 K

输出 训练好的多类分类器 C

1) 初始化生成器 $G_1, G_2, \cdots, G_K$，判别器 D 和多类分类器 C，设置 $j = 0$
2) for epoch $j < K$ do
3) for each training iteration do
4) 根据式(8)计算类标签向量 $P_j(y^g)$
5) 基于随机噪声向量 z 和类标签向量 $P_j(y^g)$ 生成一批样本 x^g
6) 利用 (x^r, y^r) 和 (x^g, y^g) 进行对抗训练
7) 根据式(15)计算生成器 G_j 的损失
8) 根据式(18)计算判别器 D 的损失
9) 根据式(23)计算多类分类器 C 的损失
10) 更新本批次中生成器 G_j、判别器 D 和多类分类器 C 的参数
11) end for
12) 将权值参数共享给生成器 G_{j+1}
13) 更新误分类系数 $1 - \overline{p_i^{j-1,g}}$
14) 最小化式(25)以更新本轮中判别器 D 和多类分类器 C 的损失
15) $j = j + 1$
16) end for
17) 返回训练好的多类分类器 C

算法 2 CAE-GAN 异常检测（检测）

输入 待检测样本 x

输出 样本 x 所属类别

1) 将待检测样本 x 输入训练好的多类分类器 C
2) 根据式(26)计算样本 x 的异常得分
3) if S_1>0.5
4) return 正常类
5) else
6) 从 $S_2, \cdots, S_K$ 中寻找最大值 S_k
7) return 异常类 k
8) end if

3 实验设计

3.1 实验数据集

本文选择 4 个与无人机网络异常相关的开源数据集进行实验，其中 2 个数据集是专门为无人机网络安全设计的数据集：无人机攻击（UA）[43]和无人机 ADS-B（ADS-B）[44]。另外 2 个是著名的开源网络入侵数据集：NSL-KDD 和 UNSW-NB15。UA 数据集由四旋翼无人机的原始飞行日志和无人机遭受网络攻击时产生的异常数据组成，它包括具有 23 个特征的 3 种类型的飞行日志数据：安全、DoS 攻击和 GPS 欺骗。ADS-B 数据集来自 Flightradar24 网站，包含来自全球数千个地面基站的 ADS-B 数据。通过人工修改原始数据消息，模拟了噪声、注入、删除、偏移和 DoS 等 6 种异常行为。在原始变量中，选择了与异常密切相关的 5 个特征作为检测特征：经度、纬度、高度、速度和航向。

NSL-KDD 数据集包含 43 个特征和 4 种网络攻击行为：DoS、Probe、U2R 和 R2L。UNSW-NB15 数据集是新南威尔士大学发布的一个大型极端不平衡网络安全数据集，有 49 个特征和 9 种基本攻击类型。需要注意的是，上述数据集可能由具有不同初始顺序的时域信号组成，在将它们输入深度检测模

型之前，对它们进行采样、归一化等预处理操作，详细的原始数据操作参考相关原始文献。实验数据集详细信息如表 1 所示。

3.2　对比方法

本节选择了 5 种对比方法与本文方法进行比较，分别是 CVAE-GAN[32]、FL-FCA-GAN[41]、DB-CGAN[14]、AMBi-GAN[40] 和 EAL-GAN[42]，各对比方法特点和选择依据如下。

1) CVAE-GAN 旨在为未知的异常模式生成较大的重构误差，这些误差被用作异常分数，以实现不同异常类型的细粒度检测，是一种专为无人机异常检测设计的方法。

2) FL-FCA-GAN 是一种整合了生成对抗网络、联邦学习和模糊聚类的信息物理系统异常检测方法。而智能无人机网络也可以看作是一种特殊网络物理系统，因此选择该方法进行对比。

3) DB-CGAN 旨在通过数据增强解决智能物联网中深度学习的不平衡问题以及生成数据与原始数据之间的分布偏差问题。智能无人机也是一类智能终端节点，组成了一个智能化网络，与智能物联网具有相似之处。

4) AMBi-GAN 将一个使用双向长短期记忆（LSTM, long short term memory）网络和注意力机制的生成模型与工业异常检测算法相结合，能够有效捕捉时间序列之间的相关性，是一种在时间序列数据异常检测中广泛应用的方法。无人机传感和通信数据时序特征明显，因此选择该方法进行对比。

5) EAL-GAN 是一种条件 GAN 模型，具有一个生成器与多个判别器，其中异常检测通过判别器的辅助分类器实现。其设计了一种创新的集成学习损失函数，也引入了主动学习方法以降低现实世界数据的标注成本。作为一种集成式的 GAN 模型，其设计思路与本文方法有相似之处。

以上方法都在某一方面特点上与本文方法具有相似性，其中一些方法面向多分类任务设计，个别方法最初为二分类任务设计，本文采用 One-vs-All 机制来实现多类别数据异常检测。

3.3　评价指标

从以下 3 个方面将本文方法与 5 种方法进行比较，以证明其有效性和优势。首先，关注 CAE-GAN 方法的数据增扩能力。采用 KID（kernel inception distance）[45] 值来评估不同模型在分布学习和数据增扩方面的能力，越小的 KID 值代表模型的生成能力越强。

$$\mathrm{KID}=\mathrm{MMD}(P_r,P_g)=\left(E_{\substack{x_r,x_r'\sim P_r,\\ x_g,x_g'\sim P_g}}[k(x_r,x_r')-2k(x_r,x_g)+k(x_g,x_g')]\right)^{\frac{1}{2}} \tag{27}$$

其中，x_r 和 x_r' 是 2 个服从 P_r 概率分布的随机变量，x_g 和 x_g' 是 2 个服从 P_g 概率分布的随机变量，k 是多项式的核函数，可表示为

$$k(x,y)=\left(\frac{1}{d}x^Ty+1\right)^3 \tag{28}$$

其中，d 是数据 x 和 y 的维度。

其次，使用 F_1 分数、曲线下面积（AUC, area under the curve）值和 G-mean 值对异常检测性能进行了对比。AUC 值越大表明方法准确性越高。G-mean 则是数据分类领域的主要评价指标，尤其适合评估具有不同 IR 值数据集上的方法性能。该指标综合考虑了异常检测模型的敏感性和特异性。

$$\text{G-mean}=\sqrt{\frac{\mathrm{TP}}{\mathrm{TP+FN}}\frac{\mathrm{TN}}{\mathrm{TN+FP}}} \tag{29}$$

其中，TP 为真正例，FP 为假正例，TN 为真反例，FN 为假反例。

最后，采用非参数统计检验方法进行差异性分析。具体而言，首先通过 Friedman 测试来验证不同方法间的显著性差异；随后运用 Nemenyi 后续检

表 1　　**实验数据集详细信息**

数据集	样本数/个	特征维度/维	类别	占比
UA	8 657	23	Normal、DoS、GPS Spoofing	97.6%、1.7%、0.7%
ADS-B	10 000	7	Normal、Injection、DoS、Noise、Course shift、Velocity Shift	80%、8%、7%、5%、5%、5%
NSL-KDD	148 517	43	Normal、DoS、Probe、U2R、R2L	51.88%、35.95%、9.48%、2.25%、0.17%
UNSW-NB15	257 673	49	Normal、Generic、Exploits、Fuzzers、DoS、Reconnaissance、Analysis、Backdoor、Shellcode、Worms	36.09%、22.85%、17.28%、9.4%、6.35%、5.43%、1.04%、0.9%、0.59%、0.07%

验进一步区分各方法间的优劣，各方法平均排序临界值（CD, critical difference）的计算式为

$$\mathrm{CD} = q_{\alpha}\sqrt{\frac{n(n+1)}{6t}} \tag{30}$$

其中，q_{α}是 Tukey 分布的临界值，n是比较方法的数量，t是数据集的数量。

3.4　参数设置

所有实验均在 Ubuntu 18.04 操作系统、Intel Core i7-9750H 搭配 NVIDIA GeForce GTX 1650、TensorFlow 2.12.0 和 Python 3.6 的深度学习框架下进行。在模型训练阶段，批处理大小为 128，epochs 值设置为 100。采样率设置为 5%，表明每个批次中仅选择 5% 的真实数据引导样本生成。生成器、判别器和编码器使用 Adam 优化器进行梯度下降优化，生成器的学习率设置为 0.01，判别器和编码器的学习率从[0.01, 0.05]内随机选择。在对比方法中，GAN 的学习率统一设置为 0.01，批处理大小为 128，与本文方法保持相同。此外，CVAE-GAN 中自适应阈值的配置参数窗口大小为 20、步长为 1；FL-FCA-GAN 中联邦学习率为 0.001，模糊聚类阈值为 0.35；AMBi-GAN 中的双向 LSTM 网络具有 32 个隐单元；EAL-GAN 中生成器采用一个四层多层感知器，判别器采用一个三层网络。在每次实验中，80% 的可用数据被随机采样作为训练数据，其余 20% 用作测试数据。所有数据集的结果是通过 3 轮 5 折交叉验证实验获得的，异常检测性能指标值取自这些实验的算术平均值。

4　实验结果分析与讨论

4.1　数据增扩性能

使用 KID 值评估原始数据和增扩数据在分布差异方面的情况，以反映不同方法的数据增扩能力，结果如图 3 所示。本文方法在 4 个数据集上均获得了最低的 KID 值，均低于 0.5。并且在 UA 数据集上曲线收敛迅速，表明生成的数据很快就与原始数据分布几乎达到一致。这是因为 UA 数据集仅包含 3 个数据类别（其中 2 个为异常类），使 CAE-GAN

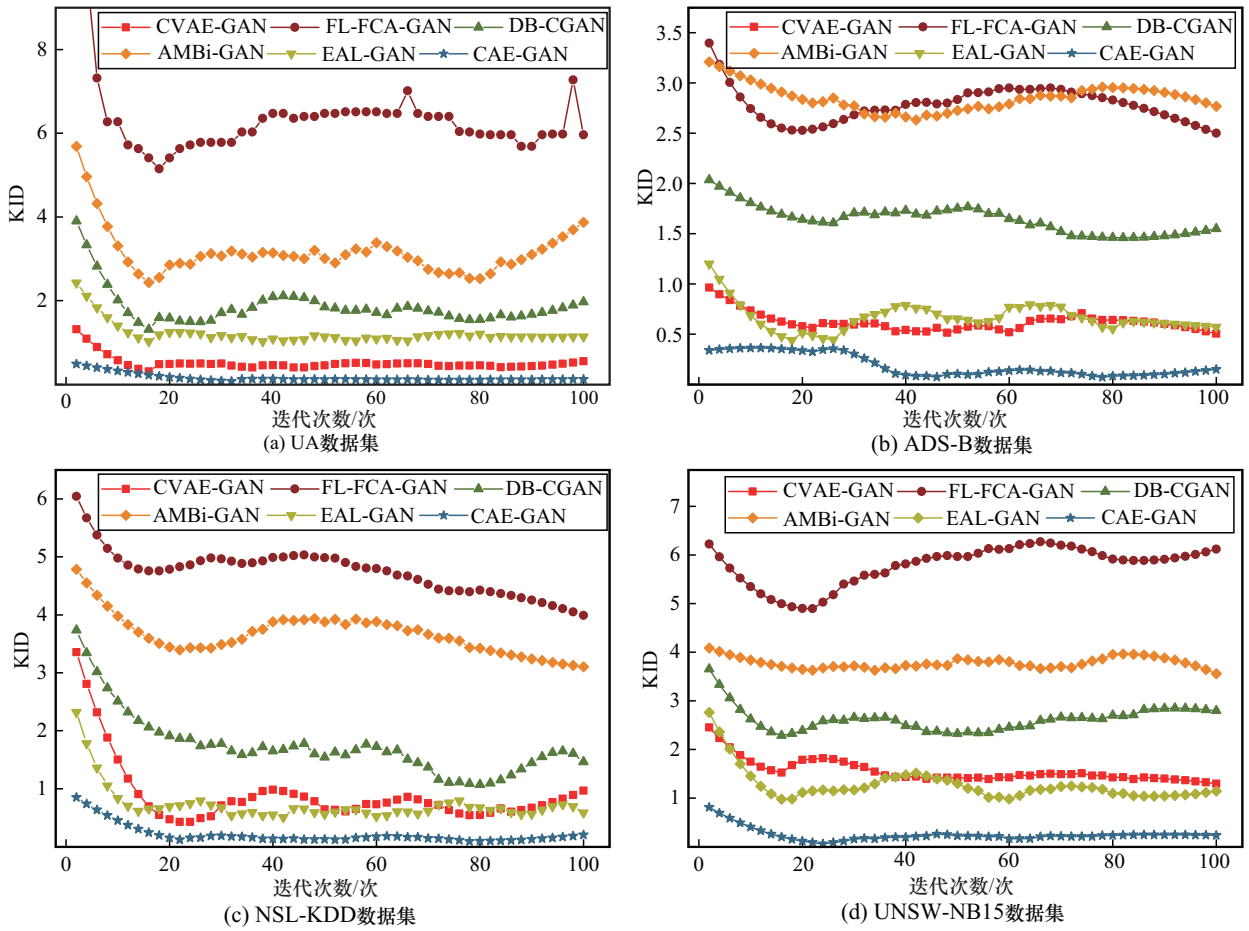

图 3　不同数据增扩方法 KID 对比

方法更容易处理。相比之下，对于其他具有更多异常类别的数据集，CAE-GAN 方法收敛速度明显变缓。

在对比方法中，CVAE-GAN 和 EAL-GAN 方法的整体性能仅次于本文方法。具体而言，CVAE-GAN 方法在 UA 和 ADS-B 数据集上表现更好，因为其是针对无人机数据设计的专用检测模型。相比之下，在 UBSW-NB15 数据集上，由于采用集成和主动学习机制，EAL-GAN 方法的检测性能优于 CVAE-GAN 方法。然而，其生成数据的稳定性有待提高，因为其 KID 值曲线存在一定的抖动，如图 3(b)~图 3(d)所示。FL-FCA-GAN 方法性能表现最差，表明它并不适合多类别异常检测任务。

4.2　异常检测性能

通过 F_1 分数可以全面比较异常检测方法的性能，特别是在不平衡学习领域中。不同异常检测方法获得的 F_1 分数如图 4 所示。结果表明，本文方法在 4 个数据集上均表现最佳。在对比方法中，CVAE-GAN 方法在 UA 数据集上排名第 2，F_1 分数比本文方法下降约 0.021。DB-CGAN 方法在 ADS-B 数据集上仅次于 CAE-GAN 方法，F_1 分数为 0.919 3。尽管 ADS-B 数据集包含 6 个数据类别，但样本数量最少的类别也占数据集总数的 5%，样本量达到 10 000。因此，分布偏差是该数据集面对的主要挑战。本文方法解决了分布偏差问题，使其能够在 ADS-B 数据集上获得更高的 F_1 分数。相比之下，NSL-KDD 和 UNSW-NB15 数据集的不平衡问题更为棘手，导致所有方法的 F_1 分数均明显下降。尽管如此，CAE-GAN 方法仍保持了较好的结果，EAL-GAN 方法排名第 2，因其集成学习机制对误分类样本更为关注。本文方法在处理具有较多异常类别的数据集时性能还有待提升。例如，在 UNSW-NB15 数据集上，CAE-GAN 方法的 F_1 分数仅为 0.838 2，与 EAL-GAN 方法相比优势并不明显。这意味着一些占比较低的异常类可能会被误分类，如占比仅为 0.7% 的 Worms。

图 5 展示了不同异常检测方法获得的平均 AUC 值。在 UA 数据集上 CVAE-GAN 方法优于本文方法，其平均性能增益约为 0.005。在 ADS-B 数据集上，与排名第 2 的 DB-CGAN 方法相比，CAE-GAN 方法 AUC 值提高了 0.059。因为 UA 数据集只有 3 个数据类别，而 ADS-B 数据集具有相对更充足的训练样本。此外，CAE-GAN 与 CVAE-GAN 或 DB-CGAN 方法之间的差异均相对较小，而相对于其他对比方法具有更显著优势。由上述分析可知，CAE-GAN 方法的动态类标签概率向量具有更为出众的数据增扩能力，所以在高度不平衡数据中性能优势明显，而在类似 ADS-B 数据集等具有充分训练样本的情况下，其性能与对比方法差异不大。

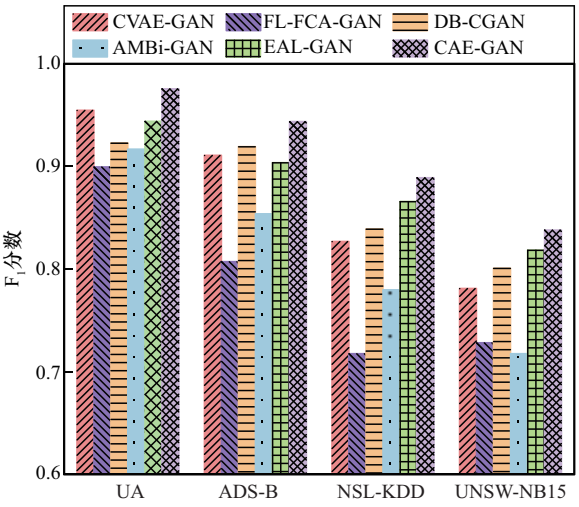

图 4　不同异常检测方法的 F_1 分数

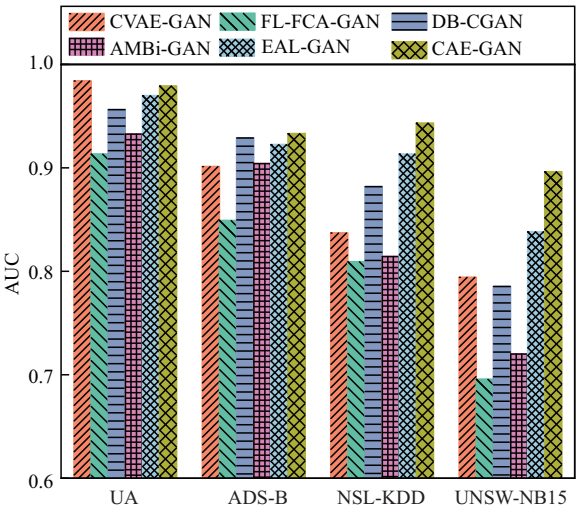

图 5　不同异常检测方法的平均 AUC 值

随着数据类型和 IR 值的增加，CAE-GAN 方法在 NSL-KDD 和 UNSW-NB15 数据集上均表现出显著的优越性能。CAE-GAN 方法相对于排名第 2 的 EAL-GAN 方法平均性能提升约为 0.03，与 FL-FCA-GAN 方法相比，AUC 增益更是超过 10%。尽管 CVAE-GAN 方法在 UA 数据集上性能表现最佳，但在复杂场景中表现未达到预期。主要原因是其侧重于无人机传感数据异常检测，当应用于 NSL-

KDD和UNSW-NB15等网络攻击导致的异常检测任务时，局限性更加突出，其架构缺乏针对这些数据的不平衡性、小样本等特征进行的优化。总体而言，本文方法不仅可以处理网络攻击导致的无人机异常检测，还可以胜任系统故障导致的无人机异常检测任务。除了整体检测性能外，还分析了CAE-GAN方法在少数异常类上的检测性能。表2展示了不同检测方法在各数据集少数异常类别上获得的AUC值。

由表2可知，本文方法在UA数据集的2个异常类中均获得了最高AUC值，在ADS-B数据集的5个异常类中的3个类别上获得了最高AUC值，在NSL-KDD数据集的4个异常类中的3个类别上获得了最高AUC值，在UNSW-NB15数据集的9个异常类中的6个类别上获得了最高AUC值。进一步分析可以发现，CAE-GAN方法对占比更少的异常类的检测结果更好。例如，ADS-B数据集中Velocity Shift类占比仅为5%，CAE-GAN方法与排名第2的CVAE-GAN方法相比AUC值提高了5%左右。类似地，UNSW-NB15数据集中Worms类占比仅为0.07%，CAE-GAN方法在该类别上的AUC值比排名第2的EAL-GAN方法提高了近10%。由上述分析可知，CAE-GAN方法由于特殊设计的多生成器架构和协同对抗训练策略，使其更适用于多类别的小样本学习场景，因此在占比较小的异常类检测中性能优势更加突出，反而在占比较大的异常类检测中性能优势并不明显。

G-mean也是不平衡学习的关键评估指标，用于评估不同异常检测方法的检测性能，结果雷达图如图6所示。CAE-GAN方法在4个数据集上都获得了最大的G-mean值，其雷达图曲线完全覆盖了对比方法的雷达图区域。其他方法在不同数据集上性能表现则不尽相同，其中FL-FCA-GAN方法的性能最差，这与图3和图4所示的结果基本一致。然而，需要注意的是，其他方法获得的G-mean值平均为0.6~0.8，并且在类别更多、不平衡和小样

表2 不同检测方法在各数据集少数异常类别上获得的AUC值

数据集	异常类别	CVAE-GAN	FL-FCA-GAN	DB-CGAN	AMBi-GAN	EAL-GAN	CAE-GAN
UA	DoS	0.844 2	0.513 5	0.734 3	0.632 8	0.783 9	**0.869 2**
	GPS Spoofing	0.793 2	0.436 4	0.795 1	0.521 4	0.752 2	**0.808 4**
ADS-B	Injection	0.943 2	0.802 9	**0.950 3**	0.853 1	0.926 1	0.916 1
	DoS	0.921 1	0.757 4	0.913 0	0.802 9	**0.936 6**	0.926 3
	Noise	0.850 3	0.721 4	0.831 2	0.779 4	0.852 1	**0.900 3**
	Course Shift	0.860 5	0.701 3	0.827 1	0.789 1	0.843 9	**0.910 2**
	Velocity Shift	0.843 9	0.717 1	0.811 4	0.753 6	0.832 1	**0.906 5**
NSL-KDD	DoS	0.877 1	0.772 5	0.910 3	0.813 1	**0.955 2**	0.951 1
	Probe	0.720 7	0.796 5	0.812 1	0.762 1	0.841 9	**0.873 3**
	U2R	0.651 0	0.662 3	0.773 4	0.581 9	0.740 9	**0.804 5**
	R2L	0.510 2	0.541 0	0.714 1	0.448 7	0.785 8	**0.813 2**
UNSW-NB15	Generic	0.933 1	0.902 5	**0.976 3**	0.921 9	0.958 3	0.952 2
	Exploits	0.850 3	0.816 0	0.678 7	0.889 1	**0.932 6**	0.921 4
	Fuzzers	0.781 4	0.696 9	0.562 8	0.762 1	0.827 1	**0.891 4**
	DoS	0.709 8	0.701 2	0.510 2	0.773 9	**0.795 2**	0.753 1
	Reconnaissance	0.663 2	0.621 7	0.789 1	0.678 5	0.779 9	**0.832 2**
	Analysis	0.503 3	0.321 4	0.431 2	0.532 1	0.665 8	**0.701 2**
	Backdoor	0.445 6	0.335 4	0.314 6	0.487 5	0.621 4	**0.669 8**
	Shellcode	0.496 5	0.256 9	0.396 5	0.387 5	0.553 2	**0.625 6**
	Worms	0.465 8	0.123 6	0.213 6	0.399 6	0.501 1	**0.598 9**

注：加粗代表性能排名第1，加下划线代表性能排名第2。

本问题更突出的UNSW-NB15数据集上G-mean值最小。可见对比方法随着检测的异常数据类别增加，G-mean值进一步下降，可能导致检测方法失效。而CAE-GAN方法获得的G-mean值在所有数据集上均大于0.9，包括最为复杂的UNSW-NB15数据集。这些结果表明，CAE-GAN方法在多类别不平衡数据异常检测的复杂任务场景下依然可以保持良好的稳定性。

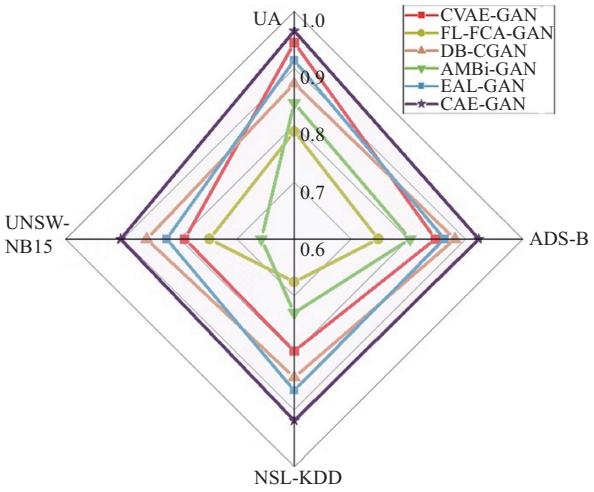

图6　不同异常检测方法在各数据集上的G-mean值

同样地，进一步分析不同方法在少数异常类检测中的G-mean值，重点关注具有更多异常类的ADS-B和UNSW-NB15这2个数据集，结果如图7所示。总体而言，CAE-GAN方法的雷达图区域在所有异常类上都很好地覆盖了其他方法，表现出了最佳的检测结果。如图7(a)所示，各方法在所有异常类上表现出的性能差异并不明显，因为ADS-B数据集中Injection、DoS、Noise、Course Shift和Velocity Shift这5个异常类别占比分别为8%、7%、5%、5%和5%，其IR值比较接近，而且不存在极端不平衡和小样本的情况，对于检测方法而言较为容易学习异常类的特点，所以各个方法在不同异常类的检测结果之间基本保持了相近的结果。但在UNSW-NB15数据集上，情况却大不相同。如图7(b)所示，各检测方法在不同异常类上的检测结果差异非常显著，Generic、Exploits、Fuzzers、DoS、Reconnaissance等占比超过5%的异常类的G-mean结果较好，均超过0.85。相比之下，Analysis、Backdoor、Shellcode和Worms等异常类的G-mean结果明显降低，基本处于0.6~0.8。这些异常类占比仅为1.04%、0.9%、0.59%和0.07%，明显少于Generic、Exploits、Fuzzers、DoS、Reconnaissance等类别，是一个更典型的高度不平衡和小样本学习场景，对检测方法而言，学习难度更大。但与其他方法相比，本文方法在这些小样本和高度不平衡类别上的性能提升却更为明显。例如，在Worms类异常检测中，CAE-GAN方法与排名第2的EAL-GAN方法相比，G-mean值提高了10%，而与FL-FCA-GAN方法相比，提高了近50%。这一方面得益于CAE-GAN方法采用动态类标签概率向量，在对抗性训练中逐渐提高少数类异常样本增扩概率，而使模型逐渐关注样本数量更少的异常类别。另一方面也因为编码器中的特征缩聚和激励模块有效地进行了特征的重新标定，使分类器能够更好地从小样本异常数据中获取有效的异常特征。

4.3　非参数统计实验

通过非参数统计实验进一步证实本文方法的优

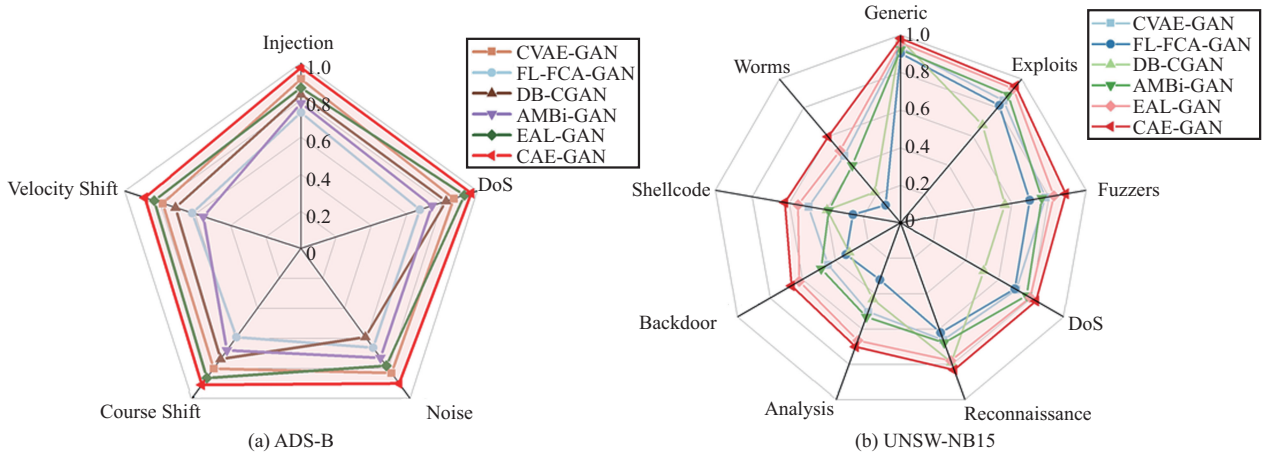

图7　ADS-B和UNSW-NB15数据集上各异常类的G-mean值

势。通常，Friedman 测试用以表明 2 种对比方法之间是否存在显著差异。换言之，通过 Friedman 测试结果来表明对比方法在多个数据集上的平均性能差异程度。在 0.05 的置信水平下，分别以 F_1 分数、AUC 值和 G-mean 值为指标进行 Friedman 测试，结果如表 3 所示。

表 3　　Friedman 测试结果

指标	F 值	p 值	假设结果（置信水平 0.05）
F_1 分数	3.04	4.22×10^{-3}	拒绝
AUC	21.2	5.01×10^{-4}	拒绝
G-mean	41.6	1.02×10^{-6}	拒绝

Friedman 测试结果的 p 值均小于置信水平 0.05，这表明 Friedman 假设被拒绝，各异常检测方法之间存在明显差异性。进一步通过 Friedman 测试排名分数直观地展示每种方法之间的差异，结果如图 8 所示。更高的排名分数表明方法具有更好的异常检测能力。可见，以 F_1 分数、AUC 值和 G-mean 值作为衡量指标，CAE-GAN 方法都获得了最高 Friedman 测试排名分数，并且与其他方法相比正向差异性非常显著。

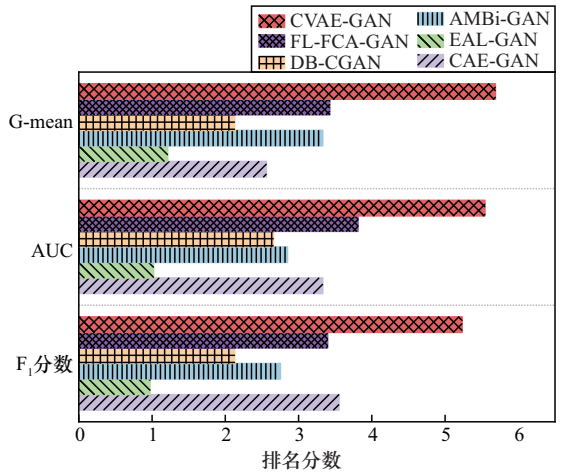

图 8　基于 Friedman 测试的排名分数

进一步进行 Nemenyi 后续检验，在 4 个数据集上比较 6 种方法时，Nemenyi 后续检验的 $\alpha = 0.05$ 的临界值 q_α 为 2.576。相应的 CD 值根据式(30)计算为 3.40，进而获得 Nemenyi 后续检验的 CD 图如图 9 所示。结果表明，在 95% 的置信水平（$\alpha = 0.05$）下，本文方法均排名第 1，且明显优于其他方法。特别是以 G-mean 作为评价指标时，与其他方法相比，CAE-GAN 方法表现出的正向差异性更为显著。

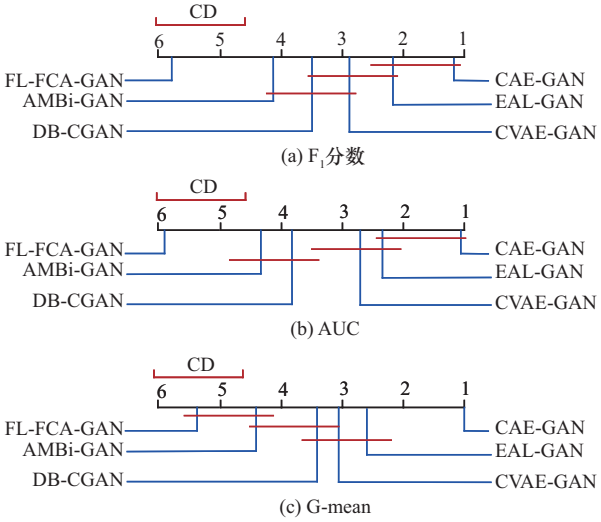

图 9　Nemenyi 后续检验的显著性差异

4.4　消融实验

为了进一步验证本文方法中各个主要功能模块的作用效果，在完整 CAE-GAN 方法基础上进行消融实验。消融实验中主要考虑 3 个主要创新工作，即类标签向量（CLV, class label vector）、编码器（En）和 CAT 机制。因此，在消融实验中，将完整的 CAE-GAN 方法与无类标签向量（CAE-GAN w/o CLV）、无编码器（CAE-GAN w/o En）和无协同对抗训练（CAE-GAN w/o CAT）3 种情况进行异常检测效果对比，消融实验结果如表 4 所示。

由表 4 可知，消融掉类标签向量、编码器或者协同对抗训练等主要功能模块后，CAE-GAN 方法在 F_1 分数、AUC 值和 G-mean 值方面均性能下降，说明本文提出的创新性工作具有显著优势。进一步，不同功能模块的影响在不同数据集上也体现出差异性。类标签向量对类别较多且不平衡问题突出的数据集影响更大，所以消融掉类标签向量后，UNSW-NB15 数据集上 F_1 分数下降了约 0.06，AUC 值下降了约 0.07，G-mean 值下降了约 0.06。而对类别较少的 UA 数据集，3 个指标的下降都非常微弱。编码器则对特征较多的数据集影响较大，如 NSL-KDD 数据集上，消融编码器后，3 个指标分别下降了约 0.08、0.06 和 0.07。而对于只有 7 个特征的 ADS-B 数据集，其影响甚微。协同对抗训练机制对所有数据集均体现出较为明显的影响，所以消融掉协同对

表 4　　消融实验结果

评价指标	检测方法	UA	ADS-B	NSL-KDD	UNSW-NB15
F_1 分数	CAE-GAN	**0.975 7**	**0.943 8**	**0.889 9**	**0.838 2**
	CAE-GAN w/o CLV	0.959 6	0.906 5	0.813 9	0.749 2
	CAE-GAN w/o En	0.929 1	0.940 6	0.807 5	0.756 3
	CAE-GAN w/o CAT	0.908 1	0.889 4	0.800 9	0.774 4
AUC 值	CAE-GAN	**0.979 2**	**0.933 2**	**0.943 3**	**0.896 3**
	CAE-GAN w/o CLV	0.970 1	0.886 1	0.853 2	0.775 3
	CAE-GAN w/o En	0.913 5	0.926 5	0.888 4	0.743 7
	CAE-GAN w/o CAT	0.900 9	0.858 5	0.867 8	0.824 6
G-mean 值	CAE-GAN	**0.965 7**	**0.922 8**	**0.918 9**	**0.903 8**
	CAE-GAN w/o CLV	0.958 6	0.875 1	0.854 6	0.797 9
	CAE-GAN w/o En	0.912 4	0.900 3	0.844 9	0.789 8
	CAE-GAN w/o CAT	0.898 1	0.858 2	0.854 6	0.840 5

抗训练功能后所有数据集上 3 个指标均明显下降。需要特殊说明的是，在实际场景下，检测数据更为复杂多变，依靠单一功能模块很难保持较好的检测效果，本文方法融合了多个创新的功能模块，更适用于复杂多变的智能无人机网络异常检测任务。

5　结束语

针对智能无人机网络异常检测中数据多类别不平衡的问题，本文提出了基于协同对抗增强生成模型的多类异常检测方法 CAE-GAN。CAE-GAN 方法构建了一个集成多个生成器的 GAN 框架，提出了一种特殊的类标签概率向量用于多类别样本生成，以逐渐增强模型对相对少数异常类的关注，克服了多类别不平衡。设计了一个带有特征缩聚和激励模块的编码器，以辅助判别器和分类器进行样本鉴别和多类别异常检测，通过重新校准异常相关的关键特征，强化了模型从小样本中学习异常特征的能力。提出了分类器和判别器共同引导样本生成的协同对抗训练策略，并设计了一种新颖的损失函数，实现了生成数据和真实数据之间的分布偏差协同感知和稳健纠正。在 4 个数据集上将本文方法与 5 种方法进行了比较，样本增扩、异常检测和统计测试等方面的实验结果都证明了 CAE-GAN 方法的优异性能。上述工作通过融合多粒度数据增扩、深度特征提取和分布偏差缓解的一体化综合方案，为智能无人机网络异常检测，特别是多类别不平衡学习任务，提供了一种有效的生成式人工方法。该方法的有效性和先进性已在本文中得到初步验证。未来，在方法优化层面，将进一步探索 CAE-GAN 面向具有更多异常类别和更高维度数据集的检测任务的改进，也将面向多噪声、标签缺失、含有未知威胁等更复杂的无人机网络异常检测任务场景，优化该方法以扩大其适用范围。在实际应用层面，下一步将重点研究 CAE-GAN 与智能无人机网络的组网方式、数据格式、时延要求、硬件平台等实际环境的适配，以加速本文方法在低空多场景应用。

参考文献：

[1] GUPTA R, SHUKLA A, TANWAR S. BATS: a blockchain and AI-empowered drone-assisted telesurgery system towards 6G[J]. IEEE Transactions on Network Science and Engineering, 2021, 8(4): 2958-2967.

[2] 吴一全, 童康. 基于深度学习的无人机航拍图像小目标检测研究进展[J]. 航空学报, 2025, 46(3): 181-207.
WU Y Q, TONG K. Research advances on deep learning-based small object detection in UAV aerial images[J]. Acta Aeronautica et Astronautica Sinica, 2025, 46(3): 181-207.

[3] SANGAIAH A K, YU F N, LIN Y B, et al. UAV T-YOLO-rice: an enhanced tiny yolo networks for rice leaves diseases detection in paddy agronomy[J]. IEEE Transactions on Network Science and Engineering, 2024, 11(6): 5201-5216.

[4] HUANG C H, CHEN W T, CHANG Y C, et al. An edge and trustworthy AI UAV system with self-adaptivity and hyperspectral imaging for air quality monitoring[J]. IEEE Internet of Things Journal, 2024, 11(20): 32572-32584.

[5] 孙长银, 袁心, 王远大, 等. 具身智能自主无人系统技术[J]. 自动化学报, 2025, 51(4): 762-777.
SUN C Y, YUAN X, WANG Y D, et al. Embodied intelligence autonomous unmanned systems technology[J]. Acta Automatica Sinica, 2025, 51(4): 762-777.

[6] 张钹, 朱军, 苏航. 迈向第三代人工智能[J]. 中国科学: 信息科学, 2020, 50(9): 1281-1302.
ZHANG B, ZHU J, SU H. Toward the third generation of artificial intelligence[J]. Scientia Sinica (Informationis), 2020, 50(9): 1281-1302.

[7] 李国旗, 洪晟, 兰雪婷, 等. 多旋翼无人机系统的信息安全参考模型[J]. 信息网络安全, 2022, 22(3): 10-19.
LI G Q, HONG S, LAN X T, et al. The security reference model of the multi-rotor UAV system[J]. Netinfo Security, 2022, 22(3): 10-19.

[8] 邓余婉祺, 王越, 杨超, 等. 无人机语义安全研究综述[J]. 计算机学报, 2025, 48(6): 1495-1515.
DENG Y W Q, WANG Y, YANG C, et al. A survey on semantic safety of unmanned aerial vehicle[J]. Chinese Journal of Computers, 2025, 48(6): 1495-1515.

[9] ZHANG Y Z, LI S B, ZHANG A S, et al. FW-UAV fault diagnosis based on knowledge complementary network under small sample[J]. Mechanical Systems and Signal Processing, 2024, 215: 111418.

[10] ZHONG J, ZHANG Y J, WANG J Y, et al. Unmanned aerial vehicle flight data anomaly detection and recovery prediction based on spatio-temporal correlation[J]. IEEE Transactions on Reliability, 2022, 71(1): 457-468.

[11] 胡天柱, 沈玉龙, 任保全, 等. 基于内存增强自编码器的轻量级无人机网络异常检测模型[J]. 通信学报, 2024, 45(4): 13-26.
HU T Z, SHEN Y L, REN B Q, et al. Lightweight anomaly detection model for UAV networks based on memory-enhanced autoencoders[J]. Journal on Communications, 2024, 45(4): 13-26.

[12] ZHAO Y F, LI J, SONG Z Y, et al. Language-inspired relation transfer for few-shot class-incremental learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 47(2): 1089-1102.

[13] CHENG X, SHI F, ZHANG Y, et al. FRAME: feature rectification for class imbalance learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2025, 37(3): 1167-1181.

[14] ZHOU X K, HU Y Y, WU J Y, et al. Distribution bias aware collaborative generative adversarial network for imbalanced deep learning in industrial IoT[J]. IEEE Transactions on Industrial Informatics, 2023, 19(1): 570-580.

[15] 于季弘, 林子砚, 叶能, 等. 基于三方生成对抗网络的隐蔽通信方法[J]. 通信学报, 2023, 44(11): 225-236.
YU J H, LIN Z Y, YE N, et al. Covert communication method based on tripartite generative adversarial network[J]. Journal on Communications, 2023, 44(11): 225-236.

[16] LI C X, XU K, ZHU J, et al. Triple generative adversarial networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(12): 9629-9640.

[17] HASSAN A U, MEMON I, CHOI J. Real-time high quality font generation with conditional font GAN[J]. Expert Systems with Applications, 2023, 213: 118907.

[18] WANG S D, LIU Z B, JIA Z, et al. Fault detection for UAVs with spatial-temporal learning on multivariate flight data[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 2529517.

[19] LI T T, HONG Z, CAI Q M, et al. BisSiam: bispectrum Siamese network based contrastive learning for UAV anomaly detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(12): 12109-12124.

[20] YANG L, LI S B, LI C J, et al. A survey of unmanned aerial vehicle flight data anomaly detection: technologies, applications, and future directions[J]. Science China Technological Sciences, 2023, 66(4): 901-919.

[21] DENG H L, LU Y, YANG T, et al. Unmanned aerial vehicles anomaly detection model based on sensor information fusion and hybrid multimodal neural network[J]. Engineering Applications of Artificial Intelligence, 2024, 132: 107961.

[22] BRULIN P Y, KHENFRI F, RIZOUG N. Generating fault databases through simulated and experimental multi-rotor UAV propulsion systems[J]. IEEE Transactions on Vehicular Technology, 2024, 73(4): 4671-4682.

[23] 唐立, 郝鹏, 任沛阁, 等. 基于改进孤立森林算法的无人机异常行为检测[J]. 航空学报, 2022, 43(8): 326789.
TANG L, HAO P, REN P G, et al. UAV abnormal behavior detection based on improved iForest algorithm[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(8): 326789.

[24] 顾兆军, 王婧煜, 王家亮, 等. 基于时空关联分析的无人机飞行数据异常检测[J]. 西安电子科技大学学报, 2025, 52(4): 33-45.
GU Z J, WANG J Y, WANG J L, et al. Spatial-temporal correlation-based anomaly detection in UAV flight data[J]. Journal of Xidian University, 2025, 52(4): 33-45.

[25] 张永清, 卢荣钊, 乔少杰, 等. 一种基于样本空间的类别不平衡数据采样方法[J]. 自动化学报, 2022, 48(10): 2549-2563.
ZHANG Y Q, LU R Z, QIAO S J, et al. A sampling method of imbalanced data based on sample space[J]. Acta Automatica Sinica, 2022, 48(10): 2549-2563.

[26] 顾兆军, 刘婷婷, 高冰, 等. 基于GAN-Cross的工控系统类不平衡数据异常检测[J]. 信息网络安全, 2022, 22(8): 81-89.
GU Z J, LIU T T, GAO B, et al. Anomaly detection of imbalanced data in industrial control system based on GAN-cross[J]. Netinfo Security, 2022, 22(8): 81-89.

[27] FARSHIDVARD A, HOOSHMAND F, MIRHASSANI S A. A novel two-phase clustering-based under-sampling method for imbalanced classification problems[J]. Expert Systems with Applications, 2023, 213: 119003.

[28] WANG R F, QIU H, JIANG G Q, et al. Class-imbalanced spatial-temporal feature learning for blade icing recognition of wind turbine[J]. IEEE Transactions on Industrial Informatics, 2024, 20(8): 10249-10258.

[29] TANG J J, LI Y, HOU Z J, et al. Robust two-stage instance-level cost-sensitive learning method for class imbalance problem[J]. Knowledge-Based Systems, 2024, 300: 112143.

[30] ABDELMONEM S, ELREEDY D, SHAHEEN S I. CIRA: class imbalance resilient adaptive Gaussian process classifier[J]. Knowledge-Based Systems, 2024, 304: 112500.

[31] 陆克中, 陈超凡, 蔡桓, 等. 面向概念漂移和类不平衡数据流的在线分类算法[J]. 电子学报, 2022, 50(3): 585-597.
LU K Z, CHEN C F, CAI H, et al. Online classification algorithm for concept drift and class imbalance data stream[J]. Acta Electronica Sinica, 2022, 50(3): 585-597.

[32] LI C J, LUO K, YANG L, et al. A zero-shot fault detection method for UAV sensors based on a novel CVAE-GAN model[J]. IEEE Sensors Journal, 2024, 24(14): 23239-23254.
[33] ZHANG C, TAN K C, LI H Z, et al. A cost-sensitive deep belief network for imbalanced classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(1): 109-122.
[34] CROITORU F A, HONDRU V, IONESCU R T, et al. Diffusion models in vision: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(9): 10850-10869.
[35] 刘伟欣, 管晔玮, 霍嘉荣, 等. 一种基于安全多方计算的快速 Transformer 安全推理方案[J]. 计算机研究与发展, 2024, 61(5): 1218-1229.
LIU W X, GUAN Y W, HUO J R, et al. A fast and secure transformer inference scheme with secure multi-party computation[J]. Journal of Computer Research and Development, 2024, 61(5): 1218-1229.
[36] ZHENG M, LI T, ZHU R, et al. Conditional Wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification[J]. Information Sciences, 2020, 512: 1009-1023.
[37] DLAMINI G, FAHIM M. DGM: a data generative model to improve minority class presence in anomaly detection domain[J]. Neural Computing and Applications, 2021, 33(20): 13635-13646.
[38] ZHU G Y, ZHOU K, LU L, et al. Partial discharge data augmentation based on improved Wasserstein generative adversarial network with gradient penalty[J]. IEEE Transactions on Industrial Informatics, 2023, 19(5): 6565-6575.
[39] LUO Y X, YANG Z W. DynGAN: solving mode collapse in GANs with dynamic clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(8): 5493-5503.
[40] KONG F H, LI J Q, JIANG B, et al. Integrated generative model for industrial anomaly detection via bidirectional LSTM and attention mechanism[J]. IEEE Transactions on Industrial Informatics, 2023, 19(1): 541-550.
[41] CHEN Z, DUAN J, KANG L, et al. Supervised anomaly detection via conditional generative adversarial network and ensemble active learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(6): 7781-7798.
[42] 王坤, 付钰, 段雪源, 等. 基于深度学习的 SDN 异常流量分布式检测方法[J]. 通信学报, 2024, 45(11): 114-130.
WANG K, FU Y, DUAN X Y, et al. Distributed abnormal traffic detection method for SDN based on deep learning[J]. Journal on Communications, 2024, 45(11): 114-130.
[43] PARK K H, PARK E, KIM H K. Unsupervised fault detection on unmanned aerial vehicles: encoding and thresholding approach[J]. Sensors, 2021, 21(6): 2208.
[44] 丁建立, 邹云开, 王静, 等. ADS-B 异常检测模型研究[J]. 航空学报, 2019, 40(12): 323220.
DING J L, ZOU Y K, WANG J, et al. ADS-B anomaly detection model based on deep learning[J]. Acta Aeronautica et Astronautica Sinica, 2019, 40(12): 323220.
[45] BINKOWSKI M, SUTHERLAND D J, ARBEL M, et al. Demystifying MMD GANs[J]. arXiv Preprint, arXiv: 1801.01401, 2018.

[作者简介]

隋嵩（1987– ），男，吉林省吉林市人，博士，中国民航大学讲师、硕士生导师，主要研究方向为民航智能设备与系统、安全评估、异常检测。

马春燕（1999– ），女，山东滨州人，中国民航大学硕士生，主要研究方向为民航智能系统安全检测。

龙岭春（2002– ），女，湖南邵阳人，中国民航大学硕士生，主要研究方向为民航智能系统安全检测。

顾兆军（1966– ），男，黑龙江哈尔滨人，博士，中国民航大学教授、博士生导师，主要研究方向为民航网络安全、网络安全检测与评估。

刘佳佳（1988– ），女，山西大同人，博士，中北大学讲师、硕士生导师，主要研究方向为智能设备与系统。

丁磊（1995– ），男，天津人，广州大学博士生，主要研究方向为深度学习、异常检测。