中北大学学报(自然科学版)
JOURNAL OF NORTH UNIVERSITY OF CHINA(NATURAL SCIENCE EDITION)
2026年 第47卷 第4期(总第228期)
Vol. 47 No. 4 2026(Sum No. 228)

文章编号：1673-3193(2026)04-0541-12 J. North Univ. China, Nat. Sci. Ed., 2026, 47(4): 541-552.

基于元学习与VAE-GAN的META-VAEGAN-DNN网络入侵检测方法

王 璐[1], 赵利辉[1*], 安 洋[1], 符 赫[2], 常家瑜[1]

(1. 中北大学 软件学院, 山西 太原 030051; 2. 中北大学 半导体与物理学院, 山西 太原 030051)

摘 要： 针对网络入侵检测中数据集不平衡导致少数类攻击识别困难的问题，本文提出一种基于元学习与变分自编码器生成对抗网络(VAE-GAN)的网络入侵检测方法META-VAEGAN-DNN。首先，构建META-VAEGAN数据增强模型以缓解数据集不平衡的问题；其次，提出差异化样本处理策略，并针对少数类样本设计渐进式质量筛选机制，在提升整体样本质量的同时确保少数类样本的有效生成；最后，通过结合通道注意力与多层残差连接的DNN模型对生成的平衡数据集进行分类。基于NSL-KDD数据集的实验表明，META-VAEGAN-DNN模型在网络攻击检测中的准确率、$F1$分数和AUC分别达到了89%，90%和98%；其中U2R类别的精确率和召回率分别提升至98%和99%；R2L类别的精确率和召回率分别提升至99%和73%，显著优于传统方法。在CIC-IDS-2017数据集上的辅助验证进一步证明了模型的通用性。

关键词： 变分自编码器；生成对抗网络；元学习；入侵检测

中图分类号： TP393.08 **文献标识码：** A **doi:** 10.62756/jnuc.issn.1673-3193.2025.09.0002

引用格式： 王璐, 赵利辉, 安洋, 等. 基于元学习与VAE-GAN的META-VAEGAN-DNN网络入侵检测方法[J]. 中北大学学报(自然科学版), 2026, 47(4): 541-552.

Wang Lu, Zhao Lihui, An Yang, et al. A meta-learning and VAE-GAN based META-VAEGAN-DNN method for network intrusion detection[J]. Journal of North University of China(Natural Science Edition), 2026, 47(4): 541-552.

A Meta-Learning and VAE-GAN Based META-VAEGAN-DNN Method for Network Intrusion Detection

Wang Lu[1], Zhao Lihui[1*], An Yang[1], Fu He[2], Chang Jiayu[1]

(1. School of Software, North University of China, Taiyuan 030051, China;
2. School of Semiconductor and Physics, North University of China, Taiyuan 030051, China)

Abstract: Class imbalance in network intrusion detection makes minority attack classes particularly challenging to identify. To address this issue, we proposed META-VAEGAN-DNN, a novel intrusion detection framework that integrated meta-learning with a variational autoencoder generative adversarial network (VAE-GAN). A META-VAEGAN-based augmentation model was first developed to alleviate class imbalance, followed by a differentiated sample processing strategy with a progressive quality filtering mechanism designed to enhance the generation of minority-class instances. The augmented data were then classified using a deep

收稿日期： 2025-09-05

基金项目： 山西省基础研究计划青年科学研究项目(202203021212114)

作者简介： 王 璐(1998—), 女, 硕士生, 主要从事网络安全领域的研究。

* **通信作者：** 赵利辉(1979—), 男, 副教授, 博士, 主要从事网络测试与网络安全方向的研究。E-mail: leehwi@nuc.edu.cn。

neural network (DNN) enhanced with channel attention and multi-layer residual connections. Experiments on the NSL-KDD dataset demonstrate that META-VAEGAN-DNN achieves 89% accuracy, 90% *F*1-score, and 98% AUC. Notably, the precision and recall of the U2R class reach 98% and 99%, respectively, while those of the R2L class reach 99% and 73%, substantially outperforming conventional methods. These results highlight the effectiveness of META-VAEGAN-DNN in improving the detection of minority attack categories and enhancing overall intrusion detection performance. Supplementary validation on the CIC-IDS-2017 dataset further demonstrates the generalizability of the model.

Key words: variational autoencoder; generative adversarial network; meta-learning; intrusion detection

0 引 言

随着网络技术的快速发展，电子设备数量激增，网络环境变得愈加复杂，网络安全问题也愈加严重，如隐私泄露、恶意攻击等事件频发，给社会和经济带来了巨大损失。因此，如何有效防范网络攻击已成为当前亟待解决的重要课题。

为有效防范网络攻击，入侵检测系统(Intrusion Detection System，IDS)得到广泛应用，其通过主动防御策略来有效检测入侵并及时响应，在应对多种网络威胁时发挥了重要作用[1]。IDS通常可以分为基于主机的IDS和基于网络的IDS。前者通过采集和分析系统日志、应用日志、端口调用、系统调用和审计记录检测入侵行为；后者则通过采集和分析网络数据流量、网络数据包以及协议检测入侵行为[2]。基于机器学习(Machine Learning，ML)的IDS在过去几十年中取得了显著进展，如贝叶斯[3]、支持向量机[4]等。随着攻击方式的不断进化，传统机器学习模型由于无法有效捕捉深层特征，已经无法满足入侵检测的需求，而深度学习(DL)因其强大的特征提取能力，目前在IDS领域得到广泛应用[5]。如卷积神经网络[6]、循环神经网络[7]、生成式对抗网络[8]等。然而，由于网络入侵检测训练数据不足以及数据集类别不平衡[9]，大部分深度学习模型面临特征提取不足的问题。近年来，针对网络入侵检测训练的公开数据集不断增加，但模型对少数攻击类别的检测准确率仍然较低。其原因在于网络入侵检测数据集中部分攻击类别的样本数量仍然较少，导致模型在训练过程中更侧重于学习多数类样本，难以充分提取少数类样本的特征，进而产生过拟合问题。

1 相关工作

随着相关研究的不断深入，针对数据集不平衡问题，研究者们提出了多种数据增强模型。王华忠等[10]使用条件Wasserstein生成对抗网络(Conditional Wasserstein Generative Adversarial Networks，CWGAN)进行数据不平衡处理，通过引入Wasserstein距离来衡量真实样本与生成样本的概率分布差异，避免了GAN的模式崩溃问题；Yang等[11]提出一种结合GAN和条件编码器的模型，其借助条件编码聚合结构，在维度转换阶段优先区分不同攻击类别，并通过复合损失函数从多维度约束训练，确保新生成样本的质量。Kang等[12]提出一种基于VAE和GAN的跨类别生成增强策略，该模型除引入变分GMM外，还通过分类迁移预训练与针对性微调的结合，有效捕捉了多数类及少数类攻击的关键特征，同时增设辅助分类器以强化生成数据的辨别性。Yuan等[13]提出了一种名为B-GAN的数据均衡方法。B-GAN在生成器与鉴别器模块均采用LSTM网络模型来更有效地捕捉数据中的时序特征与序列规律，生成了具备高可信度的异常样本。

上述模型从不同的角度出发，虽在一定程度上解决了数据集不平衡问题，但普遍忽视了对少数类样本的针对性优化，导致模型对少数类攻击检测的准确率偏低。此外，原始网络流量数据的高维特性使得少数类样本生成难度增加。

为解决上述问题，本文提出了基于元学习与VAE-GAN的网络入侵检测方法META-VAEGAN-DNN，通过在低维潜在空间中引入元学习动态优化机制，实现了对少数类攻击样本的高质量增强。主要工作包括：

1) 针对入侵检测数据集不平衡以及原始高维流量数据导致的少数类样本生成困难的问题，提出一种融合VAE-GAN与元学习的数据增强模型META-VAEGAN，采用VAE作为生成器，通过编码器实现高维到低维的特征映射从而降低生成过程的复杂度。判别器为多任务学习模型，一方面进行生成数据的真假判断，另一方面进行生成

数据的数据分类，同时模型基于元学习设计了元权重控制器，动态调整对抗训练中的权重，使模型更关注少数类攻击样本的生成。

2）针对网络流量数据样本分布差异大的特点，本文设计了一种差异化样本处理策略，在META-VAEGAN模型的新样本生成过程中，针对不同样本采取不同的处理方法，从而进一步提升样本质量。同时，为了解决少数类别样本生成质量不佳问题，设计了渐进式质量筛选机制。该机制基于多任务判别器的类别概率分布，动态筛选生成的少数类别样本，在有效解决数据不平衡问题的同时可保证少数类样本的生成质量。

3）利用残差网络和注意力机制结合的分类模型对新生成的平衡数据集进行分类测试。

2　META-VAEGAN-DNN模型

2.1　模型整体框架

为缓解入侵检测数据集类别不平衡与少数类样本难以生成的问题，本文提出了META-VAEGAN-DNN模型，将变分自编码器VAE[14]、生成对抗网络GAN与元学习机制[15]三者融合。VAE通过编码器将数据映射至连续、可采样的潜在空间并由解码器重建，但其以最小化重构误差为目标，易生成趋于多模态均值的样本，难以捕捉少数类攻击的特征。GAN通过生成器与判别器的对抗学习提升样本细节与真实性，但在入侵检测数据高度不平衡场景下，少数类样本稀缺导致判别器学习不足，生成器易发生模式崩溃。针对上述局限，本文由VAE通过先验对齐构建连续且可采样的潜在空间，由GAN的对抗训练增强生成样本的细节与真实性，并引入元学习机制自适应地提升模型对少数类样本的关注，从而在缓解数据集不平衡的同时，促进少数类样本的生成数量与质量同步提升。

本文提出的META-VAEGAN-DNN模型主要包括数据增强模块、差异化样本处理模块和分类模块三部分，如图 1 所示。

经过数据增强和少数类别质量筛选后的平衡数据集将用于训练检测模型和优化模型参数。在训练过程中，基于训练集获得最优参数后，使用测试集对模型性能进行评估。

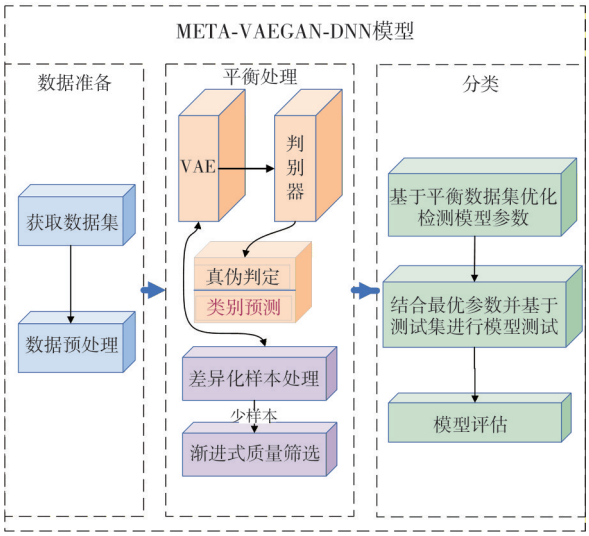


图 1　META-VAEGAN-DNN模型总体框架

Fig. 1　Overall framework of META-VAEGAN-DNN

2.2　META-VAEGAN数据增强模块

为解决入侵检测领域中数据集类别分布不均衡导致的少数类攻击样本特征学习不足的问题，本文构建了META-VAEGAN数据增强模型，通过对抗生成网络机制合成高质量的攻击样本，从而有效提升检测模型对少数类攻击的学习能力。META-VAEGAN模型的整体结构如图 2 所示。

META-VAEGAN模型在VAE-GAN模型上进行了两点重要优化：1）判别器设计为多任务网络，同时执行真假判别以及分类任务；2）设计了元权重控制器，旨在优化少数类别样本的生成效果。

2.2.1　多任务判别器

为了应对传统单一任务判别器对少数类别样本关注不足的问题，本研究对VAE-GAN模型进行了改进，设计了多任务判别器，其模型结构如图 3 所示。

该判别器在保持真假判别功能的基础上进行类别预测，其损失函数采用了多任务学习的策略，公式为

$$L_D^{\text{weighted}} = -E_{(x,y)\sim P_{\text{data}}}[\beta(y;x;t)\log D(x,y)] - E_{(z,y)\sim p_z p_y}[\log(1-D(G(z,y),y))], \quad (1)$$

式中：E为数学期望；$D(x,y)$为判别器对输入数据x且类别为y的判别概率；$G(z,y)$为生成器输出的样本；$\beta(y;x;t)$为新引入的类别加权因子。

$$\beta(y;x;t) = (1-\lambda(t))\beta\_prior(y) + \lambda(t)\beta\_meta(y;x;t), \quad (2)$$

$$\beta\_prior(y) = clip\left(\frac{r\_y}{\bar{r}}, w\_\min, w\_\max\right), \quad (3)$$

$$r\_y = (N\_\max / N\_y)^{\gamma} (\gamma \in [0.5, 2.0]), \quad (4)$$

$$\bar{r}=\frac{1}{k}\sum_{c=1}^{k}r_c, \tag{5}$$

式中：$\lambda(t)=t/T\in[0,1]$，t为当前epoch；T为总epoch数；$\beta_prior(y)$为先验统计权重；N_y为类别y的样本数量；$N_\max$为最大类别的样本数量；γ为不平衡敏感度系数，实验中采用$\gamma=1$的默认设定；$\beta_meta(y;x;t)$为动态元学习权重，由元权重控制器输出，该因子根据样本的类别来动态调整判别任务的权重。针对少数类样本，$\beta(y)$会增加，从而提升判别器对这些样本的关注度。

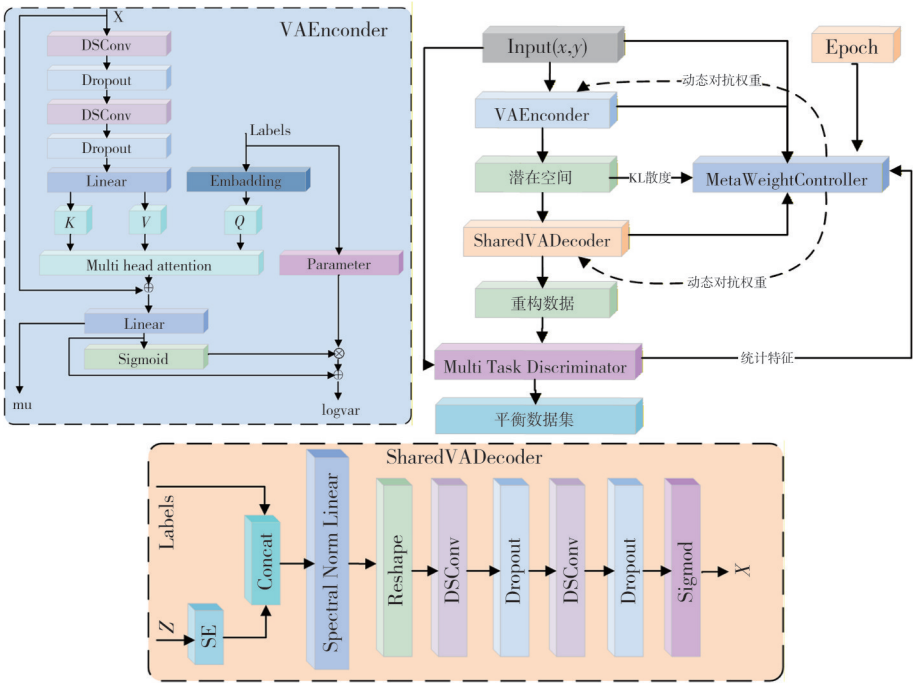


图2 META-VAEGAN模型结构

Fig. 2 Structure of the META-VAEGAN model

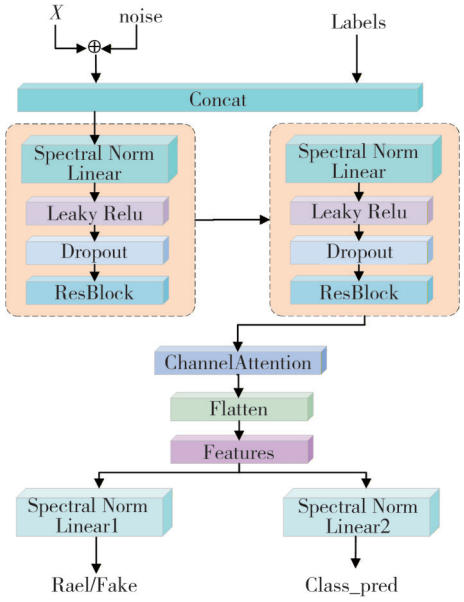


图3 MultiTaskDiscriminator模型结构图

Fig. 3 Structure of the MultiTaskDiscriminator model

对于输入的样本x及其标签y，首先通过共享特征提取网络提取共同特征，然后通过两个独立的分支网络(一个负责真假判断，另一个负责类别预测)产生不同输出，其中，类别预测分支会输出未归一化的类别分数，最后通过softmax函数转换为归一化的类别概率，此概率指标在训练中不仅是样本质量评估的重要指标，而且为后续差异化样本处理策略提供了支撑。

2.2.2 元权重控制器

在生成模型中，不同样本的重构误差、KL散度以及对抗损失在训练过程中具有不同的重要性，如果统一使用固定的权重系数，容易导致模型对某些少数类别生成能力不足。因此，本文设计了一个元权重控制器，用于动态自适应地分配每个样本的损失权重，并在训练过程中逐步调整，以平衡重构与生成对抗过程的稳定性，其模型结构如图4所示。

元权重控制器是一个多层感知机(MLP)，输入为当前样本的统计特征和类别向量以及epoch进度，输出自适应权重。元权重控制器的目标函数为

$$L_{\text{meta}}(\theta)=(1-\rho)E_i[-S_i\omega_i(\theta)]+\rho,$$
$$E_{i:y_i\in\mathrm{R}}[-S_i\omega_i(\theta)]-\lambda_c E_i[c_i]+\lambda_f E_i[d_i]+\lambda_2\|\theta\|_2^2, \tag{6}$$

式中：ρ为权重平衡系数，用于调节全体样本与少数样本之间的损失权重；E为数学期望，用于对样本分布下的损失进行平均；S_i为判别器对生成样本的真假评分；c_i为分类最大类别概率；d_i为真假特征距离；类别概率惩罚系数λ_c与特征距离系数λ_f采用对称设计，均设置为0.1。该参数配置参考了同类研究的经验值，既能有效防止模型过拟合，又可保持特征多样性。L_2正则系数λ_2设置为最小量级0.001，可以在约束参数范数的同时保证元学习主目标不被削弱。$\omega_i(\theta)$为样本权重函数，即上文提到的动态元学习权重。

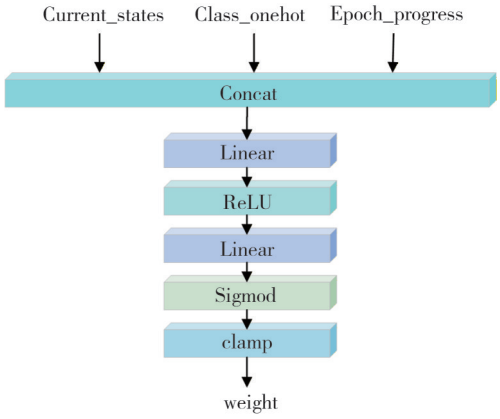


图 4　MetaWeightController模型结构

Fig. 4　Structure of the MetaWeightController model

对样本(x_i, y_i)，构造包含7项统计信号的统计特征向量$s(x_i)$，表示为

$$s(x_i)=[\ell_{rec}, \ell_{kl}, |s_i|, t/T, c_i, H(p_i), d_i], \quad (7)$$

式中：ℓ_{rec}和ℓ_{kl}分别为样本的重建误差和KL散度，用来衡量编码质量与潜在空间的不确定性；$|s_i|$为判别器真假评分的幅值，反映对抗训练的难度；$H(p_i)$为分类概率熵；d_i为真假特征的欧式距离，反映分布的偏移程度。这些特征能有效识别训练不足的样本，少数类样本因模型欠拟合表现出更高的重建误差与KL散度、更低的判别器评分以及更大的分布偏移，元权重控制器通过学习这些统计特征，能够自动识别训练不足的少数类样本并动态上调其损失权重。

基于$s(x_i)$、类别向量e_{y_i}与训练进度t/T，权重函数定义为

$$\omega_i(\theta)=\beta\_meta(y; x; t)=clip(0.3\cdot\sigma(MLP([s(x_i), e_{y_i}, t/T])), 0.01, 0.5)。 \quad (8)$$

该函数通过显式输入类别向量实现类条件权重分配，使少数类学习到更大权重；$s(x_i)$促使MLP在少数类样本上输出更高权重；训练进度t/T使权重随训练动态优化；同时截断到[0.01, 0.5]保证权重稳定且有效覆盖所有样本，截断区间的设置将在第4节进行讨论。该机制本质上是通过元学习隐式来估计样本梯度的权重，相当于对偏斜训练分布的密度比校正。少数类样本在加权判别损失中获得更高权重系数，使其在参数更新中产生更大的梯度贡献，提升模型对少数类样本的关注度，从而有效缓解了类别不平衡导致的生成器模式崩溃问题。

2.3　差异化样本处理策略

2.3.1　差异化样本处理原理

针对META-VAEGAN数据增强过程中生成样本质量不均的问题，本文提出了一种差异化样本处理策略，通过对不同样本特性施加不同处理方法有效提升了生成样本的整体质量和多样性，其结构图如图5所示（以NSL-KDD数据集为例）。

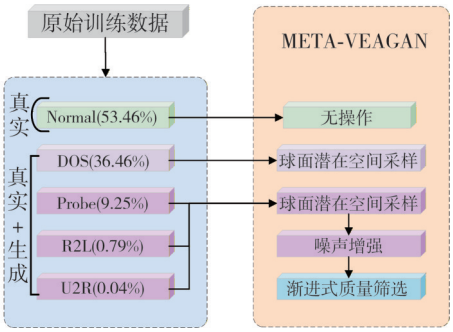


图 5　差异化样本处理模块结构

Fig. 5　Structure of the differentiated sample processing module

首先对NSL-KDD数据集进行类别分布统计分析，对于Normal类别，由于真实样本充足，直接使用原始真实样本，无需生成操作。对于DoS、Probe、R2L和U2R类别，采用真实样本与生成样本混合策略，利用VAE模型学习到的潜在空间，执行球面潜在空间采样。球面潜在空间采样的核心思想是将潜在向量限制在单位球面上，即向量的模为1。这样做的目的是确保潜在向量在高维空间中的分布更加均匀，从而提高生成样本的多样性[16]。针对少数类别Probe、R2L和U2R，由于其样本数量稀少，额外增加了噪声增强，目的是增强少数类别样本的多样性。

2.3.2　渐进式质量筛选机制

虽然球面潜在空间采样与噪声增强策略有助于增强生成样本的多样性，但这些生成的样本中仍可能存在与目标类别分布不一致的低质量样

本。因此，针对少数类别样本，本文设计了一种渐进式质量筛选机制，用于逐步过滤低质量样本。

该机制以多任务判别器输出的类别概率为核心评估指标，将生成样本划分为3个区间，划分原则如下：

1) $C_{\text{score}}(x_g, y_c) > T_{\text{high}}(c, e)$：高可信度区间。高可信度样本因其类别一致性高直接被保留。

2) $T_{\text{low}} < C_{\text{score}}(x_g, y_c) \leqslant T_{\text{high}}(c, e)$：中等可信度区间。对中等可信度样本采用概率采样策略，依据判别器给出的可信度评分，结合预设阈值计算保留概率，灵活筛选样本。

3) $C_{\text{score}}(x_g, y_c) \leqslant T_{\text{low}}$：低可信度区间。低可信度样本因其类别一致性不足被过滤。

$T_{\text{high}}(c, e)$为随类别C与训练轮次e动态调整的高可信度阈值，T_{low}为固定的低可信度阈值(本文设为0.05)。

可信度阈值采用基础阈值结合训练轮次衰减的动态调整策略。DoS、Probe、R2L、U2R类别基础阈值分别设为0.5，0.3，0.1，0.05，通过训练轮次相关计算调整阈值，适配训练阶段需求。

模型训练初期，生成样本质量较低，使用高阈值易导致样本流失；训练后期，生成样本质量提升，需提高筛选标准。为此，引入训练轮次衰减因子$\alpha(e)$对基础阈值进行动态调整，高可信度阈值计算公式为

$$\alpha(e) = 1 - \left(\frac{e}{E}\right)^2, \tag{9}$$

$$T_{\text{high}}(c, e) = T_{\text{base}}(c) + \alpha(e), \tag{10}$$

式中：e为当前训练轮次；E为总训练轮次；$\alpha(e)$为二次衰减因子。为了防止训练后期低质量样本混入，设置了全局约束$T_{\text{high}}(c, e) \geqslant 0.02$。

2.4 DNN分类模块

DNN分类模块用于执行入侵检测任务，其结构如图6所示。该模块采用包含多层残差结构和通道注意力机制的全连接网络架构。在训练阶段，模块以均衡后的训练集为输入，在测试阶段，模型对测试样本进行分类预测，评估其在各入侵类别上的检测性能，从而验证数据增强方法对检测能力的提升效果。

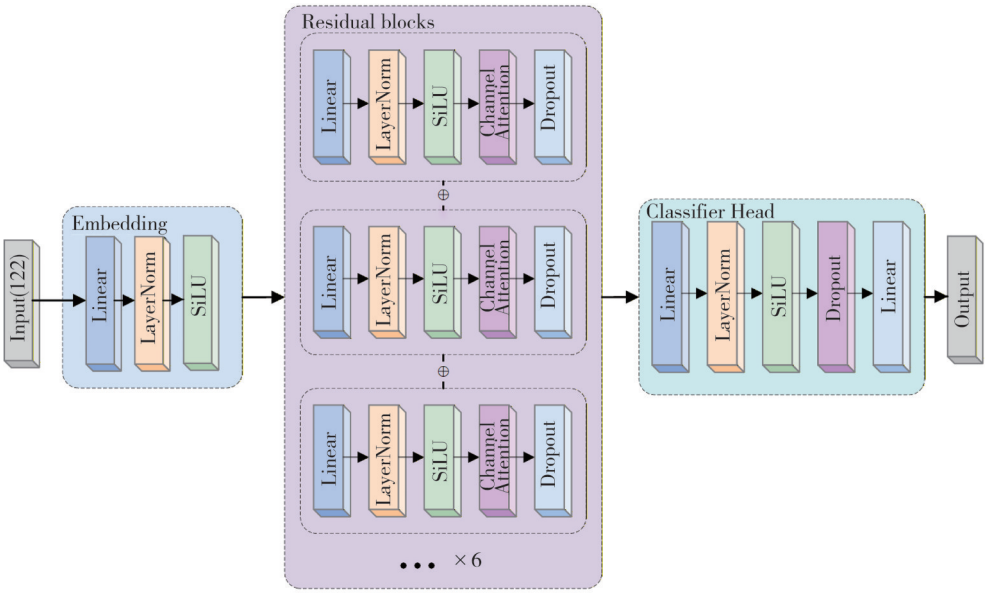


图6 DNN分类模块结构

Fig. 6 Structure of the DNN classification model

3 基于META-VAEGAN-DNN的网络入侵检测方法

本文提出的网络入侵检测方法采用META-VAEGAN模块结合差异化样本处理策略对少数类攻击样本进行数据增强，通过渐进式质量筛选机制确保生成样本的质量，最终利用DNN分类器完成入侵检测任务。整体入侵检测流程如图7所示。

3.1 数据预处理

数据预处理分为2步：

Step 1：完成数据清洗与特征工程，包括添加列名、删除冗余列、处理缺失值、字符型特征编码、攻击类型归类及标签数值化；

Step 2：进行数据标准化与归一化处理，按8∶2划分训练集和测试集，转换为PyTorch张量供模型训练。

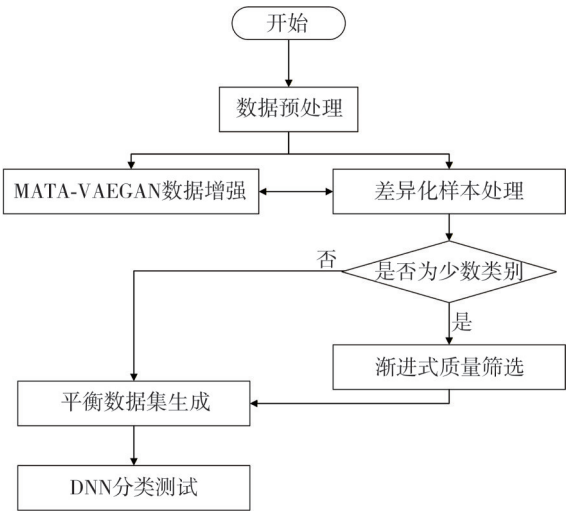


图 7　检测流程

Fig. 7　Detection process

3.2　平衡处理

利用META-VAEGAN模型并结合差异化样本处理策略，解决数据集中的类别不平衡问题。

3.3　少数类别质量筛选

利用渐进式质量筛选算法，逐步过滤少数类别中无效的低质量样本，生成平衡且高质量的数据集。

3.4　分类

使用多层残差结构和通道注意力机制的全连接网络DNN对生成的平衡数据集进行分类，实现网络攻击检测。

4　实验结果及分析

4.1　数据集

本文采用NSL-KDD数据集[17]作为基准数据集，并在CIC-IDS2017数据集[18]进一步评估模型性能。NSL-KDD数据集由Tavallaee等[19]于2009年提出，共包含37种攻击类型。根据攻击特性，数据集中的类别特征分为Normal、DoS、Probe、R2L和U2R五类。文献[20]介绍了各类攻击的功能特征。数据集构成如表1所示，可以看出数据集存在严重的类别不平衡问题，R2L和U2R攻击样本数量较少。

CIC-IDS2017数据集是由加拿大通信安全机构和加拿大网络安全研究院合作开发的，数据共计5 d，其中星期一仅包含正常流量，其余天数包含了各种攻击。攻击类型包括Botnet、DDos、Dos、FTP-Parator（FTP-Pat）、SSH-Parator（SSH-Pat）、PortScan（PS）和WebAttack（web）七种。本次实验选取其中5种具有代表性的攻击类型（Botnet、Benign、DDoS、DoS、Web）进行模型性能评估。

表 1　NSL-KDD数据集

Tab. 1　NSL-KDD dataset

原始类别标签	标签化后类别标签	样本数量/条	样本占比/%
Dos	0	53 387	35.95
Normal	1	77 054	51.88
Probe	2	14 077	9.48
R2L	3	3 880	2.61
U2R	4	119	0.08

表 2　CIC-IDS2017数据集

Tab. 2　CIC-IDS2017 dataset

原始类别标签	标签化后类别标签	样本数量/条	样本占比/%
Botnet	0	1 966	0.07
Benign	1	2 270 397	80.28
DDoS	2	128 027	4.53
DoS	3	252 661	8.93
FTP-Pat	4	7 938	0.28
PS	5	158 930	5.62
SSH-Pat	6	5 897	0.21
Web	7	2 180	0.08

4.2　实验设置与评价指标

为保证对比实验的公平性，本文对所有方法均采用统一的数据预处理流程；在与不同数据类别均衡方法的对比实验中，所有方法均采用相同架构的DNN分类器进行入侵检测；此外，所有实验均在相同硬件环境下运行，并在相同原始测试集上采用统一评估指标进行评估。

4.2.1　实验环境

框架：核心框架为PyTorch2.0.1，用于构建和训练模型。

GPU：NVIDIA Tesla P100。

数据集：NSL-KDD、CIC-IDS2017。

4.2.2　模型参数

1）VAE（3.84×10^6参数）：VAEncoder参数量约为2.08×10^6，SharedVADecoder参数量约为1.76×10^6；采用Adam优化器；初始学习率为1×10^{-4}；结合CosineAnnealingLR调度器将学习率逐步降至1×10^{-6}。

2）判别器（0.89×10^6参数）：Adam优化器；学习率为4×10^{-5}。

3）元权重控制器（0.48×10^3参数）：Adam优

化器；学习率为1×10^{-5}。

4）DNN(3.68×10^{6}参数)：Lookahead优化器包装的AdamW优化器，学习率为0.001。

5）对抗损失权重配置：元学习自适应权重通过元权重控制器动态预测，权重约束范围为[0.01,0.5]。下界0.01确保对抗训练的最小强度，防止权重退化至零导致元学习机制失效；上界0.5基于实验调优，在该值下模型能够平衡对抗与重构目标，保证生成质量。实验结果如表3所示，在[0.1,1]区间选取7个值进行测试，可见上界在0.5取值处模型效果最佳。

表3 不同对抗权重上界对模型性能的影响

Tab.3 The impact of different adversarial weight upper bounds on model performance

λ_{max}	准确率/%	精确率/%	$F1$/%	召回率/%
0.1	76	78	73	69
0.3	83	85	82	80
0.4	87	89	87	85
0.5	**91**	**92**	**90**	**89**
0.6	87	90	88	86
0.7	84	87	84	82
1.0	79	82	78	75

KL散度自适应权重的公式为

$$\beta=sigmoid(L\_recon)。\tag{11}$$

在重构误差较大时自动增强正则化强度，以保证潜在空间的连续性。

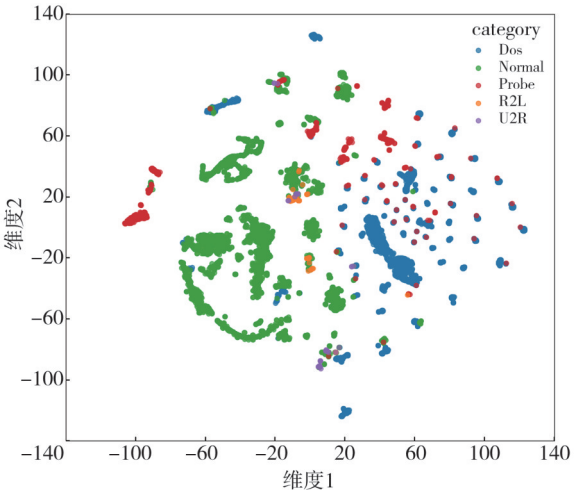
4.2.3 评价指标

评估指标包括准确率、精确率、召回率、$F1$分数和AUC[21]。AUC(Area Under Curve)用于衡量模型的综合性能，ROC曲线展示分类器在不同阈值下的真阳性率与假阳性率的变化，AUC值越大表示模型性能越好[22]。AUC在类别不平衡情况下能提供客观评估，且不受决策阈值的影响，是网络入侵检测性能比较的重要工具[23]。

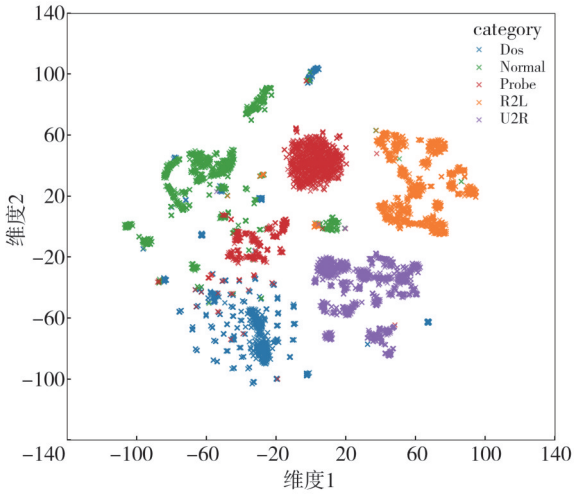
4.3 结果分析

1）首先使用T-SNE降维技术[24]对原始数据与增强后数据进行可视化对比，以NSL-KDD数据集为例的结果见图8。

考虑到计算效率，采用了分层采样策略，确保每类至少100个样本，总样本量为8 000。图8显示，增强前Normal类样本占主导，U2R类几乎不可见；增强后各类别形成明显分离的簇，特别是R2L和U2R类样本显著增加，且DoS和Probe类的聚类结构更为紧凑。结果证明，META-VAEGAN-DNN模型有效提升了少数类样本的数量。



(a) 数据增强前



(b) 数据增强后

图8 数据增强前后的T-SNE可视化对比

Fig.8 The T-SNE visualization comparison before and after data augmentation

2）本文使用的分类算法为深度神经网络模块(DNN)，为了验证算法的性能，本文同DT等模型进行了对比，以NSL-KDD数据集为例的结果如表4所示。可以看出，DNN的分类性能最佳，各项指标均优于其他的分类算法。

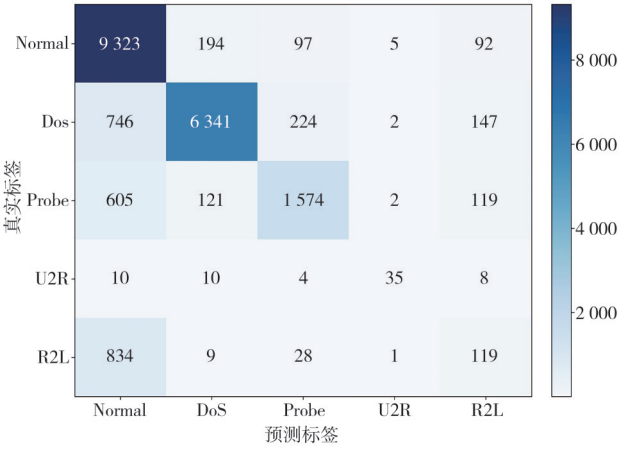
表4 分类算法对比结果

Tab.4 Comparison of classification algorithms %

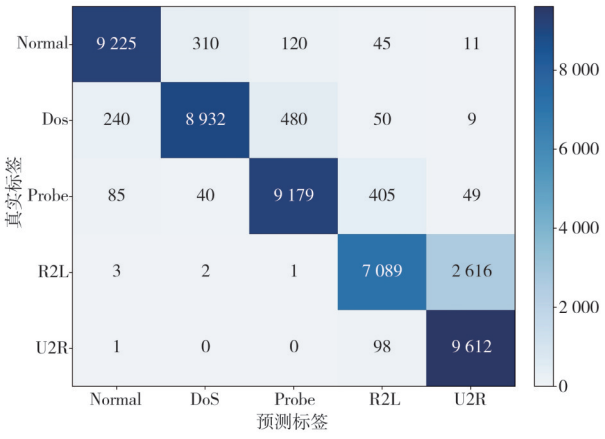
分类算法	准确率	精确率	$F1$	召回率
DT	75	80	72	75
KNN	74	78	70	74
LR	74	79	71	74
NB	53	69	44	53
RF	76	82	72	76
SVM	77	80	73	77
DNN	**79**	**83**	**75**	**79**

3）为了验证模型在极少数类别上的泛化能力，对比了DNN与META-VAEGAN-DNN在两个数

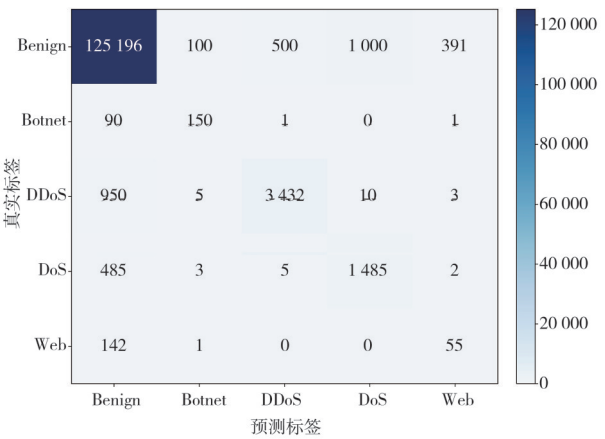
据集上的混淆矩阵，结果如图 9 所示。可以看出，DNN在少数类别上表现较差，大量样本被误判为多数类。META-VAEGAN-DNN的混淆矩阵呈现出明显的对角线高值且非对角线低值的特征，少数类别的召回率和精确率均显著提升。实验表明，虽然对极少数类别进行了大量样本生成，但模型在测试集上仍保持较好的精确率，主要混淆仅发生在特征相似的类别间，证明模型具有良好的泛化能力，质量筛选机制有效避免了过拟合风险。



(a) DNN模型/NSL-KDD数据集



(b) META-VAEGAN-DNN模型/NSL-KDD数据集



(c) DNN模型/CIC-IDS2017数据集

(d) META-VAEGAN-DNN模型/CIC-IDS2017数据集

图 9　两种模型在两个数据集上的混淆矩阵

Fig. 9　Confusion matrix of two model on two datasets

4）将META-VAEGAN-DNN与其他不同数据类别均衡方法的模型进行性能比较，通过分析不同模型在异常检测任务中的表现来验证混合模型及网络架构有效性，结果如表 5 所示。

由表 5 可以看出，META-VAEGAN-DNN在NSL-KDD与CIC-IDS2017两个数据集上均表现较好。在NSL-KDD上，相比其他类均衡方法，准确率提升 5 百分点～16 百分点，*F*1 分数提升 5 百分点～14 百分点，AUC 提升 7 百分点～15 百分点，在CIC-IDS2017上同样也有明显优势。实验表明，采用均衡样本测试的各种DNN模型的表现都比仅使用原始测试集的基础DNN更好，证明了类均衡方法对解决类不平衡问题的有效性。

表 5　META-VAEGAN与其他类均衡方法的比较

Tab. 5　Comparison of META-VAEGAN with other class equalization methods　%

数据集	模型	准确率	*F*1	AUC
NSL-KDD	DNN	75	72	89
	SMOTE-DNN[25]	79	77	86
	GAN-DNN[8]	79	76	86
	CGAN-DNN[26]	78	75	88
	ROS-DNN[27]	79	76	88
	ANASYN-DNN[28]	78	75	83
	CCWGAN-DNN[29]	86	85	91
	NBW[30]	85	84	90
	META-VAEGAN-DNN	**91**	**90**	**98**
CIC-IDS2017	DNN	72	69	86
	SMOTE-DNN	76	74	83
	GAN-DNN	77	74	84
	CGAN-DNN	75	72	85
	ROS-DNN	76	73	85
	ANASYN-DNN	75	72	81
	CCWGAN-DNN	84	83	90
	NBW	83	82	89
	META-VAEGAN-DNN	**92**	**89**	**96**

5）对不同类均衡方法在各类别上的性能表现进行了对比实验，结果如表 6 和表 7 所示。

表6 不同类均衡方法在NSL-KDD数据集各类别上的精确率与召回率比较

Tab. 6 Comparison of precision and recall for each class under different class balancing methods on the NSL-KDD dataset

模型	精确率/%				
	Normal	DOS	Probe	U2R	R2L
DNN	81	95	81	78	25
SMOTE-DNN	71	95	79	23	96
GAN-DNN	67	95	72	68	97
CGAN-DNN	67	94	84	63	83
ROS-DNN	71	95	82	12	93
ANASYN-DNN	71	96	76	14	72
CCWGAN-DNN	85	95	90	85	92
NBW	80	94	88	75	88
META-VAEGAN-DNN	**93**	**96**	**94**	**98**	**99**
模型	召回率/%				
	Normal	DOS	Probe	U2R	R2L
DNN	96	85	65	52	12
SMOTE-DNN	97	85	70	34	14
GAN-DNN	92	81	77	25	5
CGAN-DNN	96	79	75	21	2
ROS-DNN	96	84	71	36	13
ANASYN-DNN	96	88	67	40	6
CCWGAN-DNN	93	89	85	60	20
NBW	94	87	80	50	10
META-VAEGAN-DNN	**95**	**92**	**93**	**99**	**73**

由表6和表7可知，META-VAEGAN-DNN在各类别上的精确率和召回率相较其他方法都有明显提升，尤其在少数类上表现更为突出。结果表明，META-VAEGAN-DNN能够有效解决类不平衡问题，在保持对多数类良好识别能力的同时，显著提升了对少数类样本的检测性能。

6）将本文所提方法与近年来前沿的网络入侵检测算法进行对比实验，结果如表8所示。

由表8可以看出，META-VAEGAN-DNN能在保持高召回率的同时显著提升检测的精确率。在与其他先进的入侵检测算法进行比较时，该模型在准确率、精确率、$F1$分数及召回率上均表现最佳；在检测效率方面，META-VAEGAN-DNN在NSL-KDD数据集上的检测时间为2.8 s，在CIC-IDS2017数据集上为9.2 s，相比DSC-Inception-BiLSTM等复杂模型展现出了较好的实时检测性能。模型性能提升主要得益于META-VAEGAN数据增强模型有效缓解了类不平衡问题，而轻量化的DNN分类器架构确保了检测效率。

表7 不同类均衡方法在CIC-IDS2017数据集各类别上的精确率与召回率比较

Tab. 7 Comparison of precision and recall for each class under different class balancing methods on the CIC-IDS2017 dataset

模型	精确率/%				
	Benign	Botnet	DDoS	DoS	Web
DNN	98	58	87	59	12
SMOTE-DNN	70	84	93	90	55
GAN-DNN	69	85	92	89	58
CGAN-DNN	67	83	90	87	52
ROS-DNN	70	84	92	89	54
ANASYNvDNN	69	83	91	88	48
CCWGAN-DNN	82	88	94	92	75
NBW	78	86	93	90	68
META-VAEGAN-DNN	**90**	**92**	**96**	**94**	**87**
模型	召回率/%				
	Benign	Botnet	DDoS	DoS	Web
DNN	98	62	78	75	28
SMOTE-DNN	95	68	82	79	35
GAN-DNN	91	65	80	77	30
CGAN-DNN	94	60	75	72	25
ROS-DNN	94	65	78	76	32
ANASYN-DNN	95	63	76	74	28
CCWGAN-DNN	92	78	88	85	65
NBW	93	72	84	81	58
META-VAEGAN-DNN	**93**	**88**	**92**	**90**	**84**

表8 META-VAEGAN与其他先进方法的对比

Tab. 8 Comparison of the performance of the class equalization method across categories

数据集	分类算法	准确率/%	精确率/%	$F1$/%	召回率/%	检测时间/s
NSL-KDD	CNN-LightGBM[31]	82	63	64	70	0.6
	CWGAN[32]	79	79	76	78	1.2
	SEMI-GRU[33]	82	85	83	82	2.5
	Anomal-E[34]	85	83	84	85	3.8
	DSC-Inception-BiLSTM[35]	88	86	87	88	4.5
	META-VAEGAN-DNN	**91**	**92**	**90**	**89**	**2.8**
CIC-IDS2017	CNN-LightGBM	80	65	67	72	2.3
	CWGAN	78	76	75	77	5.1
	SEMI-GRU	81	84	82	83	9.6
	Anomal-E	84	82	83	84	14.2
	DSC-Inception-BiLSTM	87	85	86	87	16.8
	META-VAEGAN-DNN	**92**	**89**	**89**	**90**	**9.2**

5 结 论

本文针对入侵检测数据集中存在的类别不平衡问题，提出了一种融合VAE-GAN与元学习的数据增强模型META-VAEGAN，并结合差异化样本处理策略与渐进式质量筛选机制，有效提升

了少数类攻击样本的生成质量。同时，借助残差网络与注意力机制结合的分类模型，对平衡后的数据集进行分类测试，并从准确率、精确率、$F1$分数、AUC和召回率五个指标上进一步验证了所提方法的优势。实验结果表明，META-VAEGAN-DNN在少数攻击类别上，不仅兼顾了高精确率和高召回率，而且提升了检测性能，为解决入侵检测任务中的类不平衡问题提供了一种有效方案，有效提升了对少数类网络攻击识别的效果。本文仅在静态数据集上进行实验，未来可考虑将该方法应用于实时入侵检测系统，并在平衡实时性和计算效率的同时，提升入侵检测性能。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors declare no relevant conflict of interests.

参考文献：

[1] He J X, Wang X D, Song Y F, et al. Network intrusion detection based on conditional Wasserstein variational autoencoder with generative adversarial network and one-dimensional convolutional neural networks [J]. Applied Intelligence, 2023, 53(10): 12416-12436.

[2] Wang S, Wang Z L, Zhou T, et al. THREATRACE: detecting and tracing host-based threats in node level through provenance graph learning[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 3972-3987.

[3] 闫海涛，张之义，朱晓明，等. 基于WOA-XGBoost模型的网络入侵检测[J]. 计算机测量与控制，2023，31(3)：127-133.
Yan Haitao, Zhang Zhiyi, Zhu Xiaoming, et al. Network intrusion detection based on WOA-XGBoost model[J]. Computer Measurement & Control, 2023, 31(3): 127-133. (in Chinese)

[4] Hamid Y, Sugumaran M. A t-SNE based non linear dimension reduction for network intrusion detection [J]. International Journal of Information Technology, 2020, 12(1): 125-134.

[5] 刘广睿，张伟哲，李欣洁. 基于边缘样本的智能网络入侵检测系统数据污染防御方法[J]. 计算机研究与发展，2022，59(10)：2348-2361.
Liu Guangrui, Zhang Weizhe, Li Xinjie. Data contamination defense method for intelligent network intrusion detection systems based on edge examples [J]. Journal of Computer Research and Development, 2022, 59(10): 2348-2361. (in Chinese)

[6] Liu Z H, Liu S Q, Zhang J. An industrial intrusion detection method based on hybrid convolutional neural networks with improved TCN[J]. Computers, Materials & Continua, 2024, 78(1): 411-433.

[7] Imrana Y, Xiang Y, Ali L, et al. CNN-GRU-FF: a double-layer feature fusion-based network intrusion detection system using convolutional neural network and gated recurrent units [J]. Complex & Intelligent Systems, 2024, 10(3): 3353-3370.

[8] Chalé M, Cox B, Weir J, et al. Constrained optimization based adversarial example generation for transfer attacks in network intrusion detection systems [J]. Optimization Letters, 2024, 18(9): 2169-2188.

[9] 冯英引，师智斌. 不平衡数据下基于CNN的网络入侵检测[J]. 中北大学学报（自然科学版），2021，42(4)：318-324.
Feng Yingyin, Shi Zhibin. Network intrusion detection based on CNN on unbalanced data [J]. Journal of North University of China (Natural Science Edition), 2021, 42(4): 318-324. (in Chinese)

[10] 王华忠，田子蕾. 基于改进CGAN算法的工控系统入侵检测方法[J]. 信息网络安全，2023，23(1)：36-43.
Wang Huazhong, Tian Zilei. Intrusion detection method of ICS based on improved CGAN algorithm[J]. Netinfo Security, 2023, 23(1): 36-43. (in Chinese)

[11] Yang Y, Liu X Y, Wang D L, et al. A CE-GAN based approach to address data imbalance in network intrusion detection systems [J]. Scientific Reports, 2025, 15: 7916.

[12] Kang F Y, Feng T, Lin J Q. VAE-GAN-guided cross-class generation: a class imbalance data augmentation method for network intrusion detection[J]. Electronics, 2025, 14(11): 2103.

[13] Yuan L X, Yu S Y, Yang Z B, et al. A data balancing approach based on generative adversarial network [J]. Future Generation Computer Systems, 2023, 141: 768-776.

[14] Kingma D P, Welling M. Auto-encoding variational bayes [PP/OL]. Vl. arXiv(2013-12-20)[2025-09-05]. https://arxiv.org/pdf/1312.6114v1.

[15] Yang A, Lu C, Li J, et al. Application of meta-learning in cyberspace security: a survey [J]. Digital Communications and Networks, 2023, 9(1): 67-78.

[16] Habeck M, Hasenpflug M, Kodgirwar S, et al. Geodesic slice sampling on the sphere[PP/OL]. Vl. arXiv(2023-01-19) [2025-09-05]. https://arxiv.org/abs/

2301.08056.

[17] Cholakoska A, Shushlevska M, Todorov Z, et al. Analysis of machine learning classification techniques for anomaly detection with NSL-KDD data set[M]//Data Science and Intelligent Systems. Cham: Springer International Publishing, 2021: 258-267.

[18] Sharafaldin I, Lashkari A H, Ghorbani A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization [C]//Proceedings of the 4th International Conference on Information Systems Security and Privacy, 2018: 108-116.

[19] Tavallaee M, Bagheri E, Lu W, et al. A detailed analysis of the KDD CUP 99 data set[C]//IEEE Symposium on Computational Intelligence for Security & Defense Applications, 2009: 1-6.

[20] Abdelkhalek A, Mashaly M. Addressing the class imbalance problem in network intrusion detection systems using data resampling and deep learning[J]. The Journal of Supercomputing, 2023, 79(10): 10611-10644.

[21] Shanmugam V, Razavi-Far R, Hallaji E. Addressing class imbalance in intrusion detection: A comprehensive evaluation of machine learning approaches [J]. Electronics, 2024, 14(1): 69.

[22] 余长宏, 许孔豪, 张泽, 等. 基于分割点改进孤立森林的网络入侵检测方法[J]. 计算机工程, 2024, 50(6): 148-156.
Yu Changhong, Xu Konghao, Zhang Ze, et al. Improving network intrusion detection methods in isolated forests based on split points[J]. Computer Engineering, 2024, 50(6): 148-156. (in Chinese)

[23] Díaz-Coto S, Corral-Blanco N O, Martínez-Camblor P. Two-stage receiver operating-characteristic curve estimator for cohort studies[J]. The International Journal of Biostatistics, 2021, 17(1): 117-137.

[24] Huang H Y, Wang Y F, Rudin C, et al. Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization[J]. Communications Biology, 2022, 5(1): 719.

[25] Wongvorachan T, He S, Bulut O. A comparison of undersampling, oversampling, and smote methods for dealing with imbalanced classification in educational data mining[J]. Information, 2023, 14(1): 54.

[26] Mirza M, Osindero S. Conditional generative adversarial nets[J]. Computer Science, 2014: 2672-2680.

[27] Johnson J M, Khoshgoftaar T M. Survey on deep learning with class imbalance[J]. Journal of Big Data, 2019, 6(1): 27.

[28] He H, Bai Y, Garcia E A, et al. ADASYN: adaptive synthetic sampling approach for imbalanced learning [C]//2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008: 1322-1328.

[29] 倪志伟. 基于数据生成与特征降维的网络入侵检测研究[D]. 南京: 南京信息工程大学, 2025.

[30] 黄迎春, 邢秀祺. 融合CNN-GRU和Transformer的网络入侵检测方法[J]. 火力与指挥控制, 2025, 50(6): 21-27.
Huang Yingchun, Xing Xiuqi. Networkintrusion detection method based on fusion of CNN-GRU and transformer [J]. Fire Control and Command Control, 2025, 50(6): 21-27. (in Chinese)

[31] 魏明军, 闫旭文, 纪占林, 等. 基于CNN与LightGBM的入侵检测研究[J]. 郑州大学学报(理学版), 2023, 55(6): 35-40.
Wei Mingjun, Yan Xuwen, Ji Zhanlin, et al. Research on intrusion detection based on CNN and LightGBM[J]. Journal of Zhengzhou University (Natural Science Edition), 2023, 55(6): 35-40. (in Chinese)

[32] 贺佳星, 王晓丹, 宋亚飞, 等. CWGAN-DNN: 一种条件Wasserstein生成对抗网络入侵检测方法[J]. 空军工程大学学报(自然科学版), 2021, 22(5): 67-74.
He Jiaxing, Wang Xiaodan, Song Yafei, et al. CWGAN-DNN: An intrusion detection method based on conditional wasserstein generative adversarial network [J]. Journal of Air Force Engineering University (Natural Science Edition), 2021, 22(5): 67-74. (in Chinese)

[33] Li H T, Wang R M, Dong W Y, et al. Semi-supervised network traffic anomaly detection method based on GRU [J]. Computer Science, 2023, 50(3): 380-390.

[34] Caville E, Lo W W, Layeghy S, et al. Anomal-E: a self-supervised network intrusion detection system based on graph neural networks[J]. Knowledge-Based Systems, 2022, 258: 110030.

[35] 李聪聪, 袁子龙, 滕桂法. 基于深度学习的时空特征融合网络入侵检测模型研究[J]. 信息安全研究, 2025, 11(2): 122-129.
Li Congcong, Yuan Zilong, Teng Guifa. Research on deep learning-based spatio-temporal feature fusion network intrusion detection model[J]. Journal of Information Security Research, 2025, 11(2): 122-129. (in Chinese)