Wuhan University
Journal of Natural Sciences

2026, Vol.31 No.1, 025-034

Article ID 1007-1202(2026)01-0025-10 DOI https://doi.org/10.1051/wujns/2026311025

Cite this article: SU Zhilong, SHEN Zhidong, SUN Hui. An Augmentation Method for Small-Sample Imbalanced Industrial IoT Detection Data[J]. *Wuhan Univ J of Nat Sci*, 2026, 31(1): 25-34.

An Augmentation Method for Small-Sample Imbalanced Industrial IoT Detection Data

□ **SU Zhilong[1], SHEN Zhidong[1†], SUN Hui[2†]**

1. School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, Hubei, China;

2. Zhongnan Hospital, Wuhan University, Wuhan 430071, Hubei, China

Abstract: IoT devices are highly vulnerable to cyberattacks due to their widespread, distributed nature and limited security features. Intrusion detection can counter these threats, but class imbalance between normal and abnormal traffic often degrades model performance. We propose a novel multi-generator adversarial data augmentation method that blends the strengths of TMG-GAN(Tabular Multi-Generator Generative Adversarial Network) and R3GAN (Re-GAN). Our approach uses multiple class-specific generators to create diverse, high-quality synthetic samples, improving training stability and minority-class detection. A dual-branch discriminator-classifier enhances authenticity and class prediction, while feature similarity and decoupling techniques ensure clear class separation. Experiments on TON-IoT and Edge-IIoTset datasets show our method outperforms existing techniques like hybrid sampling, SNGAN (Spectral Normalization GAN), and TMG-GAN, achieving higher detection accuracy and better minority-class recall for imbalanced IoT intrusion detection.

Key words: Internet of Things (IoT); intrusion detection system; generative adversarial networks; class imbalance; data augmentation

CLC number: TP393

0 Introduction

With the rapid advancement of communication and sensor technologies, the Internet of Things (IoT) is increasingly integrated into daily life and industrial applications. However, this growth brings significant security challenges. IoT devices, constrained by limited computational resources and weak defenses, are often deployed in physically accessible environments, making them prime targets for cyberattacks. Network Intrusion Detection Systems (NIDS) are critical in IoT security. Machine learning and deep learning-based anomaly detection methods can learn normal traffic patterns and identify deviations, enhancing detection of unknown threats[1].

A prevalent challenge for machine learning-based NIDS is data imbalance, where normal traffic dominates while attack traffic is scarce and unevenly distributed. Traditional solutions include data-level techniques like oversampling[2], under-sampling[3], SMOTE (Synthetic Minority Over-sampling Technique)[4], and its variants (e. g., SMOTETomek), as well as algorithm-level approaches like cost-sensitive learning[5] and ensemble methods. Recently, deep learning techniques, particularly Genera-

Received date: 2025-06-20

Foundation item: Supported by the Key R&D Projects in Hubei Province (2025BAB018, 2022BAA041) and Wuhan University Comprehensive Undergraduate Education Quality Reform Project

Biography: SU Zhilong, male, Master candidate, research direction: Cyber Security. E-mail: szlwhu@whu.edu.cn

† Corresponding author. E-mail: shenzd@whu.edu.cn; sunh189@whu.edu.cn

tive Adversarial Networks (GANs)[6] such as CTGAN[7] and TAC-GAN[8], have shown promise in generating high-quality synthetic minority class samples to address imbalance. However, current NIDS still face limitations in handling imbalanced traffic distributions, particularly in detecting low-frequency attacks.

This study proposes a multi-generator GAN based on TMG-GAN[9] (Tabular Multi-Generator GAN) and R3GAN[10] (Re-GAN), utilizing cosine similarity of high-dimensional generator features to decouple different class generators, reducing inter-class overlap. Relative pairing loss and R_1/R_2 gradient penalties stabilize training, preventing mode collapse. Additionally, a classifier is introduced to enhance class discrimination, avoiding inter-class confusion in generated samples. Experiments on the TON-IoT and Edge-IIoTset datasets, using F_1-score and Recall as metrics, compare the proposed method against oversampling, SNGAN (Spectral Normalization GAN), and TMG-GAN, demonstrating superior performance.

1 Overview of Imbalanced Learning

1.1 Sample-Based Imbalanced Learning

Sample-based methods address data imbalance by adjusting the proportion of different class samples, primarily through oversampling, which increases minority class samples, and undersampling, which reduces majority class samples. Traditional random oversampling duplicates minority class samples to achieve balance but risks overfitting due to mechanical replication, failing to capture intrinsic data relationships. To address this, Chawla *et al*[4] proposed the SMOTE algorithm, which generates synthetic minority samples by using k-nearest neighbors to define local structures in the feature space and applying linear interpolation, effectively mitigating the scarcity of minority samples, as shown in Fig. 1.

Undersampling achieves balance by selectively removing majority class samples, avoiding the noise introduced by synthetic samples but risking information loss. Random undersampling, while simple, may discard critical samples, disrupting the original data distribution. Density-based methods like Edited Nearest Neighbors (ENN) optimize majority class structure by retaining boundary samples and removing redundant ones, preserving discriminative information. However, both oversampling and undersampling artificially alter the original data distribution, potentially leading to models learning incorrect statistical patterns. This is particularly problematic in scenarios like industrial fault diagnosis, where synthetic or removed samples may introduce diagnostic biases.

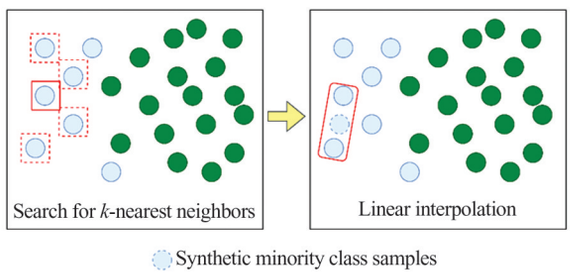

Fig.1 Diagram of the SMOTE algorithm[4]

1.2 Imbalanced Learning Based on Generative Adversarial Networks

GANs offer a novel approach to handling imbalanced data through adversarial learning. A GAN consists of a generator, which learns the latent feature distribution of minority class samples to produce realistic synthetic data, and a discriminator, which distinguishes real samples from generated ones. Through adversarial training, both components improve, enabling the generator to produce high-quality minority samples to balance the dataset. The adversarial process can be described by equation (1):

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

where p_{data} denotes the real data distribution, $D(x)$ represents the discriminator's output for real samples (approaching 1 for strong discrimination), and $z\sim p_z(z)$ is noise sampled from a prior distribution, with $G(z)$ as the generated sample. $V(D,G)$ is the value function in GANs, describing the minimax game between generator G and discriminator D. The discriminator aims for $D\left(G(z)\right)\approx 0$, identifying fake samples, while the generator strives for $D\left(G(z)\right)\approx 1$, making generated samples indistinguishable from real ones.

Compared with traditional oversampling, GANs excel at capturing complex nonlinear feature relationships, producing semantically meaningful samples, particularly in high-dimensional data like images and text. For example, in drone aerial imaging, cloud occlusion often causes data loss. Wei *et al*[11] proposed a two-stage DCGAN-based method for thick cloud region content

generation, incorporating Bag of Words (BoW) algorithms and an affine network to improve DCGAN's structure and loss function, outperforming baseline methods in semantic accuracy and visual quality.

In imbalanced learning, GAN variants enhance performance. Conditional GANs (CGANs) [12] incorporate class labels to generate targeted samples, addressing multi-class imbalance. Wasserstein GANs (WGANs) [13] use Wasserstein distance to stabilize training and improve sample quality. Auxiliary Classifier GANs (ACGANs)[14] extend the discriminator to classify categories, incorporating labels into both generator and discriminator to ensure generated samples align with specific classes and are realistic. This enhances class-specific sample generation, improving model learning in imbalanced scenarios.

However, these technical innovations do not fully mitigate GANs' inherent limitations. Mode collapse remains a critical issue, where the generator may overfit to specific minority class features, producing samples that lack diversity and fail to cover the full feature space of real samples. In financial fraud detection, for instance, if the generator only captures a few typical fraud patterns, the synthetic samples may not reflect the complexity of evolving fraud tactics, limiting the model's ability to detect novel fraud types.

Additionally, training instability is a key constraint for GAN applications[15]. An overly strong discriminator can cause vanishing gradients in the generator, while a weak discriminator leads to poor-quality samples. This delicate balance is highly sensitive to training parameters and data preprocessing. In practice, GAN training for imbalanced data often requires regularization techniques like gradient penalties or spectral normalization, increasing model tuning complexity and computational demands. Inter-class overlap further exacerbates the aforementioned issues[16]. When data from different classes show significant overlap in the feature space, the discriminator struggles to clearly distinguish the boundary between real samples and generated ones, causing the generator to fall into local optima during training.

1.3 Imbalanced Learning Based on Diffusion Models

Diffusion models[17] have emerged as a powerful approach for data generation, offering an alternative to GANs for addressing class imbalance in tabular datasets. TabDDPM[18] is a diffusion-based generative framework specifically designed for tabular data which can handle mixed data types consisting of numerical and categorical features. By iteratively refining noisy data through a series of denoising steps, TabDDPM generates realistic and diverse samples that enhance the representation of underrepresented classes. This approach improves the robustness of intrusion detection systems and other classifiers in imbalanced settings, such as IoT environments.

2 Proposed Approach

2.1 Research Motivation

Compared with traditional oversampling, conventional GANs achieve better results in imbalanced learning but still face challenges like mode collapse and inter-class overlap. This study proposes a novel model integrating R3GAN's adversarial loss mechanism with TMG-GAN's multi-generator architecture, as shown in Fig. 2. The approach employs a dual optimization strategy: a multi-generator structure with classification loss captures fine-grained feature distributions of different attack types, enhancing class specificity; and R3-loss with gradient penalties significantly improves training stability. This model reduces inter-class overlap while enhancing intrusion detection performance. Experiments were conducted on the TON_IoT [19] and Edge-IIoTset [20] datasets, comparing the proposed method against other data augmentation techniques.

2.2 Proposed Methodology

2.2.1 TMG-GAN

TMG-GAN, proposed by Ding *et al*[9], enhances GAN performance for imbalanced data through multiple generators, a classifier, and a feature extractor to separate class-specific generators. Its implementation is as follows:

1) Multi-Generator Structure: Each class uses a dedicated generator to handle diverse data types simultaneously.

2) Classifier Integration: A classifier is added to output class labels alongside authenticity judgments. The feature extractor, shared by the discriminator and classifier, splits into two branches: one for the discriminator (judging sample authenticity) and one for the classifier (determining sample class).

3) Feature Extractor: High-dimensional feature vectors are extracted from the discriminator's final layer. During generator training, cosine similarity between high-dimensional features of real x_k and generated $\widetilde{x_k}$

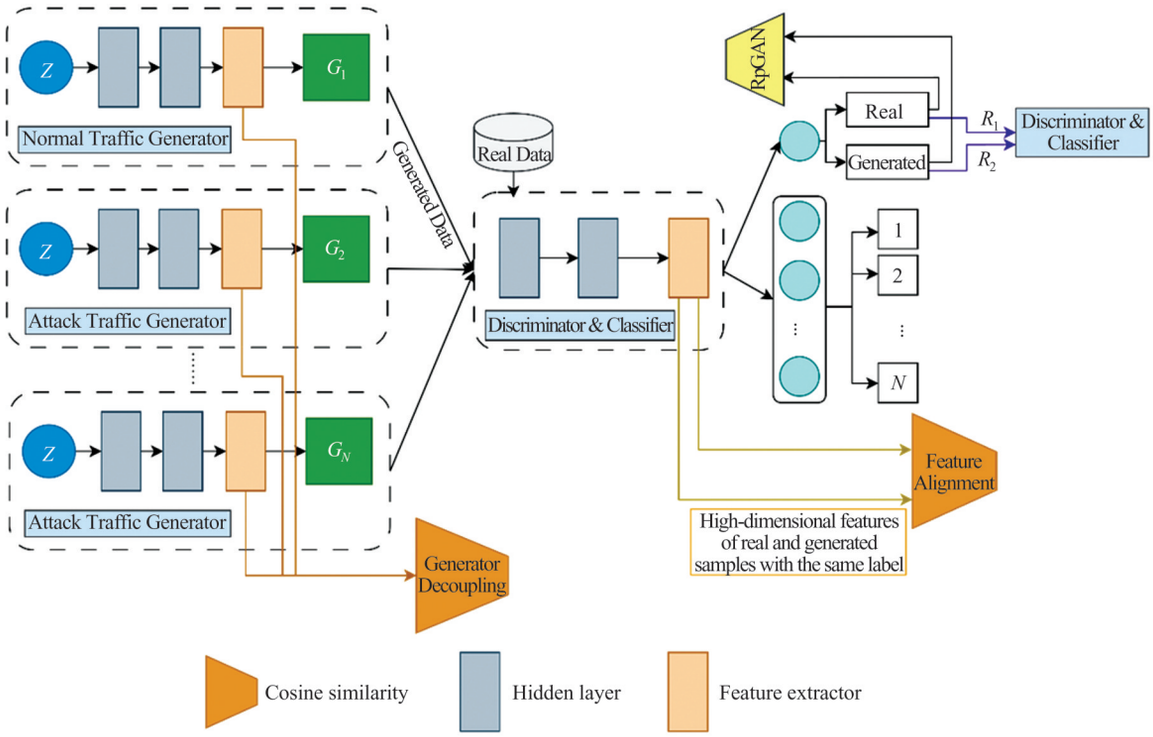

Fig.2 Multi-generator generative adversarial network based on R3-Loss

samples for class k is computed and maximized to align generated samples with real ones, as shown in equation (2):

$$\uparrow O_k\left(\boldsymbol{F}\left(\widetilde{x_k}\right),\boldsymbol{F}\left(x_k\right)\right)=\left|\frac{\boldsymbol{F}\left(\widetilde{x_k}\right)\boldsymbol{F}\left(x_k\right)}{\left\|\boldsymbol{F}\left(\widetilde{x_k}\right)\right\|\left\|\boldsymbol{F}\left(x_k\right)\right\|}\right|,\quad k\in\{1,\cdots,N\},\tag{2}$$

where $\boldsymbol{F}(x)$ and $\boldsymbol{F}\left(\widetilde{x_k}\right)$ are feature vectors, and the goal is to increase cosine similarity for each class $k\in\{1,\cdots,N\}$.

4) Generator Decoupling: To reduce inter-class overlap, cosine similarity between high-dimensional features of samples from different generators ($k\neq j$) is minimized, as shown in equation (3):

$$\downarrow O_k\left(\boldsymbol{F}\left(\widetilde{x_k}\right),\boldsymbol{F}\left(\widetilde{x_j}\right)\right)=\frac{1}{N-1}\sum_j\left|\frac{\boldsymbol{F}\left(\widetilde{x_k}\right)\boldsymbol{F}\left(\widetilde{x_j}\right)}{\left\|\boldsymbol{F}\left(\widetilde{x_k}\right)\right\|\left\|\boldsymbol{F}\left(\widetilde{x_j}\right)\right\|}\right|.\tag{3}$$

This encouraging distinct feature learning and separating generated samples in the feature space.

The loss functions of TMG-GAN are summarized as:

discriminator loss = WGAN loss + classification loss;

generator loss = WGAN loss + classification loss + feature similarity + decoupling loss.

This approach generates high-quality samples with minimal inter-class overlap.

2.2.2 R3GAN

Proposed by Huang *et al*[10], R3GAN integrates the relativistic pairing GAN (RpGAN) loss function with specialized gradient penalties to address mode collapse and non-convergence issues previously mitigated by ad-hoc techniques. RpGAN, introduced by Jolicoeur-Martineau *et al*[21] in 2019, contributes a novel loss function. The general GAN loss is:

$$L_{\theta,\psi}=E_{z\sim p_z}\left[f\left(D_\psi\left(G_\theta(z)\right)\right)\right]+E_{x\sim p_D}\left[f\left(-D_\psi(x)\right)\right].\tag{4}$$

RpGAN uses the softplus function: $f(t)=-\log\left(1+\mathrm{e}^{-t}\right)$, and computes the loss based on the difference between discriminator outputs for real and generated samples:

$$L(\theta,\psi)=E_{z\sim p_z,\ x\sim p_D}\left[f\left(D_\psi\left(G_\theta(z)\right)-D_\psi(x)\right)\right].\tag{5}$$

Unlike traditional GANs, which can only distinguish real from fake samples, RpGAN amplifies the gap between them by feeding their output difference into the loss function. R3GAN further incorporates zero-centered

gradient penalties for real data R_1 and generated data R_2, with ablation studies confirming their optimality, naming the model R3GAN and its loss R3-loss.

2.2.3 Model fusion design

Building on TMG-GAN and R3GAN, this study proposes a fused model, illustrated in Fig. 3, which systematically optimizes architecture and training strategies. The model retains TMG-GAN's multi-generator structure, assigning a generator per class, and uses a shared discriminator-classifier with dual branches. It incorporates R3GAN's training mechanisms as follows:

1) Loss Function Replacement: TMG-GAN's WGAN loss is replaced with RpGAN's loss to better amplify the real-fake sample gap.

2) Gradient Penalties: R_1 and R_2 penalties stabilize training, suppress mode collapse, and enhance decision boundary clarity.

The discriminator loss is:

$$L_{\mathrm{D}}=L_{\mathrm{RpGAN\text{-}D}}+\lambda_{\mathrm{c}}L_{\mathrm{Classify}}+\lambda_1 R_1+\lambda_2 R_2, \tag{6}$$

where $L_{\mathrm{RpGAN\text{-}D}}$ is the RpGAN loss, L_{Classify} is the classification loss, and $\lambda_{\mathrm{c}},\lambda_1,\lambda_2$ are weight hyperparameters. The calculation formula for $L_{\mathrm{RpGAN\text{-}D}}$ is as follows.

$$L_{\mathrm{RpGAN\text{-}D}}=E_{z\sim p_z,x\sim p_{\mathrm{D}}}\left[f\left(D(x)-D\left(G(z)\right)\right)\right], \tag{7}$$

where p_z is the normal distribution, p_{D} is the real data distribution, $D(x)$ is the output of discriminator, $G(z)$ is the generated example of generator.

The generator loss is:

$$L_{\mathrm{G}}=L_{\mathrm{RpGAN\text{-}G}}+\lambda_{\mathrm{c}}L_{\mathrm{Classify\text{-}G}}+\lambda_{\mathrm{f}}L_{\mathrm{feature}}+\lambda_{\mathrm{d}}L_{\mathrm{decouple}}, \tag{8}$$

where $L_{\mathrm{RpGAN\text{-}G}}$, $L_{\mathrm{Classify\text{-}G}}$, L_{feature}, and L_{decouple} represent RpGAN generator loss, classification loss, feature similarity loss, and decoupling loss, respectively, with corresponding weights. The calculation formula for $L_{\mathrm{RpGAN\text{-}G}}$ is as follows:

$$L_{\mathrm{RpGAN\text{-}G}}=E_{z\sim p_z,x\sim p_{\mathrm{D}}}\left[f\left(D\left(G(z)\right)-D(x)\right)\right]. \tag{9}$$

The differences between the proposed model, TMG-GAN, and R3GAN are shown in Table 1.

The overall process, shown in Fig. 3, involves training the generative model, augmenting minority attack data to 50% of the dataset using class-specific generators, and evaluating the augmented data quality with an intrusion detection model.

Table 1 Differences between models

Model design	TMG-GAN	R3GAN	The proposed model
GAN variant	WGAN	RpGAN	RpGAN
Gradient penalty strategy	—	zero-centered gradient penalties (R_1 and R_2)	zero-centered gradient penalties (R_1 and R_2)
Multi-generator structure	Multi-generator structure	—	Multi-generator Ssructure
Generator decoupling	Generator decoupling	—	Generator decoupling
Classifier Branch	Classifier branch	—	Classifier branch

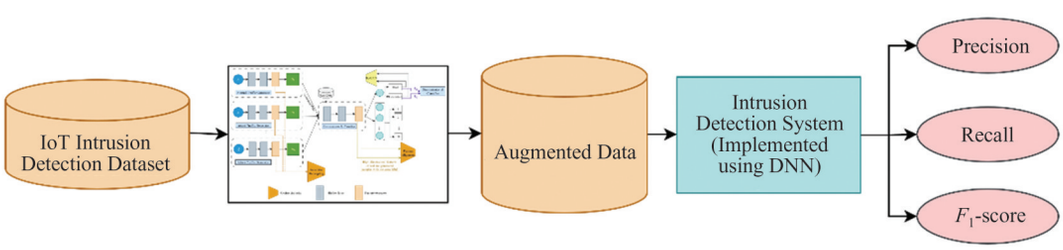

Fig.3 Overall process of data augmentation

3 Experimental Setup

3.1 Datasets

This study uses the TON-IoT and Edge-IIoTset datasets, both recent IoT intrusion detection datasets reflecting the long-tail distribution typical of network traffic: abundant normal traffic and scarce attack traffic.

TON-IoT[19], developed by the University of New South Wales, Australia, supports AI-based security applications in Industry 4.0. It integrates heterogeneous data sources, including IoT sensor telemetry, multi-OS audit logs, and network traffic. Table 2 shows its class distribution.

Table 2 Sample distribution of TON-IoT

Attack category	Number of samples
Normal	119 611
Injection	19 849
Password	15 248
DDoS	11 962
XSS	8 740
DoS	6 448
Scanning	4 493
Backdoor	1 966
MITM	1 039
Ransomware	260
Total	189 616

Edge-IIoTset[20], released in 2022, is a novel dataset for IoT and Industrial IoT (IIoT) security, collected using platforms like ThingsBoard, OPNFV, specifically incorporating industrial protocols like Modbus TCP/IP and Mosquitto MQTT. It includes 15 attack types, with this study focusing on the top 10 by frequency. From the original 1.8 million records, 10% were systematically sampled for efficiency. Table 3 details its class distribution.

Data preprocessing addressed continuous and categorical features. Categorical features were handled using one-hot encoding, transforming each category into a binary vector with a single 1 and remaining 0 s. To mitigate scale and unit differences, data standardization was applied, converting features to a standard normal distribution (mean 0, standard deviation 1), as shown in equation (10):

Table 3 Sample distribution of Edge-IIoTset

Attack category	Number of samples
Normal	136 566
DDoS_UDP	12 166
DDoS_ICMP	6 858
SQL_injection	5 135
Vulnerability_scanner	5 013
DDoS_TCP	4 987
Password	4 825
DDoS_HTTP	4 731
Uploading	3 658
Backdoor	2 401
Total	186 340

$$x' = \frac{x - \bar{x}}{\sigma}, \tag{10}$$

where x is the original feature value, $\bar{x}$ is the feature mean, σ is the standard deviation, and x' is the standardized value. This eliminates adverse effects of scale differences on model training.

3.2 Baseline Setup

Three baseline models were selected for comparison: SMOTETomek[4], SNGAN[22], and TMG-GAN[9] (also serving as an ablation study).

1) SMOTETomek combines SMOTE ' s oversampling, generating synthetic minority samples via k-nearest neighbor interpolation, with undersampling to balance class distribution.

2) SNGAN employs spectral normalization on discriminator weights, constraining the Lipschitz constant to enhance training stability and generate high-quality, diverse minority samples that preserve data structure.

3) TMG-GAN, described in Section 2.2.1, is not reiterated here.

3.3 Experimental Configuration

For fairness and reproducibility, all experiments were conducted under identical hyperparameter settings. The configuration of data augmentation model are as Table 4.

Besides, intrusion detection system (IDS) is im-

Table 4 Configuration of data augmentation model

Hyperparameter	Value
Epochs	2 000
Batch size	256
Learning rate	0.000 2
Optimizer	Adam
Number of discriminator updates per iteration (for GANs)	5
Number of generator updates per iteration (for GANs)	1

plemented as a Deep Neural Network (DNN) because of the remarkable performance of DNNs[23]. For this model, the hyperparameter are listed in Table 5 and the architecture settings are shown in Fig. 4.

Table 5 Configuration of deep neural network

Hyperparameter	Value
Epochs	100
Batch size	256
Learning rate	0.001
Optimizer	Adam

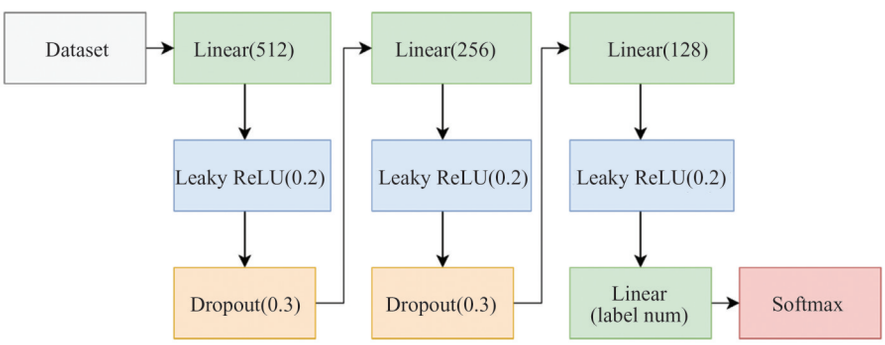

Fig.4 Architecture of deep neural network

4 Experimental Results

Using the augmented datasets, this study conducted multi-class intrusion detection experiments with a deep neural network, comparing the performance of unaugmented data, SMOTETomek, SNGAN, TMG-GAN, and the proposed model across Precision, Recall, and F_1-score metrics.

Bar charts illustrate the performance of the five models on the TON-IoT and Edge-IIoTset datasets, with overall Precision, Recall, and F_1-score comparisons shown in Fig. 5 and Fig. 6 (best values in bold). On the TON-IoT dataset, unaugmented data yielded lower scores: Precision (74.01%), Recall (83.93%), and F_1-score (78.34%), indicating severe performance degradation due to data imbalance. While SMOTETomek, SNGAN, and TMG-GAN improved performance, the proposed model achieved the highest scores: Precision (98.53%), Recall (98.51%), and F_1-score (98.51%), demonstrating its effectiveness in data augmentation and model training.

On the Edge-IIoTset dataset, unaugmented data recorded Precision (87.10%), Recall (89.74%), and F_1-score (87.97%). Although SMOTETomek, SNGAN, and TMG-GAN showed improvements, the proposed model outperformed them with Precision (98.41%), Recall (98.38%), and F_1-score (98.39%), highlighting its strong generalization and ability to address data imbalance.

Table 6 compares TMG-GAN and the proposed model's Precision, Recall, and F_1-score for each attack class. The proposed model achieved overall scores of 98.53%, 98.51%, and 98.51%, surpassing TMG-GAN's 97.80%, 97.65%, and 97.67%, respectively. It excelled in detecting complex attacks, with Precision improving by nearly 5% to 99.96% for password attacks and Recall reaching 99.97% for SQL_injection attacks. For DDoS_UDP and DDoS_ICMP, the proposed model's F_1-score significantly outperformed TMG-GAN. The model proposed in this paper may not outperform TMG-GAN on certain metrics, but due to the use of R3-loss, it achieves higher overall scores compared to TMG-GAN. Its balanced performance across attack types, with less

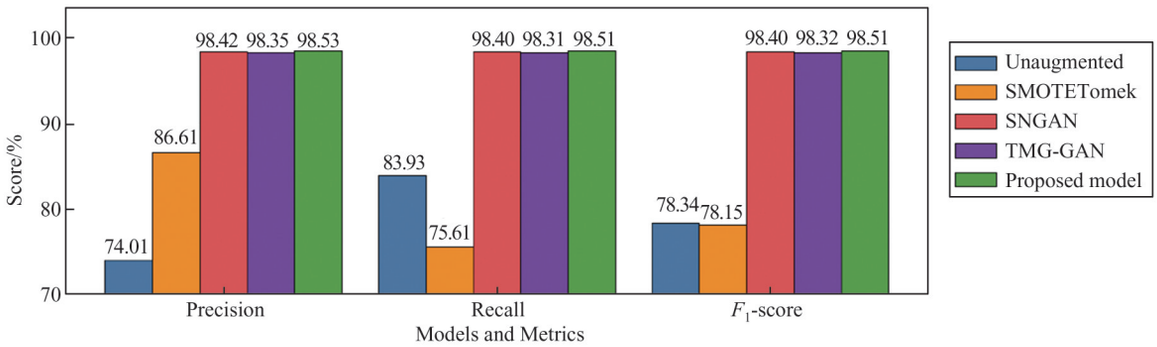

Fig.5 TON-IoT dataset performance comparison of various models

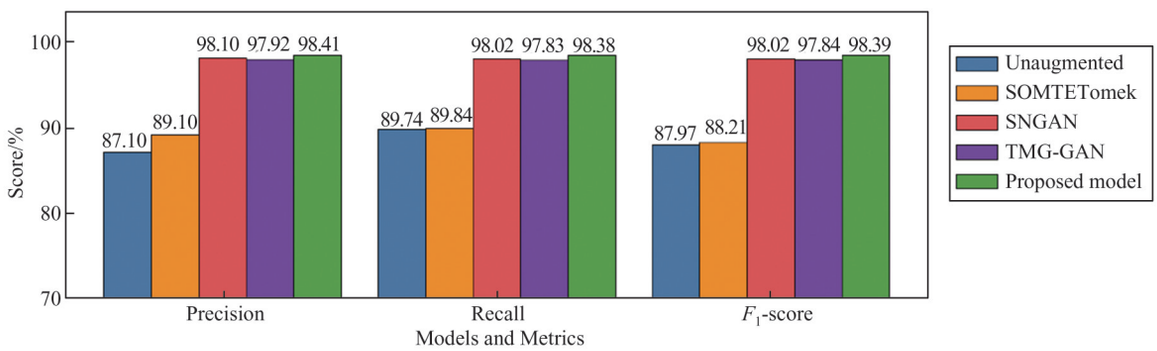

Fig.6 Edge-IIoTset dataset performance comparison of various models

Table 6 Detection performance of Edge-IIoTset in each attack category

%

Attack category	TMG-GAN			Proposed model		
	Precision	Recall	F_1-score	Precision	Recall	F_1-score
Normal	**99.99**	**100**	**100**	96.03	97.98	96.99
DDoS_UDP	92.86	**100**	96.30	**98.98**	**99.37**	**99.17**
DDoS_ICMP	**100**	92.54	96.13	99.02	**97.11**	**98.05**
SQL_injection	96.36	97.56	96.96	**99.84**	**99.97**	**99.91**
Vulnerability_scanner	**99.71**	**99.07**	**99.39**	98.11	98.23	98.17
DDoS_TCP	**100**	**100**	**100**	99.54	97.80	98.66
Password	94.35	96.91	95.62	**99.96**	**98.31**	**99.13**
DDoS_HTTP	**100**	94.85	**97.36**	96.65	**98.06**	97.35
Uploading	97.26	**98.60**	97.92	**99.84**	97.16	**98.48**
Backdoor	**98.68**	98.80	**98.74**	96.08	**99.79**	97.90
Average	97.80	97.65	97.67	**98.41**	**98.38**	**98.39**

Note: The best performance in each metric is highlighted in bold.

fluctuation than TMG-GAN, ensures higher detection consistency and reliability, reducing false negatives and positives in practical cybersecurity scenarios.

Figure 7 illustrates the loss curves of TMG-GAN and the proposed model on the TON-IoT dataset. TMG-GAN's loss, without R3-loss, exhibits significant fluctuations, while the proposed model's loss is notably smoother, indicating improved training stability.

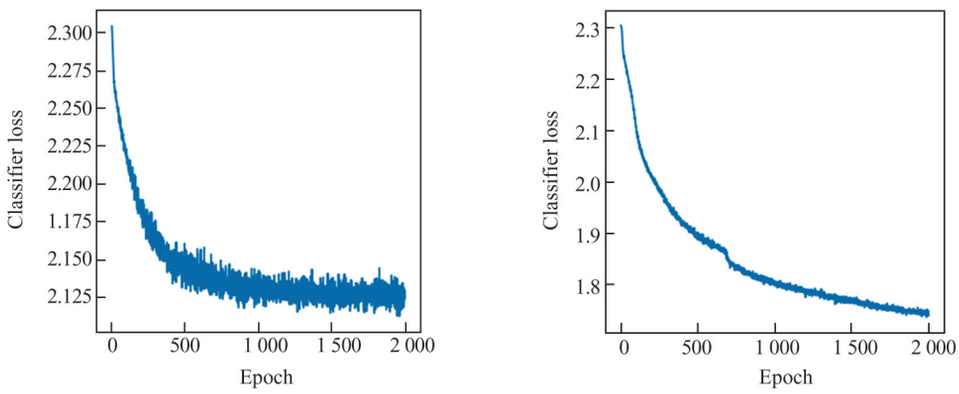

Fig.7 Classifier loss variation of TMG-GAN and the proposed model

5 Conclusion

This study addresses performance degradation in intrusion detection systems due to class imbalance in IoT datasets by introducing a novel data augmentation framework based on a multi-generator adversarial network enhanced with R3-loss. The framework integrates TMG-GAN's class-specific generators for targeted sample generation and inter-class feature decoupling with R3GAN's relativistic adversarial training, incorporating RpGAN loss and dual gradient penalties to ensure stability and prevent mode collapse. It features a dual-branch discriminator-classifier and cosine-similarity-driven feature optimization to produce high-quality, diverse, and class-consistent synthetic samples, enhancing minority class representation and sharpening class boundaries. Evaluations on TON-IoT and Edge-IIoTset datasets show superior performance over baselines like SMOTE-Tomek, SNGAN, and TMG-GAN, particularly in detecting complex attacks.

Future work could explore lightweight model architectures to reduce computational demands, dynamic training strategies for adaptive learning in varying IoT scenarios, and cross-domain applicability to extend the framework beyond industrial IoT, such as to consumer or edge computing environments, thereby improving real-world deployment and efficiency.

References

[1] Jiang J C, Ma H T, Ren D E, *et al*. A survey of intrusion detection research on network security[J]. *Journal of Software*, 2000, **11**(11): 1460-1466(Ch).

[2] Tao X M, Zheng Y J, Chen W, *et al*. SVDD-based weighted oversampling technique for imbalanced and overlapped dataset learning[J]. *Information Sciences*, 2022, **588**: 13-51.

[3] Zhang R, Zhang Z Q, Wang D. RFCL: A new undersampling method of reducing the degree of imbalance and overlap[J]. *Pattern Analysis and Applications*, 2021, **24**(2): 641-654.

[4] Chawla N V, Bowyer K W, Hall L O, *et al*. SMOTE: Synthetic minority over-sampling technique[J]. *Journal of Artificial Intelligence Research*, 2002, **16**: 321-357.

[5] Wang Z, Chu X, Li D D, *et al*. Cost-sensitive matrixized classification learning with information entropy[J]. *Applied Soft Computing*, 2022, **116**: 108266.

[6] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets[C]//*Neural Information Processing Systems*, 2014: 27.

[7] Xu L, Skoularidou M, Cuesta-Infante A, *et al*. Modeling tabular data using conditional GAN[C]//*Proceedings of the 33rd International Conference on Neural Information Processing Systems*. 2019: 7335-7345.

[8] Ding H W, Chen L Y, Dong L, *et al*. Imbalanced data classification: A KNN and generative adversarial networks-based hybrid approach for intrusion detection[J]. *Future Generation Computer Systems*, 2022, **131**: 240-254.

[9] Ding H W, Sun Y, Huang N N, *et al*. TMG-GAN: Generative adversarial networks-based imbalanced learning for network intrusion detection[J]. *IEEE Transactions on Information Forensics and Security*, 2023, **19**: 1156-1167.

[10] Huang N, Gokaslan A, Kuleshov V, *et al*. The GAN is dead; long live the GAN! A modern GAN baseline[J]. *Advances in Neural Information Processing Systems*, 2024, **37**: 44177-44215.

[11] Wei Z, Li C L, Shen Y A, *et al*. Thick cloud region content generation of UAV image based on two-stage model[J]. *Chi-*

nese Journal of Computers, 2021, **44**(11): 2233-2247(Ch).

[12] Mirza M, Osindero S. Conditional generative adversarial nets[EB/OL]. [2014-12-28]. https://arXiv preprint arXiv: 1411.1784.

[13] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks[C]//*International Conference on Machine Learning.* New York: ACM, 2017, **70**: 214-223.

[14] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs[C] //34*th International Conference on Machine Learning.* Sydney: ICML, 2017, **6**: 2642-2651.

[15] Chen F J, Zhu F, Wu Q X, *et al*. A survey of generative adversarial networks and their applications in image generation [J]. *Chinese Journal of Computers*, 2021, **44**(2): 347-369 (Ch).

[16] Vuttipittayamongkol P, Elyan E, Petrovski A. On the class overlap problem in imbalanced data classification[J]. *Knowledge-Based Systems*, 2021, **212**: 106631.

[17] 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.

[18] Kotelnikov A, Baranchuk D, Rubachev I, *et al*. TabDDPM: Modelling tabular data with diffusion models[C]//*Proceedings of the* 40*th International Conference on Machine Learning*. Hawaii: PMLR, 2023: 17564-17579.

[19] Moustafa N. A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets[J]. *Sustainable Cities and Society*, 2021, **72**: 102994.

[20] Ferrag M A, Friha O, Hamouda D, *et al*. Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning [J]. *IEEE Access*, 2022, **10**: 40281-40306.

[21] Jolicoeur-Martineau A. The relativistic discriminator: A key element missing from standard GAN[C]//*International Conference on Learning Representations*. New Orleans: OpenReview.net, 2019: 8622-8647.

[22] Miyato T, Kataoka T, Koyama M, *et al*. Spectral normalization for generative adversarial networks[C]//*International Conference on Learning Representations*. Vancouver: OpenReview.net, 2018: 376-401.

[23] Xu K Y Z, Lu Y J, Wang Z Y, *et al*. A survey of adversarial examples in computer vision: Attack, defense, and beyond [J]. *Wuhan Univ J of Nat Sci*, 2025, **30**(1): 1-20.

小样本非平衡工业物联网检测数据扩增方法

苏之龙[1]， 沈志东[1†]， 孙慧[2†]

1.武汉大学 国家网络安全学院，湖北 武汉 430072

2.武汉大学 中南医院，湖北 武汉 430071

摘要：物联网设备由于其广泛分布和安全机制有限，极易成为网络攻击的目标。入侵检测能够缓解这些威胁，但正常与异常流量之间的类别不平衡常常导致模型性能下降。为此，本文结合TMG-GAN和R3GAN的优势，提出了一种新颖的多生成器对抗数据增强方法。该方法采用多个类别特定的生成器来生成多样且高质量的合成样本，从而提升训练稳定性和少数类检测能力。双分支判别器-分类器结构同时增强了样本真实性判别与类别预测，而特征相似性和解耦机制则确保了清晰的类别分离。在TON-IoT和Edge-IIoTset数据集上的实验结果表明，本文的方法优于现有的混合采样、SNGAN和TMG-GAN，在检测精度和少数类召回率方面均取得了更好的表现，有效应对了物联网入侵检测中的类别不平衡问题。

关键词：物联网；入侵检测系统；生成对抗网络；类别不平衡；数据增强

□