ISSN 2616-5775 Vol. 7, Issue 4: 36-44, DOI: 10.25236/AJCIS.2024.070406

Industrial Internet Intrusion Detection Method Based on VAE-WGAN-GP Data Enhancement

Yahui Wang[1,a,*], Zhiyong Zhang[1]

[1]*Information Engineering College, Henan University of Science and Technology, Luoyang, China*
[a]*wangyahuihappy@163.com*
[*]*Corresponding author*

Abstract: *Deep learning has played a significant role in intrusion detection. However, deep learning-based intrusion detection methods require a large amount of annotated data for model training. In the real world, the types of intrusion data that are of concern often belong to minority classes that lack labels. This imbalance creates an imbalanced dataset for intrusion detection, where normal data significantly outweighs attack data. Class imbalance can lead to biased decision boundaries, resulting in increased classification errors for attack data. In the face of imbalanced data, we propose a data augmentation model based on VAE-WGAN-GP. VAE-WGAN-GP combines variational autoencoders (VAE) and Wasserstein generative adversarial networks (WGAN) with gradient penalty (GP), creating a deep learning generative model. We augment the minority class data using this model to balance the dataset. Finally, we demonstrate significant improvements in multi-class intrusion detection using multiple classifiers by applying our data augmentation model to a traditional internet dataset and an industrial control system network dataset.*

Keywords: *Intrusion detection system; Variational autoencoder; Industrial Internet; Generative adversarial network; Data augmentation*

1. Introduction

The rapid development of the industrial Internet has brought great convenience to the industry by enabling real-time data collection, analysis, and remote control to improve production efficiency, reduce costs, optimize resource utilization, and enhance operational management [1]. It has found widespread applications in industries such as manufacturing, energy, transportation, and agriculture. However, the convergence of information technology and operational technology has blurred the boundaries between the security of industrial manufacturing and the external Internet [2]. This means that once the industrial Internet is attacked, the consequences can be highly destructive. The 2015 Ukraine power grid attack is a typical example where attackers used malicious software called "Black Energy" to successfully infiltrate the computer systems of power companies and disrupt infrastructure, leading to widespread power outages. This case demonstrates the serious threat of network attacks to infrastructure [3] and the unreliability of traditional intrusion detection systems. Therefore, it is necessary to develop and apply new methods and technologies to protect the security of the industrial Internet.

The widespread application of deep learning has had a significant impact on areas such as speech and image recognition and has introduced new concepts into fields like intrusion detection [4-5]. By combining deep learning with intrusion detection, it is possible to improve detection accuracy and reduce the risks of false positives and false negatives, simplifying the problem of intrusion detection [6]. In intrusion detection, anomaly-based detection can effectively identify attack behaviors, and deep learning can detect and recognize potential network intrusions by learning the pattern differences between normal network traffic and abnormal intrusion behavior [7]. By using deep learning algorithms, more intelligent and accurate intrusion detection systems can be built, providing strong support for enhancing the security of the industrial Internet.

However, deep learning models require a large amount of labeled data for training, and in reality, there is a lack of large-scale labeled data for new types of attacks. The data used for model training often suffers from extreme class imbalance, where there is an imbalance in class distribution between a large number of normal class samples and a small number of abnormal class samples. Therefore, deep learning-based intrusion detection inevitably faces the challenges brought by imbalanced learning [8-9]. To effectively address this problem, it is necessary to augment the minority class samples that lack labels.

ISSN 2616-5775 Vol. 7, Issue 4: 36-44, DOI: 10.25236/AJCIS.2024.070406

Currently, data augmentation based on Variational Autoencoders (VAE) [10] and Generative Adversarial Networks (GAN) [11] has been widely applied in the field of deep learning. These methods can generate data by learning the distribution of real data. Based on VAE-Wasserstein GAN with Gradient Penalty (VAE-WGAN-GP), we propose a data balancing model that can augment a small number of labeled samples to improve the recognition of the minority class in intrusion detection. The main contributions of this paper are as follows: (1) We propose a data augmentation model based on VAE-WGAN-GP, which can intelligently learn real information and generate synthetic but credible training data to balance the training dataset and improve the performance of deep learning-based intrusion detection models. (2) We test the intrusion detection system based on VAE-WGAN-GP data augmentation using datasets from both traditional Internet and industrial control networks.

The remaining sections of this paper are organized as follows: Section 2 introduces related work; Section 3 presents the detailed structure and working mechanism of the prototype network model constructed in this paper; Section 4 describes the datasets and experimental setup used in the study and analyzes the experimental results. Finally, Section 5 summarizes the paper.

2. Related Work

2.1. Deep Learning-Based Intrusion Detection

Wang et al. [12] proposed a novel intrusion detection model based on a one-dimensional convolutional autoencoder (1DCAE) and Support Vector Data Description (SVDD) to identify attack behaviors in Industrial Control Systems (ICS). The deep 1DCAE filters redundant features to obtain a low-dimensional representation of the raw data while preserving key features for intrusion detection. The model then utilizes the traditional anomaly detection classifier SVDD to detect attack behaviors. Khan et al. [13] introduced an IDS model based on LSTM autoencoders to identify anomalous behaviors in Industrial Internet of Things (IIoT) within ICS. This model combines the advantages of network pattern statistical properties with the powerful training architecture of DL-based autoencoders, providing an operational learning and detection system for IIoT and ICS networks. Krithivasan et al. [14] proposed an anomaly detection method for ICS networks based on a hypergraph-centered PCA technique combined with convolutional neural networks (CNN). This method reduces data dimensionality using PCA and identifies anomalies using hypergraph-centered CNN. Safari et al. [15] presented an industrial intrusion detection method based on a four-layer fully connected neural network and established a behavior model. During each data operation, online data is compared to the real data of the behavior prediction model to detect anomalies in the operation of rotating machinery. Sun et al. [16] proposed a model based on Bayesian networks (BN) and temporal automata (TA) theory. This model utilizes probabilistic temporal automata to simulate the regular behavior of time series and establishes the dependencies between sensors and actuators using Bayesian networks.

However, the excellent performance of these detection models relies on a large number of labeled samples, while in reality, we need to focus on new attacks that often lack labels. Therefore, it is necessary to design a method to improve the detection rate of minority class intrusions.

2.2. Intrusion Detection Based on Imbalanced Learning

In industrial Internet applications based on deep learning, the given dataset often exhibits significant differences among different types of samples, resulting in data imbalance. This means that the data for certain classes in the training dataset is much smaller or much larger than other types of data [17]. Therefore, it is necessary to find a method to balance the data and reduce its impact on the accuracy of the detection model. To address the data imbalance problem more effectively, some researchers have focused on the data level and employed data augmentation techniques to increase the samples for certain data types. Zhou et al. [18] proposed a Distribution Bias-aware Collaborative Generative Adversarial Network (DB-CGAN) model for imbalanced deep learning in industrial IoT. By introducing complementary classifiers into the basic GAN model, they constructed an integrated data augmentation framework to effectively increase the number of samples in the minority class. Mbow et al. [19] proposed a hybrid method to handle the imbalance problem, which combines Synthetic Minority Over-sampling Technique (SMOTE) and under sampling using Tomek links to reduce noise. Fu et al. [20] addressed the data imbalance issue by applying the Adaptive Synthetic Sampling (ADASYN) method to augment the minority class samples, achieving a relatively balanced state of the samples and improving the detector's performance. Liu et al. [21] introduced a Difficult Set Sampling Technique (DSSTE) algorithm to tackle

ISSN 2616-5775 Vol. 7, Issue 4: 36-44, DOI: 10.25236/AJCIS.2024.070406

class imbalance by amplifying and reducing the continuous attributes of minority samples and synthesizing new samples to increase the quantity of minority samples, thus achieving relative balance in the data samples. Huang et al. [22] proposed an Imbalanced Generative Adversarial Network (Imbalanced GAN) to address class imbalance. They introduced imbalance data filters and convolutional layers in the GAN to generate representative instances for minority classes. Ding et al. [23] presented a tabular data sampling method to solve the imbalanced learning problem. They utilized a table-augmented classifier generative adversarial network model (TACGAN) to oversample attack samples and achieve class balance in the data samples.

Although the aforementioned studies have achieved promising results in intrusion detection with imbalanced learning, most of the research focuses on traditional internet datasets. Therefore, it is necessary to find a practical and effective method to obtain sufficient training data specifically for industrial Internet of Things, which remains a significant challenge in IIoT applications.

3. Method Design

3.1. Data processing

There is a large amount of discrete and continuous data present in both real-world Internet datasets and industrial control system datasets. These data can impact the convergence speed and detection accuracy of the models, necessitating preprocessing of the data before model training. The goal is to achieve homogeneity and eliminate heterogeneous data, making the model more easily learnable and convergent.

Data normalization: Due to significant distribution differences in feature attribute data within both datasets, it is necessary to perform data normalization as a preliminary step. In this study, the Z-score normalization method is employed to address the imbalance in feature data distribution. By normalizing the data, it is uniformly scaled to the range of [0, 1]. This bounded range is desirable in many deep learning algorithms as it helps prevent certain features from dominating others due to differences in their original scales. The calculation formula is shown in Equation (1), where μ represents the mean of the overall population data samples, and σ represents the standard deviation of the overall population data samples.

$$x_{normalization} = \frac{x-\mu}{\sigma} \tag{1}$$

One-hot encoding: One-hot encoding is a popular technique used to encode discrete datasets. It effectively converts categorical variables into a binary format that machine learning models can handle. In our proposed model, we utilize one-hot encoding to handle the discrete features in the dataset. By employing one-hot encoding, each discrete feature value is transformed into a separate binary column. This encoding scheme enables the model to easily differentiate between different categories within the same feature, which is advantageous for the proposed model. It allows the model to capture the relationships and patterns present in the data more effectively. Furthermore, one-hot encoding simplifies the representation of discrete features, contributing to improved training efficiency. By converting categorical variables into binary columns, the model can handle the data more efficiently, reducing computational complexity and speeding up the training process.

3.2. VAE-WGAN-GP model

3.2.1. VAE Module

Variational Autoencoder (VAE) and Autoencoder (AE) share similarities in their structure, consisting of an encoder and a decoder. However, they have significant differences in handling latent representations. An autoencoder maps the input data to a deterministic latent representation, while the encoder part of a variational autoencoder maps the input data to a distribution in the latent space, typically represented by mean and variance parameters. This makes the latent representation a probability distribution, introducing latent variability to the model. During the training process, the variational autoencoder learns by minimizing a combination loss function of reconstruction error and maximizing the similarity between the encoder output and the prior latent distribution. This probabilistic latent representation enables the variational autoencoder to exhibit greater flexibility and diversity in generating new samples, providing a powerful framework for generative models. The architecture of the VAE is illustrated in Figure 1.

ISSN 2616-5775 Vol. 7, Issue 4: 36-44, DOI: 10.25236/AJCIS.2024.070406

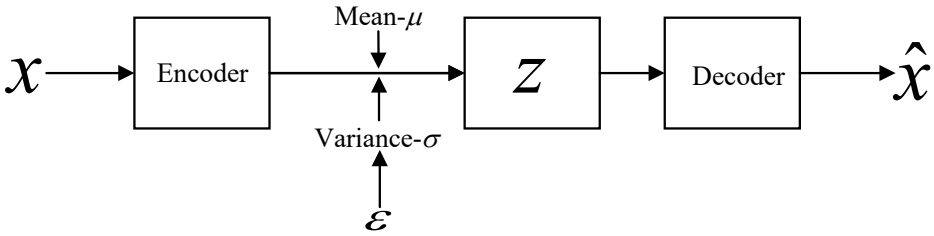

Figure 1: VAE structure diagram.

In the VAE module, it consists of an encoder and a decoder, serving as the generator in the generative adversarial network. In Figure 1, x represents the input raw data, and the goal of VAE is to learn the latent distribution $Q(z|x)$ of the data. To achieve this goal, VAE assumes that the distribution of its latent variable z follows a Gaussian distribution. The process of reconstructing generated data x_{rec} involves sampling the variable z from the latent distribution $P(z)$, where the mean μ and variance σ of the latent distribution $P(z)$ are obtained from the encoder, while noise ε is introduced in the variance σ. Then, the decoder decodes the latent variable z to generate the distribution $P(x|z)$ of the original data x. Finally, the Kullback-Leibler (KL) divergence is used as a regularization term to constrain the distance between $Q(z|x)$ and $P(z)$, and the optimization objective is to maximize the evidence lower bound (ELBO), as shown in Equation (2):

$$\mathrm{ELBO} = \mathrm{E}[\log \mathrm{P}(\mathrm{x}|\mathrm{z})] - \mathrm{KL}(\mathrm{Q}(\mathrm{z}|\mathrm{x}) \parallel \mathrm{P}(\mathrm{z})) \tag{2}$$

Here, $E[log\ P\ (x|z)]$ represents the reconstruction error of input data x given the latent variable , and $KL(Q(z|x) \parallel P(z))$ represents the KL divergence of the latent variable. Since the KL divergence is non-negative, maximizing the ELBO is equivalent to minimizing the KL divergence and maximizing the reconstruction error.

To compute the ELBO, we need to estimate the posterior distribution $Q(z|x)$ of the latent variable and the generative distribution $P(x|z)$ of the original data. Specifically, the input data x is mapped through the encoder network to the mean μ and variance σ of the latent variable space. Then, the latent variable z is resampled from the latent variable space, and the decoder decodes the latent variable into the distribution of the original data. The process is shown in Equation (3):

$$\mathrm{z} \sim \mathrm{Q}(\mathrm{z}|\mathrm{x}) = \mathrm{N}(\mu, \sigma^2) \sim \mathrm{P}(\mathrm{x}|\mathrm{z}) \tag{3}$$

$$\mathrm{N}(\mu, \sigma^2) = \sigma * \varepsilon + \mu \tag{4}$$

The mean μ and variance σ are the outputs of the encoder. To ensure continuity, we apply a parameterized correction as shown in Equation (4), where ε represents the noise.

3.2.2. WGAN-GP Module

GAN is a deep learning model composed of a Generator and a Discriminator. The core idea of GAN is to train the Generator through an adversarial process to generate realistic data samples.

The Generator utilizes a random vector sampled from the latent variable z to generate synthetic samples x_{rec} . Subsequently, the generated samples, along with real data, are inputted to the Discriminator to distinguish between real and fake data. The optimization objective of GAN is to minimize the probability that the Discriminator cannot differentiate between generated and real data. The corresponding loss function is shown in Equation (5):

$$\min_{\mathrm{G}} \max_{\mathrm{D}} \mathrm{V}(\mathrm{D}, \mathrm{G}) = \mathrm{E}_{\mathrm{x}\sim \mathrm{P}_{\mathrm{data}}(\mathrm{x})}[\log \mathrm{D}(\mathrm{x})] + \mathrm{E}_{\mathrm{z}\tilde{\mathrm{p}}_{\mathrm{z}}(\mathrm{z})}[\log(1 - \mathrm{D}(\mathrm{G}(\mathrm{z})))] \tag{5}$$

Here, $E_{x\sim P_{data}(x)}[log\ D\ (x)]$ represents the logarithm probability of the sample , which is sampled from the real data distribution $P_{data(x)}$, being classified as real by the Discriminator. The objective of the Discriminator is to maximize this value, correctly identifying real data as real. On the other hand, $E_{z\tilde{p}_z(z)}[log(1 - D(G(z)))]$ represents the logarithm probability of the sample (z) , generated by the Generator, being classified as fake by the Discriminator. The objective of the Generator is to minimize this value, making the generated samples more likely to pass the Discriminator's detection, i.e., the probability of the generated samples being classified as real data should be closer to 1.

Despite the remarkable performance of GAN in many applications, its training process often encounters issues such as instability, mode collapse, and vanishing gradients. To overcome these challenges, the Wasserstein distance has been proposed as an improvement strategy for GAN.

ISSN 2616-5775 Vol. 7, Issue 4: 36-44, DOI: 10.25236/AJCIS.2024.070406

The introduction of the Wasserstein distance makes it easier to handle gradient-related issues during the training of generative models, particularly avoiding the gradient explosion and vanishing problems caused by the KL divergence in traditional GANs. WGAN achieves this by redefining the loss functions of the Generator and the Discriminator, approximating the computation of the Wasserstein distance, and thereby improving the training stability and generation quality of GAN. The distance formula is shown in Equation (6):

$$W(p,q) = \inf_{\gamma \in \prod(p,q)} E_{(x,y)\sim\gamma} [\| x - y \|] \tag{6}$$

Here, $\prod(p,q)$ denotes the set of all joint distributions of γ, with marginal distributions represented by p and q. The variable (x,y) represents pairs of samples from γ, and $\| x-y \|$ denotes the distance between sample pairs. Subsequently, on the foundation of WGAN, the GP is introduced to address the issues of gradient vanishing and mode collapse within WGAN, thereby further enhancing the performance and stability of the model.

Finally, the corresponding loss function for WGAN-GP is represented as Equation (7):

$$L_{WGAN-GP} = E_{x\sim P_{data}(x)}[D(x)] - E_{z\sim \tilde{p}_z(z)}[D(G(z))] + \lambda E_{\hat{x}\sim p_{\hat{x}}}[(\| \nabla_{\hat{x}} D(\hat{x}) \|_2 - 1)^2] \tag{7}$$

Where $E_{x\sim P_{data}(x)}[D(x)] - E_{z\sim \tilde{p}_z(z)}[D(G(z))]$ represents the Wasserstein distance, $\lambda E_{\hat{x}\sim p_{\hat{x}}}[(\| \nabla_{\hat{x}} D(\hat{x}) \|_2 - 1)^2]$ is the gradient penalty term employed to enforce gradient continuity, $\hat{x}$ denotes random interpolations between samples from real and generated data, and λ is the weight parameter for the gradient penalty.

3.2.3. VAE-WGAN-GP

Finally, we employ the VAE network as the generator model in the GAN network, fully leveraging the outstanding feature extraction capabilities of the VAE to capture the latent space distribution of complex input data. By sampling from the latent space distribution, we reconstruct the data and optimize the reconstruction error and KL divergence to ensure that the generator learns the features of the true data distribution. Additionally, we adopt the optimization training strategy of WGAN-GP to enhance the stability of the training process, improve the quality of generated samples, and strengthen the model's generalization capability. The overall architecture of the model is illustrated in Figure 2.

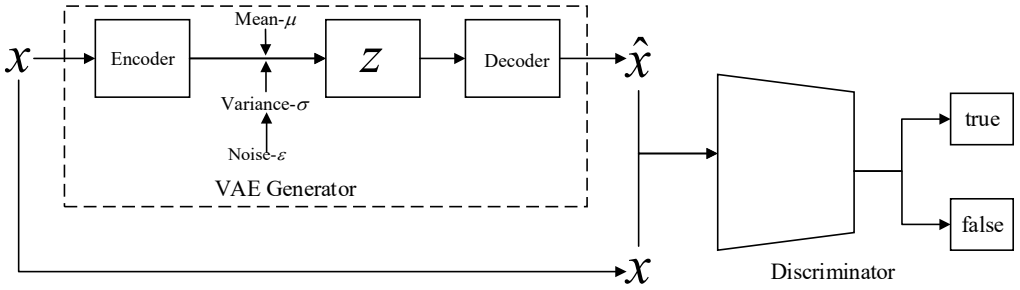

Figure 2: VAE-WGAN-GP Architecture Diagram.

The encoder of the VAE generator maps input data x to a latent vector z, while the decoder maps the latent vector z back to the input space, generating data $\hat{x}$ similar to the input data x. The discriminator's task is to assess whether the generated data is close to real data. The overall loss function of the model is shown in Equation (8):

$$L_{VAE-WGAN-GP} = L_{VAE} + L_{WGAN-GP} \tag{8}$$

Where $L_{VAE} = -E_{Q_\varphi(z|x)}[log\, P_\theta(x|z)] + D_{KL}(Q_\varphi(z|x) \| P_\theta(z))$, $L_{WGAN-GP}$ as shown in Equation (7).

3.3. Intrusion Detection Process based on VAE-WGAN-GP

In the intrusion detection based on data augmentation, the process is illustrated in Figure 3. We initially preprocess the dataset, including normalization and one-hot encoding, to ensure the accuracy and availability of the data. Subsequently, the preprocessed data is divided into training and testing sets. The minority class in the training set is augmented using VAE-WGAN-GP. The classifier is then trained on the balanced dataset, and the performance of the classifier is evaluated using the testing set.

ISSN 2616-5775 Vol. 7, Issue 4: 36-44, DOI: 10.25236/AJCIS.2024.070406

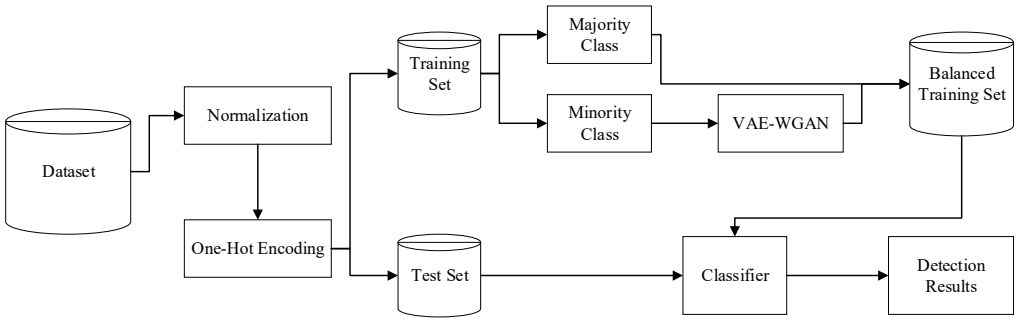

Figure 3: Intrusion Detection Framework Based on Data Augmentation.

4. Experiment and Analysis

In this section, we have designed and conducted experiments to evaluate the effectiveness of the proposed model in comparison to similar works. This is done to demonstrate the efficacy of the model proposed in this paper.

4.1. Dataset and Experimental Design

To assess the effectiveness of the intrusion detection model we constructed in the field of industrial Internet, we opted for two representative public datasets: CIC-IDS 2017 and GasPipeline. These datasets are widely utilized in the intrusion detection domain. The CIC-IDS 2017 dataset comprises extensive records of network traffic, while the GasPipeline dataset captures real industrial network data.

In both of these datasets, individual classes represent extreme minority classes. The purpose of the experiment is to verify whether our data augmentation system can effectively generate conditionally valid sample data in real-world scenarios, thereby improving the recognition accuracy of minority classes. To achieve this, we augment the minority class samples using the model, then train and test multiple classification models on the enhanced balanced dataset. Finally, we assess the improvement in intrusion detection performance for minority classes after the augmentation.

Table 1: Statistical Data for CIC-IDS 2017.

Type	Type Number	Sample Count	Sample Ratio
BENIGN	0	2271320	80.3189%
DoS (Hulk,GoldenEye,slowloris,Slowhttptest)	4	251712	8.9011%
PortScan	8	158804	5.6157%
DDoS	3	128025	4.5272%
FTP-Patator	5	7935	0.2806%
SSH-Patator	9	5897	0.2085%
BotNet	1	1956	0.0692%
Web Attack Brute Force	2	1507	0.0533%
Web Attack XSS	11	652	0.0231%
Infiltration	7	36	0.0013%
Web Attack Sql Injection	10	21	0.0007%
Heartbleed	6	11	0.0004%

In the specific experimental design, we initially preprocess the dataset, including data cleaning and feature extraction, to ensure the accuracy and availability of the data. Subsequently, we divide the preprocessed data into training and testing sets and augment the minority class in the training set using multiple data augmentation models. We then train multiple classifiers on the augmented dataset and evaluate the intrusion detection model before and after data augmentation using the testing set. We compare accuracy, recall, and other metrics of the intrusion detection model after enhancing the data with different augmentation models.

The details of the two datasets are shown in Tables 1 and 2.

The CIC-IDS 2017 dataset is an intrusion detection dataset generated by the Canadian Institute for Cybersecurity. We merged types with the same attack categories, resulting in a dataset encompassing 11 attack methods. The statistical details are outlined in Table 1. This is a large-scale and imbalanced dataset

ISSN 2616-5775 Vol. 7, Issue 4: 36-44, DOI: 10.25236/AJCIS.2024.070406

comprising 78 features and 11 attack types.

The GasPipeline dataset comprises real data collected from the SCADA system of the natural gas pipeline testing platform at the University of Mississippi. This dataset consists of 26 features and 7 attack types. The specific statistical details are presented in Table 2.

Table 2: Statistical Data for GasPipeline.

Type	Type Number	Sample Count	Sample Ratio
Normal	6	61156	63.0351%
CMRI	0	15466	15.9412%
MPCI	3	7637	7.8717%
Recon	7	6805	7.0141%
NMRI	5	2763	2.8479%
DOS	1	1837	1.8934%
MSCI	4	782	0.8060%
MFCI	2	573	0.5906%

4.2. Experimental Environment and Evaluation Metrics

Experimental Environment Hardware: Windows 11 operating system, Intel I5 10400 CPU, 16GB RAM. Software: Python 3.9, including libraries such as pytorch, numpy, pandas, Sklearn, etc. To effectively evaluate the validity of our model, we employ the following evaluation metrics:

Precision, calculated as shown in Equation (9):

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{9}$$

Recall, calculated as follows:

$$\text{recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{10}$$

F1-Score, calculated as follows:

$$\text{F1} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{11}$$

Where TP represents the number of correctly classified positive samples, TN is the number of correctly classified negative samples, FN is the number of incorrectly classified negative samples, and FP is the number of incorrectly classified positive samples. To mitigate the impact of imbalanced data, we utilize Macro avg to optimize the three evaluation metrics, avoiding excessive focus on larger categories and ensuring our experimental results have good interpretability for multi-class classification.

4.3. Experimental Results and Analysis

In order to compare the superiority of our model, we simultaneously utilized the original data (RAW Data, RAW), SMOTE, and the WGAN data augmentation model to augment the training dataset as control groups. Additionally, we employed three classical classifiers, XGBoost (XGB), Random Forest (RF), and Decision Tree (DT), to demonstrate the performance improvement of our model in multi-class classification.

Table 3: Experimental Results of VAE-WGAN-GP on the CIC-IDS 2017 Dataset.

Classifier / Enhanced model	Macro avg Precision(%)			Macro avg Recall(%)			Macro avg F1-Score(%)		
	XGB	RF	DT	XGB	RF	DT	XGB	RF	DT
RAW	90.89	84.31	82.45	62.66	74.09	79.56	66.2	76.92	81.06
SMOTE	93.17	83.34	80.69	73.21	74.55	82.38	73.62	77.95	81.82
WGAN	91.84	84.21	81.65	71.45	75.72	80.87	71.38	79.12	81.69
VAE-WGAN-GP	94.26	89.35	85.12	84.93	77.55	85.23	85.41	83.94	85.58

Table 4: Experimental Results of VAE-WGAN-GP on the Gas Dataset.

Classifier / Enhanced model	Macro avg Precision(%)			Macro avg Recall(%)			Macro avg F1-Score(%)		
	XGB	RF	DT	XGB	RF	DT	XGB	RF	DT
RAW	60.87	59.47	59.84	60.12	61.03	60.67	61.49	60.68	59.31
SMOTE	62.32	61.42	61.82	61.35	62.15	62.39	61.82	61.35	61.15

ISSN 2616-5775 Vol. 7, Issue 4: 36-44, DOI: 10.25236/AJCIS.2024.070406

WGAN	78.71	77.57	77.49	73.32	72.78	70.97	73.23	73.98	69.81
VAE-WGAN-GP	83.74	89.05	81.32	85.46	86.14	84.98	84.28	83.64	83.07

From Tables 3 and 4, it can be observed that across different classifiers, our model effectively improves the performance of classifiers by enhancing the balanced dataset through data augmentation. In the best-case scenario, the proposed VAE-WGAN-GP model shows a significant improvement compared to the SMOTE and WGAN data augmentation models. On the CIC-IDS 2017 dataset and GAS dataset, the Macro avg Precision reaches 94.26% and 83.74%, respectively. This strongly indicates that our model can enhance the performance of intrusion detection systems in the context of class imbalance.

5. Conclusion

In this paper, we propose an intrusion detection system based on the VAE-WGAN-GP data augmentation system. VAE-WGAN-GP is a deep learning generative model that combines VAE and WGAN, utilizing gradient penalty. This model is adept at generating more realistic and continuous samples. By incorporating the KL divergence term from VAE and the gradient penalty from WGAN-GP, it helps alleviate the issue of mode collapse, making the generator more stable. This approach effectively addresses the problem of imbalanced datasets by augmenting samples of minority classes through data augmentation in scenarios with imbalanced data types. Furthermore, we conducted computer simulation experiments on the public network datasets CIC-IDS2017 and GasPipeline. The experimental results indicate that, compared to traditional data augmentation models, VAE-WGAN-GP generates samples that balance authenticity and category attributes. Moreover, the adversarial training process is stable and significantly enhances the recognition performance of intrusion detection classifiers, especially against attacks on minority class samples.

References

[1] Park, K.J., Kim, J., Lim, H. and Eun, Y., 2014. Robust path diversity for network quality of service in cyber-physical systems. IEEE Transactions on Industrial Informatics, 10(4), pp.2204-2215.
[2] Kou, L., Ding, S., Rao, Y., Xu, W. and Zhang, J., 2022. A lightweight intrusion detection model for 5G-enabled industrial Internet. Mobile Networks and Applications, 27(6), pp.2449-2458.
[3] Yang, Y., Wu, L., Yin, G., Li, L. and Zhao, H., 2017. A survey on security and privacy issues in Internet-of-Things. IEEE Internet of things Journal, 4(5), pp.1250-1258.
[4] Malik, S., Amin, J., Sharif, M., Yasmin, M., Kadry, S. and Anjum, S., 2022. Fractured elbow classification using hand-crafted and deep feature fusion and selection based on whale optimization approach. Mathematics, 10(18), p.3291.
[5] Abu-Khzam, F.N., Abd El-Wahab, M.M., Haidous, M. and Yosri, N., 2022. Learning from obstructions: An effective deep learning approach for minimum vertex cover. Annals of Mathematics and Artificial Intelligence, pp.1-12.
[6] Sayour, M.H., Kozhaya, S.E. and Saab, S.S., 2022. Autonomous robotic manipulation: Real-time, deep-learning approach for grasping of unknown objects. Journal of Robotics, 2022.
[7] Wang, J., Li, P., Kong, W. and An, R., 2022. Unknown Security Attack Detection of Industrial Control System by Deep Learning. Mathematics, 10(16), p.2872.
[8] Khan, I.A., Keshk, M., Pi, D., Khan, N., Hussain, Y. and Soliman, H., 2022. Enhancing IIoT networks protection: A robust security model for attack detection in Internet Industrial Control Systems. Ad Hoc Networks, 134, p.102930.
[9] Krithivasan, K., Pravinraj, S. and VS, S.S., 2020. Detection of cyberattacks in industrial control systems using enhanced principal component analysis and hypergraph-based convolution neural network (EPCA-HG-CNN). IEEE Transactions on Industry Applications, 56(4), pp.4394-4404.
[10] Krawczyk, B., 2016. Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence, 5(4), pp.221-232.
[11] Zhou, X., Hu, Y., Wu, J., Liang, W., Ma, J. and Jin, Q., 2022. Distribution bias aware collaborative generative adversarial network for imbalanced deep learning in industrial IoT. IEEE Transactions on Industrial Informatics, 19(1), pp.570-580.
[12] Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J. and Ahmad, F., 2021. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Transactions on Emerging Telecommunications Technologies, 32(1), p.e4150.
[13] Tomczak, J. and Welling, M., 2018, March. VAE with a VampPrior. In International conference on artificial intelligence and statistics PMLR. (pp. 1214-1223).
[14] Engelmann, J. and Lessmann, S., 2021. Conditional Wasserstein GAN-based oversampling of

ISSN 2616-5775 Vol. 7, Issue 4: 36-44, DOI: 10.25236/AJCIS.2024.070406

tabular data for imbalanced learning. Expert Systems with Applications, 174, p.114582.
[15] Safari, M., Parvinnia, E. and Haddad, A.K., 2021. Industrial intrusion detection based on the behavior of rotating machine. International Journal of Critical Infrastructure Protection, 34, p.100424.
[16] Sun, Y., Wang, G., Yan, P.Z., Zhang, L.F. and Yao, X., 2021, December. Industrial Control System Attack Detection Model Based on Bayesian Network and Timed Automata. In International Conference on Big Data (pp. 79-92). Cham: Springer International Publishing.
[17] Mbow, M., Koide, H. and Sakurai, K., 2021, November. An intrusion detection system for imbalanced dataset based on deep learning. In 2021 Ninth International Symposium on Computing and Networking (CANDAR) IEEE. (pp. 38-47).
[18] Bao, F., Deng, Y., Kong, Y., Ren, Z., Suo, J. and Dai, Q., 2019. Learning deep landmarks for imbalanced classification. IEEE transactions on neural networks and learning systems, 31(8), pp.2691-2704.
[19] Liang, W., Hu, Y., Zhou, X., Pan, Y., Kevin, I. and Wang, K., 2021. Variational few-shot learning for microservice-oriented intrusion detection in distributed industrial IoT. IEEE Transactions on Industrial Informatics, 18(8), pp.5087-5095.
[20] Fu, Y., Du, Y., Cao, Z., Li, Q. and Xiang, W., 2022. A deep learning model for network intrusion detection with imbalanced data. Electronics, 11(6), p.898.
[21] Liu, L., Wang, P., Lin, J. and Liu, L., 2020. Intrusion detection of imbalanced network traffic based on machine learning and deep learning. IEEE access, 9, pp.7550-7563.
[22] Huang, S. and Lei, K., 2020. IGAN-IDS: An imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks. Ad Hoc Networks, 105, p.102177.
[23] Ding, H., Chen, L., Dong, L., Fu, Z. and Cui, X., 2022. Imbalanced data classification: A KNN and generative adversarial networks-based hybrid approach for intrusion detection. Future Generation Computer Systems, 131, pp.240-254.