RCALAD: Regularized Complete Adversarially Learned Anomaly Detection

Abstract

The present study provides a comprehensive adversarial framework for anomaly detection in real-world problems based on adversarial neural networks' capability to model complex high-dimensional data. The proposed regularized complete adversarially learned anomaly detection (RCALAD) model, utilizes a generator network to reconstruct the input data and then calculates the anomaly score to determine anomalies. In the proposed model, it is attempted to use the information of a complete cycle in order to improve the training process by adding a new discriminator to the structure. Moreover, to increase the distance between the anomalous sample and its reconstruction, a further supplementary distribution in the input space is used to bias all the reconstructions toward the normal data distribution. In addition to the proposed model, two new anomaly scores are presented in this study. Experimental results demonstrate the superiority of the RCALAD model to other state-of-the-art models in the field of anomaly detection.

Keywords: Anomaly detection, Machine learning, Generative adversarial networks, Reconstruction error, Anomaly scores.

1. Introduction

Discovering dissimilar instances and rare patterns is one of the most essential tasks in real-world data. Such samples are referred to as anomalies, and identifying such instances is called anomaly detection [1]. Data anomalies play a crucial role in a variety of applications and are a vital part of any dataset. i.e. an unusual traffic pattern on a computer network can mean hacking the computer and transferring data to unauthorized destinations. Abnormal behaviors in credit card transactions may indicate fraudulent economic activities [2], or abnormality in an MRI image may indicate the presence of a malignant tumor [3]. Despite the existence of statistical and machine learning-based methods, designing effective models for detecting anomalies in complex high-dimensional data space is still a major challenge [4].

Generative adversarial networks are able to overcome this challenge and model the distribution of real-world data, which have high complexity and dimension; this results in a promising performance in the field of anomaly detection. In the generative adversarial network, a generator network is contrasted with a discriminator network; the discriminator attempts to differentiate between the real data and data which are produced by the generator network. Training phase of generator and discriminator network are simultaneously. The generator network G records the distribution of the data, and the discriminator D estimates the likelihood of generating the sample in G. The objective function for the G generator network is maximizing the error probability of the D network. This platform leads to a two-player game like mini-max games [5].

The ability of adversarial neural networks to model natural images has been proven [6,7], and their usage is increasing in processing speech and text [4], as well as medical images [8]. In the present article, an efficient and effective method based on adversarial generator networks is proposed, which is particularly designed for anomaly detection. Like many learning-based algorithms, there are two main steps: training and testing. In the training part, like other adversarial frameworks, we train the generator and the discriminator networks so that both parts are updated in turn. Here, the joint distribution of the parameters in the network is used to stabilize the training of the adversarial structure. The encoder E is trained along with the discriminator and generator network to apply the inverse mapping of the input samples to the latent space.

Most of the previous have focused only on the density estimation of the normal data, and there is no obligation for the weak reconstruction of the anomalous data [8]. Indeed, it has been assumed in previous works that if the network learns the norm data distribution, it will necessarily have weak reconstruction for abnormal data, But in practice, it is observed that it is not always the case. In the RCALAD model, to focus on reconstructing abnormal samples as weakly as possible, the distribution $\sigma(x)$ is added to the proposed adversarial structure so that the encoder and generator are biased towards the normal data distribution.

Experiments have been carried out on a various range of imagery and tabular dataset with high dimensions, and the outcomes have demonstrated that the proposed model performs better than other baseline models. The main contributions of this paper are summarized as follows:

- Introduce complete cycle consistency problem and utilize D_{xxzz} to settle this challenge.
- Bias the encoder and the discriminator towards the normal manifold via supplementary distribution σ(x).
- Define two new anomaly scores specified for various types of datasets.

2. Related works

Anomaly detection, also known as novelty detection and outlier detection, has been widely studied as reviewed in [9] [10] [11]. The previous methods that have been used in this field are generally divided into two categories: representation learning and generative model.

Methods which are based on representation learning solve the problem of anomaly detection by extracting main characteristics or learning a map from normal data. One-class support vector machine finds the marginal boundary around the normal data [12]. The isolation forest method is one of the classic machine learning methods. In this method, the tree is built with randomly chosen features. The anomaly score in this model is the average distance to the root [13]. Deep support vector data description (DSVDD) finds a hypersphere to enclose the representation of normal samples [14]. Liu and Gryllias constructed frequency domain features using cyclic spectral analysis and used them in svdd framework. This method has been proven robust against outliers and can achieve a high detection rate for bearing anomaly detection [15]. In [16], researchers presented a new approach to identify imagery anomalies by training the model on normal images altered by geometric transformation. In this model, the classifier calculates the anomaly score using softmax statistics.

Usually, generative models attempt to learn the reconstruction of the data and use this reconstruction to identify anomalous samples [17]. For instance, autoencoders model the normal data distribution, and the reconstruction error is used as the anomaly score [18,19]. Deep structured energy-based models (DSEBM) learn an energy-based model and map each sample to an energy score [20]. Deep autoencoding Gaussian mixture model (DAGMM) estimates a mixed Gaussian distribution by using an encoder for normal samples [21]. Recently, adversarial neural networks have been used in anomaly detection. For example, this structure has been used to identify anomalies in medical images. In [8], the inverse mapping to the latent space was performed using the recursive backpropagation mechanism. In [22], which is a continuation of the previous work, the mapping to the latent space was done by the encoder network in order to reduce the computational complexity. In [23], the proposed model was based on BiGAN. Also, the inverse mapping from the input data space to the latent space was within the scope of the responsibilities of the encoder. Unlike the standard GAN structure, where the discriminator takes only the real image, and the generated image of the generator network as input, the representation of these images in the latent space was also considered as input to the discriminator network. In the article, anomaly detection was performed by adding D_{zz} discriminator to the adversarial structure introduced to stabilize the training process. The deep convolutional autoencoder model is a classic autoencoder model, in which the encoder and decoder have a convolutional structure. The anomaly score in this model is l2-norm of reconstruction errors [24].

3. Preliminaries

The adversarial generator networks were first proposed in 2014 by Goodfellow et al [5]. In these networks. The generator network is placed against the discriminator network. These networks are trained on an M set of $\{x^{(i)}\}_{i=1}^{M}$ unlabeled samples. The generative sub-network maps selected samples from the latent space z to the input data space. The discriminator sub-network attempts to distinguish between the real data $x^{(i)}$ and the data produced by the generator network G. These two sub-networks compete with each other; The generator network G tries to imitate the distribution of the input data, while the discriminator network tries to distinguish between the real samples and the data produced by the generator network. In the training

phase, the G generator network and the D discriminator network are alternatively optimized using gradient descent.

The distribution of the input data is represented as q(x), and p(z) is considered as the generator network in the latent space of Z. GAN network training is done by finding a discriminator and a generator that can solve the saddle point problem as: $\min_G \max_D V_{GAN}(D,G)$.

The $V_{GAN}(D,G)$ function is defined as:

$$V_{GAN}(D,G) = E_{x\sim q(x)}[log(D(x))] + E_{z\sim p(z)}[log(1-D(G(z)))]$$

Solving this problem concludes that the generating distribution is equal to the true data distribution. The global optimal discriminator will be obtained only if $P_G(x) = q(x)$. By P_G, we mean the distribution learned by the generator network. It has been proved in [5]. In the adversarially learned inference (ALI) article [25], it has been attempted to model the joint distribution of encoder as q(x,z)=q(x)e(z|x) and the distribution of generator network as $p(x,z) = p(z)p(x|z)$ using the encoder E. Here, $e(z|x)$ is learned by the encoder network. The objective function of the ALI model is as follows:

(3) $$\min_{G,E} \max_D V_{ALI}(D,G) = E_{q(x,z)}[log\, D(x,E(x))] + E_{p(x,z)}[log\,(-D(G(z),z)]$$

where D represents the discriminator network, taking x and z as input, and its output value specifies with what probability the current inputs originate from the $q(x,z)$ distribution. Encoder, generator network, and discriminator are in their optimal state only if $q(x,z) = p(x,z)$. This has been proved in [25].

Although p and q distributions are apparent, in practice and during model training, they are not necessarily converging to the optimal point. This issue in [26] was attributed to the problem of cycle consistency which was defined as $G(E(x)) \approx \hat{x}$. In [26], a framework called ALICE was proposed to solve the above problem by adding the discriminator D_{xx} to the ALI network structure. The objective function of this model is as follows:

(4) $$\min_{E,G} \max_{D_{xz},D_{xx}} V_{ALICE} = V_{ALI} + E_{x\sim q(x)}[log\ D_{xx}(x,x) + log\ 1 - D_{xx}(x,G(E(x)))]$$

This work demonstrates that using a discriminator D_{xx} can achieve the best reconstruction for the input data [26]. In [4], a conditional distribution was applied to the baseline ALICE model with the addition of another discriminator to stabilize the training process. To deep into the detail, a discriminator D_{zz} is added to the model to ensure the cycle consistency in the latent space, which tries to make the latent space variable and its reconstruction as analogous as possible. By assembling the block proposed in [4] in ALICE framework, the cost function of the ALAD model will finally be as follows:

(5)

$$min_{G,E} max_{D_{xz},D_{xx},D_{zz},} V_{ALAD} = V_{ALICE} + \mathbb{E}_{z\sim p(z)}[log(D_{zz}(z,z)) + \mathbb{E}_{z\sim p(z)}[log(-D_{zz}(z,G(E(z))))]$$

In the present article, it is declared that the training model will be stabilized by adding Lipschitz to the discriminator of the GAN model. Moreover, it is shown in practice that with spectral normalization of the weight parameters, we will improve the network's performance [4].

Although the idea of ALAD helps stabilize the cycle, still the latent and input space variables are scrutinized x in two independent spaces, and the inherent dependence between the variables is ignored. More precisely, the x and xˆ variables are checked in a separate process from the z and zˆvariables, while the reconstruction process of these two pairs of data is along with each other and affects each other directly. To model this dependence, we define a complete cycle and a new discriminator that utilizes this complete cycle information.

Moreover, another problem of this model is the careless assumption of the necessity of weak reconstruction for anomalous samples. In fact, in all the previous models, this assumption has been implied that if the model is trained with normal data, it will necessarily have a poor reconstruction for the anomalous data, while there is no constraint to bias the model towards generating poor reconstructions of anomalous samples. In the proposed RCALAD model, an attempt has been made to address this weakness. Also, this assumption should not be applied negligently. Here, by using the envelope distribution of σ(x), the model

bias all the reconstructions toward the normal data distribution and, with this procedure, tries to reduce the distance between the input and the network reconstruction for normal data while increasing the distance for anomalous data.

4. The proposed model

In this section, the aforementioned problems are scrutinized. First, the problem of complete cycle consistency and its solution method will be described; then, we will examine the issue of the necessity of weak reconstruction, and the final proposed model designed to solve both problems will be introduced. At the end of this section, two new anomaly scores are presented based on the proposed model.

4.1. Complete consistency cycle

As mentioned before, in the ALAD model, the consistency of the cycle for the input data and latent space variable is examined in two independent processes. Indeed, the process of approaching the reconstruction of the latent space variable, i.e., zˆto the variable z itself, and approaching the reconstruction of the input data, i.e., xˆto x itself, are done separately. It should be noted here that the variable z is an example of the Gaussian distribution that is given as an input to the generator network and has nothing to do with the input mapping in the latent space.

The complete cycle consistency issue (CCC) declares that for each variable of x in the input space if the encoder first estimates the inverse mapping to the latent space, which equals $E(x) = z_x$Then, the obtained representation is entered into the generator network to generate the reconstruction of the network from the input variable of $G(z_x) = G\big(E(x)\big) = \hat{x}$. Finally, the same reconstruction is given to the encoder network in order to calculate the reconstruction in the latent space, that is, $E(\hat{x}) = E\big(G(z_x)\big) = \hat{z}_{\hat{x}}$ˆ, it is logically expected from any reconstruction-based network that the two variables x and xˆas well as the two variables $\hat{z}_{\hat{x}}$ and z_x have the least possible difference. That is, the CCC problem is defined in such a way that, in any model based on reconstruction, for each input data and its mapping in the latent space, the reconstruction provided by the network for both variables should have a minimum error and maximum similarity with them.

In the ALAD model, the similarity between the input data and its reconstruction, as well as the z similarity and its reconstruction, were examined independently and in two separate cycles. It was assumed that they are independent, while we know these two cycles are entirely dependent on each other, and the assumption of independence is not valid in these two issues. To solve this problem, it has been recommended in the present article to model the dependency by examining the variables in the CCC in the new discriminator D_{xxzz} and using the information flow in this chain to improve network training for anomaly detection in the best possible way. The difference between the input of D_{xxzz} and the input of D_{zz} used in the ALAD model is represented in Figure 1.

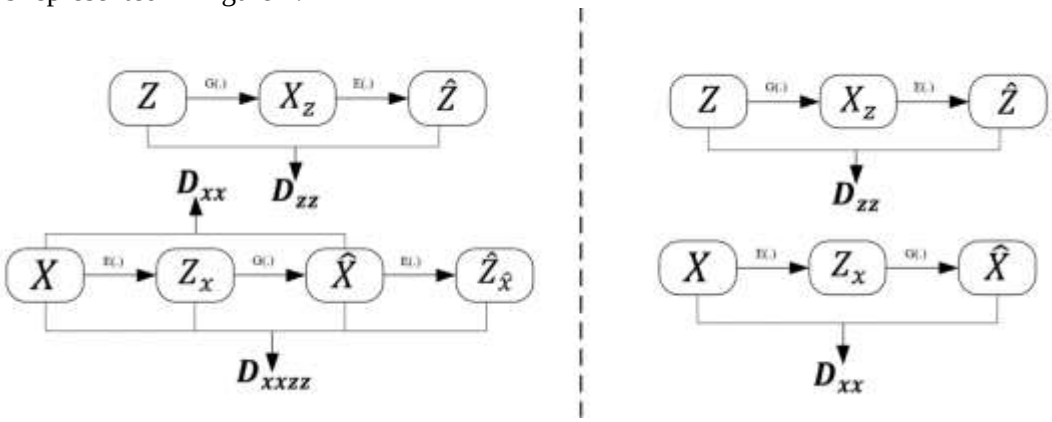

Figure 1: Using the variables of input data space and latent space in the cycle consistency of the ALAD network (right side) and the information of a complete cycle consistency in the proposed model (left side);

As can be seen in Figure 1, the ALAD model did not use the information of a complete cycle. In order to use the available information in a complete cycle, a new variable called $\hat{z}_{\hat{x}}$ is introduced. To calculate this variable, the inverse mapping of the input data of x is provided to the generator network and the resulting inverse mapping is calculated again using the decoder. Hence, the complete cycle results from the process of transforming the input data.

In order to ensure the condition of complete cycle consistency, the new D_{xxzz} discriminator is used with the joint input. It is noteworthy that the effectiveness of the joint discriminators has already been proven once in ALIGAN. Actually, when adding the encoder network to the GAN framework, two procedures can be scrutinized. The first mode is adding an independent discriminator to train the coder, and the second mode is changing the input of the discriminator from a single input mode to a joint input mode. Therefore, it is proved that the second mode performs better and obtains better results. According to the same idea, the input of the joint discriminator D_{xxzz} extracts the most information for model training.

This discriminator uses the quadruple of (x, x, z_x, z_x) as the actual data and the quadruple of $(x, G(E(x)), z_x, E(G(z_x)))$ as fake data. This discriminator attempts to make the input x and its reconstruction provided by the network as well as the inverse mapping of the input image in the latent space and its reconstruction by the encoder as close as possible to each other, so that a complete stable loop is provided and the model is trained and stabilized better.

4.2 Necessity of weak reconstruction

In reconstruction-based models, it has always been assumed that if the training and reconstruction of the normal data are properly done, the reconstruction of abnormal data will necessarily be weak and different from the input data. However, experiments indicate it is not always the case, and the reconstructed anomalous sample is slightly different from the input sample. Hence, it won't be easy to recognize it as an abnormal sample. In fact, in none of the previous models, there is no requirement or control condition to bias the model towards producing poor reconstruction for anomalous samples.

The reason for this phenomenon is the weak mapping from the input data space to the latent space. Because in the training phase, the encoder only learns the mapping of the normal samples to the latent space and, as a result, the corresponding space of z for the normal samples is well modeled, but in the test phase, given the fact that the model has not yet seen the rest of the space, including abnormal examples, it may map it onto an unknown point of the latent space. That is, in this case, there is no information regarding the anomalous data. One solution for this issue is mapping all the input space to the latent norm subspace. It means that to cover the data space as much as possible, the supplementary distribution called σ(x) is used. By sampling this function and learning the network toward producing the reconstruction of the normal data class, the network learns to reconstruct the norm data class for a relatively more expansive range of inputs. In this case, if the input data is abnormal, the model is trained to produce a reconstruction close to the normal data class; as a result, a suitable distance is created between the input data and its reconstruction. This distance is regarded as an appropriate criterion for detecting abnormal samples. In Figure 2, you will be familiar with how this training routine works.

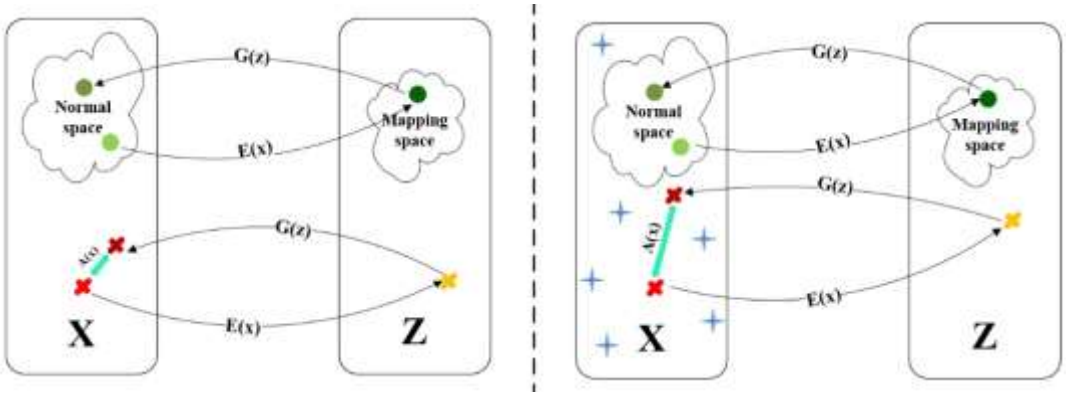

Figure 2: Effect of the presence of σ(x) distribution in the model training process. In this figure, x represents the input data space and z represents the latent space. Samples are mapped from the input data space to the latent space by the generator G, and the E encoder is responsible for performing the inverse mapping. Green circles show normal samples red crosses represent abnormal samples and blue stars represent samples generated by the σ(x) distribution. The turquoise-colored line shows the value of the abnormality score. As can be seen in Figure 5-3, if σ(x) (on the left side of the figure) is not present in the training process, the abnormality score for abnormal samples is lower than when this distribution is used. On the right, the distribution of σ(x) has biased the model towards the reconstruction of all samples, both abnormal and normal.

4.3. RCALAD model

In this section, by combining both ideas presented in the previous sections, i.e., using the new variable of $\hat{z}_{\hat{x}}$ in the D_{xxzz} discriminator, as well as using the σ(x) distribution and adding them to the basic model [4], the main model proposed by RCALAD is introduced. In this network, issues of complete consistency cycle and the necessity of weak reconstruction are addressed simultaneously and it has been attempted to provide a comprehensive, practical and compatible framework for all anomaly detection problems. The outline of the proposed model can be seen in Figure 3.

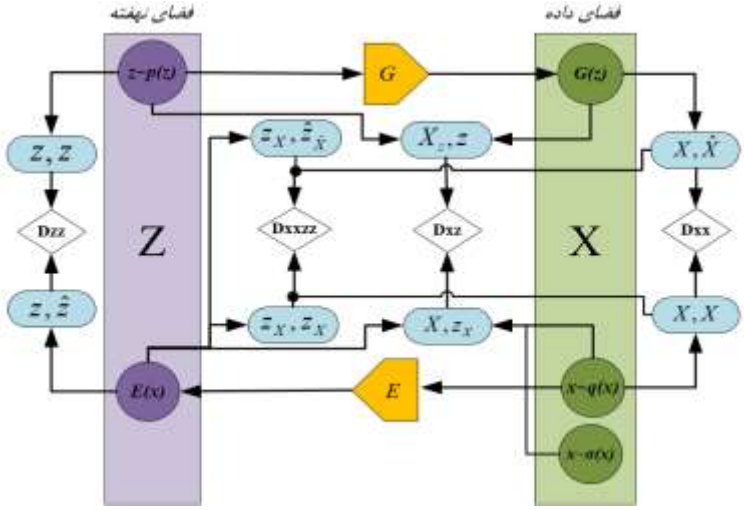

Figure 3: Overall structure of the RCALAD model

As in Figure 3, an encoder combined with the generator network is trained in the general structure of the adversarial neural network to reduce the time complexity. The inverse mapping from the input data space to the latent space is obtained simply by embedding the E encoder in the proposed structure. Here, a simultaneous discriminator network called D_{xz} is used to train both generator and encoder networks simultaneously. This discriminator checks whether the input variable pair belongs to the distribution of the input data x and its corresponding point in the latent space of E(x) or is generated by the generator network of G(z) and sampling from the latent space of z. In order to satisfy the condition of cycle consistency and in the input data space, Dxx and Dzz discriminators are used so each sample and its corresponding reconstruction can be improved and modeled independently. The D_{xxzz} is introduced to use all the information in a complete cycle. That is, in addition to examining both variables and reconstructing them

in the same space, their quadruple distribution is used in the process of detecting anomalous samples so that the network has access to the state of the input data during successive mappings and more information is available to distinguish the data. This network is responsible for determining between quadruple samples of (x, x, z_x, z_x) and $(x, G(E(x)), z_x, E(G(z_x)))$ and tries to extract x and the reconstruction provided by the network as well as to make the mapping of the input image in the latent space of z_x and the reconstruction of the output of the generating network by the coder of $E(G(z_x))$ as close as possible. The σ(x) block is added to this model to cover the maximum latent space. Adopting this block leads to generating new samples in the input data space and mapping them to the latent space corresponding to normal data distribution. Finally, the objective function of the proposed model is as follows:

(7)

$$\min_{G,E}\max_{D_{xxzz},D_{xz},D_{xx},D_{zz}} V_{RCALAD}\left(D_{xxzz}, D_{xz}, D_{xx}, D_{zz}, E, G\right)$$
$$= V_{ALAD} + \mathbb{E}_{x\sim\sigma(x)}\left[\log\left(1 - D_{xz}\left(x, E(x)\right)\right)\right] + \mathbb{E}_{x\sim q(x)}\left[\log D_{xxzz}(x, x, E(x), E(x))\right]$$
$$+ \mathbb{E}_{\boldsymbol{x}\sim\boldsymbol{q}(x)}\left[1 - log\, D_{xxzz}\left(x, G\left(E(x)\right), E(x), E\left(G\left(E(x)\right)\right)\right)\right]$$

4.4 Anomaly detection

The major aim of presenting the proposed model in this study is to detect anomalies based on input data reconstruction. In this model, the ultimate goal is accurate and similar reconstruction for the norm data and weak and different reconstruction for the abnormal sample. One of the key elements in anomaly detection is the definition of the anomaly score for calculating the distance between the input sample and the reconstruction provided by the network [4]. In some models, such as DCAE, the expression of anomaly scores that were used in previous models is as follows:

$$A_{L_1}(x) = \|x - \hat{x}\|_1$$
$$A_{L_2}(x) = \|x - \hat{x}\|_2$$
$$A_{Logits}(x) = \log\left(D_{xx}(x, \hat{x})\right)$$
$$A_{Features}(x) = \|f_{xx}(x, x) - f_{xx}(x, \hat{x})\|_1$$

Here, the raw output of the discriminators is named logit, and the output of the latent layer before the logit layer is called feature. Since the D_{xxzz} adds the ability of extracting new information to the model which has not been introduced in any of the previous scores, there is a strong need to define the new anomaly scores to use this ability. In this paper, two new anomaly scores are introduced based on the information mentioned in the D_{xxzz}:

The first anomaly score presented in this paper is called $A_{fm}(x)$. In this score, the feature space in the D_{xxzz} discriminator is used to calculate the distance between samples and reconstruct them. For this purpose, the logit output of the second last layer is used as a feature. The anomaly score used is defined as follows, using the soft reconstruction error of and according to the following equation:

(8)

$$A_{fm}(x) = \left\| f_{xxzz}\left(x, x, E(x), E(x)\right) - f_{xxzz}\left(x, G\left(E(x)\right), E(x), E\left(G\left(E(x)\right)\right)\right)\right\|_1$$

In this equation, f (·) represents the activation function of the second last layer in the D_{xxzz} differentiating structure. The concept used in the definition of this score is using the level of discriminator confidence on the quality of the reconstructions provided by the network. If it is performed well, the sample belongs to the trained data of the network or the normal data distribution. Thus, the higher the value of this criterion, the greater the difference in reconstructions and the higher the possibility of the input data's abnormality.

The second point in this article is presented with the aim of maximizing the use of information in the model for anomaly detection. In this section, the A_{all} criterion is defined. This score is made up of the sum of the output of all three discriminators, including D_{xx}, D_{zz} and D_{xxzz}. In fact, since all the discriminators in the

proposed model are trained only based on the normal samples and the reconstruction for all the input data space is biased towards the norm data space, it is expected that reconstructed image of the anomalous sample as well as its representation in the latent space produced by the encoder are very different from the input, and the discriminators in the model can easily identify these anomalous inputs. The mathematical expression of this criterion is given in the following equation:

(9)

$$\boldsymbol{A_{all}(x) = D_{xxzz}(x, \hat{x}, z_x, \hat{z}_{\hat{x}}) + D_{xx}(x, \hat{x}) + D_{zz}(z_x, \hat{z}_{\hat{x}})}$$

Now, the issue that should be scrutinized is whether the criterion A_{all} contains enough information to distinguish normal from anomalous data or not. Generally speaking, the answer to this question is yes, since, during the training phase, the discriminators learn to pay attention to the difference between the pairs of (x, x) and $(x, \hat{x})$ as well as the pairs of (z_x, z_x) and $(z_x, \hat{z}_{\hat{x}})$: that is, the farther $\hat{x}$ from x or $\hat{z}_{\hat{x}}$ from z_x, the easier it is for the discriminators to recognize it. In the proposed model, by adding the distribution of σ(x) and biasing all the reconstruction towards the normal data distribution, the reconstruction error for abnormal data is increased and the discriminators' output can be considered a reliable criterion for abnormality detection. Finally, the recommended anomaly scores can be viewed according to the algorithm below:

Algorithm 1: Process of calculating anomaly scores

Algorithm 1 Regularized Complete Adverarially Learned Anomaly Detection

Input $\boldsymbol{x \sim p_{X_{\text{Test}}}(x), E, G, D_{xx}, D_{zz}, D_{xxzz}, f_{xxzz}}$ where $\boldsymbol{f_{xxzz}}$ is the feature layer of , $\boldsymbol{D_{xxzz}}$

Output $\boldsymbol{A_{all}(x)}$, $\boldsymbol{A_{fm}(x)}$, **where A is the anomaly score**

procedure INFERENCE

$\boldsymbol{z_x \leftarrow E(x)}$ Encode samples, Construct latent Embedding

$\boldsymbol{\hat{x} \leftarrow G(z_x)}$ Reconstruct samples

$\boldsymbol{\hat{z}_{\hat{x}} \leftarrow E(\hat{x})}$ Reconstruct latent Embedding

$\boldsymbol{A_{fm}(x) \leftarrow \|f_{xxzz}(x, x, z_x, z_x) - f_{xx}(x, \hat{x}, z_x, \hat{z}_{\hat{x}})\|_1}$

$\boldsymbol{A_{all}(x) \leftarrow D_{xxzz}(x, \hat{x}, Z_x, \hat{z}_{\hat{x}}) + D_{xx}(x, \hat{x}) + D_{zz}(Z_x, \hat{z}_{\hat{x}})}$

return $\boldsymbol{A_{all}(x)}$, $\boldsymbol{A_{fm}(x)}$

end procedure

5. Experiments

In this section, we evaluate the performance of the proposed model in comparison with the prominent models in the field of anomaly detection, which were examined in detail in Section 2. To test the models on a fair basis, the reported outcomes for all the implemented models are based on tabular data obtained from the average of ten runs, and, for each class of imagery data, it is based on the average of three runs. The anomaly score used in tabular data is A_{all} score and, for imagery data, it is A_{fm} score. The reason for choosing this score will be discussed further in Section 5.6. Moreover, the ALAD model is implemented and the results of the best anomaly score A_{fm} are reported. For other models, the available results are adopted from [27].

5.1 Datasets

In order to evaluate the performance of the proposed model and scrutinize its efficiency from different viewpoints, various data sets with diverse characteristics are used. The proposed method is tested on the available imagery and tabular datasets. For tabular datasets, four datasets, including kddcup99, arrhythmia, thyroid and musk are used. Kddcup99 is a dataset related to network penetration. Arrhythmia is a medical collection related to cardiac arrhythmia with 16 classes. Also, thyroid is a three-class data related to thyroid disease. The musk dataset was created to classify six classes on molecular musk. In the introduced data set, 20, 15, 2.5 and 3.2 percent of the data are anomalous samples, respectively. Hence, in the test phase, after calculating the anomaly score, the proportion of the data that has the highest anomaly score is classified as anomaly.

Assessing the proposed model on these data sets is performed by calculating F1, Recall and Precision criterion. Two datasets of CIFAR10 and SVHN are considered for the imagery dataset. Both of these datasets have ten classes and, each time, one class is considered as the normal class and the other nine classes as the abnormal class. The criterion used to evaluate the model on imagery data is area under the receiver operating curve (AUROC). For all the data which have been used, 80% of the data are selected as training data and 20% as test data. 25% of training data is selected as validation data. It is noteworthy that, in the training phase, all the anomalous samples are eliminated from the training data.

5.2. Experiments on the tabular datasets

Table 1: Output results of the proposed model in comparison with the basic models on the tabular data set.

	KDDCUP			Arrhythmia			Thyroid			Musk		
Model	Prec.	Recall	F_1	Prec.	Recall	F_1	Prec.	Recall	F_1	Prec.	Recall	F_1
IF	92.16	93.73	92.94	51.47	54.69	53.03	**70.13**	**71.43**	**70.27**	47.96	47.72	47.51
OC-SVM	74.57	85.23	79.54	53.97	40.82	45.18	36.39	42.39	38.87	–	–	–
DSEBMr	85.12	64.72	73.28	15.15	15.13	15.10	4.04	4.03	4.03	–	–	–
DSEBMe	86.19	64.46	73.99	46.67	45.65	46.01	13.19	13.19	13.19	–	–	–
AnoGAN	87.86	82.97	88.65	41.18	43.75	42.42	44.12	46.87	45.45	3.06	3.10	3.10
DAGMM	92.97	94.22	93.69	49.09	50.78	49.83	47.66	48.34	47.82	–	–	–
ALAD	94.27	**95.77**	95.01	50.00	53.13	51.52	22.92	21.57	22.22	58.16	59.03	58.37
DSVDD	89.81	94.97	92.13	35.32	34.35	34.79	22.22	23.61	23.29	–	–	–
RCALAD	**95.36**	95.62	**95.49**	**58.82**	**62.50**	**60.60**	53.76	51.53	52.62	**62.96**	**63.33**	**63.14**
error bar	0.28	0.29	0.28	6.6	6.8	5.8	4.3	2.7	2.8	5.06	2.53	2.62

Evaluation results of the proposed RCALAD model on tabular data of kddcup99, arrhythmia, thyroid and musk are summarized in Table 1. The structures used in the generator, discriminator and encoder networks are all fully connected layers with nonlinear activation functions. It should be noted that, in this step, N(0, I) distribution is used as $\sigma(x)$.

In order to make a clear comparison between different models, the error bar is used in the last row of Table 1. As can be seen in this table, the proposed model has a successful performance on the arrhythmia and musk datasets, compared to other models also on KDD dataset our model is the best according to F1 criteria, but on thyroid dataset secures the second place due to the extraordinary performance of the IF model. The reason for this phenomenon can be attributed to the nature of the data in this dataset since there are various features in this data set, only a few of them are informative; therefore, the results of classic models such as IF, which are based on feature selection, are better. An idea to improve the proposed model results is to use models such as IF in the pre-processing stage to select more effective characteristics for training the model.

5.3. Experiments on the imagery datasets

In this section, the performance of the proposed model on CIFAR10 and SVHN imagery data is scrutinized in two separate tables.

Table 2: Output results of the proposed model compared to the basic models on the CIFAR10 dataset.

Normal	DCAE	DSEBM	DAGMM	IF	AnoGAN	ALAD	RCALAD
Airplane	59.1 ± 5.1	41.4 ± 2.3	56.0 ± 6.9	60.1 ± 0.7	67.1 ± 2.5	64.7 ± 2.6	$\mathbf{72.8 \pm 0.8}$
auto.	$\mathbf{57.4 \pm 2.9}$	57.1 ± 2.0	56.0 ± 6.9	50.8 ± 0.6	54.7 ± 3.4	45.7 ± 0.8	50.2 ± 0.3
Bird	48.9 ± 2.4	61.9 ± 0.1	53.8 ± 4.0	49.2 ± 0.4	52.9 ± 3.0	67.0 ± 0.7	$\mathbf{72.6 \pm 0.2}$
Cat	58.4 ± 1.2	50.1 ± 0.4	51.2 ± 0.8	55.1 ± 0.4	54.5 ± 1.9	59.2 ± 1.1	$\mathbf{64.2 \pm 0.9}$
Deer	54.0 ± 1.3	73.2 ± 0.2	52.2 ± 7.3	49.8 ± 0.4	65.1 ± 3.2	72.7 ± 0.6	$\mathbf{74.9 \pm 0.5}$
Dog	$\mathbf{62.2 \pm 1.8}$	60.5 ± 0.3	49.3 ± 3.6	58.5 ± 0.4	60.3 ± 2.6	52.8 ± 1.2	60.1 ± 1.1
Frog	51.2 ± 5.2	68.4 ± 0.3	64.9 ± 1.7	42.9 ± 0.6	58.5 ± 1.4	69.5 ± 1.1	$\mathbf{75.3 \pm 0.4}$
Horse	58.6 ± 2.9	53.3 ± 0.7	55.3 ± 0.8	55.1 ± 0.7	$\mathbf{62.5 \pm 0.8}$	44.8 ± 0.4	56.6 ± 0.2
Ship	$\mathbf{76.8 \pm 1.4}$	73.9 ± 0.3	51.9 ± 2.4	74.2 ± 0.6	75.8 ± 4.1	73.4 ± 0.4	$\mathbf{77.5 \pm 0.3}$
Truck	$\mathbf{67.3 \pm 3.0}$	63.6 ± 3.1	54.2 ± 5.8	58.9 ± 0.7	66.5 ± 2.8	43.2 ± 1.3	52.6 ± 0.6
Mean	59.4	60.3	54.4	55.5	61.8	59.3	**65.7**

Table 3: Output results of the proposed model compared to the basic models on the SVHN dataset.

Normal	OCSVM	DSEBMr	DSEBMe	IF	ANOGAN	ALAD	RCALAD
0	52.0 ± 1.6	56.1 ± 0.2	53.4 ± 1.8	53.0 ± 0.6	57.3 ± 0.4	58.7 ± 0.9	$\mathbf{60.4 \pm 0.1}$
1	48.6 ± 5.3	52.3 ± 0.9	52.1 ± 0.3	51.2 ± 0.9	57.0 ± 0.8	$\mathbf{62.8 \pm 1.7}$	59.2 ± 0.3
2	49.7 ± 7.7	51.9 ± 0.8	51.8 ± 0.4	52.3 ± 0.1	53.1 ± 0.4	$\mathbf{55.2 \pm 2.3}$	54.9 ± 0.1
3	50.9 ± 1.4	51.8 ± 0.4	51.7 ± 0.5	52.2 ± 0.3	52.6 ± 0.4	53.8 ± 3.3	$\mathbf{55.8 \pm 1.9}$
4	48.4 ± 5.2	52.5 ± 0.1	52.4 ± 0.2	49.1 ± 0.6	53.9 ± 0.5	58.0 ± 0.1	$\mathbf{58.5 \pm 0.2}$

5	51.1 ± 2.6	52.4 ± 2.3	52.3 ± 2.6	52.4 ± 0.8	52.8 ± 0.1	56.1 ± 0.9	$\mathbf{56.2 \pm 0.4}$
6	50.1 ± 3.9	52.1 ± 1.8	52.2 ± 1.8	51.8 ± 0.2	53.2 ± 0.0	57.4 ± 0.6	$\mathbf{59.4 \pm 0.5}$
7	49.6 ± 1.3	53.4 ± 0.9	55.3 ± 1.1	52.0 ± 0.4	55.0 ± 0.0	$\mathbf{58.8 \pm 0.3}$	58.0 ± 0.4
8	45.0 ± 4.2	51.9 ± 0.3	52.5 ± 0.6	52.3 ± 0.8	52.2 ± 0.7	$\mathbf{55.2 \pm 0.4}$	$\mathbf{56.1 \pm 0.5}$
9	52.5 ± 3.9	55.8 ± 1.7	52.7 ± 1.4	53.7 ± 0.6	53.1 ± 0.1	57.3 ± 0.6	$\mathbf{58.3 \pm 0.2}$
Mean	50.2	52.9	52.4	51.6	54.0	57.3	**57.7**

As in Tables 2 ad 3, the proposed model has significantly improved the result on CIFAR10 dataset. The results obtained on the majority of classes have the best performance and, on other classes, the outcomes are comparable with other models. The performance of the proposed model is also suitable on the SVHN dataset and, in addition to being superior in seven classes compared to other models, it also performs the best in the average of the whole classes.

5.4 ablation studies

In this section, we scrutinize the performance of each component added to the basic model on both types of datasets. In these experiments, the average results of the model are repeated in the presence and absence of the discriminator of D_{xxzz} and the supplementary distribution $\sigma(x)$.

Table 4: Effects of the various proposed sections in improving the results of the imagery data

Model	AUROC
CIFAR-10	
Baseline (ALAD)	0.593 ± 0.017
Baseline + D_{xxzz} (CALAD)	0.634 ± 0.018
Baseline + $\sigma(x)$ (RALAD)	0.642 ± 0.012
Baseline + D_{xxzz} + $\sigma(x)$ (RCALAD)	$\mathbf{0.657 \pm 0.016}$
SVHN	
Baseline (ALAD)	0.573 ± 0.016
Baseline + D_{xxzz} (CALAD)	0.576 ± 0.014
Baseline + $\sigma(x)$ (RALAD)	0.568 ± 0.018
Baseline + D_{xxzz} + $\sigma(x)$ (RCALAD)	$\mathbf{0.577 \pm 0.019}$

Table 5: Effects of the various proposed sections in improving the results of the imagery data

Model	Precision	Recall	F1 score
KDD99			
Baseline (ALAD)	0.942 ± 0.008	$\mathbf{0.957 \pm 0.006}$	0.950 ± 0.007
Baseline + D_{xxzz} (CALAD)	$\mathbf{0.959 \pm 0.004}$	0.957 ± 0.007	$\mathbf{0.958 \pm 0.005}$
Baseline + $\sigma(x)$ (RALAD)	0.943 ± 0.005	0.955 ± 0.004	0.949 ± 0.004
Baseline + D_{xxzz} + $\sigma(x)$ (RCALAD)	0.953 ± 0.007	0.956 ± 0.005	0.954 ± 0.006
Arrhythmia			
Baseline (ALAD)	0.500 ± 0.049	0.531 ± 0.047	0.515 ± 0.048
Baseline + D_{xxzz} (CALAD)	0.574 ± 0.021	0.605 ± 0.022	0.575 ± 0.021
Baseline + $\sigma(x)$ (RALAD)	0.546 ± 0.035	0.565 ± 0.039	0.555 ± 0.037
Baseline + D_{xxzz} + $\sigma(x)$ (RCALAD)	$\mathbf{0.588 \pm 0.42}$	$\mathbf{0.625 \pm 0.41}$	$\mathbf{0.606 \pm 0.41}$
Thyroid			
Baseline (ALAD)	0.229 ± 0.067	0.215 ± 0.067	0.222 ± 0.067
Baseline + D_{xxzz} (CALAD)	0.529 ± 0.071	$\mathbf{0.518 \pm 0.075}$	0.523 ± 0.073
Baseline + $\sigma(x)$ (RALAD)	0.431 ± 0.039	0.457 ± 0.043	0.443 ± 0.041
Baseline + D_{xxzz} + $\sigma(x)$ (RCALAD)	$\mathbf{0.537 \pm 0.054}$	0.515 ± 0.057	$\mathbf{0.526 \pm 0.055}$
Musk			
Baseline (ALAD)	0.500 ± 0.068	0.531 ± 0.070	0.515 ± 0.069
Baseline + D_{xxzz} (CALAD)	0.574 ± 0.026	0.605 ± 0.027	0.575 ± 0.026
Baseline + $\sigma(x)$ (RALAD)	0.546 ± 0.051	0.565 ± 0.051	0.555 ± 0.051
Baseline + D_{xxzz} + $\sigma(x)$ (RCALAD)	$\mathbf{0.629 \pm 0.011}$	$\mathbf{0.633 \pm 0.016}$	$\mathbf{0.631 \pm 0.013}$

According to Tables 4 and 5, the proposed RCALAD model achieves the highest efficiency in the presence of both parts. In scrutinizing the role of the D_{xxzz} discriminator, this discriminator has improved the accuracy on the CIFAR10 dataset to an optimal level but did not make significant improvement on the SVHN dataset. Concerning the role of $\sigma(x)$ distribution, this distribution performed well on the CIFAR10 dataset and improved the AUROC criterion. However, the SVHN dataset reduced the AUROC criterion by a small amount compared to the base model. Still, its presence in the final model led to extracting new information and a more comprehensive view.

5.5 Evaluating the adequacy of the D_{xxzz} discriminator

By adding the D_{xxzz} discriminator, is there a need for D_{xx} and D_{za} discriminators or not? To answer this question correctly, you should observe the following table. In fact, in this section, in addition to the abovementioned question, the result of adding D_{xxzz} discriminator in basic models such as ALI and ALICE is investigated.

Table 6: Assessing the performance of the model in the presence or absence of each of the components

Model	D_{zz}	D_{xx}	D_{xxzz}	Prec.	Recall	F_1
KDD99						
ALAD	√	√	×	0.942 ± 0.008	$\mathbf{0.957 \pm 0.006}$	0.950 ± 0.007
ALI + D_{xxzz}	×	×	√	0.938 ± 0.007	0.951 ± 0.010	0.944 ± 0.009
ALI + D_{zz} + D_{xxzz}	√	×	√	0.946 ± 0.005	0.955 ± 0.004	0.950 ± 0.004
ALICE + D_{xxzz}	×	√	√	0.941 ± 0.005	0.954 ± 0.008	0.947 ± 0.006
CALAD	√	√	√	$\mathbf{0.959 \pm 0.004}$	0.957 ± 0.007	$\mathbf{0.958 \pm 0.005}$
RCALAD	√	√	√	0.953 ± 0.007	0.956 ± 0.005	0.954 ± 0.006
Arrhythmia						
ALAD	√	√	×	0.500 ± 0.049	0.531 ± 0.047	0.515 ± 0.048
ALI + D_{xxzz}	×	×	√	0.522 ± 0.054	0.529 ± 0.049	0.525 ± 0.052
ALI + D_{zz} + D_{xxzz}	√	×	√	0.571 ± 0.033	0.582 ± 0.028	0.576 ± 0.031
ALICE + D_{xxzz}	×	√	√	0.543 ± 0.052	0.561 ± 0.044	0.551 ± 0.048
CALAD	√	√	√	0.574 ± 0.021	0.605 ± 0.022	0.575 ± 0.021
RCALAD	√	√	√	$\mathbf{0.588 \pm 0.42}$	$\mathbf{0.625 \pm 0.41}$	$\mathbf{0.606 \pm 0.41}$

According to Table 6 and as expected from the theoretical results, adding the D_{xxzz} discriminator to the general frameworks had the highest efficiency. Subsequently, eliminating D_{xx} damages the model less since some of the information it extracts is covered with the D_{xxzz} discriminator. But considering that D_{zz} examines the similarity of z and its reconstruction in an independent cycle, it is evident that eliminating it reduces the accuracy. As can be seen, it can be concluded in this section that these three discriminators are regarded as the most effective, and the D_{xxzz} discriminator alone does not cover all aspects.

It is noteworthy that, in this part, the results are reported on the data set, and this is also the case for the tabular and imagery data sets.

5.6 Anomaly detection

In this section, the anomaly scores introduced in this article are evaluated and compared with the anomaly scores presented in previous works [4]. Here, the raw output of the discriminators is called logit, and the output of the latent layer before the logit layer is called features. Moreover, the variables in the discriminator structure of D_{xxzz} are used to calculate the anomaly score. The mathematical expression of all the available anomaly scores is as follows:

(10)

$$A_{L_1}(x) = \|x - \hat{x}\|_1$$
$$A_{L_2}(x) = \|x - \hat{x}\|_2$$
$$A_{Logits}(x) = \log\left(D_{xx}(x, \hat{x})\right)$$
$$A_{Features}(x) = \|f_{xx}(x, x) - f_{xx}(x, \hat{x})\|_1$$
$$A_{fm}(x) = \|f_{xxzz}(x, x, z_x, z_x) - f_{xxzz}(x, \hat{x}, z_x, \hat{z}_{\hat{x}})\|_1$$
$$A_{all}(x) = D_{xxzz}(x, \hat{x}, z_x, \hat{z}_{\hat{x}}) + D_{zz}(z_x, \hat{z}_{\hat{x}}) + D_{xx}(x, \hat{x})$$

Table 7: **Comparing the performance of the proposed anomaly scores with other scores on tabular data.**

Model	Precision	Recall	F1 score
KDD99			
A_{L_1}	0.9081 ± 0.0638	0.9108 ± 0.0638	0.9094 ± 0.0638
A_{L_2}	0.9011 ± 0.0155	0.9004 ± 0.0157	0.9007 ± 0.0156
A_{Logits}	0.9169 ± 0.0162	0.9168 ± 0.0164	0.9168 ± 0.0163
$A_{Features}$	0.9127 ± 0.0029	0.9177 ± 0.0039	0.9151 ± 0.0034
A_{fm}	$\mathbf{0.9327 \pm 0.0017}$	$\mathbf{0.9377 \pm 0.0017}$	$\mathbf{0.9301 \pm 0.0017}$
A_{all}	0.9231 ± 0.0018	0.9207 ± 0.0018	0.9218 ± 0.0018
Arrhythmia			

A_{L_1}	0.3529 ± 0.0148	0.3750 ± 0.0164	0.3636 ± 0.0256
A_{L_2}	0.3529 ± 0.0107	0.3750 ± 0.0108	0.3636 ± 0.0107
A_{Logits}	0.5588 ± 0.0334	0.5937 ± 0.0386	0.5757 ± 0.0359
$A_{Features}$	0.2325 ± 0.0029	0.2500 ± 0.0029	0.2424 ± 0.0029
A_{fm}	0.4411 ± 0.0013	0.4687 ± 0.0013	0.4545 ± 0.0013
A_{all}	$\mathbf{0.6176 \pm 0.0208}$	$\mathbf{0.6562 \pm 0.0221}$	$\mathbf{0.6363 \pm 0.0214}$
Thyroid			
A_{L_1}	0.4981 ± 0.0028	0.4908 ± 0.0024	0.4994 ± 0.0024
A_{L_2}	0.5011 ± 0.0330	0.5004 ± 0.0318	0.5007 ± 0.0324
A_{Logits}	0.4969 ± 0.0142	0.4968 ± 0.0144	0.4968 ± 0.0143
$A_{Features}$	0.5127 ± 0.0119	0.5177 ± 0.0119	0.5151 ± 0.0119
A_{fm}	0.5227 ± 0.0083	0.5123 ± 0.0083	0.5174 ± 0.0083
A_{all}	$\mathbf{53.76 \pm 0.0029}$	$\mathbf{51.53 \pm 0.0029}$	$\mathbf{52.62 \pm 0.0029}$
Musk			
A_{L_1}	0.5979 ± 0.0103	0.5931 ± 0.0109	0.5954 ± 0.0106
A_{L_2}	0.6008 ± 0.0021	0.6018 ± 0.0028	0.6013 ± 0.0024
A_{Logits}	0.5868 ± 0.0124	0.5897 ± 0.0127	0.5882 ± 0.0125
$A_{Features}$	0.5824 ± 0.0011	0.5883 ± 0.0019	0.5883 ± 0.0015
A_{fm}	0.6111 ± 0.0481	0.6187 ± 0.0468	0.6148 ± 0.0474
A_{all}	$\mathbf{62.96 \pm 0.0013}$	$\mathbf{63.33 \pm 0.0013}$	$\mathbf{63.14 \pm 0.0013}$

As in Table 7, on the tabular data, the raw output of the D_{xxzz} discriminator (A_{all}) has the best results compared to other anomaly scores. According to Table 8, the amount of scores on the imagery data is as follows:

Table 8: Comparing the performance of the proposed anomaly scores with other scores on image data

Anomaly Score	AUROC
SVHN	
A_{L_1}	0.5778 ± 0.0141
A_{L_2}	0.5636 ± 0.0251
A_{Logits}	0.5369 ± 0.0785
$A_{Features}$	0.5763 ± 0.0367
A_{fm}	$\mathbf{0.5778 \pm 0.0161}$
A_{all}	0.5768 ± 0.0251
CIFAR-10	
A_{L_1}	63.41 ± 0.0321
A_{L_2}	63.27 ± 0.0782
A_{Logits}	62.97 ± 0.0643
$A_{Features}$	63.12 ± 0.0368
A_{fm}	$\mathbf{65.73 \pm 0.0194}$

As is clear in Table 8, the performance of the feature-based score (A_{fm}) on imagery data is significant. This difference in the performance of the scores can be owing to the difference in the number of features of these two types of datasets. Given the fact that the number of features on tabular data is less than those of imagery data, the D_{xxzz} discriminator is able to extract and recognize abnormal samples properly. However, in the imagery data set, the output of the latent layer before the logit layer contains richer information to distinguish between normal and abnormal data. This improves the performance of the A_{fm} score.

6. Conclusion

The work presented in this paper introduces a new model for anomaly detection based on adversarial neural networks. In the proposed model, an encoder was used for the inverse mapping of the input data space, and the discriminator D_{xx} was employed to satisfy the condition of cycle consistency. In order to stabilize the training process of the adversarial generator network, discriminator D_{zz} was adopted in the adversarial structure of the model. To use the information of a complete cycle in the proposed model, the $\hat{z\hat{x}}$ variable was introduced. Consequently, the discriminator D_{xxzz} was employed to benefited from maximum information flow in the proposed model.

Moreover, supplementary distribution σ(x) was utilized to bias the network output toward the normal data distribution. The outcomes of the experiments demonstrated the effectiveness of the proposed model in the field of anomaly detection, as well as its superiority over other state-of-the-art models on tabular and imagery datasets. In spite of the impressive and brilliant results of the proposed RCALAD model, this model has a robustness problem when dealing with a variety of anomaly classes, like other GAN-based models. This issue could be improved by applying robustness methods that exist in [28, 29] this works.

References

[1] X. Shu, L. Cheng, and S. J. Stolfo, "Anomaly Detection as a Service."

[2] C. Jiang, J. Song, G. Liu, L. Zheng, and W. Luan, "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3637–3647, 2018, doi: 10.1109/JIOT.2018.2816007.

[3] X. Dai and M. Bikdash, "Distance-based outliers method for detecting disease outbreaks using social media," *Conf. Proc. - IEEE SOUTHEASTCON*, vol. 2016-July, 2016, doi: 10.1109/SECON.2016.7506752.

[4] H. Zenati, M. Romain, C. S. Foo, B. Lecouat, and V. Chandrasekhar, "Adversarially Learned Anomaly Detection," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, vol. 2018-Novem, pp. 727–736, 2018, doi: 10.1109/ICDM.2018.00088.

[5] I. Goodfellow *et al.*, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, Oct. 2014, pp. 2672–2680, doi: 10.1109/ICCVW.2019.00369.

[6] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.*, pp. 1–16, 2016.

[7] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, 2018, doi: 10.1109/MSP.2017.2765202.

[8] T. Schlegl, P. Seeb, S. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery," *Int. Confrence Inf. Process. Med. Imaging*, vol. 2, pp. 146–157, 2017, doi: 10.1007/978-3-319-59050-9.

[9] R. Kaur and S. Singh, "A survey of data mining and social network analysis based anomaly detection techniques," *Egypt. Informatics J.*, vol. 17, no. 2, pp. 199–216, 2016, doi: 10.1016/j.eij.2015.11.004.

[10] A. Zimek, E. Schubert, and H. Kriegel, "REVIEW A Survey on Unsupervised Outlier Detection in High-Dimensional Numerical Data," *Signal Processing*, vol. 99, pp. 215–249, 2012, doi: 10.1002/sam.

[11] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, 2014, doi: 10.1016/j.sigpro.2013.12.026.

[12] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Piatt, "Support vector method for novelty detection," *Adv. Neural Inf. Process. Syst.*, no. January, pp. 582–588, 2000.

[13] L. Ruff *et al.*, "Deep one-class classification," *35th Int. Conf. Mach. Learn. ICML 2018*, vol. 10, pp. 6981–6996, 2018.

[14] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," *6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc.*, pp. 1–15, 2018.

[15] I. Golan and R. El-Yaniv, "Deep anomaly detection using geometric transformations," *Adv. Neural Inf. Process. Syst.*, vol. 2018-Decem, no. NeurIPS, pp. 9758–9769, 2018.

[16] I. Golan and R. El-Yaniv, "Deep anomaly detection using geometric transformations," *Adv. Neural Inf. Process. Syst.*, vol. 2018-Decem, no. NeurIPS, pp. 9758–9769, 2018.

[17] Z. Yang, I. S. Bozchalooi, and E. Darve, "Regularized Cycle Consistent Generative Adversarial Network for Anomaly Detection."

[18] D. T. Nguyen, Z. Lou, M. Klar, and T. Brox, "Anomaly detection with multiple-hypotheses predictions," *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 8418–8432, 2019.

[19] S. Pidhorskyi, R. Almohsen, D. A. Adjeroh, and G. Doretto, "Generative probabilistic novelty detection with adversarial autoencoders," *Adv. Neural Inf. Process. Syst.*, vol. 2018-Decem, no. Nips, pp. 6822–6833, 2018.

[20] S. Zhai, Y. Cheng, W. Lu, and Z. Zhang, "Deep structured energy based models for anomaly detection," *33rd Int. Conf. Mach. Learn. ICML 2016*, vol. 3, pp. 1742–1751, 2016.

[21] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, 2019, doi: 10.1016/j.media.2019.01.010.

[22] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, “Efficient GAN-Based Anomaly Detection,” 2018, [Online]. Available: http://arxiv.org/abs/1802.06222.

[23] V. Dumoulin *et al.*, “Adversarially learned inference,” *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*, pp. 1–18, 2017.

[24] C. Li *et al.*, “ALICE : Towards Understanding Adversarial Learning for Joint Distribution Matching arXiv : 1709 . 01215v2 [stat . ML] 5 Nov 2017,” no. Nips, pp. 1–22, 2017.

[25] F. Tony Liu, K. Ming Ting, and Z.-H. Zhou, “Isolation Forest ICDM08,” *Icdm*, 2008, [Online]. Available: https://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/icdm08b.pdf%0Ahttps://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/icdm08b.pdf?q=isolation-forest.

[26] B. Zong *et al.*, “Deep autoencoding Gaussian mixture model for unsupervised anomaly detection,” *6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc.*, pp. 1–19, 2018.

[27] A. Makhzani and B. Frey, “Winner-take-all autoencoders,” *Adv. Neural Inf. Process. Syst.*, vol. 2015-Janua, pp. 2791–2799, 2015.

[28] T. Salimans, I. Goodfellow, V. Cheung, A. Radford, and X. Chen, “Improved Techniques for Training GANs,” *In Advances in Neural Information Processing Systems* pp. 2234–2242,2016.

[29] R. Chalapathy, A.K. Menon, and S. Chawla, ” Robust, deep and inductive anomaly detection. ” In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* , pp. 36-51, 2017.