Multimedia Tools and Applications (2023) 82:20197–20214
https://doi.org/10.1007/s11042-022-14306-9

Two-stage anomaly detection for positive samples and small samples based on generative adversarial networks

Caie Xu[1] · Dandan Ni[2] · Bingyan Wang[1] · Mingyang Wu[1] · Honghua Gan[2]

Received: 9 June 2022 / Revised: 14 October 2022 / Accepted: 10 December 2022 /
Published online: 9 January 2023

Abstract

Anomaly detection approaches based on generative adversary networks usually directly input the image into the generator for reconstruction. As a result, the results of anomaly detection are not ideal. This paper proposes a novel anomaly detection model based on a two-stage generative adversarial network to improve the results. It consists of feature extraction and anomaly detection networks. The former combines a convolutional neural network and multi-scale feature extraction to study latent code. The latent code from the former model instead of the original image is fed to the generator of the anomaly detection module. The experiment shows the proposed method outperforms several existing anomaly detection methods with multiple datasets. Additionally, the quantitative result indicates the proposed model optimizes anomaly detection performance and improves by 8.8% and 19.2% on both the liver CT image medical dataset and the CIFAR10 public dataset respectively compared to the baseline of the skip-GANomaly model.

Keywords Anomaly detection · Small samples · Positive samples · Generative adversarial networks

✉ Honghua Gan
hgan@zju.edu.cn

Caie Xu
caiexu@163.com

Dandan Ni
nidd@zju.edu.cn

Bingyan Wang
amz950411@163.com

Mingyang Wu
1200205120@zust.edu.cn

[1] School of Information and Electronic Engineering, Zhejiang University of Science and Technology, No.318 Liuhe Road, Hangzhou, 310023, Zhejiang, China

[2] College of Computer Science and Technology, Zhejiang University, No. 38, Zheda Road, Hangzhou, 310013, Zhejiang, China

1 Introduction

Artificial intelligence was applicated to different fields such as recommendation system [5, 34], face recognition [30], machine translation [3], medicine [6, 20, 21] and so on. It has brought not only huge economic benefits for various industries but also brought great changes to our lives. Artificial intelligence's gradual penetration into our daily life, its application in various fields and scenarios is also known and used by people.

The use of various AI techniques for anomaly detection is also important research in the field of AI. Anomaly detection is a technology to identify abnormal or illogical samples from a large number of samples. It has important applications in industrial quality inspection, video surveillance, fraud detection, medical diagnosis, and other fields. Some traditional anomaly detection methods [32], are often based on statistics [22], rules [28], nearest neighbors [7] or clustering [15] to judge the characteristic of data. However, in practice, a common feature of anomaly detection datasets is that the normal data occupies the vast majority, while the abnormal data is very small. The two are in a highly unbalanced state. In addition, there are many types of abnormal data. For example, for medical images, the shape of abnormal lesions may even change over time, so the type of abnormal data cannot be determined. In this situation, the above-mentioned traditional methods are generally ineffective in the task of anomaly detection.

Later, anomaly detection methods based on deep learning [8, 9, 12, 19] gradually emerged. In terms of feature extraction, compared with some traditional methods, deep learning methods can capture richer semantics and nonlinear relationships between features. [14, 29, 36] also proposed anomaly detection based on generative adversarial networks. They let the generative model only learn the distribution of normal data, and at test time distinguish between two types of data based on the size of the difference between the generated image and the original image. The premise of these models is that abnormal samples not only have different distributions in the image space than normal samples but also have large encoding differences in the low-dimensional latent space. This ensures that the reconstructed image retains the important features of the original image. Therefore, the process of mapping an image into a low-dimensional latent space becomes particularly important. The above methods usually take the image as the input of the generator and only obtain the low-dimensional latent encoding of the image through several layers of downsampling operations in the generator. The latent code obtained by this method may not fully capture the feature distribution of the image, thus affecting the quality of the reconstructed image.

This paper proposes a novel anomaly detection method. It first passes the image through a feature extraction module consisting of multi-scale convolutional streams and convolutional neural networks. Then, the acquired image features are input into the generative adversarial network with an Attention Gate, and the normal samples and abnormal samples are distinguished according to the difference between the generated image and the original image. Compared to directly feeding the image into the generative network, the image features are better able to capture the image details, which can lead to the better generation of accurate and high-quality images. Similar to [1], our method also learns the distribution of images and the latent space through the generator and discriminator. In the image feature extraction network, we also use convolution kernels of different scales to extract image features separately and then fuse them. This approach not only ensures low-dimensional local features but also ensures high-dimensional global features of the image. The main contributions of this research include the following:

- We propose a new two-stage anomaly detection model that adds an image feature extraction network to the anomaly detection GAN. The image features extracted by convolutional neural networks and multi-scale convolution kernels are input into the generator of anomaly detection GAN to generate reconstructed images.
- In the image feature extraction module, we combine convolutional neural networks and multi-scale convolution kernels to perform feature extraction and fusion on input images, in order to obtain better low-dimensional latent codes.
- In the anomaly detection GAN, we modified the ordinary UNet-based generator to a generator with Attention Gate to perform a more detailed reconstruction of the region of interest.
- This method achieves the best results on both the medical dataset we provide and the public dataset CIFAR10.

The overall framework of this paper is as follows: Section 1 is the Introduction, which briefly introduces the related applications of anomaly detection and our proposed method. Section 2 is Related Works, which mainly introduces the work related to anomaly detection. Section 3 is the Proposed Method, which provides a detailed introduction to our proposed method. Section 4 is Experiments, and Sections 5 and 6 are Discussion and Conclusion.

2 Related works

Anomaly detection is an important research part of machine learning at present, and it has corresponding practical applications in medicine, chip, video, and other fields. As more and more relevant people begin to study anomaly detection, the existing technical solutions in this field have also increased significantly. Traditional anomaly detection methods are often based on some basic algorithms. In [28], a RIPPER classification algorithm is proposed, which uses logical rules to describe some temporal states obtained after using the clustering algorithm. Nong et al. [22] propose an anomaly detection method based on chi-square statistics. Boriah et al. [7] investigated the performance of various similarity measures in the context of outlier detection. A FindCBLOF technique is proposed in [15] which assigns an anomaly score called Cluster-Based Local Outlier Factor to each data instance. This score is then used to capture the size of the cluster to which the data instance belongs and the distance of the instance from the cluster's center of gravity.

Later, with the continuous development of deep learning, researchers began to propose anomaly detection methods based on deep learning. Some methods use a knowledge distillation model for anomaly detection [4, 27]. They distinguish normal images from abnormal images by using the differences in image features learned by the teacher model and the student model. There are also some methods that use generative models for anomaly detection, such as [13, 19]. AutoEncoder maps the original data to the low-dimensional feature space through the encoder and then uses the decoder to restore the data from the low-dimensional space to the original space. The training goal is to restore the input data as much as possible. Various AutoEncoders have been proposed to learn more efficient feature representations. For example, [33] utilizes pre-defined contaminated data for reconstruction to enhance the anti-interference of the model, [26] enhances the robustness of the model by adding a penalty term to the activation function of the encoder, and [17] prevent overfitting of the model by adding a regularization term to the prior distribution of the samples. These improved AutoEncoders all improve the data reconstruction performance to a certain extent.

However, these methods cannot avoid the feature information about abnormal data that the model may learn, resulting in a bias in the final learned information.

Generative Adversarial Networks (GANs for short) [13] is an unsupervised deep learning framework for evaluating generative models through an adversarial process, which is mainly based on the competition of two networks in a zero-sum game framework. GAN consists of two parts of the network, called the generator and the discriminator. The network architecture diagram of GAN is shown in Fig. 1 below. Among them, the goal of the generator is to capture the data distribution, learn image features from the original data, and generate images that are as similar as possible to real images. The goal of the discriminator is to judge whether the sample comes from the dataset or the generator, which is usually regarded as a binary classifier. Through the optimization of the zero-sum game framework, both networks can enhance their predictive ability until a balance is reached.

Thomas et al. found that the Generative Adversarial Network architecture was able to generate normal data better than AutoEncoder. They proposed the [29] model, one of the earliest GAN-based anomaly detection models. After that, Zhao et al. in [36] mainly improved the discriminator. They treat the discriminator as a function of energy, assigning lower energy to regions near the data stream and relatively higher energy to other regions. This is used to distinguish whether the data is normal or abnormal. The use of GAN algorithms for anomaly detection on time series datasets was studied in [14], where Wasserstein GAN was used to learn the normal distribution of the data, and stacked encoders were used for anomaly detection. Akcay et al. [1, 2] are two anomaly detection models successively proposed by Samet et al. [1] is a convolutional neural network with the skip-connected encoder-decoder structure proposed on the basis of [2], which can better capture the spatial distribution of normal data in images. These models use only normal data to train the generator and discriminator during training, and use both normal and abnormal data during testing. Then use the bias between the reconstructed data and the original data to distinguish normal data from abnormal data, so as to realize anomaly detection. Among these GAN-based anomaly detection methods, the ways in which the output of the discriminator is used to distinguish the two types of data can also be roughly classified into two categories. One is used by models such as [31], which directly uses the discriminator to classify the input data, and the final output is the label. The other is used by models such as [2, 35], which propose the concept of anomaly score. They divide two types of data according to the size of the value.

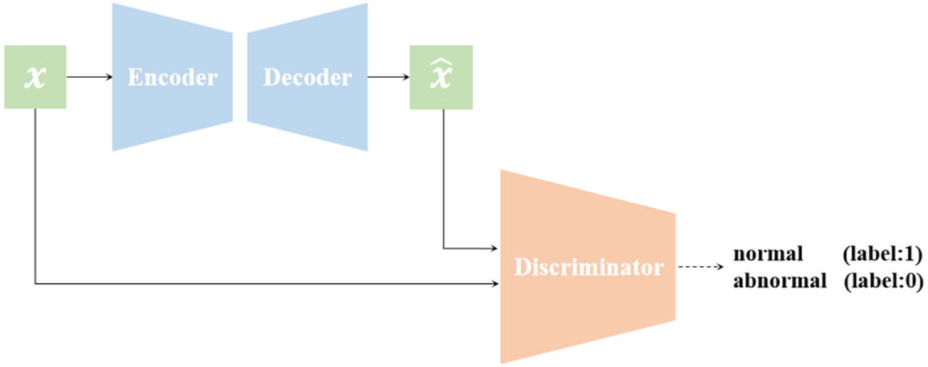


Fig. 1 The network architecture diagram of GAN

Some current methods to improve the performance of GAN networks usually focus on optimizing the generator. Oktay et al. [23] proposed the concept of Attention Gate(AG for short) on the basis of ordinary UNet, and each downsampling layer is directly connected to the corresponding upsampling layer through AG. Huang et al. [16] is based on the idea of full-scale skip connections. It adds the downsampling output of each layer of UNet to the corresponding upsampling and lowers dimensional layers. At the same time, each upsampling layer is also connected to all the upsampling layers above it to retain the feature information of each scale. Fu et al. [11] based on a dual attention mechanism, combining channel attention and spatial attention for image reconstruction. Li et al. [18] proposed to use convolution kernels of different scales to extract image features to ensure more accurate low-dimensional feature information. Song et al. [31] proposed a two-stage method for anomaly detection using an attention map and data hard augmentation. In this way, an attention map is used to concentrate the reconstruction range in the useful area, and useless areas such as background are removed to prevent unrelated regions have an impact on anomaly detection.

3 Proposed method

The proposed network mainly consists of two stages: the image feature extraction stage to obtain multi-scale image features, and the anomaly detection GAN stage to learn the feature distribution to achieve anomaly detection. The overall implementation details are explained in the following sections.

3.1 The overall proposed model

The overall network architecture and the detail table of the proposed model are shown in Figs. 2 and 3. The image feature extraction stage consists of two convolutional neural networks and parallel convolutional streams at three scales. For the input image x, it first obtains the initially extracted image feature F_{first} through a convolutional neural network. Then, through three parallel convolution streams with convolution kernels of 1x1, 3x3, and 5x5, the features of different scales are fused in pairs. After a 1x1 convolution kernel, the quantity is adjusted to obtain further image features. Finally, the final image feature F_{final} is obtained through a convolutional neural network again.

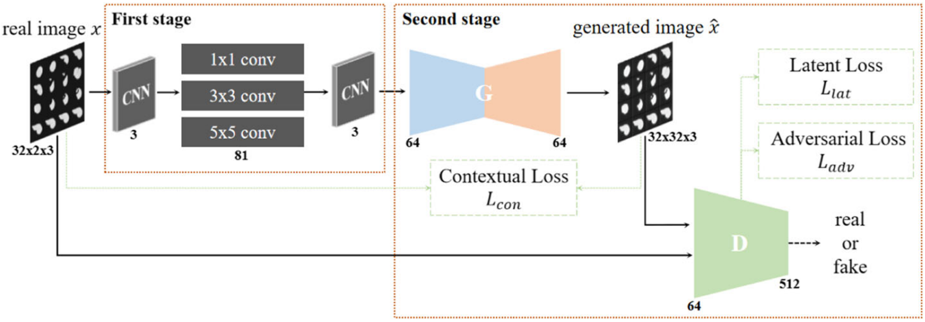
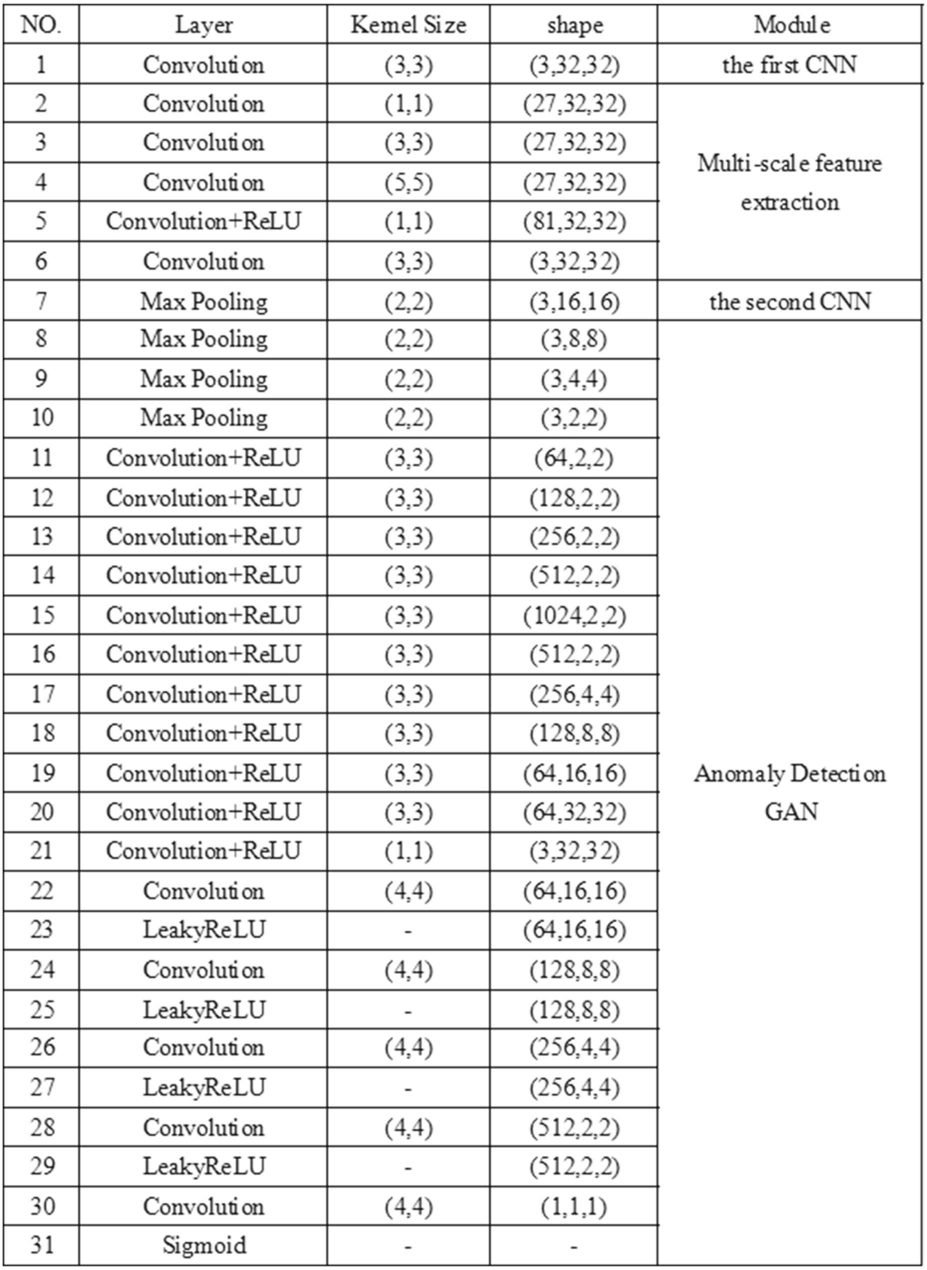


Fig. 2 The overall network architecture of the proposed model. The input image is the original image from the dataset. Then it goes through the first stage of multi-scale feature extraction. Then input the obtained features into G to get the corresponding generated image. Finally, the original image and the generated image are input into D to judge real or fake



NO.	Layer	Kemel Size	shape	Module
1	Convolution	(3,3)	(3,32,32)	the first CNN
2	Convolution	(1,1)	(27,32,32)	Multi-scale feature extraction
3	Convolution	(3,3)	(27,32,32)	
4	Convolution	(5,5)	(27,32,32)	
5	Convolution+ReLU	(1,1)	(81,32,32)	
6	Convolution	(3,3)	(3,32,32)	
7	Max Pooling	(2,2)	(3,16,16)	the second CNN
8	Max Pooling	(2,2)	(3,8,8)	Anomaly Detection GAN
9	Max Pooling	(2,2)	(3,4,4)	
10	Max Pooling	(2,2)	(3,2,2)	
11	Convolution+ReLU	(3,3)	(64,2,2)	
12	Convolution+ReLU	(3,3)	(128,2,2)	
13	Convolution+ReLU	(3,3)	(256,2,2)	
14	Convolution+ReLU	(3,3)	(512,2,2)	
15	Convolution+ReLU	(3,3)	(1024,2,2)	
16	Convolution+ReLU	(3,3)	(512,2,2)	
17	Convolution+ReLU	(3,3)	(256,4,4)	
18	Convolution+ReLU	(3,3)	(128,8,8)	
19	Convolution+ReLU	(3,3)	(64,16,16)	
20	Convolution+ReLU	(3,3)	(64,32,32)	
21	Convolution+ReLU	(1,1)	(3,32,32)	
22	Convolution	(4,4)	(64,16,16)	
23	LeakyReLU	-	(64,16,16)	
24	Convolution	(4,4)	(128,8,8)	
25	LeakyReLU	-	(128,8,8)	
26	Convolution	(4,4)	(256,4,4)	
27	LeakyReLU	-	(256,4,4)	
28	Convolution	(4,4)	(512,2,2)	
29	LeakyReLU	-	(512,2,2)	
30	Convolution	(4,4)	(1,1,1)	
31	Sigmoid	-	-	

Fig. 3 The detail table of the proposed model

The anomaly detection GAN part mainly includes a generator and a discriminator. The main function of the generator is to generate the reconstructed image x_{re} according to the image feature F_{final} obtained in the previous stage. Among them, the generator part also adds the Attention Gate module, which focuses on image reconstruction of the region of interest. The role of the discriminator can be divided into the training phase and the testing phase. The role of the discriminator in the training phase is to distinguish whether the input image is from the generator or the original dataset, and the role of the discriminator in the testing phase is to output an anomaly score used to distinguish normal and abnormal data(used in [1]).

3.1.1 Feature extraction module

3.1.2 Multi-scale feature fusion

Taking image features at different scales and fusing them [18] has been shown to yield better-reconstructed images. The overall architecture diagram of the feature extraction module is shown in Fig. 3. In our model, three-scale convolution kernels of 1x1, 3x3, and 5x5 are selected for multi-scale feature fusion. This method can help to learn global features as well as local features. In addition, we also use the residual structure to integrate the final extracted features and the preliminary extracted features to prevent the loss of features as much as possible.

3.1.3 Convolutional neural network

There is a convolutional neural network before and after the multi-scale feature fusion module, as shown in Fig. 4, both of which have the same structure and are also used to extract image features. The main function of the first convolutional neural network is to perform preliminary feature extraction on the input image and convert high-dimensional data into low-dimensional encoding for subsequent further convolutional feature extraction. The main function of the second convolutional neural network is to extract the image features after multi-scale feature fusion again so that the extracted overall features are more complete and stable. This convolutional neural network structure mainly includes three layers, namely Conv layer, InstanceNormalization layer, and ReLU layer.

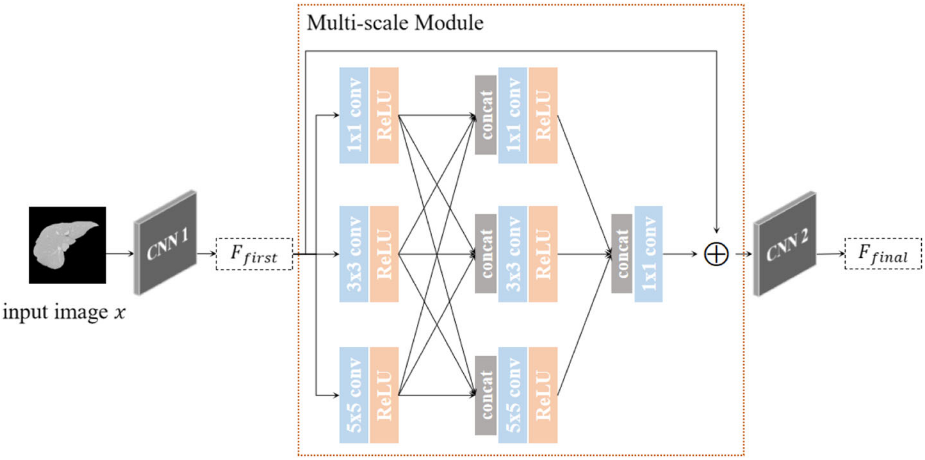


Fig. 4 The overall architecture diagram of the first stage. Among them, the part framed by the dotted line is the Multi-scale module. During this stage, we get the final features from the input image

3.2 Anomaly detection GAN

Unlike the previous GAN-based anomaly detection methods, we do not use the entire image as the input of the generator, but the image features obtained by the input image after passing through the above feature extraction module as the input of the generator of this anomaly detection GAN. The input image features are passed through a series of down-sampling layers and then connected with the corresponding up-sampling layers through the Attention Gate structure to generate the final reconstructed image. Through the AG structure, the reconstruction of the region of interest can be focused on in the process of reconstructing the image, so that the key region of the reconstructed image is closer to the corresponding region of the original image. The discriminator part adopts the discriminator structure of DCGAN [25]. The network architecture diagram of the anomaly detection GAN module is shown in Fig. 5.

In the training phase, since only normal data is used for training, the generator needs to learn the feature distribution of normal data to generate reconstructed images that are as similar as possible to the original images. The discriminator needs to learn to distinguish the original image from the reconstructed image. In the testing phase, the generator obtains the feature distribution of normal data and abnormal data to generate corresponding reconstructed images. The discriminator computes an anomaly score based on the context loss and latent loss between the input original image and the reconstructed image. Since the training phase, the generator only learns the feature distribution of normal images. Therefore, when an abnormal image is an input, the bias between the reconstructed image generated by the generator and the original image is large, that is, the corresponding anomaly score value is also large. Finally, it is judged whether the input image is normal or abnormal according to the anomaly score output by the discriminator.

3.3 Loss function

As mentioned in the previous section, only normal images are used during training, while both normal and abnormal images are used during testing. Our training goal is to expect the model to be as accurate as possible for the reconstruction of normal images in both

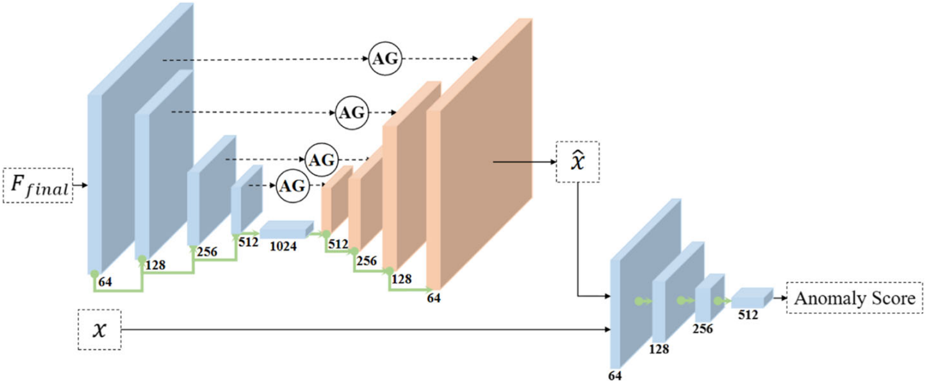


Fig. 5 The network architecture diagram of the anomaly detection GAN module,i.e. the second stage. In this stage, the features obtained in the previous stage are reconstructed into images by G with Attention Gate. Then the original image and the reconstructed image are input into D together to get the anomaly score

the image space and the low-dimensional latent space. During testing, we believe that the generator only learns the feature distribution of normal images during training. Therefore, when an abnormal image is an input, the reconstruction effect of the generator is relatively poor, resulting in a large difference between the generated image and the original image. Based on these purposes, we propose the following three loss functions to achieve.

1. Adversarial Loss

   To maximize the reconstruction power of the generator for normal images as possible, we use the adversarial loss mentioned in [1]. The calculation formula for this loss is shown in (1). It means that the generator G wants to make the difference between the original image x and the generated image $\hat{x}$ as small as possible. At the same time, the discriminator D wants to make the difference between x and $\hat{x}$ as large as possible.

$$\mathcal{L}_{adv} = \underset{x \sim p_x}{\mathbb{E}} [\log D(x)] + \underset{x \sim p_x}{\mathbb{E}} [\log(1 - D(\hat{x}))] \tag{1}$$

2. Contextual Loss

   In order to make the bias between the reconstructed image of the normal image and the original image as small as possible, we use contextual loss. The formula for calculating this loss is shown in (2). It represents the L1 loss between the input image x and the reconstructed image $\hat{x}$, ensuring that the model can make the original image and the generated image as similar as possible.

$$\mathcal{L}_{con} = \underset{x \sim p_x}{\mathbb{E}} \left\| x - \hat{x} \right\|_1 \tag{2}$$

3. Latent Loss

   On top of the above two loss functions, a latent loss is also used to make x and $\hat{x}$ as similar as possible. The calculation formula for this loss is shown in (3). It represents the latent representation constructed from the image features extracted from the last convolutional layer of the discriminator.

$$\mathcal{L}_{lat} = \underset{x \sim p_x}{\mathbb{E}} \left\| f(x) - f(\hat{x}) \right\|_2 \tag{3}$$

The overall loss function of the model is the weighted sum of the three losses above, as shown in (4).

$$\mathcal{L}_{con} = w_{adv}\mathcal{L}_{adv} + w_{con}\mathcal{L}_{con} + w_{lat}\mathcal{L}_{lat} \tag{4}$$

3.4 Anomaly score

In order to distinguish normal data from abnormal data, we adopt the metric of anomaly score, which was also used in [1]. For any test image x, its anomaly score calculation formula is as follows:

$$A(x) = \lambda\mathcal{L}_{con}(x) + (1 - \lambda)\mathcal{L}_{lat}(x) \tag{5}$$

Among them, $\mathcal{L}_{con}()$ represents the reconstruction score between the input image and the generated image in (2) above. $\mathcal{L}_{lat}()$ is the latent representation score between the input image and the generated image in (3) above. λ is a weight coefficient that controls the importance of the two functions. During the experiment, we mainly adjusted the parameters several times to ensure that this anomaly score distinguishes normal data from abnormal data as accurately as possible.

For each input test image x, the discriminator outputs an anomaly score based on the original and generated images. The images of the entire test set D_{test} can be corresponding to a set of anomaly score vectors $A = A_i : A(x_i), x_i \in D_{test}$. In order to visually distinguish the two types of data, we normalize these scores, using the following formula, so that the range of all values becomes [0,1]. For normal class images, since the difference between the original image and the generated image is small. Therefore, the corresponding anomaly score value is also small, that is, the final normalized value of these samples will approach 0. Conversely, for abnormal images, the difference between the original image and the generated image is large because the generator does not learn its feature distribution. Therefore, the corresponding anomaly score values are also larger, so the final normalized value of these samples will approach 1. In this way, we set the threshold to 0.5. Those smaller than this threshold are considered as normal images, otherwise, they are considered as abnormal images.

$$A'(x) = \frac{A(x) - \min(A)}{\max(A) - \min(A)} \tag{6}$$

4 Experiments

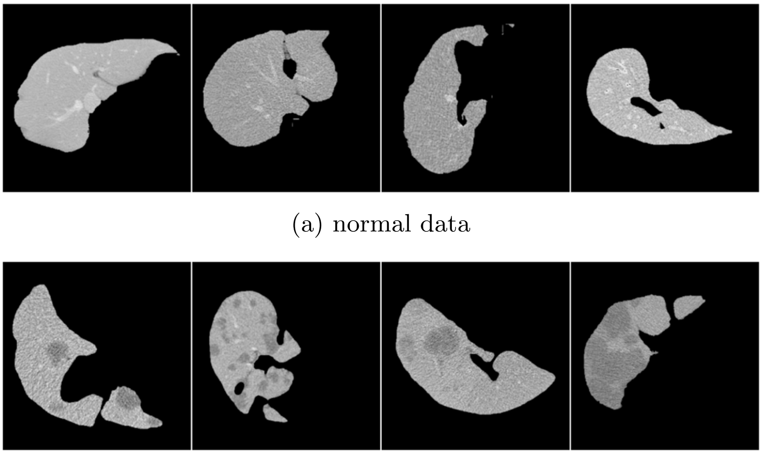
4.1 Dataset and evaluation metrics

Datasets: To verify the performance of the proposed model, we perform validation on a medical dataset and a public computer vision dataset. **Medical dataset:** This is a CT image of the liver of several patients provided by the Department of Hepatobiliary and Pancreatic Surgery of the Second Affiliated Hospital of Zhejiang University School of Medicine. The total amount of data is about 5k. The image content is mainly the partial CT image of the patient's liver. Among them, the color of the normal area is light gray, and the lesion area is displayed as dark gray in the CT image. The used liver CT image data is shown in Fig. 6 below, where Fig. 6(a) represents a normal liver CT image, and Fig. 6(b) represents a cancerous liver CT image. In the experiment, we take normal CT images as normal data and lesion CT images as abnormal data. In the data preprocessing stage, we performed a 4-fold crossover process on the dataset. In addition, due to the small amount of data, when loading data, we also use a relatively basic random rotation [-5,5] degrees for data augmentation. **CIFAR10 Dataset:** It is a dataset containing ten categories of natural images such as automobiles, airplanes, dogs, birds, etc. It contains 50k training images and 10k test images. In the experiments, we used one of the categories as abnormal data and the other nine categories as normal data for a total of ten experiments. Data will be made available on reasonable request.

Evaluation metrics: The performance of the model is evaluated by the area under the Receiver Operating Characteristic (ROC) curve. ROC is a curve with the true positive rate (TPR) as the vertical axis and the false positive rate (FPR) as the horizontal axis. Using the metric AUC, the predictive performance of the model can be judged (also used in previous work [2, 29]).

4.2 Implementation details

For the model loss L represented by (4) in 3.4, we use the Adam optimizer to optimize it. The initial learning rate lr is set to 1e-3. For the three weight parameters in L, we set them as $L_{adv} = 1$, $L_{con} = 50$, $L_{lat} = 1$. The number of training times for the model was initially set to 40 epochs(on the medical dataset) and 15 epochs(on the CIFAR10 dataset). In practice, during the training process, we found that the model has learned enough information and



(a) normal data

(b) abnormal data

Fig. 6 The example of liver CT image dataset

showed better prediction performance fewer times. Therefore, we save the models trained in each epoch, test each saved model in the testing phase, and save the network parameters of the best model. The overall model is implemented using PyTorch (v1.8.1, Python3.7, CUDA11.1, cuDNN8.0.5). Experiments were performed on NVIDIA GeForce RTX 2080 Ti.

In the following sections, we will mainly introduce two categories of experiments we conduct on the dataset, namely ablation experiments and comparative experiments. Ablation experiments are validated on both the liver CT image medical dataset and the CIFAR10 public dataset, respectively. The purpose of ablation experiments is to demonstrate the effectiveness of the key components of the proposed model. At first, we conduct experiments on the skip-GANomaly[1] model and use it as the baseline. Then we retrain the proposed model with the same network parameters for each ablation experiment by removing or adding different key components of the proposed model. Comparative experiments are also conducted on the above two datasets. The purpose is to verify that our proposed model performs better than some existing models on different types of datasets.

4.3 Ablation experiment

To investigate the effectiveness of each component in the proposed method, we conduct ablation experiments. We choose skip-GANomaly [1] as the baseline model, and the experiments are performed on the 4-fold cross-validated liver CT image medical dataset and the public dataset CIFAR10, respectively. The average AUC results based on the medical dataset are shown in Table 1 below, and the average AUC results based on the CIFAR10 dataset are shown in Table 2 below.

The **Baseline** experiments are performed without changing any skip-GANomaly [1] model structure. In the experiment **Stage 2 with AG**, an Attention Gate is added to the structure of the generator. The purpose of this experiment is to verify that the Attention Gate

Table 1 AUC results of ablation experiments on medical datasets

Model	+Attention gate	+CNN	+Multi-scale	AUC
Baseline [1]				0.5652
Stage 2 with AG	✓			0.6160
Stage 1 with CNN		✓		0.5612
Stage 1 with multi-scale			✓	0.5705
Proposed model	✓	✓	✓	0.6534

can focus attention on the region of interest in the image. In the reconstruction process, the influence of irrelevant regions such as background on the reconstructed image is minimized, so that the reconstructed image can retain as much important feature information of useful regions as possible. The experimental results show that after adding the Attention Gate, the AUC metric of the model is improved by 5.1% and 19.1% on the two datasets, respectively. The experiment **Stage 1 with CNN** is to replace the image that was originally input directly into the generator with the image features extracted by passing the image through a convolutional neural network. The purpose of this experiment is to verify that using image features instead of images to input into the generator can get better-reconstructed images. Experimental results show that replacing images with image features as input into the generator improves the model's AUC by 2.1% on the CIFAR10 dataset, but decreases by 0.4% on the medical image dataset. We speculate that it may be due to the fact that ordinary convolution extraction cannot obtain enough feature information for finer images such as medical images, resulting in a slight decrease in the prediction accuracy. The experiment **Stage 2 with multi-scale** is to replace the image with the image features obtained after multi-scale feature extraction. This experiment is also to verify that image features can reconstruct better-generated images than images, and that image features extracted by multi-scale convolution kernels can better extract global and local features. Experimental results show that the metrics of the model are improved by 0.5% and 5.7% on the two datasets, respectively. The experiment **Proposed model** is based on the model proposed in this paper. According to the quantitative results in Tables 1 and 2, it can be proved that our model has a significant improvement in anomaly detection performance compared to skip-GANomaly [1], and improved by 8.8% and 19.2% on the two datasets, respectively. In addition, from the perspective of the performance improvement effects of the three contributions, adding an Attention Gate to the generator part has the most obvious effect on the performance improvement of the model.

Table 2 AUC results of ablation experiments on CIFAR10 dataset

Model	+Attention gate	+CNN	+Multi-scale	AUC
Baseline [1]				0.731
Stage 2 with AG	✓			0.922
Stage 1 with CNN		✓		0.752
Stage 1 with multi-scale			✓	0.788
Proposed model	✓	✓	✓	0.923

4.4 Comparative experiment

In addition, we also conduct comparative experiments on these two different datasets. In order to ensure as much as possible that apart from the model itself, no other objective factors will affect the experimental results. When we conduct experiments on each type of data set, we ensure that the values of relevant parameters remain unchanged.

4.4.1 Experimental results on liver CT dataset

We still use AUC as the evaluation metric for model performance. During training, only normal liver CT images were used; and during testing, normal liver CT images and cancer CT images were used. Based on the 4-fold crossover dataset, we train and validate the model four times, and use the average AUC value of the four times as the final experimental result. We selected several models [1, 2, 27] to compare with our proposed model. The total AUC result of the models is shown in Fig. 7 below.

Figure 8 is a comparison of the reconstruction effect of skip-GANomaly [1] and the generator of the proposed model for normal images. We randomly choose 8 corresponding liver CT images from the test result images for display. Among them, the first row shows the original liver CT image in the dataset, the second row shows the image reconstructed by the generator in skip-GANomaly [1], and the last row shows the image reconstructed by the generator in our proposed model. For the three small images in a column, the following two

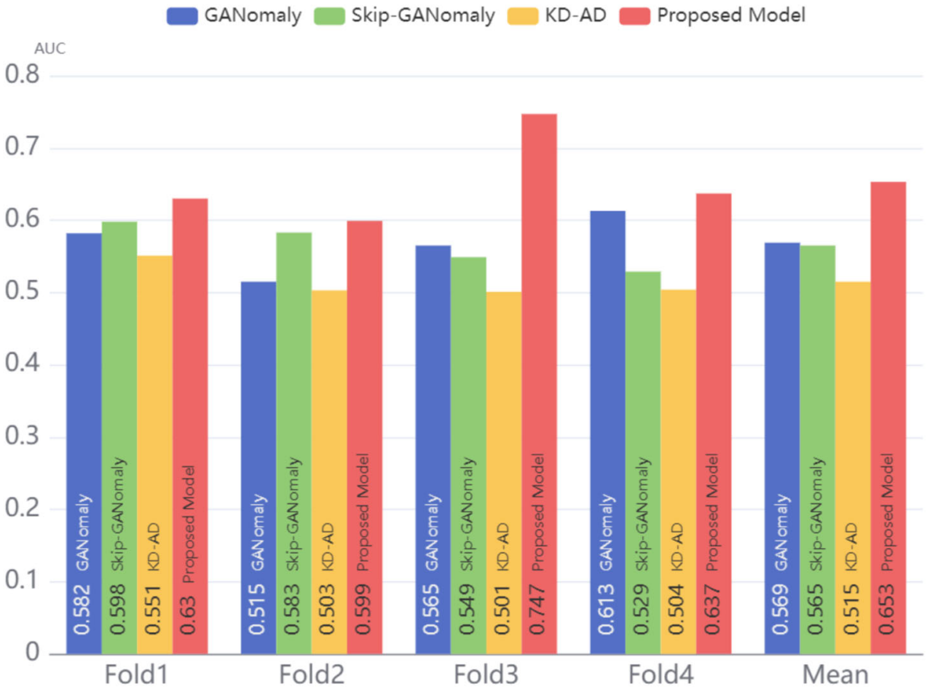


Fig. 7 The total AUC result of the models. The abscissa represents the 4-fold dataset and the average result, respectively, and the ordinate represents the value of AUC. Among them, the blue axis represents GANomaly, the green axis represents skip-GANomaly, the yellow axis represents the KD_AD model, and the red axis represents our proposed model

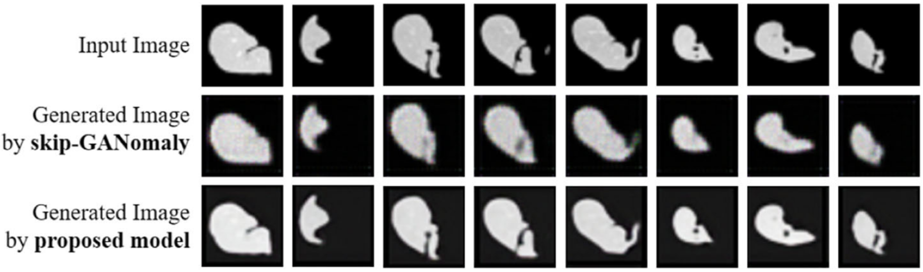

Input Image
Generated Image by **skip-GANomaly**
Generated Image by **proposed model**

Fig. 8 Comparison of the generator reconstruction performance. The first row is the original image, the second row is the generated image by the generator of skip-GANomaly, and the third row is the generated image by the generator of our proposed method. It can be found from the figure that the image reconstructed by the proposed method is more similar to the original image, and the gap between the two is smaller. However, the image reconstructed by skip-GANomaly is relatively blurry, and the reconstruction effect for small areas is poor

images are the reconstruction results of the first image by the generators of the two models. By comparison, it is obvious that the proposed model reconstructs images better than skip-GANomaly [1] for normal CT images. The stronger the reconstruction performance of the generator for normal images, the smaller the value of the anomaly score obtained from the original image and the generated image, so that the normal and abnormal images can be better distinguished.

4.4.2 Experimental results on the CIFAR10 dataset

In order to make the experimental results more objective and persuasive, in addition to the experiments on the medical image dataset, we also conduct related experiments on the public dataset CIFAR10. In each experiment, one type of data in the data set is regarded as abnormal data, and the rest of the data are regarded as normal data, so we conduct a total of ten model training and validation. Table 3 below shows the quantitative comparison results of our proposed model and some existing anomaly detection models under the AUC indicator. As can be seen from the table, the AUC values obtained by our proposed model are significantly improved compared to the existing models.

In addition to using AUC as an evaluation metric, we also plot the distribution of anomaly scores using the anomaly scores output by the discriminator. As shown in Fig. 9 below. The

Table 3 The AUC results on the CIFAR10 dataset

Model	Plane	Car	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck	Mean
AnoGAN [29]	0.671	0.547	0.529	0.545	0.651	0.603	0.585	0.625	0.758	0.665	0.618
OCGAN [24]	0.757	0.531	0.640	0.620	0.723	0.620	0.723	0.575	0.820	0.554	0.656
GANomaly [2]	0.633	0.631	0.510	0.587	0.593	0.628	0.683	0.605	0.616	0.617	0.610
SkipGANomaly [1]	0.797	0.953	0.448	0.607	0.602	0.615	0.931	0.788	0.659	0.907	0.731
KD-AD [27]	0.905	0.905	0.797	0.772	0.867	0.914	0.890	0.868	0.915	0.889	0.872
IGD [10]	0.906	0.979	0.839	0.823	0.886	0.899	0.909	0.964	0.969	0.948	0.912
Proposed	0.999	0.936	0.922	0.853	0.986	0.767	0.999	0.813	0.994	0.956	0.923

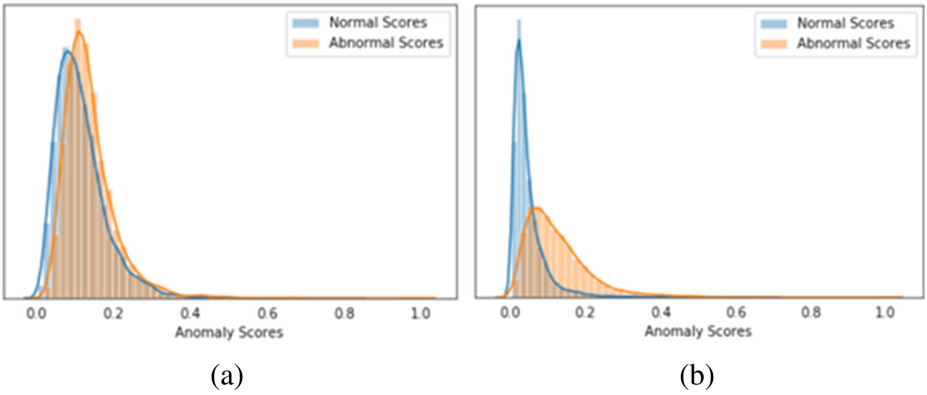


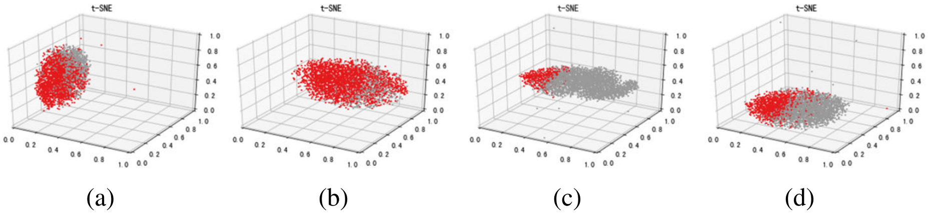
Fig. 9 Comparison of abnormal score distributions between Skip-GANomaly and the proposed model when [abnormal class=cat]

Fig. 9(a) is the result of the anomaly score distribution of skip-GANomaly [1] when [abnormal class=cat]. It can be seen that the overlapping area of the blue part and the orange part is large, which means that the discriminator has a poor distinguishing effect on normal and abnormal data. Figure 9(b) is the result of the anomaly score distribution of the proposed model when [abnormal class=cat]. The overlapping parts of blue and orange are significantly less, and the blue area(ie, normal data) is mainly concentrated in the part close to 0, indicating that the discriminator can distinguish the two types of data better. The purpose of the discriminator is to distinguish the two types of data according to the size of the output anomaly score, that is, the less the overlap between the two types of data, the better the discriminator's discriminative effect. As can be seen from the figure, the overlapping part of the figure on the right is significantly smaller than the figure on the left.

In order to further visually demonstrate the discriminator's ability to distinguish between normal data and abnormal data, we extracted the features and labels of the last convolutional layer in the discriminator D, and used a three-dimensional t-SNE graph to show their distribution results, as shown in Fig. 10 below. Among them, Fig. 10(a) and (b) represent the distribution results of the skip-GANomaly [1] discriminator for two types of data in the original data and the generated data, respectively. Figure 10(c) and (d) represent the distribution results of the discriminator in our proposed method for two classes of data in the original and generated data, respectively. The red dots in the figure represent the distribution of normal data, and the gray dots represent the distribution of abnormal data. It can be seen from the results in the figure that in the proposed method, the red points (ie, normal data) are gathered at positions close to 0, while the gray points (ie, abnormal data) are gathered at positions close to 1. Compared with the results in Fig. 10(a) and (b), the distribution positions of the two types of points in Fig. 10(c) and (d) are more clearly distinguished. Therefore, it can be shown that the proposed method is more effective in distinguishing normal and abnormal data.

5 Discussion

Some anomaly detection methods proposed at present, such as [1, 2], basically input the original image directly into the generator. The encoder in the generator is used to extract



(a) (b) (c) (d)

Fig. 10 3D t-SNE results of normal and abnormal features extracted from the last convolutional layer of the discriminator. The small Fig. 10(a) and (b) are the distribution results of skip-GANomaly. It can be seen that the two types of data represented by the red and gray points are still mixed together chaotically. This shows that the network does not distinguish the two types of data well. The small Fig. 10(c) and (d) are the distribution results of our proposed model. It can be seen that the red and gray points are clustered together to form two distinct parts. In addition to this, the red dots representing the normal data are clustered near 0

the features of the image, and the decoder is used to transform the latent expression into image-level data. The image features extracted by this method are likely to lack feature information of different scales, so the reconstructed image cannot be generated accurately. In [1], the skip-connection structure is used to directly connect the output of different layers of the encoder with the corresponding decoder layer, which to a certain extent integrates the feature information of different scales. However, compared with our proposed method, its prediction accuracy is still lower than ours.

6 Conclusion

This paper proposes a novel image anomaly detection method that includes two stages image feature extraction and anomaly detection. In the former stage, a convolutional neural network and three parallel multi-scale convolutional streams are adopted to extract image features from the original image. In the anomaly detection stage, the generator reconstructs the corresponding generated image given a latent code from the first stage, and then the discriminator judges whether the image is normal or abnormal. The experimental results show that the proposed architecture is significantly better than the baseline anomaly detection methods on both the liver CT image dataset and the CIFAR10 dataset.

Declarations

We confirm that we have read, understand, and agreed to the submission guidelines, policies, and submission declaration of the journal, and all authors of the manuscript have no conflict of interest to declare. The manuscript is the author's original work, and the manuscript has not received prior publication and is not under consideration for publication elsewhere. On behalf of all Co-Authors, I bear full responsibility for the submission.

References

1. Akcay S, Atapour-Abarghouei A, Breckon TP (2019) Skip-ganomaly: skip connected and adversarially trained encoder-decoder anomaly detection IEEE. https://doi.org/10.1109/IJCNN.2019.8851808

2. Akcay S, Atapour-Abarghouei A, Breckon TP (2018) Ganomaly: semi-supervised anomaly detection via adversarial training. https://doi.org/10.1007/978-3-030-20893_39
3. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. Computer Science
4. Bergmann P, Sdea Fauser M (2020) Uninformed students: student-teacher anomaly detection with discriminative latent embeddings. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
5. Bhatti UA, Huang M, Wu D, Zhang Y, Mehmood A, Han H (2018) Recommendation system using feature extraction and pattern recognition in clinical care systems. Enterprise Information Systems, pp 1–23
6. Bhatti UA, Yuan L, Yu Z, Nawaz SA, Xiao S (2021) Predictive data modeling using sp-knn for risk factor evaluation in urban demographical healthcare data. Journal of Medical Imaging and Health Informatics 11(1):7–14
7. Boriah S, Chandola V, Kumar V (2008) Similarity measures for categorical data: a comparative evaluation. In: Proceedings of the SIAM International Conference on Data Mining, SDM 2008, April 24-26, Atlanta, Georgia, USA
8. Chalapathy R, Menon AK, Chawla S (2018) Anomaly detection using one-class neural networks. https://doi.org/10.48550/arXiv.1802.06360
9. Chalapathy R, Chawla S (2019) Deep learning for anomaly detection: a survey. https://doi.org/10.48550/arXiv.1901.03407
10. Chen Y, Tian Y, Pang G, Carneiro G (2021) Deep one-class classification via interpolated gaussian descriptor. https://doi.org/10.48550/arXiv.2101.10043
11. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H (2018) Dual attention network for scene segmentation. https://doi.org/10.48550/arXiv.1809.02983
12. Geert L, Thijs K, Babak E, Bejnordi A, Arindra A (2017) A survey on deep learning in medical image analysis medical image analysis
13. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial networks
14. Haloui I, Gupta JS, Feuillard V (2018) Anomaly detection with wasserstein gan. https://doi.org/10.48550/arXiv.1812.02463
15. He Z, Xu X, Deng S (2003) Discovering cluster-based local outliers. Pattern Recogn Lett 24(9-10):1641–1650. https://doi.org/10.1016/S0167-8655(03)00003-5
16. Huang H, Lin L, Tong R, Hu H, Wu J (2020) Unet 3+: a full-scale connected unet for medical image segmentation IEEE. https://doi.org/10.1109/ICASSP40776.2020.9053405
17. Kingma DP, Welling M (2014) Auto-encoding variational bayes, arXiv.org. https://doi.org/10.48550/arXiv.1312.6114
18. Li J, Fang F, Mei K, Zhang G (2018) Multi-scale residual network for image super-resolution. In: 15th european conference, munich, germany, september 8-14, 2018, proceedings, part viii, Springer, Cham
19. Michelucci U (2022) An introduction to autoencoders. https://doi.org/10.48550/arXiv.2201.03898
20. Milletari F, Navab N, Ahmadi SA (2016) V-net fully convolutional neural networks for volumetric medical image segmentation
21. Nawaz SA, Li J, Bhatti UA, Bazai SU, Zafar A, Bhatti MA, Mehmood A, Ain QU, Shoukat MU (2021) A hybrid approach to forecast the covid-19 epidemic trend. PLOS ONE vol 16
22. Nong Y, Qiang C (2001) An anomaly detection technique based on a chi-square statistic for detecting intrusions into information systems
23. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, Mcdonagh S, Hammerla NY, Kainz B (2018) Attention u-net: learning where to look for the pancreas. https://doi.org/10.48550/arXiv.1804.03999
24. Perera P, Nallapati R, Bing X (2019) Ocgan: one-class novelty detection using gans with constrained latent representations IEEE
25. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. Computer ence
26. Rifai S, Vincent P, Muller X, Glorot X, Bengio Y (2013) Contracting auto-encoders
27. Salehi M, Sadjadi N, Baselizadeh S, Rohban MH, Rabiee HR (2020) Multiresolution knowledge distillation for anomaly detection. https://doi.org/10.48550/arXiv.2011.11108
28. Salvador S, Chan PK, Brodie J (2003) Learning states and rules for time series anomaly detection. Seventeenth International Florida Artificial Intelligence Research Society Conference
29. Schlegl T, Seebck P, Waldstein SM, Schmidt-Erfurth U, Langs G (2017) Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. Springer Cham. https://doi.org/10.1007/978-3-319-59050-9_12

30. Schroff F, Kalenichenko D, Philbin J (2015) Facenet: a unified embedding for face recognition and clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
31. Song JW, Kong K, Park YI, Kang SJ (2021) Attention map-guided two-stage anomaly detection using hard augmentation. https://doi.org/10.48550/arXiv.2103.16851
32. Varun C, Arindam B, Vipin K (2009) Anomaly detection: a survey. Acm Computing Surveys. https://doi.org/10.48550/arXiv.1901.03407
33. Vincent P, Larochelle H, Bengio Y, Manzagol PA (2008) Extracting and composing robust features with denoising autoencoders. https://doi.org/10.1145/1390156.1390294
34. Zeeshan Z, Ain QU, Bhatti UA, Memon WH, Ali S, Nawaz SA, Nizamani MM, Mehmood A, Bhatti MA, Shoukat MU (2021) Feature-based multi-criteria recommendation system using a weighted approach with ranking correlation Intelligent data analysis pp 25-4
35. Zenati H, Foo CS, Lecouat B, Manek G, Chandrasekhar VR (2018) Efficient gan-based anomaly detection Computer Vision and Pattern Recognition. https://doi.org/10.48550/arXiv.1802.06222
36. Zhao J, Mathieu M, Lecun Y (2016) Energy-based generative adversarial network. https://doi.org/10.48550/ arXiv:1609.03126

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.