Consistency Ensured Bi-directional GAN for Anomaly Detection

Kyosuke Komoto, Hiroaki Aizawa and Kunihito Kato

February 5, 2020

Consistency Ensured Bi-directional GAN for Anomaly Detection

Kyosuke Komoto[1], Hiroaki Aizawa[1] and Kunihito Kato[1]

[1] Gifu University, 1-1, Yanagido, Gifu, 501-1193, Japan
komoto@cv.info.gifu-u.ac.jp

Abstract. Anomaly detection is a challenging and fundamental issue in computer vision tasks. In recent years, GAN (Generative Adversarial Networks) based anomaly detection methods have achieved remarkable results. But the instability of training of GAN could be considered that decreases the anomaly detection score. In particular, Bi-directional GAN has the following two causes that make the training difficult: the lack of consistency of the mutual mapping between the image space and the latent space, and the difficulty in conditioning by the latent variables of the image. Here we propose a novel GAN-based anomaly detection model. In our model, we introduce the consistency loss for ensuring mutual mappings. Further, we propose introducing the projection discriminator as an alternative of concatenating discriminator in order to perform efficient conditioning in the Bi-directional GAN model. In experiments, we evaluate the effectiveness of our model in a simple dataset and real-world setting dataset and confirmed that our model outperforms the conventional anomaly detection methods.

Keywords: Anomaly Detection, Generative Adversarial Networks, Projection Discriminator.

1 Introduction

Anomaly detection is one of the most important issues in many situations and hence has been studied in a broad range of fields including industry, fraud detection and medical applications [1]. However, in actual anomaly detection situations, adequate anomalous data samples are often difficult to obtain since they rarely tend to appear as compared to normal samples of which a lot exist. Moreover, innumerable anomalous patterns make it impossible to define and engineer anomalous features. Because of these kinds of reasons, using a supervised machine learning method is extremely limited for anomaly detection. Therefore, anomaly detection in real-world settings needs to model the distribution of normal data for anomaly detection in an unsupervised manner without abnormal data.

Generative Adversarial Networks (GAN) [2] is one of the generative models consisting of two neural networks: the generator and the discriminator. The generator learns to generate realistic images for fooling the discriminator, whereas the discriminator tries to discriminate between real images and generated images. The generator trained with an adversarial manner allows us to generate images that appear like real images. In other words, GAN has succeeded in modeling complex and high-dimensional data such as images [3]. Some anomaly detection methods using GAN have been proposed that model normal data distribution using this characteristic of GAN [4]. However, GAN is widely known to be difficult to train and unstable [5]. According to this reason, instability could decrease the accuracy of the GAN-based anomaly detection method. In this work, we introduced a restriction in order to ensure consistency in the image space and the latent space in Bi-directional GAN (Bi-GAN) based anomaly detection method [6]. Further, we propose using the projection discriminator [7] in Bi-GAN model for combining an image and a latent vector. Therefore, it has made it possible to stabilize the training of Bi-GAN, which has resulted in highly accurate anomaly detection. In the experiment, we show the effectiveness of our model by using two types of datasets.

2 Related Work

GAN-based anomaly detection. GAN is one of the generative models which is leaned in an adversarial manner. The Generator *G* is trained for the purpose of acquiring the data distribution. On the other hand, The Discriminator D estimates the probability that sample data comes from the training data distribution or not. Then, GAN has succeeded to acquire the data distribution that expresses training data. The training objective of GAN is expressed:

$$\min_G \max_D V(D,G) = E_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + E_{\boldsymbol{z}\sim p_{data}(\boldsymbol{z})}[\log(1-D(\boldsymbol{z})]. \tag{1}$$

Schlegl et al. proposed AnoGAN [8] that is the first method exploiting GAN's capacity to modeling complex and high-dimensional data. In this method, GAN is trained only normal data, and then mapping from the latent space to the image space is learned. Anomaly detection is performed using the assumption that normal images used for the training of GAN have a corresponding latent variable but an abnormal image does not have a corresponding latent variable. In the inference phase, AnoGAN requires the process of searching a latent variable corresponding to a target image, but GAN learned the mapping from the latent space to the image space and its inverse mapping is undefinable. Therefore, Schlegl et al. approximate the inverse mapping using the generator and the discriminator with the stochastic gradient descent manner. Its approximation is conducted by minimizing following equations for $\gamma = 1, 2, \ldots, \Gamma$ steps.

$$L_r(\boldsymbol{z}_\gamma) = \left\| \boldsymbol{x} - G(\boldsymbol{z}_\gamma) \right\|_1 \tag{2}$$

$$L_D(\boldsymbol{z}_\gamma) = \left\| f(\boldsymbol{x}) - f(G(\boldsymbol{z})) \right\|_1 \tag{3}$$

Eq 2 is called Residual Loss and it minimizes the difference in image space. On the other hand, Eq 3 is the feature-matching loss [9] in consideration of the Discriminator and minimizes the difference in the feature space. Function f denotes the output of the intermediate layer of the discriminator. Hence, the approximation minimizes

$$L(\boldsymbol{z}_\gamma) = (1-\lambda)\cdot L_R(\boldsymbol{z}_\gamma) + \lambda \cdot L_D(\boldsymbol{z}_\gamma)\ , \tag{4}$$

where λ is weighting parameters between the Residual Loss and the feature-matching loss. Finally, the anomaly score is defined as:

$$A_{AnoGAN}(\boldsymbol{x}) = (1-\lambda)\cdot L_R(\boldsymbol{x}) + \lambda \cdot L_D(\boldsymbol{x})\ . \tag{5}$$

However, AnoGAN enables highly accurate abnormality detection, but the approximation requires optimization using multiple times of backpropagation in every single target image inference. Thus, inference in AnoGAN is tremendously time-consuming and is impractical. To solve this problem, some methods are proposed. Schlegl et al. also proposed to train an Encoder that maps image space to latent space using the generator and the discriminator trained with GAN [10].

Bi-directional GAN-based anomaly detection. Zenati et al. proposed Efficient GAN-Based Anomaly Detection (EGBAD) [6] which is based on Bi-directional GAN (Bi-GAN) [11] intending to solve the time-consuming problem. Bi-GAN learns the mapping of the latent space to the image space the same as vanilla GAN, and learns the mapping from the image space to the latent space simultaneously by training an encoder in the adversarial training. Bi-GAN consists of 3 sub-networks: the Encoder E, the Generator G and the Discriminator D. Bi-GAN training objective is expressed as:

$$\begin{aligned}&\min_{G,E}\max_{D} V(D,G,E)\\ &= E_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[E_{\boldsymbol{z}\sim p_{E(\boldsymbol{z}|\boldsymbol{x})}}[\log D(\boldsymbol{x},\boldsymbol{z})]] + E_{\boldsymbol{z}\sim p_{data}(\boldsymbol{z})}[E_{\boldsymbol{x}\sim p_G(\boldsymbol{x}|\boldsymbol{z})}[\log(1-D(\boldsymbol{x},\boldsymbol{z}))]]].\end{aligned} \tag{6}$$

By training Bi-GAN using only normal data, we obtained the Generator and the Encoder that can perform the mutual mapping between image space and latent space according to the distribution of normal data. The Anomaly score is similar to AnoGAN and is computed by residual loss and the discrimination loss. The residual loss is:

$$L_R(\boldsymbol{x}) = \|\boldsymbol{x} - G(E(\boldsymbol{x}))\|_1\ . \tag{7}$$

The discrimination loss is defined in 2 ways. The first is the feature-matching loss and defined as:

$$L_{Df}(\boldsymbol{x}) = \| f(\boldsymbol{x}, E(\boldsymbol{x})) - f(G(E(\boldsymbol{x})), E(\boldsymbol{x})\|_1\ . \tag{8}$$

The second uses cross-entropy loss, which is the probability that the Discriminator identifies the image as a real image and is defined as:

$$L_{D\sigma}(\boldsymbol{x}) = \sigma(D(\boldsymbol{x}, E(\boldsymbol{x})), 1)\ , \tag{9}$$

where σ denotes the cross-entropy loss function.
The Anomaly score in EGBAD is defined as:

$$A_{EGBAD}(\boldsymbol{x}) = (1-\lambda)\cdot L_R(\boldsymbol{x}) + \lambda \cdot L_D(\boldsymbol{x})\ . \tag{10}$$

Owing to the simultaneous training of the Encoder, the time required in inference is greatly decreased.

Our model is based on Bi-GAN architecture. However, in contrast to original Bi-GAN architecture, our model has an additional Encoder and Generator module. These modules allow us to ensure the consistency of mutual mapping between image space and latent space. In addition to this, our model has the Projection Discriminator which enables us to efficient integration of an image and a latent variable. Fig.1 shows the comparison of AnoGAN, EGBAD and our model.

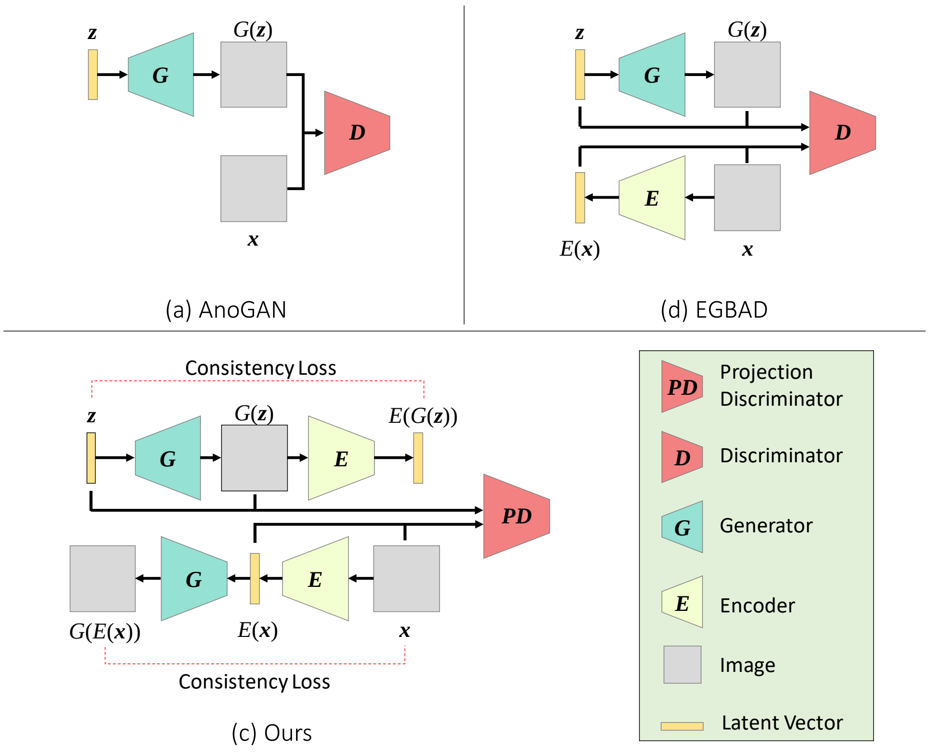

Fig. 1 The comparison of AnoGAN, EGBAD and the proposed method. In the training of AnoGAN, the Encoder that maps from images to latent variables is not trained. EGBAD and our model train the Encoder at the same time as adversarial learning. Moreover, our model has another Encoder and Generator in order to ensure the consistency of mappings.

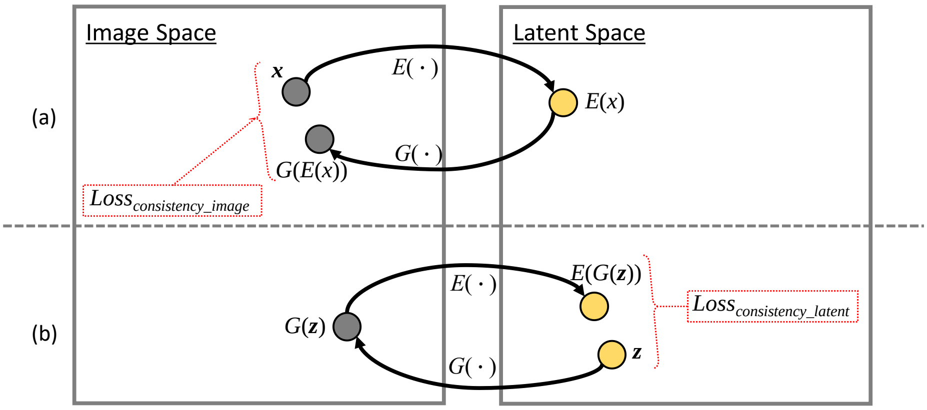

Fig. 2. The concept of consistency loss. Our model has two types of mapping: E denotes the mapping to the latent space from the image space and G denotes the mapping to the image space from the latent space. (a) shows the case of the consistency in image space. In the conventional method, it was not guaranteed that $\boldsymbol{x}$ and $G(E(\boldsymbol{x}))$ exist at exactly the same point in the image space. Therefore, by introducing Consistency Loss, we succeeded in mapping $\boldsymbol{x}$ and $G(E(\boldsymbol{x}))$ to closer points. (b) shows the consistency in the latent space, and the same is also true as in the image space, hence we introduce consistency loss to ensure that $\boldsymbol{z}$ and $E(G(\boldsymbol{z}))$ is mapped to a closer point in latent space.

3 Proposed Method

3.1 Model Overview

We adopt Bi-GAN based model which learns the mapping from latent space to image space and its inverse mapping simultaneously because GAN's modeling capacity can be fully utilized and faster anomaly detection is possible. And we attempt to introduce a constraint to ensure the consistency in these mappings for stabilization of training. The Discriminator in Bi-GAN based model requires combines an image and a latent variable. We introduce the adaptation of Projection Discriminator [7] for the combination of an image and a latent variable.

3.2 Consistency of mappings between image space and latent space

Our model has Bi-GAN based architecture which learns mappings from the image space and the latent space and its inverse mapping simultaneously. However, if the network model has a large capacity, learning by Adversarial training does not guarantee the consistency of the mapping [12]. In other words, if an image maps from image space to latent space and back again, there is no guarantee that the acquired image will be the same as the first image. Especially in GAN-based anomaly detection methods utilizing reconstruction error to compute the anomaly score, it is considered that this is a factor that greatly reduces the accuracy in abnormality detection using Bi-GAN. And the same situation could happen when a latent variable map from latent space to image space and back again. Thereby, here we introduce the Consistency Loss. Our consistency loss has 2 types of consistency in image space and consistency in latent space and is defined as:

$$Loss_{consistency\_image} = \mathbb{E}_{x \sim P_x} \|x - G(E(x))\|_1, \tag{11}$$

$$Loss_{consistency\_latnet} = \mathbb{E}_{z \sim P_z} \|z - E(G(z))\|_1. \tag{12}$$

Fig. 2 shows the conceptual image of our consistency loss. In our model, the Encoder and the Generator minimize the consistency loss and the adversarial loss simultaneously, which enables the model to learn mutual mappings between the image space and the latent space in consideration of the consistency.

3.3 Conditioning of image with latent variables in Bi-GAN architecture.

Miyato et.al propose projection discriminator [7] for conditioning class labels to images for conditional image generation tasks. Introducing projection discriminator can be enabled to generate more precise images compared to the traditional method that conditions class labels to images by concatenating [13] and the method of inferring class labels at the output layer of discriminator [14]. Teterwak et.al succeeded in conditioning semantic information in an image expansion task by introducing a projection discriminator [15]. Thus, from these studies, it is assumed that the projection discriminator is effective for conditioning images. As regards Bi-GAN, the discriminator

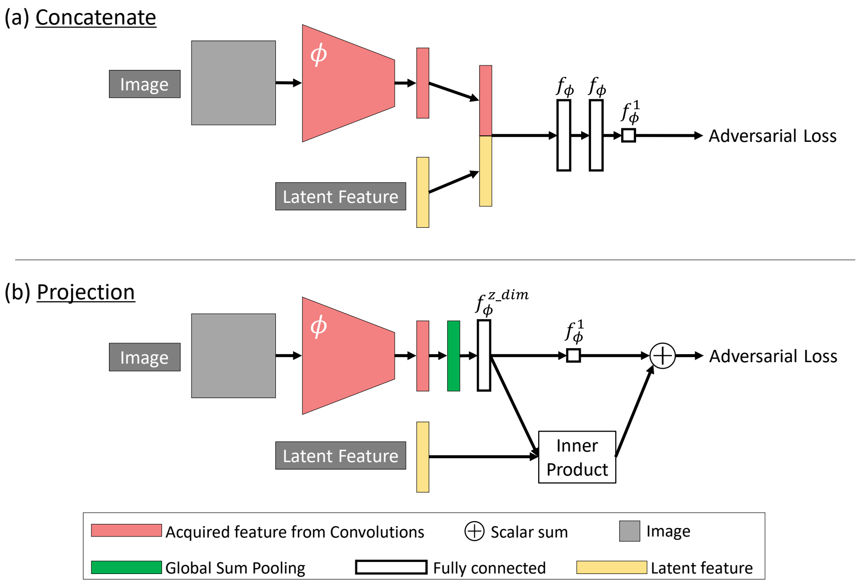

Fig. 3 The comparison of sample concatenate and projection. (a) shows traditional concatenate method which concatenates acquired feature maps and a latent variable in channel-wise. (b) shows the procedure for integrating feature maps and a latent variable by projection manner in our model.

adopts simple channel-wise concatenate for conditioning images. In our studies, we introduce a projection discriminator for the integration of images and latent variables. We try to stabilize the training of Bi-GAN, to improve the quality of generated images and hence improve the accuracy of anomaly detection task. In our model, the output $D(\boldsymbol{x}, \boldsymbol{z})$ of the Discriminator when integrating the image $\boldsymbol{x}$ and the latent variable $\boldsymbol{z}$ using the projection is shown as follows:

$$D(\boldsymbol{x},\boldsymbol{z}) = f_{\phi}^{1}(f_{\phi}^{z\_dim}(\mathrm{GSP}(\phi(\boldsymbol{x})))) + < f_{\phi}^{z\_dim}(\mathrm{GSP}(\phi(\boldsymbol{x}))), \boldsymbol{z} >, \tag{13}$$

Where ϕ is convolution operations in the Discriminator and $\phi(\boldsymbol{x})$ is feature maps acquired from the Discriminator with input image $\boldsymbol{x}$. f_{ϕ}^{n} denotes a fully connected layer with n dimensions and z_dim is the dimension of a latent variable. $<\cdot>$ is the inner product of two vectors and $\mathrm{GSP}(\cdot)$ denotes Global Sum Pooling [16]. Fig. 3 shows the comparison of simple concatenate and projection. Some traditional research using projection discriminator adopts an embedding layer but we do not adopt it, because latent variables to condition images are continuous values in our model.

3.4 Model Training

Algorithm 1 shows the training procedure for our model and Fig. 4 shows the overview of our model.

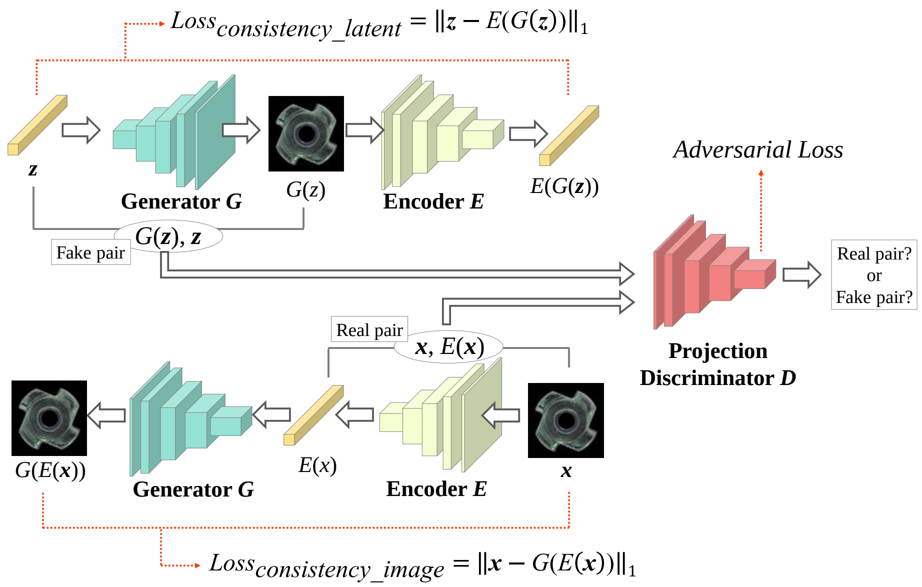

Fig. 4 Our model adopted the Bi-GAN based model and hence has the Encoder, the Generator and the Discriminator. The Discriminator has a projection mechanism. Weights of each Generator and Encoder are shared.

Our model is trained in an adversarial manner. The basic training procedure follows Bi-GAN. Additionally, the Encoder and the Generator in our model are trained by minimizing consistency loss. In our experiments, we set $\alpha = 0.9$ which weighs consistency loss and adversarial loss.

Algorithm 1 Training procedure of the proposed model

Input:

Encoder E, Generator G, Discriminator D with parameters θ_E, θ_G, θ_D

batch size M, weighting parameters α

Initialize parameters θ_E, θ_G, θ_D.

Repeat:

・Sample minibatch of M images from data distribution p_x

$$x^{(1)} \cdots x^{(M)} \sim p_x$$

・Sample minibatch of M noise samples from noise prior p_z

$$z^{(1)} \cdots z^{(M)} \sim p_z$$

・Generate images and latent vectors with the Encoder and the Generator

$$\tilde{z}^{(i)} \leftarrow E(x^{(i)}), \quad i = 1, \ldots, M$$

$$\tilde{x}^{(i)} \leftarrow G(z^{(i)}), \quad i = 1, \ldots, M$$

$$\hat{x}^{(i)} \leftarrow G(\tilde{z}^{(i)}), \quad i = 1, \ldots, M$$

$$\hat{z}^{(i)} \leftarrow E(\tilde{x}^{(i)}), \quad i = 1, \ldots, M$$

・Compute gradient for the Encoder and the Generator

$$L_{consistency} \leftarrow \frac{1}{M} \sum_{i=1}^{M} \left\| x^{(i)} - \hat{x}^{(i)} \right\|_1 + \frac{1}{M} \sum_{i=1}^{M} \left\| z^{(i)} - \hat{z}^{(i)} \right\|_1$$

$$L_{adversarial} \leftarrow -\frac{1}{M} \sum_{i=1}^{M} \log(D(x^{(i)}, \tilde{z}^{(i)})) - \frac{1}{M} \sum_{i=1}^{M} \log(1 - D(\tilde{x}^{(i)}, z^{(i)}))$$

$$L_{Encoder} \leftarrow \alpha L_{consistency} + (1-\alpha) L_{adversarial}$$

$$L_{Generator} \leftarrow \alpha L_{consistency} + (1-\alpha) L_{adversarial}$$

・Compute gradient for the Discriminator

$$L_D \leftarrow -\frac{1}{M} \sum_{i=1}^{M} \log(D(\tilde{x}^{(i)}, z^{(i)})) - \frac{1}{M} \sum_{i=1}^{M} \log(1 - D(x^{(i)}, \tilde{z}^{(i)}))$$

・Update weights of the Encoder, the Generator and the Discriminator

$$\theta_E \leftarrow \theta_E - \nabla_E L_E$$

$$\theta_G \leftarrow \theta_G - \nabla_G L_G$$

$$\theta_D \leftarrow \theta_D - \nabla_D L_D$$

Until convergence

3.5 Anomaly Score

Our model computes the anomaly score using trained Generator, Encoder and Discriminator. Anomaly score is defined by residual loss L_R and feature-matching loss L_D which considers intermediate output f of the Discriminator.

$$L_R(\boldsymbol{x}) = \|\boldsymbol{x} - G(E(\boldsymbol{x}))\|_1 \tag{14}$$

$$L_D(\boldsymbol{x}) = \| f(\boldsymbol{x}) - f(G(\boldsymbol{x}, E(\boldsymbol{x})))\|_1 \tag{15}$$

Finally, the anomaly score in our model is defined as:

$$A(\boldsymbol{x}) = (1-\lambda)\cdot L_R(\boldsymbol{x}) + \lambda \cdot L_D(\boldsymbol{x}) \ . \tag{16}$$

As regards abnormal image $\boldsymbol{x}$, $L_R(\boldsymbol{x})$ becomes larger because the Encoder is not able to acquire corresponding latent variables. For this reason, the Generator cannot reconstruct the image, and the reconstructed image differs from the original input image. Similarly, $L_D(\boldsymbol{x})$ increases for abnormal images in comparison using the feature-matching. Accordingly, $A(\boldsymbol{x})$ is expected to be large for abnormal images.

4 Experiments

We use 2 types of the dataset to evaluate our model: simple dataset MNIST and real-world setting dataset MVTec Anomaly Detection Dataset [17]. To compare traditional anomaly detection methods, we conducted the experiments using AutoEncoder [18], AnoGAN [8], EGBAD [6] and our model. All models are evaluated using the area under the receiver operating characteristic curve (AUROC).

In our model, the Encoder, the Generator and the Discriminator are optimized with Adam optimizer [19] with an initial learning rate lr = 1e−4, and momentums $\beta 1$ = 0.9, $\beta 2$ = 0.999 and are applied Batch Normalization [20] for all convolution layers. Spectral Normalization [21] is applied to convolutional layers of the Discriminator to prevent anomaly detection accuracy from decreasing due to learning instability.

4.1 MNIST Dataset

The MNIST dataset contains handwritten digit images 0 to 9. We made 10 different datasets from MNIST dataset by designating one digit as a normal class and other digits as an abnormal class. We used only normal data for training in all models and tested them with both normal and abnormal images. Fig. 5 shows the result of the experiments. In experiments using MNIST, all of the methods work well. However, our models that introduce both concatenate and projection outperform other methods.

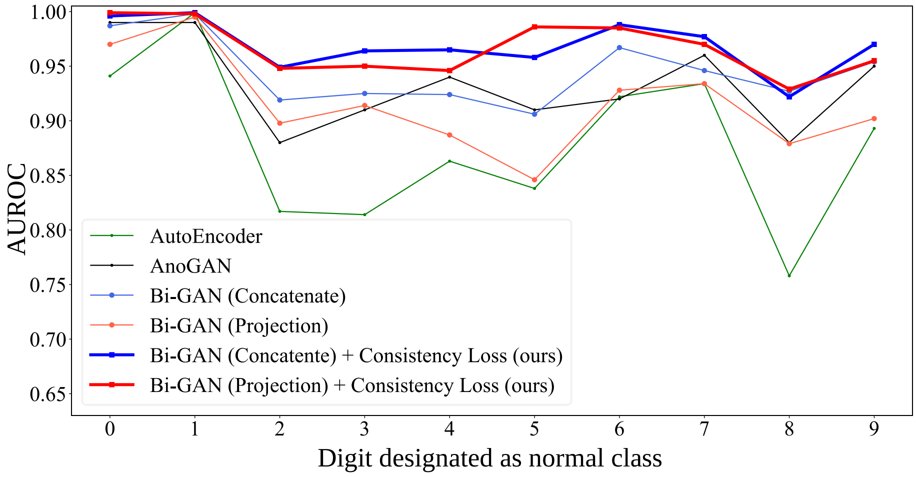

Fig. 5 The result of experiments using MNIST dataset. Each column indicates the digit that is designated as normal digits. The score of AnoGAN is obtained from [22].

4.2 MVTec Anomaly Detection Dataset

In order to evaluate anomaly detection in real-world settings dataset, we used the Metal Nut dataset in MVTec Anomaly Detection Dataset [17]. Metal Nut dataset has normal images for training and both normal and abnormal images for testing. Table 1 shows the result of anomaly detection experiments. Our model using projection results in a better score than the model using concatenate and this implies that adopting projection discriminator is effective for anomaly detection in real-world setting datasets.

Fig. 6 shows test images $\boldsymbol{x}$ and their reconstructed images $G(E(\boldsymbol{x}))$ of our model using projection discriminator. Regarding normal images, the Encoder is able to acquire latent variables corresponding to images. Therefore, the Generator is possible to reconstruct original test images from the latent variables. However, regarding abnormal images, the Generator cannot reconstruct original test images because the Encoder fails to acquire latent variables corresponding to images.

Table 1. The result of anomaly detection experiments using the Metal Nut dataset. In this table, the score of AnoGAN is obtained from [17].

Model	AUROC
AutoEncoder	0.63
AnoGAN	0.76
Bi-directional GAN (Concatenate) (EGBAD)	0.78
Bi-directional GAN (Projection)	0.77
Bi-directional GAN (Concatenate) + Consistency Loss (ours)	0.79
Bi-directional GAN (Projection) + Consistency Loss (ours)	**0.84**

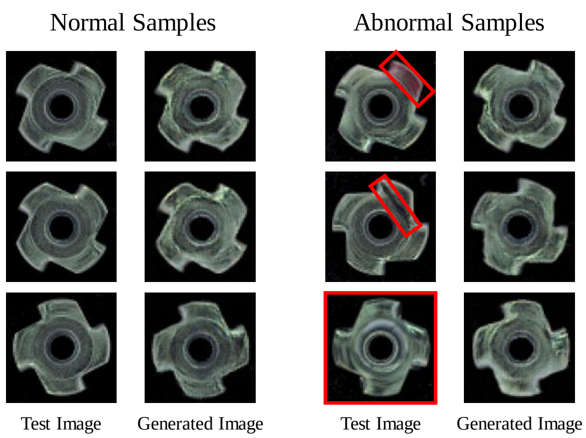

Fig. 6 Visualization of test images and generated images of normal samples and abnormal samples with anomalies highlighted in red rectangles. However our model succeeded to generate normal images, our model fails to generate abnormal samples.

5 Conclusion

We proposed a novel anomaly detection model based on Bi-directional GAN. We have the following two main contributions. First, we focused on the fact that mutual mapping between the image space and the latent space was not guaranteed in the anomaly detection method using Bi-GAN. Hence to solve this important problem, we introduce consistency loss. Second, we introduce the projection discriminator for Bi-directional GAN model to integrate images and the corresponding latent variables, which allows improvement of the quality of reconstruction and anomaly detection with high accuracy. In the experiment, we compared our model to conventional anomaly detection methods: AutoEncoder, AnoGAN and EGBAD by using MNIST dataset and the Metal Nut dataset. We confirmed that our model is able to detect anomalies and outperform conventional methods.

Acknowledgments

This work has received funding from Human resource development and research project on production technology for aerospace industry : Subsidy from Gifu Prefecture.

References

1. Chalapathy, R., Chawla, S.: Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407. (2019).
2. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems, pp 2672-2680. (2016).
3. Radford, A., Metz, L., Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In: International Conference on Learning Representations, Workshop (2016).
4. Di Mattia, F., Galeone, P., De Simoni, M., Ghelfi, E.: A survey on gans for anomaly detection. arXiv preprint arXiv:1906.11632. (2019).
5. Wiatrak, M., Albrecht, Stefano, V.: Stabilizing Generative Adversarial Network Training: A Survey. arXiv preprint arXiv:1910.00927. (2019).
6. Zenati, H., Foo, C. S., Lecouat, B., Manek, G., Chandrasekhar, V. R.: Efficient gan-based anomaly detection. In: International Conference on Learning Representations, Workshop (2018).
7. Miyato, T., Koyama, M.: cGANs with projection discriminator. In: International Conference on Learning Representations. (2018).
8. Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In International Conference on Information Processing in Medical Imaging, pp. 146-157. Springer, Cham (2017).
9. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In Advances in neural information processing systems, pp. 2234-2242. (2016).
10. Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G., Schmidt-Erfurth, U.: f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. Medical image analysis, 54, 30-44 (2019).
11. Donahue, J., Krähenbühl, P., Darrell, T.: Adversarial feature learning. In: International Conference on Learning Representations (2017).
12. Zhu, J. Y., Park, T., Isola, P., Efros, A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232 (2017).
13. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784. (2014)
14. Odena, A., Olah, C., Shlens, J.: Conditional image synthesis with auxiliary classifier gans. In Proceedings of the 34th International Conference on Machine Learning-Volume 70, pp. 2642-2651. (2017).
15. Teterwak, P., Sarna, A., Krishnan, D., Maschinot, A., Belanger, D., Liu, C., Freeman, W. T.In: Boundless: Generative Adversarial Networks for Image Extension. In Proceedings of the IEEE International Conference on Computer Vision, pp. 10521-10530 (2019).
16. Aich, S., Stavness, I.: Global Sum Pooling: A Generalization Trick for Object Counting with Small Datasets of Large Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 73-82. (2019).
17. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: MVTec AD--A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9592-9600. (2019).

18. Hinton, G. E., Salakhutdinov, R. R.: Reducing the dimensionality of data with neural networks. science, 313(5786), pp.504-507. (2006).
19. Kingma, D. P., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (2015).
20. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning-Volume 37, pp. 448-456. (2015).
21. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. In: International Conference on Learning Representations (2018).
22. Huang, C., Cao, J., Ye, F., Li, M., Zhang, Y., Lu, C.: Inverse-Transform AutoEncoder for Anomaly Detection. In: arXiv preprint arXiv:1911.10676. (2019).