

Dual-encoder Bidirectional Generative Adversarial Networks for Anomaly Detection

Teguh Budianto, Tomohiro Nakai, Kazunori Imoto, Takahiro Takimoto, Kosuke Haruki
Corporate Research and Development Center, Toshiba Corporation, Kawasaki, Japan
{teguh1.budianto, tomohiro.nakai, kazunori.imoto, takahiro.takimoto, kosuke.haruki}@toshiba.co.jp

Abstract—Generative adversarial networks (GANs) have shown promise for various problems including anomaly detection. When anomaly detection is performed using GAN models that learn only the features of normal data samples, data that are not similar to normal data are detected as abnormal samples. The present approach is developed by employing a dual-encoder in a bidirectional GAN architecture that is trained simultaneously with a generator and a discriminator network. Through the learning mechanism, the proposed method aims to reduce the problem of bad cycle consistency, in which a bidirectional GAN might not be able to reproduce samples with a large difference between normal and abnormal samples. We assume that bad cycle consistency occurs when the method does not preserve enough information of the sample data. We show that our proposed method performs well in capturing the distribution of normal samples, thereby improving anomaly detection on GAN-based models. Experiments are reported in which our method is applied to publicly available datasets, including application to a brain magnetic resonance imaging anomaly detection system.

Keywords—anomaly detection; adversarial learning; generative adversarial network; encoder, cycle consistency; latent space; unsupervised learning; unbalanced datasets

I. INTRODUCTION

Anomaly detection is a well-known problem that focuses mainly on finding abnormal data behavior that differs from a normal data distribution. Studies of the anomaly detection problem have benefitted various fields, including health care [1], video surveillance [2] [3], and image analysis [4]. Most anomaly detection problems, particularly in high-dimensional image datasets, are defined by separating abnormal samples that are visually different from the data distribution. Separating anomalies from a data distribution can be useful for improving product quality inspections in manufacturing systems [5], detecting brain tumors in medical images [6], and detecting anomalous objects in video surveillance [2] [3].

In real-world applications, there is a strong need for anomaly detection techniques that are able to handle distributions of complex high-dimensional data. In terms of data complexity, however, conventional anomaly detection methods are unsuitable for solving the aforementioned problems [7] [8]. Usually, only a small number of anomalous samples are available, which leads to collection of an imbalanced data sample. This phenomenon has led researchers to propose learning approaches in semi-supervised and unsupervised settings, such as image reconstruction-based anomaly detection systems.

Anomaly detection methods based on generative adversarial networks (GANs) have shown promising performance in capturing the distributions of high-dimensional and complex

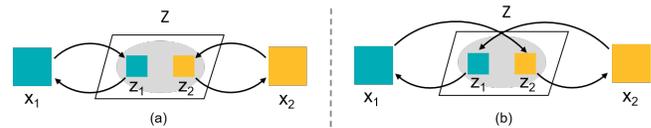


Fig. 1. Cycle consistency showing samples in data space and latent space Z . All samples in data space are represented by x . The data samples x_1 and x_2 are in the normal data distribution, and the predefined latent variables z_1 and z_2 are in latent space Z distributed in a circle. (a) Good cycle consistency that reproduces input samples consistently. (b) Bad cycle-consistency in which point x_1 maps to z_2 and reproduces x_2 instead of the original point x_1 . The depicted concept is introduced in [7].

data [1] [9]. In particular, Zenati et al. [9] proposed an anomaly detection framework employing a bidirectional GAN (BiGAN) [10] [11] that simultaneously learns an encoder, a generator, and a discriminator during training.

BiGAN is trained through bidirectional adversarial learning in which an encoder and a generator network are used to generate data both in data and latent space [7]. Working similarly to an autoencoder in data reconstruction, BiGANs generate normal and abnormal samples similar to the normal samples in order to measure the abnormality through reconstruction error. Data reproduction quality in BiGANs has shown limitations in normal sample reconstruction, resulting in high reconstruction error for normal samples. This could significantly degrade the anomaly detection performance [10] [11]. This creates an insufficient difference between samples. The condition where a model cannot reproduce samples and gives large reconstruction error is called bad cycle consistency [12]. Fig. 1 shows a conceptual image of cycle consistency.

This study assumes that bad cycle consistency can occur as a result of a model not preserving enough information of the input image. The proposed method introduces *preserved information learning* using a dual-encoder in BiGAN. The aim of preserved information learning in dual-encoder BiGAN is to significantly reduce the bad cycle consistency of GAN-based architectures. In this paper, a model employing dual-encoder BiGAN is proposed for addressing bad cycle consistency and improving anomaly detection. Toward these ends, the main goals of this paper are as follows.

- To propose a GAN-based anomaly detection technique for reducing bad cycle consistency.
- To evaluate the proposed method on several publicly available datasets and demonstrate its performance compared with state-of-the-art methods.

II. RELATED WORK

A complete review of deep learning technology for anomaly detection along with its application across various domains is comprehensively explained in [13]. Examples of existing methods include autoencoder-based anomaly detection, such as denoising autoencoder [14], robust deep autoencoder [15], and variational autoencoder [16]. Autoencoder-based methods usually detect anomalies by measuring the difference between an original sample x and its reconstruction x' as $\|x - x'\|$.

Recent anomaly detection methods employing GANs can handle the presence of anomalous samples [17]. GANs train two different networks simultaneously through a minimax game in which one network is a generator (G) that learns to generate data (e.g. images) and minimize error, and the other network is a discriminator (D) that aims to distinguish the generated data by G from the real data distribution. The first work that used GANs for anomaly detection, called AnoGAN, was proposed by Schleg et al. [1]. AnoGAN is trained using only normal samples to learn a mapping of the latent space representation. During the testing period, the latent vector that best reconstructed the test image is then searched through the latent space representation. The anomaly score in AnoGAN is defined using a combination of reconstruction loss and the difference between the intermediate discriminator feature representation of a test image and its reconstruction. Furthermore, GANomaly [18] was proposed which uses conditional GANs that jointly learn the generation of a high-dimensional image and the inference of the latent space. GANomaly frameworks consist of encoder-decoder-encoder sub-networks in the generator and a discriminator network. GANomaly defines a new anomaly score as a combination of three loss functions, namely, feature matching loss, reconstruction loss, and encoding loss. A more recent method, called Fence GAN [8], aims to generate data lying on the boundary of the normal data distribution by proposing the use of encirclement loss for the GAN loss function. In Fence GAN, the anomaly score is calculated directly using the score from the discriminator. Sabokrou et al. [19] proposed a method that is mainly composed of two networks that are trained adversarially in an unsupervised learning setting. One of the networks in this architecture learns to refine noisy input images, while the other is responsible for separating normal and abnormal sample images.

III. BACKGROUND

A. Generative Adversarial Networks

GANs consist of two networks for learning data generation. One network is a generator G that learns to generate data and to minimize error, while the other is a discriminator D that aims to distinguish generated data by G from the real data distribution. Both network G and D are trained simultaneously to minimize their loss through a two-player min-max game, formulated as

$$\min_G \max_D \mathbb{E}[\log D(x) + \log(1 - D(G(z)))] \quad (1)$$

GANs are also used to solve anomaly detection problems and defined as AnoGAN [1].

B. Efficient GAN-based Anomaly Detection

BiGAN learns an encoder E that maps input samples x to a latent representation z , and a generator G and a discriminator D are also trained at the same time. Unlike the original GAN, the discriminator D in BiGAN considers not only the input x , but also its respective latent variable z . In particular, the BiGAN training objective is defined as

$$\min_{G,E} \max_D \mathbb{E}[\log D(x, E(x)) + \log(1 - D(G(z), z))] \quad (2)$$

To generate G , D , and E , a model is trained only using normal samples and the anomaly score function $A(x)$ as in [1] is used to measure the level of abnormality. The anomaly score is based on the convex combination of reconstruction loss L_G and discriminator loss L_D .

$$A(x) = \alpha L_G(x) + (1 - \alpha) L_D(x) \quad (3)$$

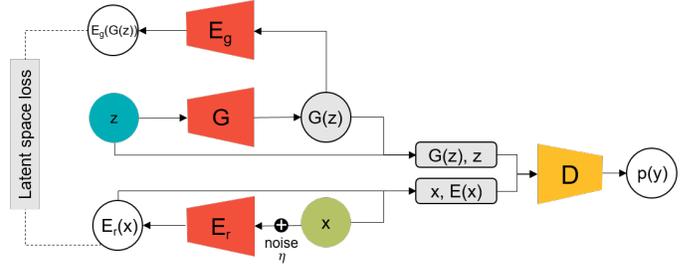


Fig. 2. Proposed method: Original method dual-encoder BiGAN architecture. For simplicity, the other losses in the proposed method are omitted from the illustration.

IV. PROPOSED METHOD

A. Dual-Encoder BiGAN

In this study, a GAN-based anomaly detection approach is proposed for handling bad cycle consistency. BiGAN forces abnormal samples to be reproduced within a normal distribution, but either the normal or abnormal samples suffer from the problem of poor reconstruction of inlier samples, making it difficult for GAN-based methods to detect outlier samples precisely. We assume that bad cycle consistency might occur when a model is unable to preserve enough information of the input image. The proposed method introduces preserved information learning employing a dual-encoder BiGAN architecture (Fig. 2). In this case, η as depicted in Fig. 2 is a Gaussian noise that is added to input sample x in order to make the proposed method more robust against corrupted samples. Furthermore, the preserved information learning in a dual-encoder BiGAN uses cycle consistency loss and latent space variable loss. In Fig. 2, $p(y)$ represents the probability that the joint input of a sample and latent variable to discriminator D comes from a real or fake sample.

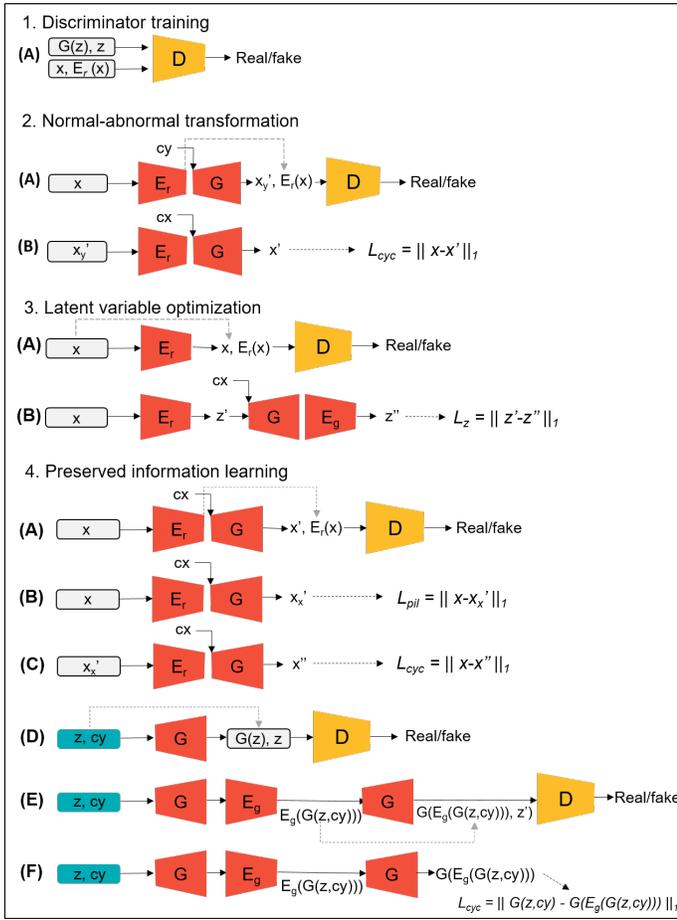


Fig. 3. Training mechanism for dual-encoder BiGAN.

The proposed method optimizes the generator G by prioritizing cycle consistency. The goal is to be able to generate an image that can be reconstructed back to its original source. It is expected that the proposed method can overcome the main problems by prioritizing the cycle consistency loss. Furthermore, an evaluation is performed on publicly available image datasets, including a real-world medical image dataset (Section V).

B. Preserved Information Learning

Fig. 3 shows the training mechanism for training all networks in the proposed method. This training mechanism is implemented simultaneously in order to train each of the networks in the dual-encoder BiGAN architecture. The concept behind the complete training scheme of the dual-encoder BiGAN is inspired by image-to-image translation methods [20] [21]. In image-to-image translation, the image transformation is done through different-domain or same-domain transformation. In contrast, the proposed method does not use any domain information, but instead uses single-class input.

The complete training scheme as shown in Fig. 3 is as follows.

- Discriminator D is trained adversarially to separate real/fake images (see Fig. 3-1). As in BiGAN, the discriminator input is a pair of samples in image space and its respective latent variable, both for real sample x and fake/reconstructed sample $G(z)$.
- In preserved information learning, the networks are conditioned using the target variable c . The target variable c is an extra information input provided to the conditioning function by feeding the real target c_x or random target c_y as an additional input layer to the networks. Both c_x and c_y are configured to control the generation of a sample corresponding to the source, whether it is from real data sample x or from generated sample.
- As shown in Fig. 3-2(A-B), a real sample x and random target c_y are regenerated through $G(E_r(x), c_y)$ as generated sample x'_y , which is then reconstructed back to x' in order to measure the loss of cycle consistency of sample x . This procedure is called normal-abnormal transformation because it employs a random target c_y (not a real normal label) that is uniformly distributed as input to encoder E_r during training (see Fig. 3-2 (A)). The input of c_x is the real label used for real normal sample x .
- The present architecture employs a dual-encoder in which the second encoder E_g is proposed in order to optimize the distance between a real sample latent variable and reconstructed latent variable in latent space. Fig. 3-3 shows the latent variable optimization in the proposed method. In particular, the output from discriminator D is also used to update the encoder E_r that appears in Eq. 11, as shown in Fig. 3-3(A). In comparison with the bottom of Fig. 3-1(A) where measurement is performed to update the discriminator D , in Fig. 3-3(A) it is used to update the encoder E_r . Furthermore, the effect of latent space optimization on dual-encoder BiGAN is shown in Fig. 8c after BiGAN (Fig. 8b) with additional encoder E_g .
- Preserved information learning (Fig. 3-4) employs input from both random variable z and latent variable $E_r(x)$ into generator G . As this is expected to reduce bad cycle consistency, we additionally support this process by prioritizing cycle consistency loss in generator G . As a result, the generator learns to enrich its ability to generate various samples and is also pushed to reconstruct input samples from $E_r(x)$ similarly to an auto-encoder. The improvement achieved by using this procedure can be seen in Fig. 8, which shows that the proposed method with preserved information learning reduced bad cycle consistency in the MNIST anomaly dataset. The generator is trained through a preserved information learning scheme as shown in Fig. 3-4 with the assistance of cycle consistency prioritization. The encoder employs an input image x and its target c . For any generated or reconstructed image, target c_y is used, where c_y is defined as a noisy random target provided to the generator for learning features from non-normal image input.

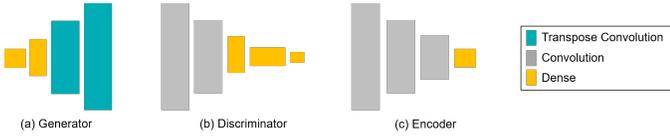


Fig. 4. Networks structure of (a) generator G , (b) discriminator D , and (c) encoders E_r and E_g .

By learning only normal features, the reconstruction in the proposed method is closed to normal samples. This mechanism therefore provides two advantages: (1) generator G reconstructs the image to normal features; and (2) discriminator D is able to measure any input x as normal or abnormal. To realize these training mechanisms, dual-encoder BiGAN losses are defined as follows.

1) *Adversarial loss*: In the proposed method, the generator learns to preserve information from the input images. To ensure that generator G is able to judge a real/fake latent variable z , the generator also receives signals from both input images and random latent variables. As mentioned earlier, the proposed method prioritizes cycle consistency, which makes it important to learn the relationship between image space and latent space. The adversarial loss of discriminator D is modified as follows.

$$\mathcal{L}_{adv}^D = \sum_{c \in \{c_y, c_x\}} \mathbb{E}[\log(1 - D(G(z, c), z)) + \log D(x, E_r(x))] \quad (4)$$

The adversarial loss of generator G is modified by adding the loss information of encoder E_g

$$\mathcal{L}_{adv}^G = \left(\sum_{c \in \{c_y, c_x\}} \mathbb{E}[\log(1 - D(G(z, c), z))] \right) + \mathbb{E}[\log(1 - D(G(E_g(G(z, c_y))), E_g(G(z, c_y))))] \quad (5)$$

2) *Prioritized cycle consistency loss*: In dual-encoder BiGAN, the generator is optimized by prioritizing cycle consistency loss. Cycle consistency loss helps the generator to preserve enough information for reconstructing the generated image back to its original.

$$\begin{aligned} \mathcal{L}_{cyc} = & \mathbb{E}_{x, c_x, c_y} \left(\left[\|G(E_r(G(E_r(x), c_y)), c_x) - x\|_1 \right] \right. \\ & + \left[\|G(E_r(G(E_r(x), c_x)), c_x) - x\|_1 \right] \\ & \left. + \left[\|G(E_g(G(z)), c_y) - G(z, c_y)\|_1 \right] \right) \end{aligned} \quad (6)$$

3) *Preserved information loss*: We modify the identity loss from [21] as *preserved information loss* to penalize generator G when learning real input images.

$$\mathcal{L}_{pil} = \begin{cases} 0, & c = c_y \\ \mathbb{E}_{x, c} [\|G(E_r(x), c) - x\|_1], & c = c_x \end{cases} \quad (7)$$

4) *Latent space loss*: We introduce a second encoder E_g for minimizing the distance between z and its latent reconstruction in latent space.

$$\mathcal{L}_z = [\|E_g(G(E_r(x), c_x)) - E_r(x)\|_1] \quad (8)$$

where \mathcal{L}_z is only introduced for the real input x while the second term in 7 is penalized through a priority parameter for cycle consistency.

5) *Full objective*: Total optimization of discriminator D , generator G , encoder E_r and E_g in Dual-encoder BiGAN is as follows.

$$\mathcal{L}_D = \mathcal{L}_{adv}^D \quad (9)$$

$$\mathcal{L}_G = \mathcal{L}_{adv}^G + \lambda_{cyc} \mathcal{L}_{cyc} + \mathcal{L}_{pil} \quad (10)$$

$$\mathcal{L}_{E_r} = \mathbb{E}[\log D(x, E_r(x))] + \mathbb{E}[\|G(E_r(x), c_x) - x\|_1] \quad (11)$$

$$\mathcal{L}_{E_g} = \mathcal{L}_z \quad (12)$$

Here, λ_{cyc} is a priority parameter of cycle consistency for generator G . In our experiments, we set $\lambda_{cyc} = 0.1$.

C. Anomaly Score

The proposed method is trained using only normal samples and employs a preserved information learning mechanism (Fig. 3-4). The anomaly score is defined as in Equation 3 where we use $\alpha = 0.1$, which has been found empirically through experiments in [1]. We also find that the discriminator score $D(x)$ can also be employed as an alternative anomaly score. In the evaluation phase, real target c_x is substituted for only normal targets because the trained model is familiar with only normal input samples.

V. EXPERIMENTS

To evaluate the proposed method, extensive experiments were conducted on publicly available datasets. In all experiments using the unsupervised learning setup, the proposed method used only normal samples to train the models. A common practical performance metric, the area under the receiver operating characteristics curve (AUROC) is used to measure the quality of interchangeability of the given scoring of the methods proposed in this study.

A. Architecture

In our experiments, we follow the network structure for all generator, discriminator, and encoders introduced in BiGAN [9] when evaluating the MNIST dataset. There are only minimal differences between the structures of our networks and the original BiGAN; these differences are due to the join input of target c and the latent space variable. In particular, Fig. 4 shows the number of layers used in the proposed method, which consist mainly of Dense, Convolution, and Convolution Transpose layers. For fair comparison, we use the same architecture for the baselines with minimal differences between each of the architectures.

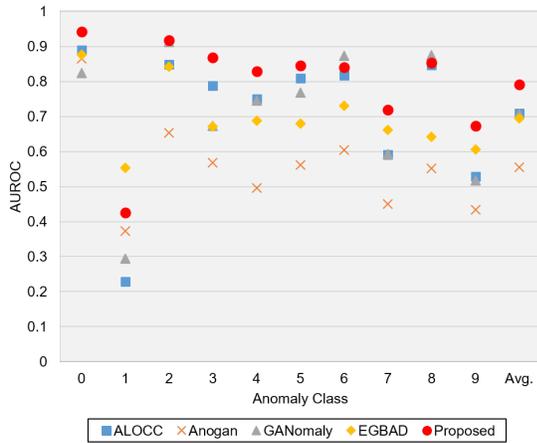


Fig. 5. AUROC results on MNIST dataset.

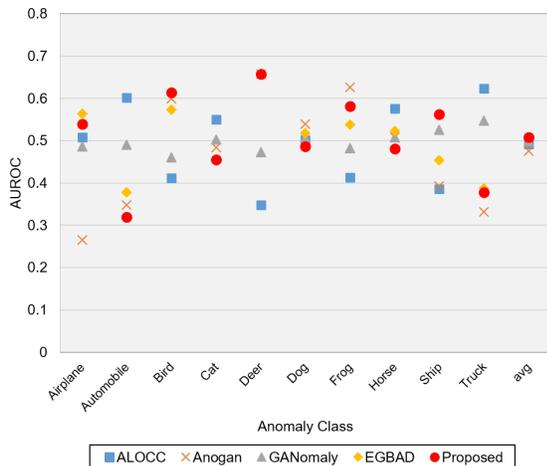


Fig. 6. AUROC results on the CIFAR10 dataset

B. Datasets

To evaluate the proposed method, experiments were conducted on three publicly available datasets. The following is a brief overview of each dataset used for evaluation of the proposed method.

1) *MNIST*: The MNIST dataset contains handwritten digits which are usually used for early stage model evaluation. The data are split between 60,000 samples in the training set and 10,000 samples in the test set. Within each set, there are 28×28 pixel grayscale images with a total of 10 output classes representing the ten digits from 0 to 9. The evaluation presented in this work was conducted on each class in which, at any time, only one class is considered as the anomaly class, and the remaining nine classes are considered together as the normal class. This means that only the normal examples in the training set were used to train the evaluated models while the training data of the class considered abnormal were ignored. The test procedure was applied to a test set that had not been seen by the trained model.



Fig. 7. Image reconstruction of anomaly digit 1. The proposed method completely reconstructs normal and abnormal samples. Fundamentally, we expect the proposed method to be unable to reconstruct any abnormal samples. This phenomenon could have the effect that the anomaly score $A(x)$ is not able to separate normal and abnormal samples.

2) *CIFAR10*: The CIFAR10 dataset [22] contains natural color 32×32 pixel images. The dataset is split into images associated with labels representing objects of 10 classes of natural images, such as image of "airplane", "automobiles", and "dog". As with the MNIST dataset, we configured the dataset for each class by using only one class as the anomaly object and trained the models only using normal data samples from the rest of the classes.

3) *BRATS 2013*: The BRATS 2013 dataset [23] [24] consists of synthetic and real images. Each image is divided into healthy samples and tumor positive samples with high-grade gliomas (HGs) and low-grade gliomas (LGs). There are 25 patients with both synthetic HG and LG images and 20 patients with real HG and 10 patients with real LG images. In this case, we are not particularly trying to segment the tumors, but rather trying to predict the separation of whether an image contains a tumor, which represents an abnormality in the sample.

C. Results

1) *Application to MNIST*: Fig. 5 shows the AUROC results obtained using the MNIST dataset, where the x-axis represents anomalous classes and the average overall performance. As shown in Fig. 5, the proposed method outperforms the state-of-the-art methods on the majority of abnormal class digits. Interestingly, the proposed method performs badly against only anomaly digit 1. This may occur due to the proposed method completely reconstructing all normal and abnormal samples as shown in Fig. 7. This causes the anomaly score for anomaly digit 1 to be very close to that of the digits in the normal class.

Fig. 8 shows the reconstruction images from generator G of the BiGAN-based methods. In the figures, digit class 0 was selected as the anomaly class. The original BiGAN evaluated in [9] (EGBAD) shows the condition in which normal samples were generated the same as other digits among normal samples. This evidence shows that BiGAN (EGBAD) suffers from bad cycle consistency, which may cause high reconstruction error of normal samples (Fig. 8b). The proposed method overcomes this shortcoming by employing a dual-encoder in the BiGAN architecture. Fig. 8d shows the results of preserved information learning in dual-encoder BiGAN. This significantly reduces the bad cycle consistency, thereby leading to improved detection performance. Preserved information learning is important for achieving the maximum capability of dual-encoder BiGAN. Based on empirical observations, dual-encoder BiGAN is not expected to reach its best performance by only adversarial training with the help of latent space

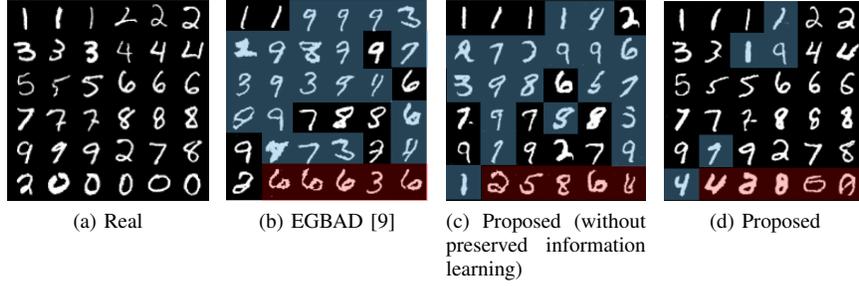


Fig. 8. Evaluation results for abnormal digits 0. Digits in blue boxes indicate models that were unable to reconstruct normal samples due to bad cycle consistency, while digits in red boxes indicate models that were unable to reconstruct abnormal images. (a) Real input images. (b) Reconstructed images by BiGAN/EGBAD. (c) Reconstructed images by proposed architecture dual-encoder BiGAN trained adversarially without preserved information learning. (d) Proposed method.

TABLE I
PERFORMANCE ON THE BRATS 2013 DATASET ACCORDING TO AUROC METRIC.

Methods	AUROC
AnoGAN	0.340279
ALOC	0.620300
GANomaly	0.845130
EGBAD + A(x)	0.286299
EGBAD + D(x)	0.786707
Proposed + A(x)	0.632038
Proposed + D(x)	0.861377

TABLE II
PERFORMANCE OF MODIFIED VERSIONS OF THE PROPOSED METHOD ON BRATS 2013 WHEN CHANGING LATENT VARIABLE SIZE.

Modification	z	AUROC
Proposed + A(x)	20	0.632038
Proposed + D(x)		0.861377
Proposed + A(x)	50	0.743657
Proposed + D(x)		0.847963
Proposed + A(x)	100	0.634741
Proposed + D(x)		0.926704

optimization (Fig. 8c). This shows the benefit of applying dual-encoder to BiGAN with a complete preserved information learning mechanism.

2) *Application to CIFAR10*: Fig. 6 shows a comparison of the performance results for the proposed method compared with the baselines for the CIFAR10 anomaly dataset. Since the images in CIFAR10 contain natural images which have more complex visual structures compared with the MNIST images, it shows that all GAN-based anomaly detection, including ours, offer fair results without providing high performance in the separation of normal and abnormal samples. On average, the proposed method is quite competitive against the baselines, particularly for the anomaly class 'deer'. The CIFAR10 dataset clearly provides a different level of difficulty and is a challenging problem for the anomaly detection task. Fig. 13 shows the reconstruction image of anomaly class 'airplane' for different latent variable sizes.

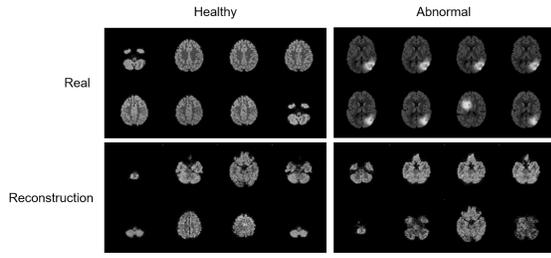
3) *Application to brain magnetic resonance imaging anomaly detection*: We employed the proposed method and

EGBAD to the brain magnetic resonance imaging problem domain, specifically the BRATS 2013 dataset, and used the anomaly score $A(x)$ and output of discriminator $D(x)$ as anomaly scoring. Proposed model with anomaly scoring $A(x)$ and $D(x)$ is presented as *proposed + A(x)* and *proposed + D(x)*, respectively. This configuration is also applied for EGBAD. The performance results obtained by the proposed method are shown in Table I. Overall, we see that the proposed method gave the highest performance on the BRATS 2013 dataset as indicated by the AUROC score (AUROC: 0.861377) and was competitive against GANomaly. Fig. 9 shows a comparison of the reconstruction with the EGBAD method for BRATS 2013. The healthy image reconstruction by the proposed method seems worse than that by EGBAD. Since our best reconstruction is obtain by $D(x)$, the reconstruction error is not very important for helping us find the normal and abnormal samples in BRATS 2013. In addition, there is a tendency for EGBAD to fall into mode collapse by qualitatively examining the reconstruction of both healthy and abnormal samples.

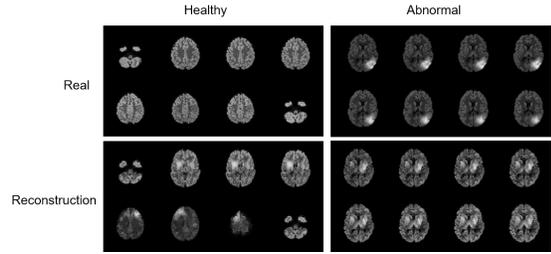
D. Modified Versions of Dual-encoder BiGAN

This section presents further studies for developing modified methods to dual-encoder BiGAN. In the evaluation on a relatively simple dataset (MNIST), dual-encoder BiGAN showed strong performance against most abnormal classes. However, the proposed method may not be able to achieve the best results due to differences in data complexity and characteristics, as shown in the experimental results for CIFAR10. The following are proposed modifications of the proposed method to achieve improved performance through selection of the architecture and training mechanism.

1) *Simple vs. complete training scheme*: The proposed method offers competitive performance through its complete training scheme. Here we propose simple versions of the proposed method: (1) without providing a random target to the generator G ; and (2) omitting step 2 (A-B) shown in Fig. 3 from the training mechanism since it is not needed when the random target is provided to generator G . The simple training scheme performs well enough on the MNIST dataset that it does not completely suffer from bad cycle consistency (see Fig. 10). While it does not show major degradation



(a) Proposed



(b) EGBAD

Fig. 9. Reconstruction of healthy and abnormal samples by (a) the proposed method and (b) EGBAD.



(a) Real (b) Complete training scheme (c) Simple training scheme

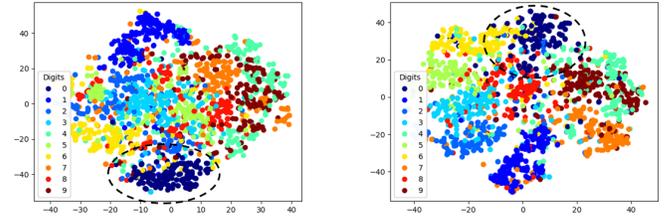
Fig. 10. Cycle consistency of different training schemes on MNIST anomaly digit 0.

on the non-complicated dataset, complete reconstruction of natural image datasets requires further improvement as in the case of CIFAR10 (see the comparison of a real sample and reconstruction in Fig. 13).

2) *Effect of latent variable size:* This section considers the effect of latent variable size on both encoder $E_r(x)$ and $E_g(x)$ and their random variables. These variables share the same size and contain the information required by the generator to regenerate an image. Our assumption is the size of z could be critical for deciding the precision of information required by the generator to translate z into image space.

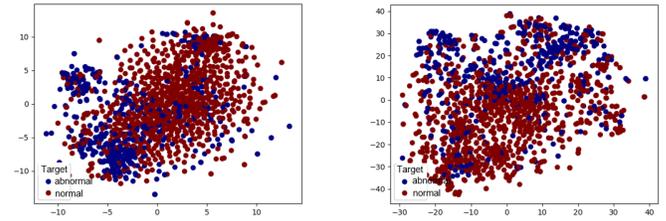
In addition, changing the latent variable size is another modification for improving the anomaly detection performance presented in Table II. This improves the AUROC performance of the proposed method from 0.861377 to 0.926704 on BRATS 2013.

3) *Latent space projection:* In MNIST, we are interested in studying the behavior of the proposed method by projecting the data onto latent space to see the data projection of a model that able to separate normal and abnormal samples. Fig. 11 visualizes this data projection of both encoders E_r and E_g



(a) Encoder E_r (b) Encoder E_g

Fig. 11. Latent space projection of encoders in the proposed method on MNIST using abnormal digit 0.



(a) Encoder E_r (b) Encoder E_g

Fig. 12. Latent space projection of encoders in the proposed method on CIFAR10 using abnormal class 'airplane'.

on the latent space. According to samples shown for both encoders, the normal and abnormal sample data is distributed to separate the anomaly sample from the whole data. This shows the linearity between the data separation in latent space and data space. When these can be separated in latent space, it might be possible to distinguish between the two groups of samples. For comparison, our argument is supported by the latent space projection of CIFAR10, for which the performance was only AUROC of 0.61 on the data space (see Fig. 12).



Fig. 13. Reconstruction image for different latent variable sizes on CIFAR10

VI. CONCLUSION

We proposed an anomaly detection method to reduce bad cycle consistency in BiGAN. This paper assumes that bad cycle consistency might occur due to limitations in the model with respect to preserving enough information from the input image. The proposed method employs a dual-encoder on BiGAN architecture and introduces a preserved information learning mechanism to solve GAN problems as well as to perform anomaly detection. Empirical evaluation on publicly available datasets and brain magnetic resonance imaging anomaly detection showed the performance of the proposed method compared with state-of-the-art methods for separating abnormal samples from the data distribution.

REFERENCES

- [1] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *International Conference on Information Processing in Medical Imaging*, pp. 146–157, Springer, 2017.
- [2] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6479–6488, 2018.
- [3] B. R. Kiran, D. M. Thomas, and R. Parakkal, "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos," *Journal of Imaging*, vol. 4, no. 2, p. 36, 2018.
- [4] M. Haselmann, D. P. Gruber, and P. Tabatabai, "Anomaly detection using deep learning based image completion," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1237–1242, IEEE, 2018.
- [5] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *Journal of Manufacturing Systems*, vol. 48, pp. 144–156, 2018.
- [6] X. Chen and E. Konukoglu, "Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders," *arXiv preprint arXiv:1806.04972*, 2018.
- [7] Y. Kim and S. Choi, "Forward-backward generative adversarial networks for anomaly detection," in *Asian Conference on Machine Learning*, pp. 1142–1155, 2019.
- [8] C. P. Ngo, A. A. Winarto, C. K. K. Li, S. Park, F. Akram, and H. K. Lee, "Fence gan: Towards better anomaly detection," *arXiv preprint arXiv:1904.01209*, 2019.
- [9] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, "Efficient gan-based anomaly detection," *arXiv preprint arXiv:1802.06222*, 2018.
- [10] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville, "Adversarially learned inference," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017.
- [11] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017.
- [12] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "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] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv preprint arXiv:1901.03407*, 2019.
- [14] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103, ACM, 2008.
- [15] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 665–674, ACM, 2017.
- [16] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *SNU Data Mining Center, Tech. Rep.*, 2015.
- [17] F. D. Mattia, P. Galeone, M. D. Simoni, and E. Ghelfi, "A survey on gans for anomaly detection.," *arXiv preprint arXiv:1906.11632*, 2019.
- [18] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "Ganomoly: Semi-supervised anomaly detection via adversarial training," in *Asian conference on computer vision*, pp. 622–637, Springer, 2018.
- [19] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli, "Adversarially learned one-class classifier for novelty detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3379–3388, 2018.
- [20] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8789–8797, 2018.
- [21] M. M. R. Siddiquee, Z. Zhou, N. Tajbakhsh, R. Feng, M. B. Gotway, Y. Bengio, and J. Liang, "Learning fixed points in generative adversarial networks: From image-to-image translation to disease detection and localization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 191–200, 2019.
- [22] A. Krizhevsky *et al.*, "Learning multiple layers of features from tiny images," *Technical Report*, 2009.
- [23] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.
- [24] M. Kistler, S. Bonaretti, M. Pfahrer, R. Niklaus, and P. Büchler, "The virtual skeleton database: an open access repository for biomedical research and collaboration," *Journal of medical Internet research*, vol. 15, no. 11, p. e245, 2013.