

Article

IWGAN: Anomaly Detection in Airport Based on Improved Wasserstein Generative Adversarial Network [†]

Ko-Wei Huang ¹, Guan-Wei Chen ¹, Zih-Hao Huang ¹ and Shih-Hsiung Lee ^{2,*}

¹ Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan

² Department of Intelligent Commerce, National Kaohsiung University of Science and Technology, Kaohsiung 82444, Taiwan

* Correspondence: shlee@nkust.edu.tw

[†] This paper is an extended version of the paper “Anomaly Detection in Airport based on Generative Adversarial Network for Intelligent Transportation System” published in 2022 IEEE International Conference on Consumer Electronics—Taiwan (ICCE-TW), Taipei, Taiwan, 6–8 July 2022.

Abstract: Anomaly detection is an important research topic in the field of artificial intelligence and visual scene understanding. The most significant challenge in real-world anomaly detection problems is the high imbalance of available data (i.e., non-anomalous versus anomalous data). This limits the use of supervised learning methods. Furthermore, the abnormal—and even normal—datasets in the airport field are relatively insufficient, causing them to be difficult to use to train deep neural networks when conducting experiments. Because generative adversarial networks (GANs) are able to effectively learn the latent vector space of all images, the present study adopted a GAN variant with autoencoders to create a hybrid model for detecting anomalies and hazards in the airport environment. The proposed method, which integrates the Wasserstein-GAN (WGAN) and Skip-GANomaly models to distinguish between normal and abnormal images, is called the Improved Wasserstein Skip-Connection GAN (IWGAN). In the experimental stage, we evaluated different hyper-parameters—including the activation function, learning rate, decay rate, training times of discriminator, and method of label smoothing—to identify the optimal combination. The proposed model’s performance was compared with that of existing models, such as U-Net, GAN, WGAN, GANomaly, and Skip-GANomaly. Our experimental results indicate that the proposed model yields exceptional performance.

Keywords: anomaly detection; generative adversarial networks; deep learning

check for
updates

Citation: Huang, K.-W.; Chen, G.-W.; Huang, Z.-H.; Lee, S.-H. IWGAN: Anomaly Detection in Airport Based on Improved Wasserstein Generative Adversarial Network. *Appl. Sci.* **2023**, *13*, 1397. <https://doi.org/10.3390/app13031397>

Academic Editors: Jan Egger and José Salvador Sánchez Garreta

Received: 2 December 2022

Revised: 31 December 2022

Accepted: 18 January 2023

Published: 20 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With recent advances in artificial intelligence (AI), several scholars have begun to invest in deep learning research. AI represents the use of mathematical methods to enable computers to mimic human intelligence by performing inferential, predictive, perceptive, and social tasks. AI has previously exhibited varying degrees of success in image recognition [1], voice recognition [2], natural language processing [3], expert systems [4], and automatic planning [5]. A common, everyday example of AI is the automatic driving system, which encompasses image recognition, object detection, target recognition, semantic segmentation, and other imaging technologies [6].

Owing to the exceptional performance achieved by deep learning and AI, many deep learning models have recently been developed [7,8]. The defining characteristic of deep learning is the use of an artificial neural network architecture to perform training and then accomplish prediction, classification, identification, and other tasks. The architecture of a neural network resembles that of the human brain, wherein it comprises neurons. These architectures are highly diverse, with examples including the multilayer perceptron

(MLP) [9], feedforward neural network (FF) [10], recurrent neural network [11], long short-term memory [11], autoencoders (AE) [12], variational AE (VAE) [13], automatic noise reduction encoders (Denoising, DAE), convolutional neural networks (CNN) [14], and generative adversarial networks (GANs) [15]. GANs, which are typically used to generate images, were proposed by Ian Goodfellow in 2014 [15], representing a breakthrough in the learning of unsupervised neural networks and rapidly becoming a prevalent research topic. The current applications of GANs include style transfer [16], as well as mammography, which is used to generate highly accurate digital breast images [17].

Although deep learning has an overwhelmingly positive impact on data classification, detection, and analysis, two challenging factors frequently manifest in the anomaly detection task: the insufficient quantity of normal and abnormal data during the training phase. These issues are difficult to address in the context of supervision learning. The two-stream neural network proposed by Waseem Ullah et al. [18] performs a two-stage model learning strategy. First, a lightweight convolutional neural network is employed on resource-constrained IoT devices to classify events as normal or abnormal. Second, the abnormal data is through bi-directional long short-term memory (BD-LSTM) for furthering their respective anomaly classification. A GAN and AE were adopted in this study to learn the form of the data latent distribution to generate a new dataset. Thus, the increase and variety of data are expected to improve the model's anomaly detection performance. The proposed method, which integrates the Wasserstein-GAN (WGAN) and Skip-GANomaly models to distinguish between normal and abnormal images, is called the Improved Wasserstein Skip-Connection GAN (IWGAN). In the experimental stage, we evaluated different hyper-parameters—including the activation function, learning rate, decay rate, training times of discriminator, and method of label smoothing—to identify the optimal combination. Consequently, the IWGAN can generate high-quality images, thereby increasing the success rate of abnormal image detection. The contributions of this study are as follows:

- The proposed IWGAN model, which combines WGAN and Skip-GANomaly, resolves the issues posed by training difficulty and model collapse.
- We optimized the training parameters of IWGAN, such as the LeakyReLU activation layer, decay learning rate, training times, and label smoothing.
- The proposed model was evaluated using the Fréchet Inception Distance (FID) and Area Under Curve (AUC) values. The experimental results indicate superior performance to that of existing models, such as U-Net, GAN, WGAN, GANomaly, and Skip-GANomaly.

The remainder of this paper is organized as follows. Section 2 serves as a summary of the related work. In Section 3, we discuss the proposed network's overall architecture, as well as the methods used to overcome the training challenges inherent to GANs. Section 4 presents the experimental results and a discussion of our work. Finally, Section 5 concludes the paper.

2. Related Works

2.1. AutoEncoder, AE

Although the AE was proposed in 1988, the calculation of high-dimensional data was complex and difficult to optimize at the time. In 2006, Hinton et al. [19] employed gradient descent as an optimization tool to produce an abstract representation of original sample features, thereby improving the reduction in feature dimensionality. Henceforth, the AE method has attracted considerable scholarly attention. The defining characteristic of an AE is the use of two networks: an encoder and a decoder. The encoder compresses the image to reduce dimensionality while retaining the main features and the decoder restores the image to its original form. At the end of training, the AE obtains a low-dimensional vector representing the input data in the hidden layer. The optimization objective is to minimize the gap between the input and reconstructed images. The overall process is therefore an unsupervised learning method for representing the features of input images.

Waseem Ullah et al. [20] used an autoencoder to extract spatially optimal features and forward them to the echo state network to obtain a single spatiotemporal information-aware feature vector. At the same time, this feature vector is fused with 3D convolution features to achieve an intelligent dual-stream convolution neural network-based framework for anomaly detection. This study shows that autoencoders can effectively learn features in anomaly events. U-Net [21,22], an AE variant proposed by Ronneberger et al. in 2015, is regarded as one of the best models for image segmentation in biomedical imaging. U-Net is based on a CNN framework that uses each pixel for classification and its defining feature is its U-shaped architecture. We adopted the AE to learn the distribution of latent data, generate a new dataset, and compare it with the experimental dataset.

2.2. Generative Adversarial Network, GAN

The GAN [23] is an unsupervised method that constructs a model through two neural networks: a generator and discriminator. The underlying concept, along with its many variations, represents one of the most innovative ideas in machine learning over the last decade. GANs are most commonly used to generate images, as in the case of CycleGAN [24] for style conversion, GauGAN [25] for automatic painting, DeepFake for face changing, and HoloGAN [26] for full-angle image generation. There are also applications in the medical [2], semiconductor [27], astronomy [28], fashion advertising, and other major fields, with an extensive scope of applicability. Within the GAN architecture, the discriminator learns to distinguish between authentic and forged images, whereas the generator attempts to generate forged images to deceive the discriminator. The two networks are trained sequentially. The present study examined the task of using a GAN for anomaly detection.

There are several problems associated with the original GAN architecture. If the discriminator is over-trained, the generator's gradient will vanish more rapidly, rendering the generator useless. Conversely, if the discriminator is poorly trained, the generator's gradient will be inaccurate. Thus, the overall network operates as intended only if the discriminator is trained well, which is difficult to ensure. Another problem inherent to the conventional GAN architecture is the potential mode collapse due to a suboptimal loss function. WGAN [29,30] can effectively solve these problems by substituting the loss function with a smoother Earth-Mover (EM) distance. WGAN yields the following improvements: (1) the last layer of the discriminator removes the sigmoid function; (2) the losses of the generator and discriminator do not affect the logarithmic operator; (3) the discriminator weight is updated in each iteration to limit the maximum and minimum weights; and (4) there is no optimizer based on momentum change, with optimizers such as RMSProp [31] and SGD [31] being used instead.

2.3. GANs in Anomaly Detection

Schlegl et al. [32] developed AnoGAN based on a deep convolutional GAN to learn the information between normal and local anomalies. Zenati et al. [33] improved the GAN encoder by introducing an extra discriminator to ensure cycle consistency. Fast Unsupervised Anomaly Detection with GAN (F-AnoGAN) [34] was based upon AnoGAN and WGAN to achieve anomaly detection. GANomaly [35], which employs the AE and GAN architectures, comprises four sub-models: an encoder, a decoder, a discriminator, and an additional encoder. Accordingly, GANomaly uses three loss functions—adversarial, contextual, and latent—to calculate the distribution of normal images, thereby learning to distinguish between normal and abnormal datasets as shown in Equation (1). Skip-GANomaly [36] employs the same principle as GANomaly, except that the generator is added to the skip connection, the extra encoder is deleted, and the last convolutional layer of the discriminator is regarded as the encoder. The loss function of Skip-GANomaly is similar to that of GANomaly.

$$\mathcal{L} = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{con}\mathcal{L}_{con} + \lambda_{lat}\mathcal{L}_{lat} \quad (1)$$

where λ_{adv} , λ_{con} , and λ_{lat} are the weights for the three losses, respectively.

3. Proposed Model

The following section discusses the proposed network architecture and core technologies as well as details regarding the internal architecture of each subnetwork. Unlike the traditional GAN architecture, IWGAN employs WGAN and collaborates with Skip-GANomaly through a fusion-network structure.

3.1. Generator Architecture

The generator can be divided into two parts joined by a skip connection: an encoder and a decoder. The skip connection bridges the deep feature of the decoder with the shallow feature of the encoder, allowing the encoder to refer to the decoder when extracting image features after convolution, which ensures a higher-quality restoration. In the small structure of the entire network, the batch normalization layer [37] is used for normalization operations so that the entire network's gradient layer does not vanish easily. LeakyReLU [38,39] was adopted as each network's activation layer in place of ReLU[40] and Adam [41] was employed as the optimizer. A dynamic learning rate strategy was implemented to halve the learning rates among 500, 750, 875, and 950 training iterations. The overall generator architecture is illustrated in Figure 1.

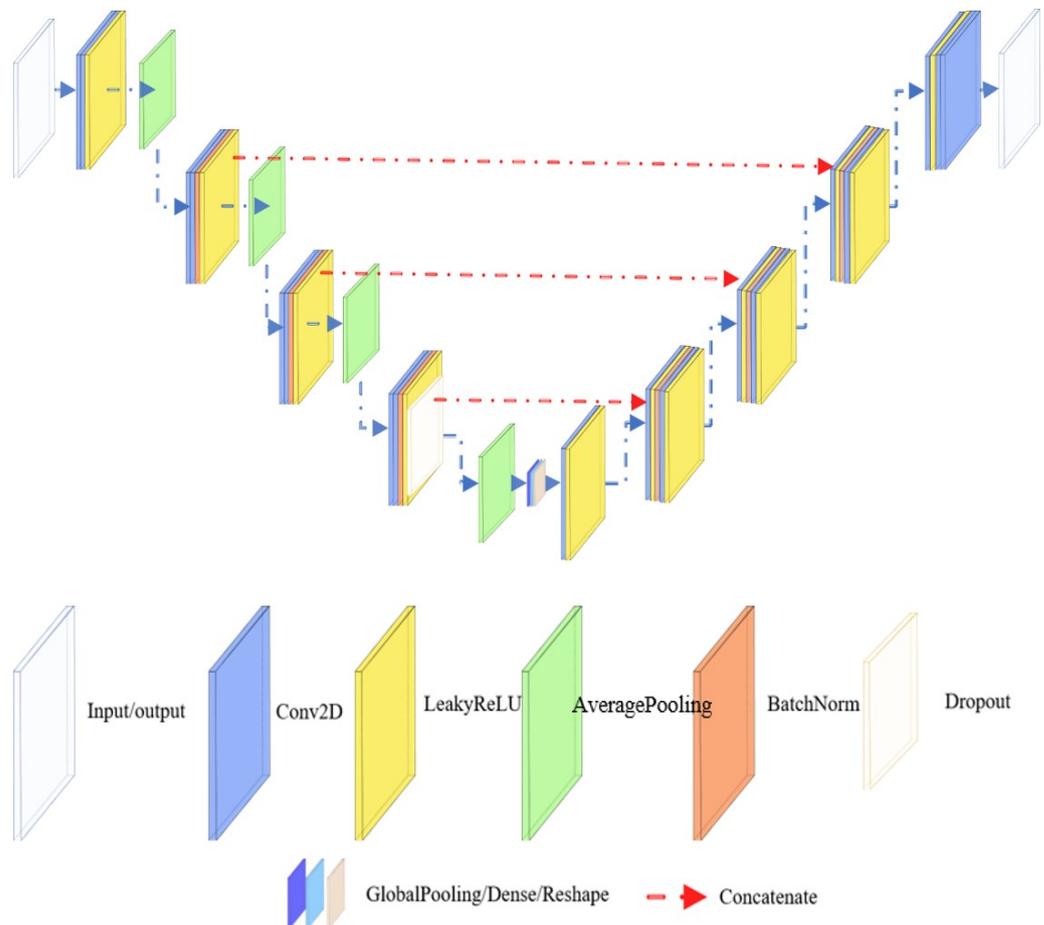


Figure 1. Generator architecture.

3.2. Discriminator Architecture

The discriminator architecture developed in this study is identical to that of the encoder, with the addition of two fully-connected layers. The first of these layers is connected to global pooling (GlobalMaxPooling2D), with 100 connection points used for feature extraction. This layer corresponds to the feature similarity between the original and generated images to optimize the model loss. The second fully-connected layer is represented by one neuron, which reflects the output of determining the falsification of

an image. The primary difference between the proposed architecture and that of WGAN is that the sigmoid layer is replaced by the LeakyReLU layer, as shown in Figure 2. In addition, we adopted the loss function used in Skip-GANomaly, which combines three loss values: adversarial, contextual, and latent. Adversarial loss is used to increase the reconstruction ability regarding normal images, contextual loss guides the model to learn contextual information and sufficiently capture the data distribution, and latent loss helps generate realistic and contextually similar images. During the training phase, the model is able to correctly reconstruct normal samples and incur a high loss for the reconstruction of abnormal samples, thereby improving the efficiency of anomaly detection.

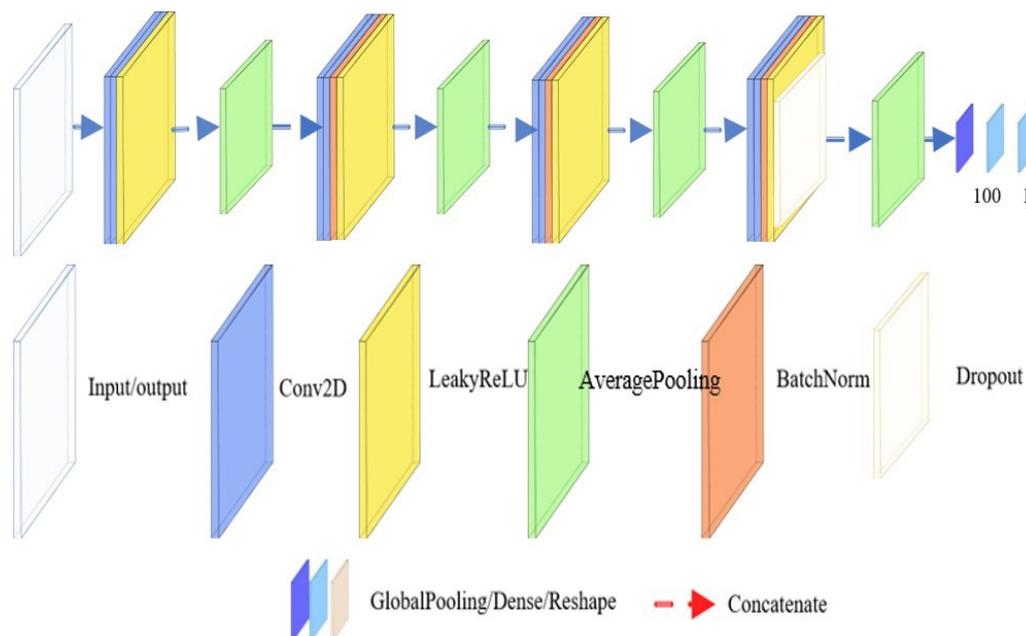


Figure 2. Discriminator architecture.

3.3. Image Normalization

We normalized all the images' pixels from a $[0, 255]$ range to a $[-1, 1]$ range, as the neural network performs a weighted inner product on each pixel of the input image during forward propagation. A wider range leads to a significant increase in computational time, causing the model to converge slowly during backpropagation. Another reason behind normalization is the distance between image samples. If the range of feature points per pixel is particularly wide, the result may be inaccurate. Therefore, the pixels were normalized [42] to improve the model accuracy.

3.4. Unilateral Label Smoothing

Hard labels may cause overfitting during training, particularly when the number of training samples is relatively small. Label smoothing [43] can enhance the model generalizability, alleviate the problem of overfitting, and serve as a preventive measure against noise. Furthermore, it increases the amount of feature information learned by the model, which is beneficial for distinguishing relationships between classes within the data. Szegedy [44] et al. demonstrated this method's effectiveness for classification using the weighted average of hard labels, as well as the uniform distribution on the labels as soft labels. As a method to improve the performance of neural networks and avoid the overconfidence of the discriminator in real samples, this approach has proven useful across many models. Therefore, the present study adopted and evaluated this approach.

3.5. Proposed Architecture

The proposed architecture integrates WGAN and Skip-GANomaly, as shown in Figure 3. WGAN avoids the various training challenges inherent to the conventional

GAN and uses the EM distance smoothness to completely resolve the vanishing gradient issue. In addition to producing satisfactory results for the anomaly detection task, SkipGANomaly uses three loss functions to improve the generator’s performance in identifying anomalous objects [36]. We also converted the hard labels of the GAN network into smooth soft labels and reduced the optimizer’s learning rate at specific iterations of the training process to determine the optimal weights of the neural network. Furthermore, LeakyReLU was applied in each activation layer to prevent gradient vanishing. In detecting anomaly data, we use the anomaly score that was proposed by [32,33]. We evaluate the new image data x as being normal or abnormal images. The anomaly score is defined by Equation (2).

$$S(x) = \lambda R(x) + (1 - \lambda)\mathcal{L}_{lat}(x) \tag{2}$$

where λ is the weight for controlling the importance of the score function, $S(x)$ is the anomaly score function, and $R(x)$ is the reconstruction score that measured the contextual similarity between the input and generated image. In the testing dataset D_{test} , we will obtain the anomaly score vector S such that $S = \{ S_i : S(x_i), x_i \in D_{test} \}$. Finally, we scale the S_i anomaly scores within the probabilistic range from 0 to 1. The hyperparameters are set as same as the reference [36].

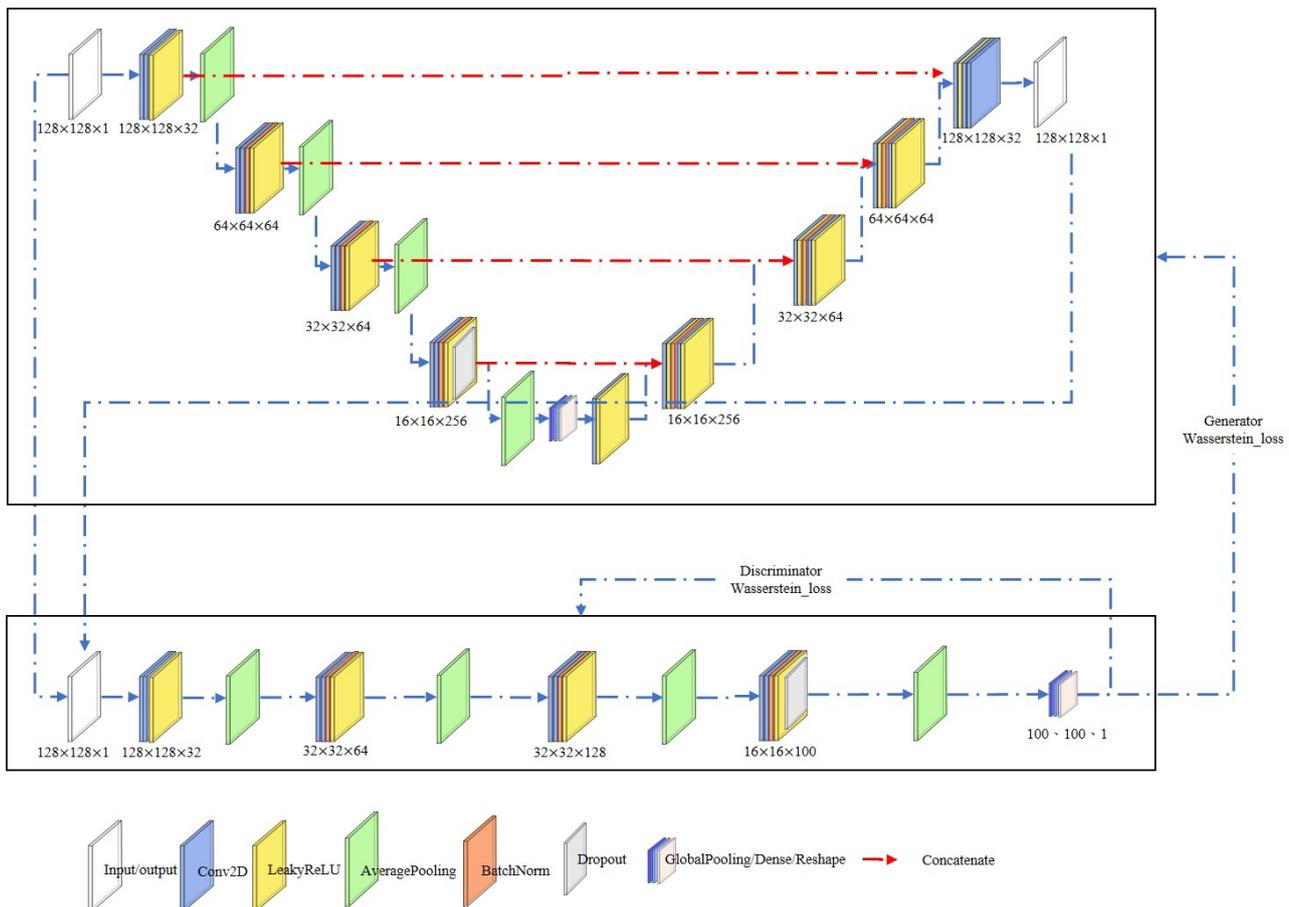


Figure 3. Proposed architecture.

4. Experiment

4.1. Environment Setup and Evaluation Metrics

The experimental environment was a Windows 10 computer with an Intel(R) Core(TM) i7-8500 CPU @ 3.20 GHz and a memory of 16.0 G. All the programs were written in the Python3.7 programming language.

An output value is referred to as a true positive (TP) if it is correctly predicted to be positive, whereas a false positive (FP) occurs when a value is incorrectly predicted to be

positive. Conversely, a false negative (*FN*) is a value incorrectly predicted to be negative, whereas a true negative (*TN*) is a value correctly predicted to be negative. Precision, recall, and *F1-score* were used as the evaluation indices in this experiment. In addition, the AUC of the receiver operating characteristics (*ROC*) was calculated using the true positive rate (*TPR*) and false positive rate (*FPR*).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

$$\text{TPR} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{FPR} = \frac{FP}{TN + FP} \quad (7)$$

4.2. Dataset

Our experiments were conducted on the GDXray+ [45] database, which comprises more than 19,407 X-ray images and is used for research and educational purposes only. The dataset encompasses five types of X-ray images: castings, welds, luggage, natural objects, and environments. Only the luggage category was considered in this study. This category includes 8150 X-ray images over a total of 77 series, such as pocket knives, pistols, and razor blades, as shown in Figure 4.



Figure 4. GDXary+ dataset.

4.3. Data Augmentation

To diversify the data, an augmentation method was applied [46], wherein the original images were randomly rotated, offset in the horizontal or vertical direction, sheared, zoomed in or out, and horizontally flipped. To avoid information loss during augmentation, any missing areas of images were filled via a neighboring interpolation, wherein the nearest pixel's value was used as a supplementary pixel. Either GAN or the augmentation method would be used to generate the new data and increase the performance of the deep neural network efficiently.

4.4. Difference between ReLU and LeakyReLU

This study evaluated the effectiveness of ReLU and LeakyReLU on the proposed model, under a fixed learning rate and the use of hard labels. Using ReLU as the activation layer, the images that generated over 900 epochs exhibited high quality in the later stages, although the model performed poorly in the early stages of training. The results are shown in Figures 5 and 6.

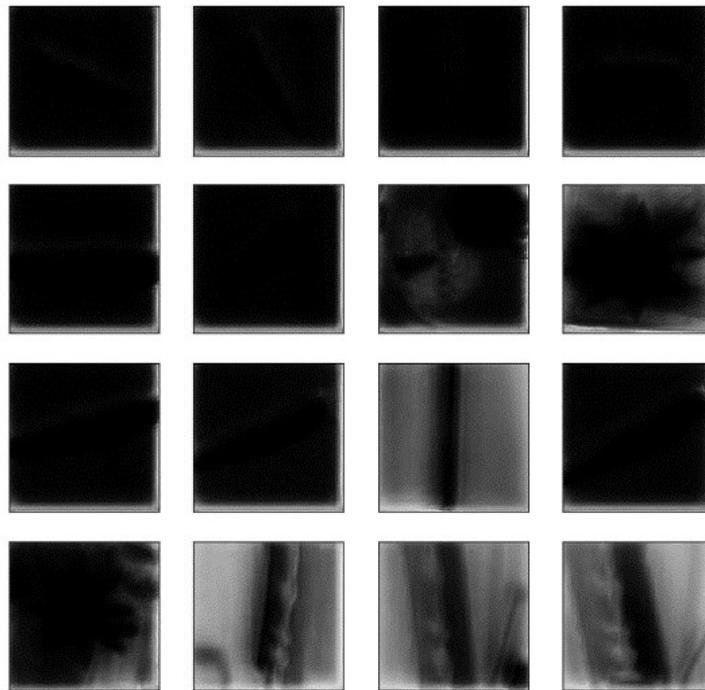


Figure 5. Image generated using ReLU activation layer after Epoch 1.

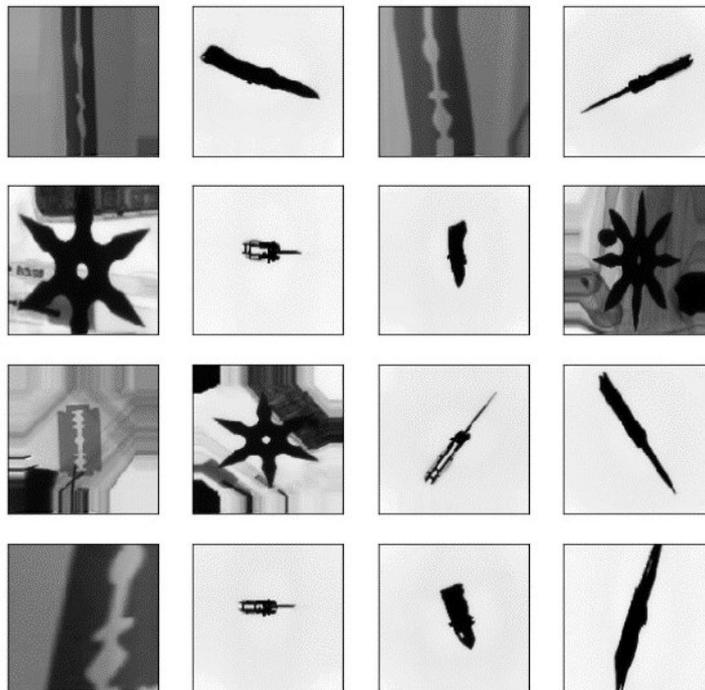


Figure 6. Image generated using ReLU activation layer after Epoch 900.

The anomaly scores were calculated using the Skip-GANomaly evaluation method after training. All the test set images contained anomalous and normal images. The scatter points shown in Figure 7 illustrate the distribution of anomaly scores for all the images. Here, the red dots represent anomalous data and the blue dots represent normal data.

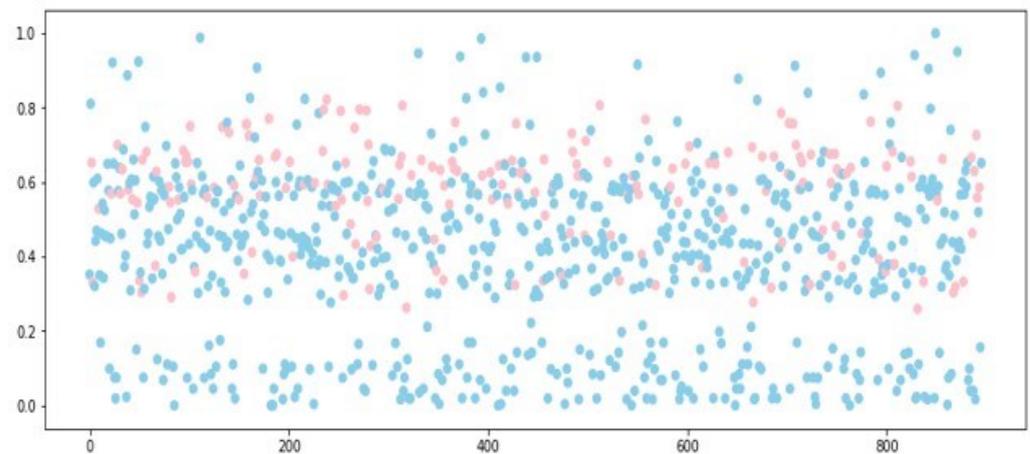


Figure 7. Anomaly score point scatter using ReLU as an activation layer.

Using LeakyReLU as the activation function, the model exhibited a similar increase in quality over 900 epochs.

The distribution of abnormal images and the distribution density of normal images are shown in Figure 8. LeakyReLU evidently performed better than ReLU. Both activation layers were used for ten training sessions to draw a box-and-whisker plot to ensure reliability, as shown in Figure 9. Table 1 shows all the AUC values corresponding to the two activation layers after 10 training cycles. Although LeakyReLU is slightly less stable than ReLU, it achieved higher maximum values, indicating superior performance.

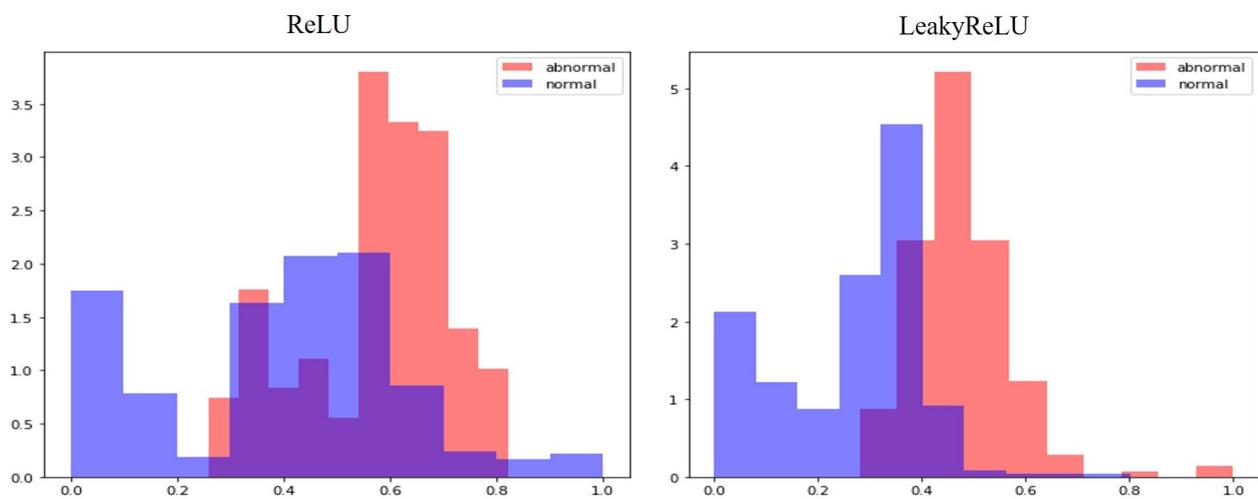


Figure 8. Distribution density map of different activation functions.

Table 1. The AUC value comparison of using LeakyReLU and RELU layers.

	std	min	Q1	Q2	Q3	Max	IQR
ReLU	0.0948	0.6100	0.7200	0.7840	0.8271	0.9048	0.1071
LeakyReLU	0.1521	0.5228	0.6444	0.7942	0.8745	0.9392	0.2301

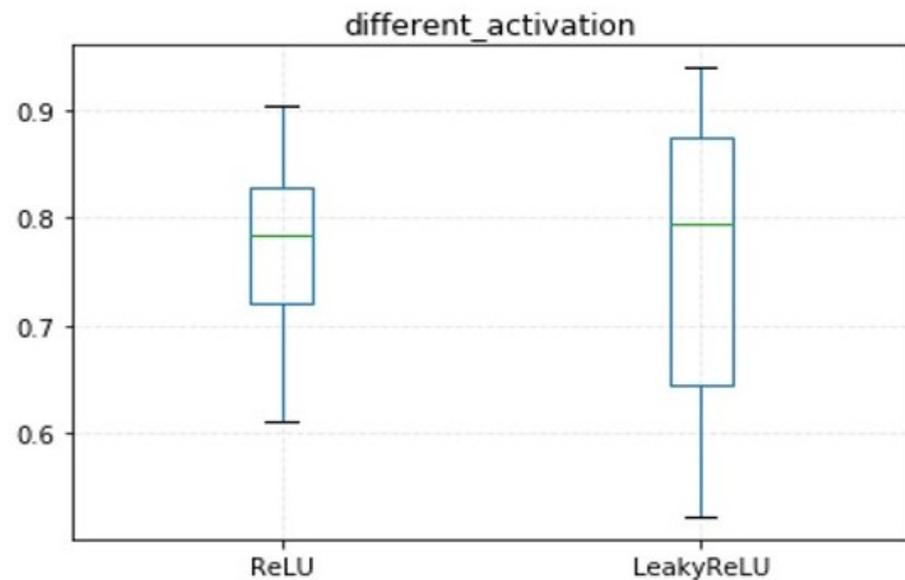


Figure 9. AUC average box and whisker after model training 10 times.

4.5. Difference in Learning Rate

To evaluate whether the learning rate decay [47] is effective within the proposed model, the learning rate decay ratio was evaluated at 0.1 and 0.5, with the ReLU activation layer and hard labels. As shown in Figure 10, the learning rate decayed by a factor of 0.1 every 100 iterations to avoid missing weights closer to the optimal convergence point. Under this decay rate, the images generated over 900 epochs indicate that the proposed model performs poorly in the early stages of the training process. However, when comparing these results with Figure 5, the learning rate decay evidently yields improved performance, as shown in Figure 11.

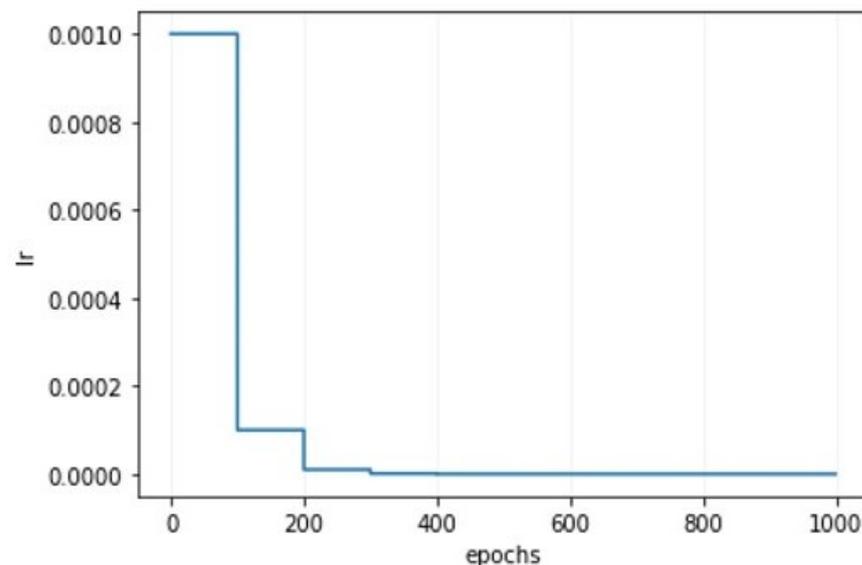


Figure 10. Learning rate with decay rate 0.1.

A learning rate decay rate of 0.5 applied every 100 iterations is shown in Figure 12. Although the network performed poorly in the early stages of the training process, it successfully generated high-quality images in later stages, as shown in Figure 13. The distribution density plots corresponding to different learning rates are shown in Figure 14. The AUC values of the ten training sessions were drawn into a box-and-whisker plot to

ensure reliability at learning rate decay rates of 0.1 and 0.5. The decay rate of 0.1 yields superior results, as shown in Figure 15.

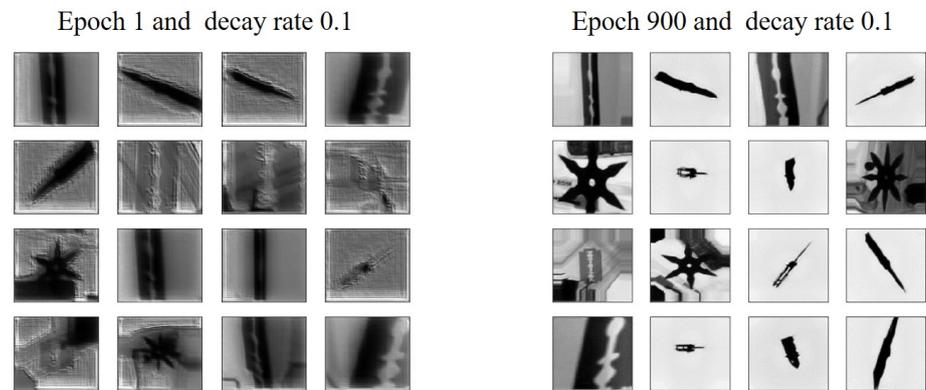


Figure 11. Generated images of learning rate with decay rate 0.1.

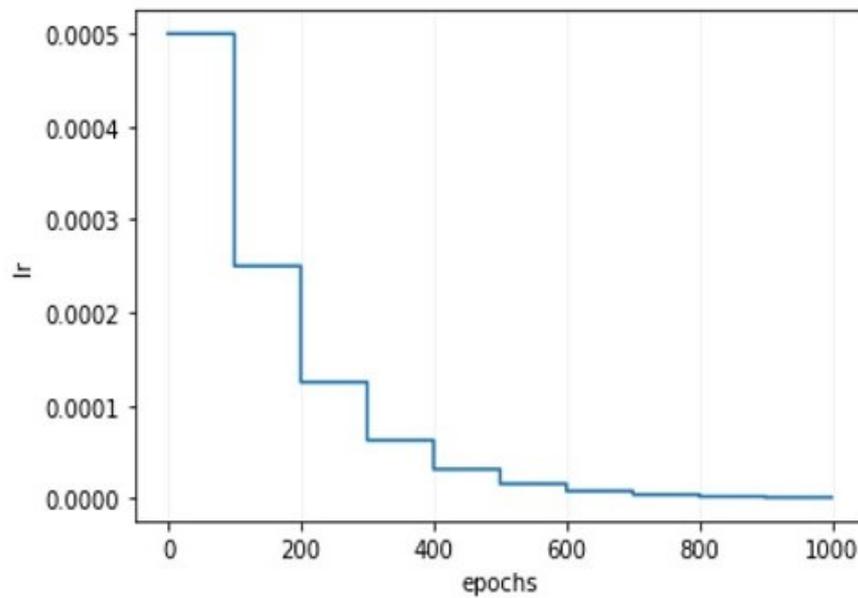


Figure 12. Learning rate with decay rate 0.5.

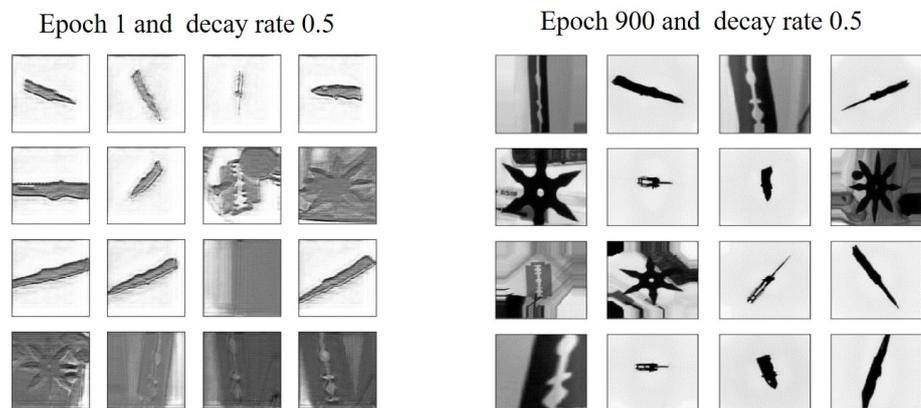


Figure 13. Generated images of learning rate with decay rate 0.5.

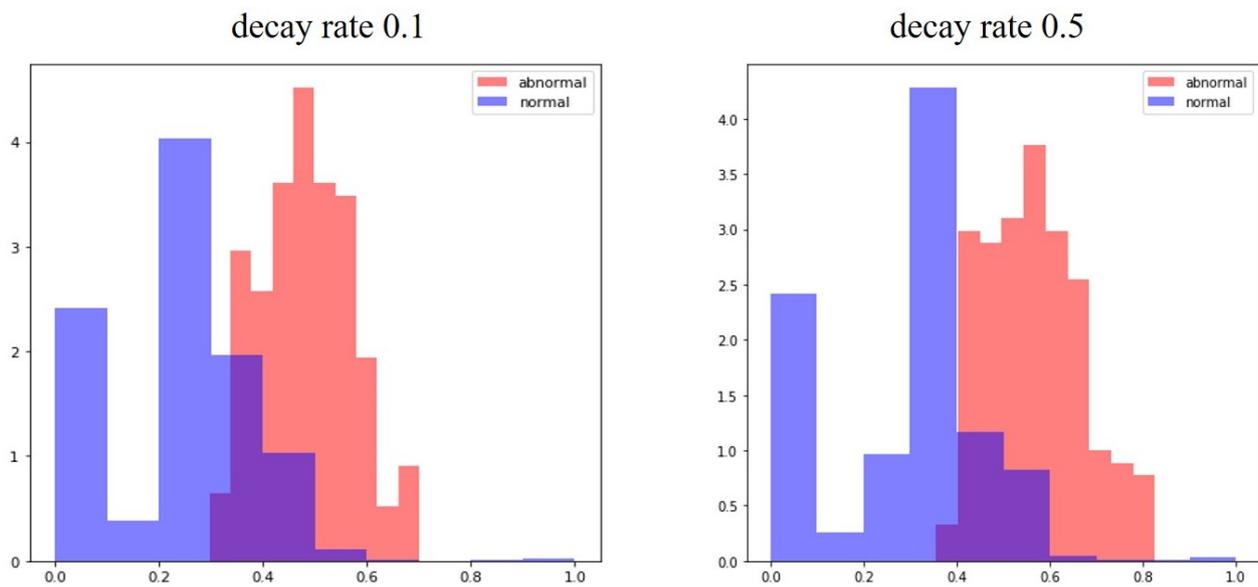


Figure 14. Distribution density map of different decay rates.

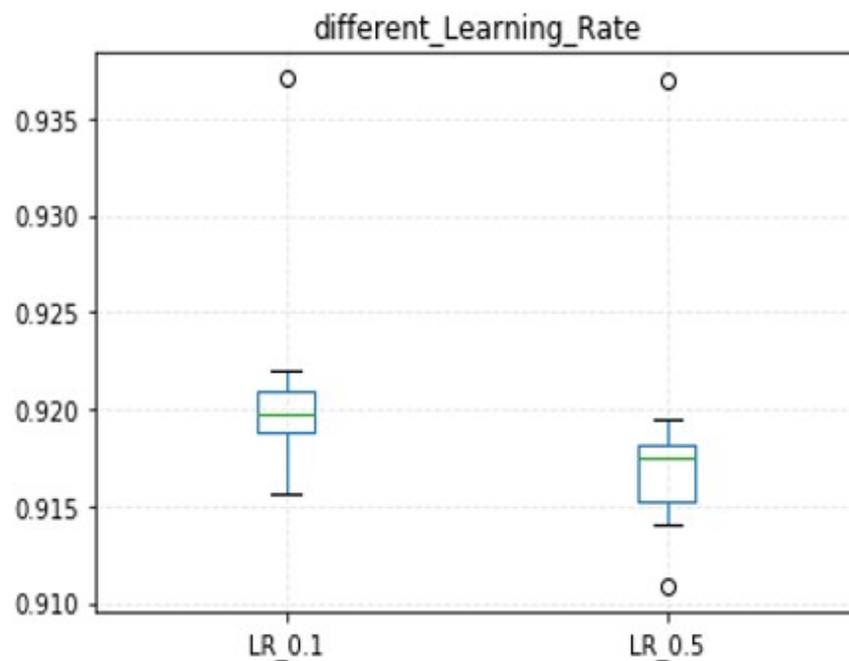


Figure 15. AUC average box and whisker after model training 10 times with different decay rates.

4.6. Training Iterations for Discriminators

We evaluated the most effective number of training iterations for the discriminator by alternating that number between 3 and 5. The five-iteration setup represents a parameter specified in the WGAN paper. Both experiments used a ReLU activation layer, fixed learning rate, and hard labels. After three training iterations, the network performed poorly in the early stages of the training process. However, the generated images were of a higher quality in the later stages. After five training iterations, although the network still performed poorly in the early training stages, it was an improvement compared to its performance after three iterations. Likewise, the network generated higher-quality images at later stages. Both iteration settings were used in 20 training sessions to draw box-and-whisker plots. The results indicate that the use of five training iterations yielded substantially improved performance, as shown in Figure 16.

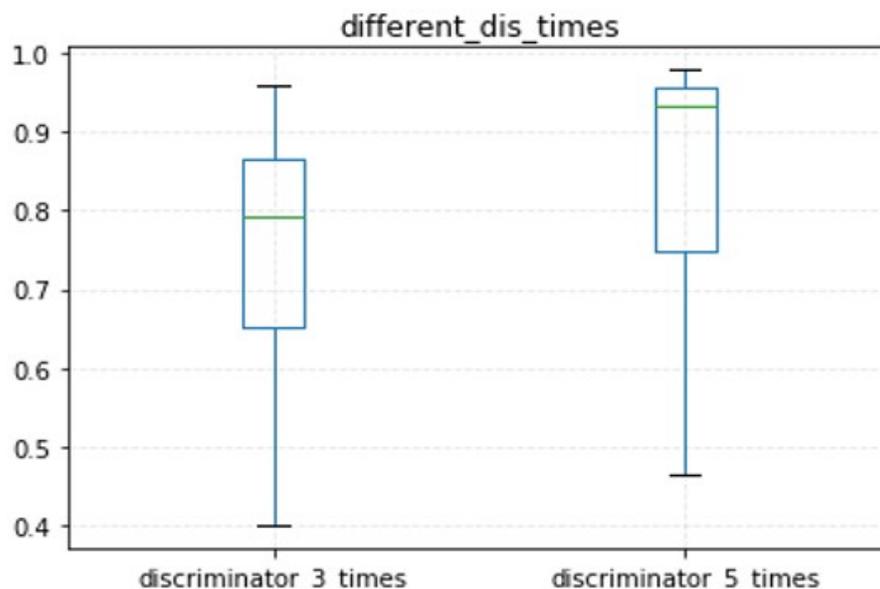


Figure 16. AUC average box and whisker after model training 20 times with different training times.

4.7. Smoothing Label

This section evaluates the impact of label smoothing in the proposed model. The experiment was performed using unilateral and bilateral label smoothing, with a ReLU activation layer and fixed learning rate being used for both settings. Under unilateral label smoothing, the images that generated over 900 epochs indicate that the network performed poorly in the early stages of training. However, the performance was improved over the case of using a simple hard label and higher-quality images were generated in later stages. Under bilateral label smoothing, although the network still performed poorly in the early stages, there was a clear improvement. Likewise, excellent-quality images were generated in the later stages. Table 2 lists all the AUC values over 10 training sessions with unilateral and bilateral training. According to the AUC values in Table 2, unilateral labels are better than double labels.

Table 2. The AUC value comparison of using unilateral and bilateral labels.

	std	min	Q1	Q2	Q3	Max	IQR
Single Label	0.1100	0.6050	0.7637	0.8127	0.8907	0.9717	0.1270
Double Label	0.1400	0.5197	0.6553	0.7995	0.8734	0.9373	0.2281

4.8. Discussion and Analysis

Although the parametric analysis discussed in the previous sections is not fully interpretable, each parameter had an impact on the overall performance during training. The box-and-whisker plots corresponding to different activation layers indicate that although ReLU is more stable than LeakyReLU, the latter produced superior maximum and average AUC values. The use of a decaying learning rate has been demonstrated to stabilize the model accuracy, with a rate of 0.1 yielding superior performance to a rate of 0.5. The discriminator was trained repeatedly to effectively reduce false positives, with five iterations of training demonstrating superior results over three iterations of training. Furthermore, the dataset labels were smoothed, with unilateral smooth labeling producing superior results compared to bilateral or hard labeling. Under optimal parameters, although the images generated in the first training epoch are of a somewhat higher quality, they are still not sufficiently good. However, high-quality images were generated after 900 epochs, as shown in Figure 17.

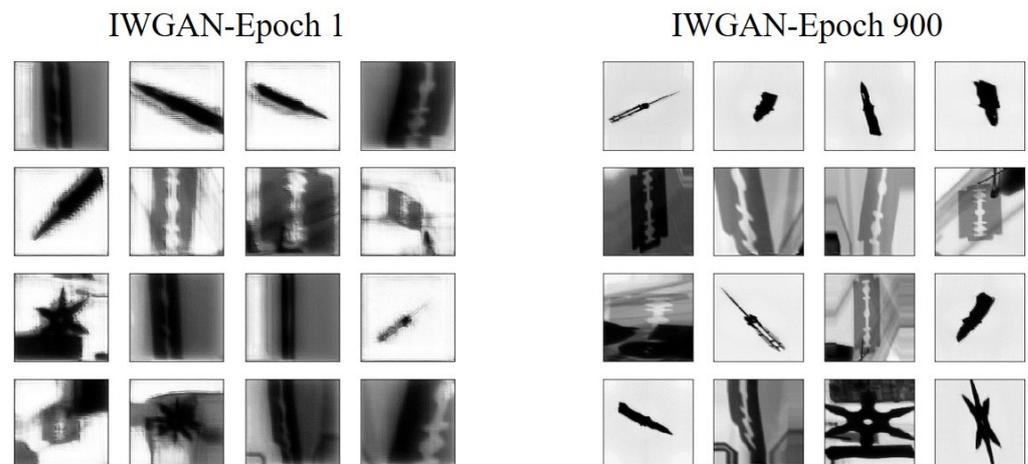


Figure 17. Generated images of IWGAN.

4.9. Evaluation

Two evaluation methods are generally used to evaluate the quality and diversity of images generated by GANs: the inception score (IS) [48] and the FID [49]. However, IS exhibits a disadvantage wherein certain types of images lead to an incorrect IS score. Accordingly, we employed the FID to evaluate the distance between the two images' data distributions, wherein a smaller FID value indicates higher quality and diversity of the generated image. We evaluated the quality of images generated using U-NET, GAN, WGAN, and GANomaly, as summarized in Table 3. The proposed model exhibits a significant improvement in performance. Table 4 lists the AUC values and F1 scores obtained by each model, likewise indicating the superior performance of IWGAN.

Table 3. FID comparison.

Methods	Fréchet Inception Distance
U-Net	116.017
GAN	325.596
WGAN	369.543
GANomaly	304.231
Skip-GANomaly	91.804
IWGAN-ReLU	73.295
IWGAN	56.421

Table 4. The AUC value and F1-score of each method.

Methods	AUC Value	F1-Score
GAN	0.79	0.84
WGAN	0.84	0.83
GANomaly	0.75	0.79
Skip-GANomaly	0.97	0.81
IWGAN-ReLU	0.69	0.81
IWGAN	0.95	0.96

For the evaluation of time complexity, this paper uses floating-point operations per second (FLOPs) for comparison, as shown in Table 5. The method proposed in this paper requires a large amount of computation in execution. Compared with Skip-GANomaly, IWGAN has additional 20% of computing resources. However, compared with Skip-GANomaly, IWGAN improves the performance by 38.5% and 19% in the evaluation indicators of the FID and F1-score, respectively. In the case of unconstrained resources, the method proposed in this paper can obtain better results of anomaly detection.

Table 5. The FLOPs of each method.

Methods	FLOPs(G)
GAN	2.98
WGAN	3.11
GANomaly	3.04
Skip-GANomaly	3.27
IWGAN-ReLU	3.92
IWGAN	3.92

5. Conclusions

The anomaly detection task involves two major challenges: insufficient data and insufficient abnormal data. This paper proposes the IWGAN network, whose architecture combines WGAN and Skip-GANomaly. The WGAN subnetwork mitigates the issues of training difficulty and mode collapse, while the excellent detection ability of Skip-GANomaly resolves the insufficient data problem. In addition, we found an optimal combination of training parameters for IWGAN: the LeakyReLU activation layer, a decay learning rate of 0.1, five training iterations, and unilateral label smoothing. The proposed model was evaluated using the FID, with experimental results exhibiting significant improvements in performance. The AUC value of the overall network for the GDXray+ discriminant gun samples reached an average of approximately 0.95, which is excellent in terms of a single-generation network. In future work, we will introduce an attention model mechanism to extend the proposed model's applicability for all anomaly detection tasks. In addition, we plan to adopt a neural architecture search to optimize the hyper-parameters (HPO) automatically.

Author Contributions: Conceptualization and methodology, K.-W.H., G.-W.C., Z.-H.H. and S.-H.L.; Formal analysis, K.-W.H. and G.-W.C.; Supervision, K.-W.H. and S.-H.L.; Project administration, K.-W.H. and S.-H.L.; Writing—original draft preparation, K.-W.H. and S.-H.L.; Writing—review and editing, S.-H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research is financially supported by the Ministry of Science and Technology of Taiwan (under grants No. 110-2222-E-992-006- and No. 111-2221-E-992-070-MY2).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: The authors would like to thank Editor and anonymous Reviewers for their valuable reviews.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Guo, Y.; Wang, H.; Hu, Q.; Liu, H.; Liu, L.; Bennamoun, M. Deep learning for 3d point clouds: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 4338–4364. [[CrossRef](#)] [[PubMed](#)]
- Wang, Z.; Chen, L.; Wang, L.; Diao, G. Recognition of Audio Depression Based on Convolutional Neural Network and Generative Antagonism Network Model. *IEEE Access* **2020**, *8*, 101181–101191. [[CrossRef](#)]
- Otter, D.W.; Medina, J.R.; Kalita, J.K. A survey of the usages of deep learning for natural language processing. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 604–624. [[CrossRef](#)]
- Islam, R.U.; Hossain, M.S.; Andersson, K. A Deep Learning Inspired Belief Rule-Based Expert System. *IEEE Access* **2020**, *8*, 190637–190651. [[CrossRef](#)]
- Lv, L.; Zhang, S.; Ding, D.; Wang, Y. Path planning via an improved DQN-based learning policy. *IEEE Access* **2019**, *7*, 67319–67330. [[CrossRef](#)]
- Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:1511.07122.
- Swarna, S.R.; Boyapati, S.; Dutt, V.; Bajaj, K. Deep Learning in Dynamic Modeling of Medical Imaging: A Review Study. In Proceedings of the 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 3–5 December 2020; pp. 745–749.

8. Zhong, G.; Zhang, K.; Wei, H.; Zheng, Y.; Dong, J. Marginal deep architecture: Stacking feature learning modules to build deep learning models. *IEEE Access* **2019**, *7*, 30220–30233. [[CrossRef](#)]
9. Wilson, E.; Tufts, D.W. Multilayer perceptron design algorithm. In Proceedings of the IEEE Workshop on Neural Networks for Signal Processing, Valais, Switzerland, 4–6 September 1994; pp. 61–68.
10. Bebis, G.; Georgiopoulos, M. Feed-forward neural networks. *IEEE Potentials* **1994**, *13*, 27–31. [[CrossRef](#)]
11. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [[CrossRef](#)]
12. Baldi, P. Autoencoders, unsupervised learning, and deep architectures. *ICML Workshop Unsupervised Transf. Learn.* **2012**, *27*, 37–49.
13. Li, X.; Zhao, Z.; Song, D.; Zhang, Y.; Niu, C.; Zhang, J.; Li, J. Variational Autoencoder based Latent Factor Decoding of Multichannel EEG for Emotion Recognition. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, 18–21 November 2019; pp. 684–687.
14. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
15. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Bengio, Y. Generative adversarial networks. *arXiv* **2014**, arXiv:1406.2661.
16. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
17. Korkinof, D.; Rijken, T.; O’Neill, M.; Yearsley, J.; Harvey, H.; Glocker, B. High-resolution mammogram synthesis using progressive generative adversarial networks. *arXiv* **2018**, arXiv:1807.03401.
18. Ullah, W.; Ullah, A.; Hussain, T.; Hussain, T.; Muhammad, K.; Heidari, A.A.; Ser, J.D.; Baik, S.W.; Albuquerque, V.H. C.D. Artificial Intelligence of Things-assisted two-stream neural network for anomaly detection in surveillance Big Video Data. *Future Gener. Comput. Syst.* **2022**, *120*, 286–297. [[CrossRef](#)]
19. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [[CrossRef](#)] [[PubMed](#)]
20. Ullah, W.; Hussain, T.; Hhan, Z.A.; Haroon, U.; Baik, S.W. Intelligent dual stream CNN and echo state network for anomaly detection. *Knowl.-Based Syst.* **2022**, *253*, 190456. [[CrossRef](#)]
21. Ali, N.; Kirchhoff, J.; Onoja, P.I.; Tannert, A.; Neugebauer, U.; Popp, J.; Bocklitz, T. Predictive Modeling of Antibiotic Susceptibility in E. Coli Strains Using the U-Net Network and One-Class Classification. *IEEE Access* **2020**, *8*, 167711–167720. [[CrossRef](#)]
22. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
23. Wang, C.; Xu, C.; Yao, X.; Tao, D. Evolutionary generative adversarial networks. *IEEE Trans. Evol. Comput.* **2019**, *23*, 921–934. [[CrossRef](#)]
24. 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, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
25. Park, T.; Liu, M.Y.; Wang, T.C.; Zhu, J.Y. GauGAN: Semantic image synthesis with spatially adaptive normalization. In Proceedings of the ACM SIGGRAPH 2019 Real-Time Live, Los Angeles, CA, USA, 28 July–1 August 2019.
26. Nguyen-Phuoc, T.; Li, C.; Theis, L.; Richardt, C.; Yang, Y.L. Hologan: Unsupervised learning of 3d representations from natural images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7588–7597.
27. Wang, J.; Yang, Z.; Zhang, J.; Zhang, Q.; Chien, W.T. K. AdaBalGAN: An improved generative adversarial network with imbalanced learning for wafer defective pattern recognition. *IEEE Trans. Semicond. Manuf.* **2019**, *32*, 310–319. [[CrossRef](#)]
28. Fussell, L.; Moews, B. Forging new worlds: High-resolution synthetic galaxies with chained generative adversarial networks. *Mon. Not. R. Astron. Soc.* **2019**, *485*, 3203–3214. [[CrossRef](#)]
29. Arjovsky, M.; Bottou, L. Towards principled methods for training generative adversarial networks. *arXiv* **2017**, arXiv:1701.04862.
30. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 214–223.
31. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2016**, arXiv:1609.04747.
32. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Schmidt-Erfurth, U.; Langs, G. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. In Proceedings of the International Conference on Information Processing in Medical Imaging (IPMI), Boone, NC, USA, 25–30 June 2017.
33. Zenati, H.; Foo, C.S.; Lecouat, B.; Manek, G.; Chan-draxekhar, V.R. Efficient Gan-Based Anomaly Detection. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
34. Park, C.; Lim, S.; Cha, D.; Jeong, J. Fv-AD: F-AnoGAN Based Anomaly Detection in Chromate Process for Smart Manufacturing. *Appl. Sci.* **2022**, *12*, 7549. [[CrossRef](#)]
35. Akcay, S.; Atapour-Abarghouei, A.; Breckon, T.P. Ganomaly: Semi-supervised anomaly detection via adversarial training. In Proceedings of the Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018; pp. 622–637.

36. Akçay, S.; Atapour-Abarghouei, A.; Breckon, T.P. Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.
37. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 448–456.
38. Lu, L.; Shin, Y.; Su, Y.; Karniadakis, G.E. Dying relu and initialization: Theory and numerical examples. *arXiv* **2019**, arXiv:1903.06733.
39. Xu, B.; Wang, N.; Chen T.; Li, M. Empirical evaluation of rectified activations in convolutional network. *arXiv* **2015**, arXiv:1505.00853.
40. Agarap, A.F. Deep learning using rectified linear units (relu). *arXiv* **2018**, arXiv:1803.08375.
41. Kingma, F.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
42. LeCun Y.A.; Bottou, L.; Orr, G.B.; Müller, K.R. Efficient backprop. In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 9–48.
43. Müller, R.; Kornblith, S.; Hinton, G. When does label smoothing help. *arXiv* **2019**, arXiv:1906.02629.
44. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
45. Mery, D.; Riffo, V.; Zscherpel, U.; Mondragón, G.; Lillo, I.; Zuccar, I.; Carrasco, M. GDXray: The database of X-ray images for nondestructive testing. *J. Nondestruct. Eval.* **2015**, *34*, 1–12. [[CrossRef](#)]
46. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 1–48. [[CrossRef](#)]
47. Smith, L.N. Cyclical learning rates for training neural networks. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; pp. 464–472.
48. Barratt, S.; Sharma, R. A note on the inception score. *arXiv* **2018**, arXiv:1801.01973.
49. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *arXiv* **2017**, arXiv:1706.08500.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.