GanNeXt: A New Convolutional GAN for Anomaly Detection*

Bowei Pu[1], Shiyong Lan *[1], Wenwu Wang[2], Caiying Yang[1], Wei Pan[1], Hongyu Yang[1], and Wei Ma[1]

[1] College of Computer Science, Sichuan University, Chengdu 610065, China
[2] University of Surrey, Guildford, GU2 7XH, United Kingdom

Abstract. Anomaly detection refers to the process of detecting anomalies from data that do not follow its distribution. In recent years, Transformer-based methods utilizing generative adversarial networks (GANs) have shown remarkable performance in this field. Unlike traditional convolutional architectures, Transformer structures have advantages in capturing long-range dependencies, leading to a substantial improvement in detection performance. However, transformer-based models may be limited in capturing fine-grained details as well as the inference speed. In this paper, we propose a scalable convolutional Generative Adversarial Network (GAN) called GanNeXt. Our design incorporates a new convolutional architecture that utilizes depthwise convolutional layers and pointwise convolutional layers as extension layers. In addition, we introduce skip connections to capture multi-scale local details. Experiments demonstrate that our proposed method achieves a 58% reduction in floating-point operations per second (FLOPs), while outperforming state-of-the-art Transformer-based GAN baselines on CIFAR10 and STL10 datasets. The codes will be available at https://github.com/SYLan2019/GanNeXt.

Keywords: Anomaly Detection· CNN · Generative Adversarial Network.

1 Introduction

Anomaly detection is an increasingly important area in the field of computer vision, and has been widely studied in numerous application domains, such as video surveillance, risk management, and damage detection. Currently, the advanced methods in this field are based on deep learning, as shown in [17,5,9]. However, the performance of these methods is limited by the lack of labeled data. On one hand, collecting abnormal images is challenging due to the imbalanced distribution between normal and abnormal data. On the other hand, defining abnormal data is difficult. In contrast, unsupervised learning algorithms are commonly used to learn the distribution of normal data. They are trained only on normal data and then tested on both normal and abnormal data. If a

* This work was funded by 2035 Innovation Pilot Program of Sichuan University, China. * Corresponding author. E-mail:lanshiyong@scu.edu.cn.

given data point is significantly different from the learned distribution based on a predefined threshold, it will be classified as an abnormal object.

In the field of computer vision, anomaly detection is a challenging task that aims to detect abnormal data caused by environmental changes, device damage or human interference. In recent years, unsupervised learning methods based on Generative Adversarial Networks (GANs) [1] have made significant progress in anomaly detection. In these methods, the generator aims to generate data by learning the normal data distribution, while the discriminator tries to distinguish the generated data from the original data. Among these methods, AnoGAN [14] is the first GAN-based anomaly detection representation learning method. To overcome the limitations of GANs, many researchers have proposed improved GAN models, such as GANomaly [1] that uses an encoder-decoder-encoder network, Skip-GANomaly [2] that uses a U-Net structure, and SAGAN [11] that uses attention modules. In recent years, with the emergence of Vision-Transformer [4], attention modules have been used to improve GAN models such as AnoTrans [20], thus further improving the accuracy of the model. However, due to the use of the transformer-based autoregressive model in its inference stage, the inference speed of the attention-based model is limited.

ConvNeXt [13,18] is a convolutional neural network architecture that draws inspiration from the ideas of Vision Transformer [4] and Swin Transformer [12], utilizing attention mechanisms and window partitioning to learn long-range spatial dependencies within images. In addition, it retains the characteristics of ResNet [6] by using residual connections and group convolutions to enhance feature representation and reduce computational complexity. Through this design, ConvNeXt not only achieves faster inference speed than Vision Transformer and Swin Transformer under the same amount of computations but also achieves higher accuracy on tasks such as image classification, object detection, and semantic segmentation.

In this paper, we propose a new framework for anomaly detection. Inspired by ConvNeXt, we design a purely convolutional network called GanNeXt with a U-shaped architecture that combines the advantages of long range information modeling by using Transformer structures and the higher inference speed of CNN networks. In addition, the special module design allows for different numbers of modules to be used at different depths of the U-Net, enabling a fine-grained network design to improve the capture of image features with deep networks. Notably, to address the issue of long-term dependencies, we propose a special skip-connection mechanism that enhances the reconstruction ability of the U-shaped network.

The experimental results show that the proposed method outperforms the state-of-the-art anomaly detection methods on external tasks such as CIFAR10 [10], STL10 [15], as well as on datasets such as LBOT [11]. The main contributions of this paper are summarized as follows:

-Our paper introduces a novel anomaly detection framework named GanNeXt. The proposed framework combines ConvNeXt modules with existing GAN-

based methods, aiming to overcome the limitations of CNN encoders in GAN-based modeling.

-Our experimental results on three datasets, namely CIFAR10, STL10, and LBOT, demonstrate that our proposed approach outperforms the state-of-the-art CNN-based methods in detecting anomalies. Moreover, our approach requires significantly fewer FLOPs compared to the most advanced methods, with a reduction of up to 58%.

-We design a new Skip-Connect method to improve the reconstruction capability of the decoder.

2 Proposed Method

2.1 Model Overview

The ConvNeXt [13,18] block inherits many important design choices from Transformers, designed to limit computational cost while increasing the receptive field width to learn global features, demonstrating performance improvements over standard ResNets [6]. By leveraging these advantages to improve the inference speed while maintaining high accuracy, we propose GanNeXt (as shown in Fig. 1), based on the general design of ConvNeXt [13], with AnoTran [20] and Skip-Gan Normaly [2]. In addition, we introduce a new type of skip connection for the CNN networks, using GanNeXt blocks (similar to ConvNeXt blocks) to replace self-attention modules, similar to the skip-attention introduced by AnoTran [20] , to enhance the feature representation of the input images. The experiment shows that compared to the skip-attention based on self-attention modules, this skip connection is more efficient. We added batch normalization at the beginning of the input, which reduces the overall shift in each batch of the input data, improving the representation learning of normal data in the generator.

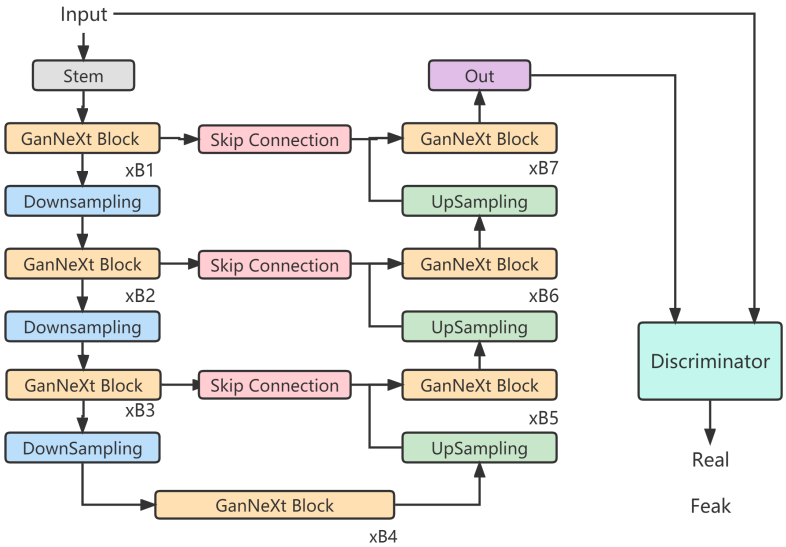
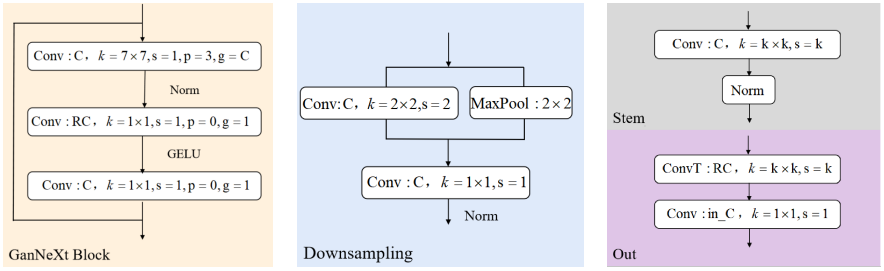


Fig. 1. The structure of our proposed model

As shown in Fig. 1, our model consists of a generator and a discriminator. The performance improvement of our model is due to the deep separable convolution based on the Transformer design, which will be described in Section 2.2. In the downsampling module, we design parallel downsampling modules to improve feature representation, as described in Section 2.3. We introduce an improved skip connection module based on GanNeXt blocks, as described in Section 2.4. The loss function and the standard for computing anomaly scores used in our model will be described in Sections 2.6 and 2.7, respectively. In our model, we use the same discriminator as AnoTran to distinguish the label of the extracted potential representation of input images.



(a) GanNeXt Block (b) Downsample Module (c) Stem and Out Module

Fig. 2. Three important modules of our proposed model

2.2 GanNeXt Block

In GanNeXt Block (see Fig. 2(a)), the first layer consists of a deep convolution layer containing a kernel size of k, followed by normalization and with C output channels. Different normalization methods are used for different tasks, including LayerNorm [3] and BatchNorm [8]. The performance of these two methods varies greatly across different tasks, which will be discussed in Section 3.3. The deep convolution enables large convolution kernels to achieve the large attention window effect of Swin-Transformers [12] while limiting computation, thereby delegating important computational tasks to the expansion layer. The expansion layer uses a dilation rate of R and is activated by GELU [7], similar to the design of the Feedforward layer [16] in Transformers. This allows the network to expand in width and effectively separate kernel size expansion from width expansion in the previous layer.

2.3 DownSampling Module

The downsampling operation is often used to reduce the size and number of parameters in feature maps, in order to decrease computational cost while maintaining model performance. In this paper, we selected parallel MaxPooling and convolution kernels of 2×2 with a stride of 2 for downsampling (see Fig. 2(b)). Downsampling through convolution enables the learning of more diversified feature representations, while max pooling emphasizes the feature representation in

the feature map. We parallelized these two methods and then used a 1x1 convolution to restore the number of channels, enabling our model to learn richer and more accurate feature representations, which is beneficial for image restoration by the decoder. Similar to GanNeXt Block, the normalization method can also be replaced.

2.4 Stem and Out

In the Stem Module (Fig. 2(c)), we employed a strategy similar to that used in [4,12]. Specifically, we used convolution with the same stride and kernel size, followed by normalization with a replaceable Norm layer.

In the Out Module (Fig. 2(c)), we designed a structure similar to the extension layers. We first performed transposed convolution to restore the output image size to match the input size. Next, we increased the channel number according to a given ratio. Finally, we used 1×1 convolution to restore the channel number to match the input.

2.5 Skip Connection

In Unet, skip connections can be used to directly transfer low-level feature map information to high-level feature maps, helping the network capture features at different scales effectively. This can improve model performance and accuracy, while mitigating the problems arising from gradient vanishing and overfitting due to information loss. AnoTran [20] found the U-Generator to be ineffective in capturing some critical local information in feature representation, and proposed the Skip Attention Connection to further improve performance. Inspired by AnoTran and SAGAN [11], we designed skip connections using GanNeXt Block (see Fig. 2(a)) in our method. Our approach is a pure convolutional structure that does not require transformations (such as reshape operation and convolution) as performed in the CBAM [19], allowing direct convolution operations. Via the GanNeXt block in the skip connections (GSC), the generator can obtain fine-grained feature representations and reduce noise and irrelevant information by preserving the original feature map information, making the model more stable and robust.

2.6 Loss Function

We employed the same loss functions Lcon, Llat, and Ladv as used in Skip-GANomaly [2] , which are formulated as follows:

$$L_{adv} = \mathbb{E}_{x\sim p_x}[\log D(x)] + \mathbb{E}_{x\sim p_z}[\log(1 - D(G(x)))] \tag{1}$$

$$L_{lat} = \mathbb{E}_{x\sim p_x} \left\| D(x) - D(G(x)) \right\|_2 \tag{2}$$

$$L_{con} = \mathbb{E}_{x\sim p_x} \left\| x - G(x) \right\|_1 \tag{3}$$

The adversarial loss, denoted as L_{adv}, is a commonly used loss function in GANs for optimizing the generator G and discriminator D. The goal is to make the generated images as realistic as possible through the classification results provided by the discriminator.

The content loss, denoted as L_{con}, is employed to improve the realism of the generated image $G(x)$. It aims to preserve contextual information in the image. The latent loss, denoted as L_{lat}, captures the difference between the latent representations of the input image x and the generated image $G(x)$. It measures the consistency between the latent representations of x and $G(x)$, denoted as $z = D(x)$ and $\hat{z} = D(G(x))$, respectively.

$$L = \lambda_{adv} L_{adv} + \lambda_{con} L_{con} + \lambda_{lat} L_{lat} \tag{4}$$

The overall objective function L is a weighted sum of L_{adv}, L_{con}, and L_{lat}, where the weights λ_{adv}, λ_{con}, and λ_{lat} are chosen empirically in our experiments.

2.7 Anomaly Scores

We compute the anomaly scores of the test images using the methods in [2,20]. The anomaly score of a test image x, denoted as $A(x)$, is calculated as $\lambda R(x) + (1-\lambda)L(x)$. Here, $R(x)$ is the reconstruction score calculated based on x and its generated image x', and $L(x)$ represents the difference between the latent representations of x and x' obtained from the discriminator. The weight λ controls the relative importance of $R(x)$ and $L(x)$ in the calculation of $A(x)$.

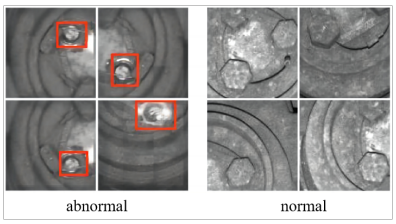
$$A(x) = \lambda R(x) + (1-\lambda)L(x) \tag{5}$$

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

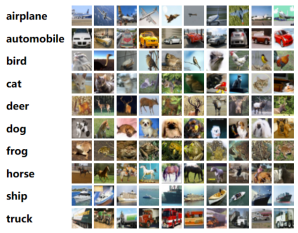
We calculate the original anomaly scores of all the test samples in the test set and represent the set of anomaly scores as a vector A. Then, we standardize each anomaly score to the range $[0, 1]$ using the same method as in [20]. Specifically, the final anomaly score $A'(x)$ of a single test sample x is computed as $(A(x) - \min(A))/(\max(A) - \min(A))$.

3 Experiment

We evaluated our model using a "leave-one-class-out" anomaly detection method with CIFAR10 [10], STL10 [15] and [11] datasets. The area under the Receiver Operating Characteristic (ROC) curve (AUC) was used as the performance metric in the evaluation. AnoTran, SAGAN, and Skip-Anomaly were used as baseline methods for performance comparison.



(a) Some image examples from the LBOT dataset, the red box contains abnormal data, which includes bolts that are either damaged or missing.



(b) Examples of the CIFAR-10 dataset.

Fig. 3. Examples in the dataset.

3.1 Dataset

CIFAR-10: Both Skip-GANomaly [2] and AnoTran [20] used the CIFAR-10 dataset and formulated a leave-one-class-out anomaly detection problem. For comparison purposes, we also utilized this dataset. Similar to [2,11,20], we divided the CIFAR-10 dataset into eight distinct categories, each of which contains 45,000 normal training samples, 9,000 normal testing samples, and 6,000 abnormal testing samples. One of the categories was defined as the anomaly class, while the remaining categories were defined as the normal classes.

STL10: The STL-10 dataset is an image recognition dataset used for developing unsupervised feature learning, deep learning, and self-supervised learning algorithms. As with CIFAR-10, the difference between them lies in the fact that STL-10 has less labeled training data per class. In addition, the images in STL-10 are of a resolution of 96×96 pixels. The training methodology used on the STL-10 dataset is similar to that used on CIFAR-10.

LBOT: The LBOT dataset, as reported by reference [11], is an image dataset used for the detection of train axle bolts in railway systems. A total of 5,000 image patches were extracted using the 128×128 overlapping sliding window method to determine the state of the train axle bolts. The dataset was divided into 4,000 training images and 1,000 test images, among which 1,000 test images comprised of 500 normal bolt images and 500 abnormal bolt images.

3.2 Training detail

We trained the model on a server with an NVIDIA GeForce RTX 3090 GPU and 24GB memory. The Adam optimization algorithm was used with an initial learning rate of $2e-4$, momentum parameters of β_1=0.5 and β_2=0.999, and a λ value of 0.9 for computing anomaly scores. In the loss function, we set λ_{adv}, λ_{con}, and λ_{lat} to 1, 50, and 1, respectively. The number of blocks per layer is discussed in Section 3.4. We used BatchNorm for training on the CIFAR10 and STL10 datasets and LayerNorm for LBOT.

3.3 Experimental Analysis

In this paper, we compare our proposed method with GANomaly [1], Skip-GANomaly [2], SAGAN [11], and AnoTran [20] models on CIFAR10, STL10 and LBOT datasets. Furthermore, we compare the effectiveness of the skip connection module, and the results show that using the GanNeXt Block with skip connections performs better on average.

Table 1. The AUC results on the CIFAR10 dataset.

Model	frog	bird	cat	deer	dog	horse	ship	truck	airplane	automobile	Average
GANomaly [1]	0.512	0.523	0.466	0.467	0.502	0.387	0.534	0.579	0.789	0.786	0.555
Skip-GANomaly [2]	0.955	0.611	0.670	0.845	0.706	0.666	0.909	0.857	0.748	0.790	0.776
SAGAN [11]	0.996	0.957	0.951	0.998	0.975	0.891	0.990	0.980	**1.000**	0.979	0.971
AnoTran [20]	**1.000**	0.944	0.960	0.999	0.968	0.949	0.999	0.990	1.000	**0.999**	0.980
Proposed w/o GSC	1.000	0.969	**0.999**	1.000	0.985	0.999	**1.000**	**0.996**	1.000	0.987	0.994
Proposed with GSC	1.000	**0.983**	0.992	**1.000**	**0.996**	**1.000**	1.000	0.996	1.000	0.997	**0.996**

Table 2. The AUC results on the STL10 dataset.

Model	bird	car	cat	deer	dog	horse	monkey	ship	truck	airplane	Average
GANomaly [1]	0.588	0.902	0.556	0.664	0.581	0.726	0.590	0.568	0.770	0.748	0.669
Skip-GANomaly [2]	0.929	**1.000**	0.963	0.996	0.859	0.947	0.979	**0.999**	0.998	0.879	0.955
SAGAN [11]	0.916	0.999	0.937	0.991	0.821	0.922	0.938	0.993	0.997	**1.000**	0.951
AnoTran [20]	0.966	1.000	0.985	0.998	0.942	0.975	0.980	0.999	0.997	1.000	0.984
Proposed w/o GSC	**0.968**	1.000	0.988	0.999	0.967	0.969	0.994	0.999	0.999	1.000	0.988
Proposed with GSC	0.964	1.000	**0.989**	**1.000**	**0.978**	**0.981**	**0.995**	0.999	**1.000**	1.000	**0.991**

On the CIFAR10 dataset, as shown in Table 1, our GanNeXt model outperforms the attention-based and traditional CNN models in all metrics. Notably, for three anomalous classes where the AUC is not 1.000, our method achieves a score of 1.000, and for bird, cat, and dog classes, we improve over the baseline method by more than 3%, with an average AUC improvement of 2%.

On the STL10 dataset, as shown in Table 2, our proposed method outperforms the baseline method in all classes except for bird anomalies. For the truck class, the AUC reaches 1.000. Our method outperforms the baseline on both the CIFAR10 and STL10 datasets, which is attributed to the fact that Transformer-based networks lack the inductive bias towards visual data and require more training data to learn image features. Our method combines the advantages of deep separable convolution and expansion layers to retain the long-term dependencies brought by the Transformer while preserving the bias dependencies brought by the convolutional layers, resulting in better performance without pretraining. By comparing different skip connection methods, we find that using the GanNeXt Block connection performs better on most datasets, demonstrating the effectiveness of our skip connection method.

Table 3 presents the results of our approach applied to the LBOT dataset, demonstrating that our method achieved comparable results to the state-of-the-art and surpassed all convolution-based models.

3.4 Computational Efficiency

We compared our method with Skip-GAnomaly, SAGAN, and AnoTran models, in terms of their computational efficiency. The FLOPs were calculated using a 32 $\times$ 32 $\times$ 3 input, and as shown in Table 4, our model reduced the FLOPs by 58% compared to the baseline methods. We believe that the local connectivity and parameter sharing of the CNN network enables efficient utilization of computational resources while reducing the computational complexity of the network in image processing. Conversely, Transformer-based CV networks require global self-attention mechanisms at each pixel position, resulting in significantly higher FLOPs compared to CNN networks, requiring more computational resources and time.

Table 3. The AUC results on the LBOT dataset.

Model	AUC
GANomaly [1]	0.900
Skip-GANomaly [2]	0.840
SAGAN [11]	0.960
AnoTran [20]	**0.996**
Proposed	0.985

Table 4. FLOPs of the models.

Model	FLOPs/M
Skip-GANomaly	226
SAGAN	708
AnoTran	283
Proposed	**121**

3.5 Ablation Study

To analyse the effectiveness of the various components of our proposed model, we compare the performance on the STL10 dataset with the different number of blocks in our model. The results are shown in Table 5, in which the "Num of Block" column indicates the number of GanNeXt Blocks in each layer. In addition, we set the same number of GanNeXt blocks for the decoder and generator at the same depth. For example, [2,2,6,4] denotes that the number of GanNeXt blocks of first layer in Fig. 1 is 2 (i.e., $B1 = B7 = 2$), similarly, $B2 = B6 = 2$, $B3 = B5 = 6$, $B4 = 4$. Based on this ablation experiment, we use the settings [2,2,6,4] as the default configuration of our proposed model.

Table 5. GanNeXt results of different number of blocks on the STL10 dataset.

Num of Blocks	bird	car	cat	deer	dog	horse	monkey	ship	truck	airplane	Average	FLOPs
[2,2,6,4]	0.964	1.000	0.989	0.999	**0.978**	**0.981**	0.995	**1.000**	1.000	1.000	**0.991**	121M
[1,1,3,2]	**0.967**	1.000	0.987	**0.999**	0.972	0.967	0.994	1.000	1.000	**1.000**	0.989	71M
[4,4,4,4]	0.959	1.000	**0.992**	0.999	0.964	0.976	0.994	1.000	1.000	1.000	0.988	126M
[3,3,3,3]	0.953	1.000	0.991	0.999	0.967	0.971	0.994	0.999	0.999	1.000	0.987	102M
[2,2,2,2]	0.956	1.000	0.99	0.999	0.97	0.98	0.994	1.000	0.999	1.000	0.989	78M
[1,1,1,1]	0.947	**1.000**	0.986	0.998	0.955	0.97	**0.995**	**1.000**	0.999	0.999	0.985	54M

4 Conclusion

In this paper, we have proposed a CNN-based image anomaly detection method that outperforms previous CNN and attention-based methods. Our method employs deep separable convolution and expansion layers to acquire long-term dependencies, similar to attention methods, while simultaneously reducing FLOPs by 58%. Furthermore, we demonstrate the effectiveness of our skip connection in mining multi-scale information. Our experiments reveal that different normalization methods have a significant impact on various tasks, and we plan to explore their effects on representation enhancement in future research.

References

1. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: Semi-supervised anomaly detection via adversarial training. In: Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14. pp. 622–637. Springer (2019)
2. Akçay, S., Atapour-Abarghouei, A., Breckon, T.P.: Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. In: 2019 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2019)
3. Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer normalization. arXiv preprint arXiv:1607.06450 (2016)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
5. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Somputer Vision and Pattern Recognition. pp. 770–778 (2016)
7. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016)
8. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. pp. 448–456. pmlr (2015)
9. LeCun, Y., Bengio, Y., Hinton, G., et al.: Deep learning. nature, 521 (7553), 436-444. Google Scholar Google Scholar Cross Ref Cross Ref p. 25 (2015)
10. Li, H., Liu, H., Ji, X., Li, G., Shi, L.: Cifar10-dvs: an event-stream dataset for object classification. Frontiers in Neuroscience **11**, 309 (2017)
11. Liu, G., Lan, S., Zhang, T., Huang, W., Wang, W.: Sagan: skip-attention gan for anomaly detection. In: 2021 IEEE International Conference on Image Processing (ICIP). pp. 2468–2472. IEEE (2021)
12. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022 (2021)

13. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11976–11986 (2022)
14. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings. pp. 146–157. Springer (2017)
15. Singh, H., Swagatika, S., Venkat, R.S., Saxena, S.: Justification of stl-10 dataset using a competent cnn model trained on cifar-10. In: 2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA). pp. 1254–1257. IEEE (2019)
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **30** (2017)
17. Wang, M., Deng, W.: Deep face recognition: A survey. Neurocomputing **429**, 215–244 (2021)
18. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S., Xie, S.: Convnext v2: Co-designing and scaling convnets with masked autoencoders. arXiv preprint arXiv:2301.00808 (2023)
19. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
20. Yang, C., Lan, S., Huang, W., Wang, W., Liu, G., Yang, H., Ma, W., Li, P.: A transformer-based gan for anomaly detection. In: Artificial Neural Networks and Machine Learning–ICANN 2022: 31st International Conference on Artificial Neural Networks, Bristol, UK, September 6–9, 2022, Proceedings, Part II. pp. 345–357. Springer (2022)