Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Exploiting DCGAN Generated Images for Improving Image Classification

Jingqi Chen

Department of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, China

2021010801023@std.uestc.edu.cn

Abstract. The power of Deep Neural Networks relies heavily on the quantity and quality of training data. Usually, it is expensive and time-consuming for people to collect and annotate data on a large scale. Traditional methods, including data augmentation, do not always have the effect, especially in some biomedical fields where some large-size anonymous datasets are generally not publicly available. The study investigates the effectiveness of Deep Convolutional Generative Adversarial Networks (DCGAN)-generated pseudo data compared to simple data augmentation techniques, such as geometric transformations and color enhancements. Various classifiers are trained and tested on both original and augmented datasets. Results indicate that while DCGAN-generated data visually resembles real images, it may not fully capture the statistical characteristics of the original data, leading to decreased classification accuracy compared to simple data augmentation, but the accuracy difference with the original dataset is less than 5% at worst. This shows that the fake data generated by DCGAN can indeed be used for training, but different datasets require network deepening, hyperparameter adjustment, etc.

Keywords: Generative adversarial networks; convolutional neural network; data augmentation; image classification.

1. Introduction

The power of Deep Neural Network (DNN) relies heavily on the quantity and quality of training data. However, due to the breakdown of collective devices, the collecting errors, and willful damages, gathering and annotating data extensively can be both costly and time-consuming. Simultaneously, for some fields like biomedical science, a great deal of information cannot be opened and demonstrated to the public because of personal privacy. The necessity of dataset augment is highlighted to avoid unwarranted loss of resources.

The primary methods of extending the dataset include transformation for Geometric, color gamut, and rotation angle. In 2022, Stable Diffusion (SD)-based models are used to construct the Guided Imagination Framework and generate realistic images with new content. Some relevant researchers are also combining transformers with Generative Adversarial Network (GAN) architectures that do not contain convolutions. For example, the wide variation in the appearance of breast lesions and normal breast structures can present challenges for computerized detection algorithms. Juhun Lee et al. used Cycle-GAN to highlight the lesions of the breast X-ray [1], and the new dataset was used to train the algorithm. The accuracy rate of new dataset in all test networks has increased. Chenyin et al. Research [2] on the expansion of medical image datasets compared the geometric transformation and expansion methods based on GANs [3] and suggested some improvement in the latter.

Deep Convolution Generative Adversarial Network (DCGAN) uses Convolutional Neural Networks (CNNs) as the generator and discriminator [4]. This structure can effectively learn the characteristics of images and is widely used in image generation, image enhancement, and other fields. DCGAN is one of the classic models of GANs. It has been widely researched and verified and has certain stability and reliability. Its open-source implementation makes it easy to use and extend, with plenty of community support and related resources available.

This work applies DCGAN to expand data, taking CIFAR-10 as an example. The pseudo results are tested by LeNet-5 [5], AlexNet [6], ResNet18 [7], and InceptionNet V1 [8], which are four verifiable classifiers. These classifiers are also trained by the original data from CIFAR-10 so that the augmented dataset can be discriminated and their accuracy can be obtained.

2. Methodology

2.1. Dataset

This project's dataset is CIFAR-10, consisting of 60000 32×32 color images in 10 classes; subsequent processes are implemented on this site. Sixty thousand images from CIFAR-10 are used to train the DCGANs' generator and discriminator. Classifiers with different structures and complexities are used to test the quality of extended training datasets.

The reason for choosing CIFAR-10 is that it is a classic dataset, and many models can already judge it with high accuracy. Due to the small image size of the cifar10 data set, DCGAN can learn better and generate pseudo data well.

All the classifiers resize the image of the CIFAR-10 dataset to the corresponding size during data preprocessing. LeNet is 32×32, Alexnet, ResNet and InceptionNet are 224×224.

2.2. DCGAN

DCGAN is an improved generative network based on GANs, incorporating Convolutional Neural Networks' (CNNs') structure [4]. In a generator, every operation starts with transpose convolution instead of a full connection layer since a full connection has long-term training due to too many parameters, and there is a high possibility of over-fitting. These two-dimension transpose convolution layers with a kernel size of 4×4 aim to ascending dimension from the sample of Gaussian noise, so except for the basic stride of 1×1 and padding of 0 from the first layer, their stride sizes are 2×2 and padding sizes are 1.

Following is Batch Normalization (BN), which is used to stabilize the training and avoid model crashes easily. If BN is applied in all layers, training may still be unstable, so it is canceled at the last layer. The generator uses Rectified Linear Unit (ReLU) to activate every layer except the last one, which uses Tanh. In discriminator, the first step is to under-sample the images with convolution layers in the reverse order of the generator. After the same part of BN, the activate function chooses LeakyReLU to avoid the vanishing gradient problem caused by ReLU.

The discriminator can judge whether the image is real and pass parameters to the update generator. After several updates, the generator can fool the discriminator and produce more realistic images. As the loss converges to a low value, a well-trained DCGAN system exists.

2.3. Classification Model

LeNet is a pioneering convolutional neural network architecture developed by Yann LeCun and his colleagues in the 1990s [5]. It consists of several convolutional and pooling operations layers followed by fully connected layers. LeNet was primarily designed for handwritten digit recognition tasks and was crucial in popularizing CNNs. It comprises a relatively small number of parameters compared to modern architectures and utilizes techniques such as local receptive fields and shared weights to achieve efficient feature extraction. While ReLUs and max-pooling work better now, they were not yet discovered at that moment. Coupled with a shallow network structure, the effect is poor for larger multi-classification tasks.

Due to their larger dimensions, ImageNet images typically contain objects that occupy more pixels and exhibit greater visual complexity than smaller and simpler MNIST images. AlexNet was designed to process the ImageNet dataset, and it gained widespread attention after winning this Challenge in 2012 [6]. AlexNet consists of five convolutional layers with max-pooling layer and three fully connected layers. It introduced several innovations, such as rectified linear units (ReLU) as activation functions and dropout regularization to prevent overfitting. AlexNet significantly advanced the field of computer vision and paved the way for deeper and more complex neural network architectures.

Shallow networks usually cannot capture data features well, so it is necessary to deepen the network depth. However, as the depth of the network increases, degradation problems arise, and the accuracy starts to decrease after reaching the maximum value. ResNet is short for Residual Network, and the main innovation of ResNet is the introduction of residual learning to address issues such as vanishing gradients and exploding gradients during the training process of deep neural networks [7]. Because it is difficult for deep neural networks to achieve "identity transformation", ResNet allows the output of a particular layer to directly skip one or more layers by introducing shortcut connections. This residual structure can solve the network problem being too deep and complex to train.

InceptionNet is a neural network architecture developed by researchers at Google [8]. It is renowned for its Inception Module, which efficiently captures features at multiple scales using parallel convolutional operations of different sizes. InceptionNet employs 22 layers and significantly reduces computational complexity compared to traditional deep networks using global average pooling instead of fully connected layers at the top. This architecture achieved state-of-the-art performance on the ImageNet dataset in 2014 while demonstrating high efficiency. InceptionNet introduced the concept of network "inception" to improve feature extraction and inspired subsequent versions with further improvements.

3. Result

In DCGAN, the optimizer uses Adam, with the initial learning rate is 0.0002 and the betas are (0.5, 0.999). After verification, DCGAN in this form has met the requirements of extended data sets. Representative generated images are demonstrated in Fig. 1.

Fig. 1 Representative generated images of airplane and cars (Figure Credit: Original).

Due to the operation time and equipment capability, LeNet-5, AlexNet, ResNet18, and InceptionNet V1 have different training epochs of 300, 200, 150, 150. All classifiers use the Adam optimizer with an initial learning rate of 0.0002. The accuracy plots from four training datasets of different classifiers are shown below.

Fifty thousand pseudo images are generated from the modeled DCGAN of 50 epochs, and 50000 pseudo images are generated from the modeled DCGAN of 100 epochs. Simultaneously, 50000 strengthened images are generated by random rotation and color enhancement. These images are added to the original 50000 training images using four classifiers to compare the accuracy. The performances are displayed in Fig. 2 and Fig. 3 respectively.

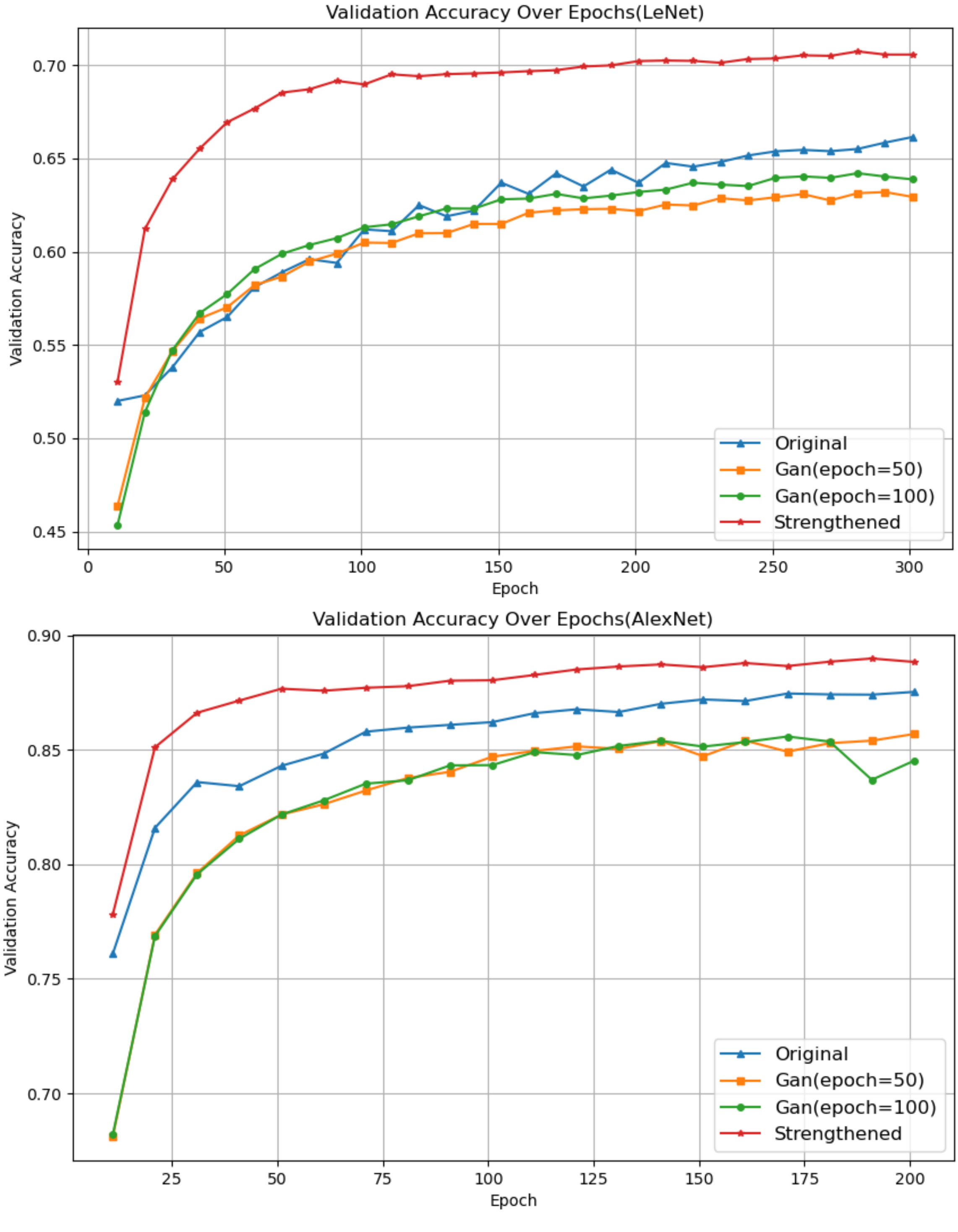

Fig. 2 Accuracy of LeNet and AlexNet (Figure Credit: Original).

For the pseudo dataset generated by DCGAN, as the epoch of GAN training increases, the classification accuracy does not improve significantly and even decreases slightly in some models.

As demonstrated in Table 1, among the four classifiers, AlexNet and InceptionNet V1 perform the best, while the performance of LeNet-5 is relatively low. Compared with the original and pseudo datasets, the accuracy difference is less than 5%. Moreover, as the network complexity increases, the accuracy difference becomes smaller and smaller.

The strengthened data has the highest accuracy in LeNet-5, AlexNet, and InceptionNet V1. However, it has the lowest accuracy rate in ResNet18.

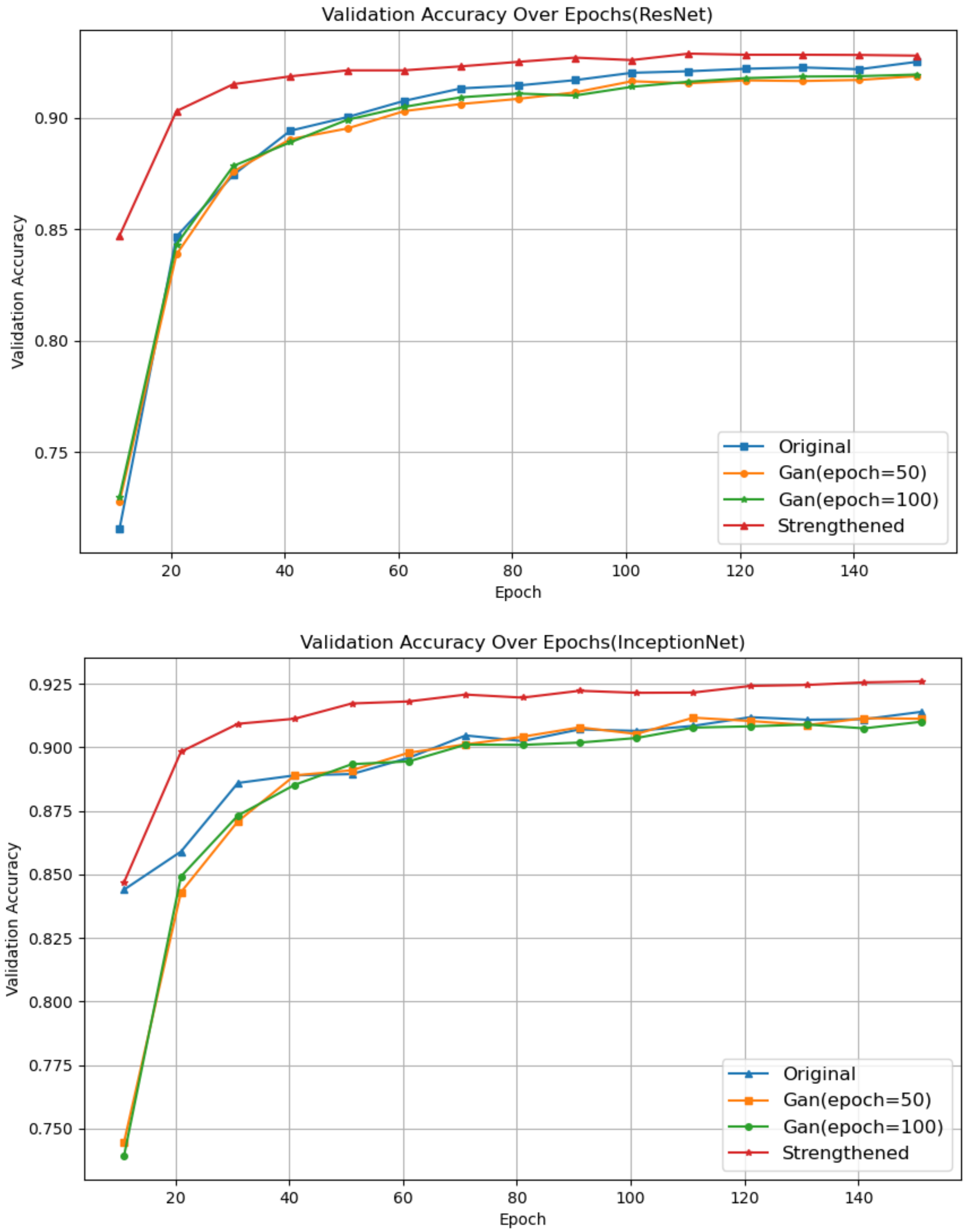

Fig. 3 Accuracy of ResNet and InceptionNet (Figure Credit: Original).

Table 1. Comparison of training datasets on four classifiers, measured by accuracy.

	LeNet-5	AlexNet	ResNet18	InceptionNet V1
Original	0.6615	0.8754	0.9251	0.9140
Gan (epoch=50)	0.6294	0.8570	0.9187	0.9113
Gan (epoch=100)	0.6388	0.8453	0.9194	0.9101
Strengthened	0.7057	0.8885	0.9279	0.9260

4. Discussion

The data generated by DCGAN may not fully capture the distribution characteristics of the original data. Although DCGAN can generate visually realistic images, they may only partially reflect the statistical characteristics of the original data, resulting in differences between the generated samples and the real data. Data augmentation technology more directly changes the representation of the original data.

At the same time, Images generated by DCGAN may contain more noise, which may interfere with the classifier's training, making it more difficult to classify correctly. Compared with simple data augmentation, the images may be blurrier or less clear on some features, thus reducing the performance of the classifiers.

Regarding the difference in the accuracy of data sets generated by DCGAN with different training times, this may be because although the samples generated by DCGAN are visually realistic, they may differ from the real data distribution in the feature space, causing the classifier to fail on these data. The performance has not significantly improved.

The project also tried a 300×300 fruit image classification dataset. However, mode collapse often occurred after increasing the number of DCGAN layers, resulting in unsatisfactory generated images. Therefore, only small-size image classification datasets such as CIFAR-10 were tested.

On the CIFAR-10 dataset, the accuracy of pseudo data generated by DCGAN is similar to the original data but worse than the strengthened data. This means that there are still many shortcomings generated by DCGAN generated. Gao Ya added Google Efficient Channel Attention-Net to Wasserstein GAN to enhance the extraction of image features and introduced residual blocks to maintain the stability of the training of the network with more convolutional layers [9]. On the basis of the original DCGAN, Huang Yingxuan increased the number of anti-convolution layers of the generator and introduced a method of feature fusion in the judgment device [10]. Using this Improved-DCGAN to expand the original data set, the model fitting time is shorter, and the generated image is clearer.

However, in GANs, mode collapse still exists to a certain extent. When generating images, noise and chessboard patterns will inevitably appear, resulting in a vague image and the poor learning effect of the classifier. These issues require multiple adjustments and testing of hyperparameters and network structure.

5. Conclusion

The experiment utilizes the CIFAR-10 dataset comprising 60,000 32×32 color images across ten classes. DCGANs are trained on this dataset, generating pseudo images. The classifiers, including LeNet-5, AlexNet, ResNet18, and InceptionNet V1, undergo training and testing on original and extended datasets. The extension generates 50,000 pseudo images from DCGAN models trained for 50 and 100 epochs and 50,000 augmented images created through random rotation and color enhancement. Training epochs for classifiers vary from 150 to 300 based on computational constraints and model complexity. The performance metrics are evaluated based on classification accuracy.

Although the convergency fields of the four classification models are not identical, the final points of plots demonstrate that the difference between original data and pseudo data is marginal, and both accuracy rates are at relatively high levels, which shows that DCGANs have a high ability to deal with data set expansion. Extended datasets exhibit high accuracy rates, showcasing the DCGANs' robustness in effectively expanding datasets. These findings underscore the potential of DCGANs in facilitating dataset expansion for improved model performance, with implications across various image classification tasks.

References

[1] Lee, Juhun, and Robert M. Nishikawa. Improving lesion detection in mammograms by leveraging a Cycle-GAN-based lesion remover. Breast Cancer Research 2024, 26(1): 21.

[2] Chen, Ying, Hongping Lin, Wei Zhang, Longfeng Feng, Cheng Zheng, Taohui Zhou, Zhen Yi, and Lan Liu. Journal of Biomedical Engineering (Shengwu Yixue Gongchengxue Zazhi), 2023, 40(1): 185-192.

[3] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. Advances in neural information processing systems, 2014: 27

[4] Radford, Alec, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint, 2015, arXiv:1511.06434.

[5] LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[6] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012: 25

[7] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[8] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[9] Gao Ya. Research on Skin Injury Data Augmentation Technology Based on Improved Generative Adversarial Network. Wuhan Textile University, 2023.

[10] Huang Yingxuan. Research on Classification of Brain Tumor MRI Image Based on Deep Learning. Beijing University of Chemical Technology, 2023.