Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 3, No. 2, 2023

Research on No-reference Image Quality Assessment Algorithm Based on Generative Adversarial Networks

Wenqing Zhao, Haoyang Chen

School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China

Abstract: Currently, with the massive generation and transmission of digital images, especially the rapid development of the Internet and mobile Internet industries where images are widely used, the study of reference-free image quality assessment has attracted much attention in academia and practical applications, and is a popular research direction in the field of computer vision. In response to the low performance of existing reference-free image quality assessment algorithms in the face of real distorted images, a reference-free image quality assessment algorithm based on generative adversarial networks is proposed. Firstly, the generator structure is changed, the U-Net structure is improved, and the channel attention mechanism SeNet structure is introduced to update the feature map after down sampling. Secondly, a feature similarity measurement system is incorporated and a dual discriminator structure is used to discriminate multiple groups of images. The FPN structure is combined in the feature extractor to produce a multi-scale feature representation. Experiments are conducted on KonIQ-10k dataset and LIVEC dataset, and the experimental results show that the algorithm exhibits good prediction accuracy as well as good generalization performance in the face of real distorted images.

Keywords: Image quality assessment; Deep learning; Generative adversarial networks (GANs); Convolutional neural networks (CNNs); Feature extraction; Multi-scale feature fusion.

1. Introduction

As an important research direction in the field of computer vision, image quality assessment aims to objectively assess the quality of images through computer algorithms. It is mainly applied in the fields of image processing, video transmission, and digital media. Reference-free image quality assessment is an algorithm that evaluates the quality of an image by using only an image itself. This algorithm is usually used for automatic image quality assessment without reference images or human subjective involvement. Therefore, the application of this algorithm on large-scale image databases is of great practical value.

Because there is no reference image as a comparison standard, NR-IQA methods can only use distorted images to extract the statistical features of distortions caused by distortion. Among the NR-IQA methods based on feature extraction, they can be broadly classified into two categories: natural statistical feature-based methods and feature learning-based methods. However, variations of natural statistical features may be presented in multiple ways. For example, in the literature [1], the main objective of this paper is to propose a reference-free image quality assessment metric based on regional mutual information, which evaluates the quality of images based on mutual information within different regions and solves the problem that traditional reference-free image quality assessment metrics are difficult to obtain high accuracy. Oszust [2] introduces a new method for blind image quality assessment that combines local feature descriptors and derivative filters, firstly, by extracting and matching the feature points in the reference image and the image to be evaluated, and calculating the variation between them. Then, the gradient information is extracted from the image using the derivative filter, and the quality of the image is evaluated by feature point matching. The literature [3] proposes an unsupervised blind vision quality assessment method based on three key factors: structure, naturalness and perception. The image is first segmented, and then the structure, naturalness and perceptual features of each region after segmentation are analyzed. Then, the quality score of each region is calculated based on these features.

The assumption underlying all these algorithms is that distortion affects certain statistical features of natural images, but these features are difficult to determine comprehensively and accurately by manual means. The literature [4] focuses on reference-free and full-reference image quality assessment using deep neural networks, proposing two deep neural network models: one for reference-free assessment and the other for full-reference assessment. Le Kang et al [5] proposed a reference-free image quality assessment model using convolutional neural networks. This model makes it possible to assess image quality without the need for reference images or specialized training datasets and allows accurate assessment under different types of distortions. The authors demonstrated the accuracy and robustness of the model by testing it against several different types of image distortions such as blur, noise, and compression. Simone Bianco et al [6] proposed a blind image quality assessment method based on convolutional neural networks, where features are extracted from the original image using a CNN and these features are fed into a support vector regression (SVR) model in which quality assessment is performed.

And the emergence of GAN has injected new ideas into IQA. A novel reference-free image quality assessment method, called Hallucinated-IQA, was proposed in the literature [7]. This method utilizes Generative Adversarial Networks (GAN) to learn the implicit representation and quality features of an image, learn the conversion from low-quality to high-quality images, and thus assess image quality. A new reference-free image quality assessment method, called RAN4IQA, was proposed in the literature [8], which combines deep convolutional neural networks and generative adversarial networks to achieve the assessment of image quality by learning loss functions. In the literature [9], the

authors proposed a new blind image quality assessment (BIQA) method, which aims to generate an image with similar visual quality to the input image. In this paper[10], a reference-free image quality assessment algorithm based on enhanced adversarial learning is proposed to improve the strength of the adversarial learning by improving the loss function and the structure of the network model, and to output a more reliable simulation "reference image" to achieve similar results as the full-reference image quality assessment method.

In order to improve the accuracy of the images output from the generative adversarial network, the prediction accuracy and generalization of the image quality assessment network are improved, and the comparison process of the human visual system can be better simulated at the same time. Firstly, we change the generator structure, improve the U-Net structure, introduce the channel attention mechanism SeNet to let the neural network pay more attention to the detailed information related to the features, get a vector with the same dimension as the number of channels, and add this vector as a weight number to the corresponding channels to generate more accurate pseudo-reference images. Secondly, a feature similarity measurement system is added, and a dual discriminator structure is used to discriminate multiple sets of images. The FPN structure is combined in the feature extractor to produce a multi-scale feature representation and all levels of feature maps with strong semantic information to better fuse the features of the reference image block and the distorted image block.

2. Related technologies and theories

2.1. Generative Adversarial Networks

Generative adversarial networks are based on the idea of adversarial training in game theory. By training two neural networks, the generator network and the discriminator network, simultaneously, the generator network can generate samples similar to the real samples, while the discriminator network can distinguish the real samples from the false samples as accurately as possible.

The objective function of the Generative Adversarial Network (GAN) consists of two parts: the loss function of the Generator and the loss function of the Discriminator. First, we look at the loss function of the Generator. The goal of the Generator is to generate samples that can fool the Discriminator, so the cross-entropy can be used as the loss function of the Generator. Its mathematical formula is:

L_G = -log(D(G(z)))

Where z is a vector sampled from the noise potential space, G(z) denotes the samples generated by the generator, and D(x) is the probability that the discriminator output sample x is the true sample

Next, we look at the loss function of the discriminator. The goal of the discriminator is to accurately distinguish between the true samples and the samples generated by the generator, so the binary cross-entropy can be used as the loss function of the discriminator. Its mathematical formula is:L_D = -[log(D(x)) + log(1-D(G(z)))]

Where x is the true sample from the dataset, D(x) is the probability that the discriminator output sample x is true, G(z) denotes the sample generated by the generator, and 1-D(G(z)) is the probability that the discriminator output sample G(z) is false.

2.2. U-net Networks

The U-Net architecture consists of an encoder and a decoder [11]. The encoder extracts high-level features from the input image by down sampling operations to decompose it into smaller representations. The decoder then up-samples by learning filters to reconstruct the size of the original image. The encoder part follows the typical pattern of a convolutional neural network, where each layer consists of two 3 × 3 convolutional layers followed by a batch normalization and a modified linear unit (ReLU) activation function, and is immediately followed by a maximum pooling operation with a step size of 2, which halves the resolution of the feature mapping and reduces the size of the space.

2.3. FSIM Algorithm

The principle of FSIM is based on the assumption that if two images are very similar in certain features, then the quality of these two images should be higher. Therefore, FSIM first calculates features such as structural information, gray value and color information of an image and uses these features as descriptors of the image. Then, a cosine distance in one feature space is used to measure the similarity between the two images.

The FSIM method consists of two steps: feature extraction and feature similarity calculation. The feature extraction stage applies wavelet transforms to the original image and the distorted image separately to extract their energy and phase information in different frequency subbands. The feature similarity calculation stage calculates the similarity between the two images based on this feature information.

2.4. Resnet Network

ResNet adds the output of each convolutional layer to the original input, i.e., a shortcut connection, so that the model can directly fit the residuals and thus learn complex feature representations more easily. Each residual block consists of two small convolutional layers and a cross-layer connection. This cross-layer connection allows the input signal to be added directly to the output of the pooling or convolutional layers without additional transformations. Thus, if the input and output of the residual block have the same dimension, a constant mapping can be achieved by simply adding the input to the output. If they have different dimensions, a linear transformation should be used to project the dimensionality of the input in order to match the dimensionality of the output, and then add them together.

3. No-Reference Image Quality Assessment Network Bon Generative Adversarial Networks

3.1. Scheme design flow

The overall flow of the image quality assessment algorithm scheme proposed in this paper is shown in Figure 2-1 below:

First, we take distorted images and undistorted original images in the dataset as inputs for training the network model. Then, the model is used to output the pseudo-reference map of the image to be tested, and the depth convolution features of the pseudo-reference map are extracted. Finally, the convolutional features of the pseudo-reference map and the distorted image to be tested are fused and fed into the trained image quality assessment regression network to obtain the assessment score of theimage.

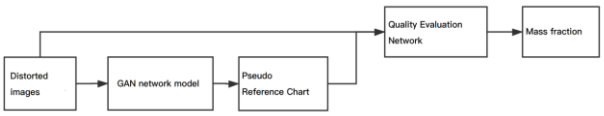

Fig.1 Overall flow chart of the program

3.2. GAN Model Design

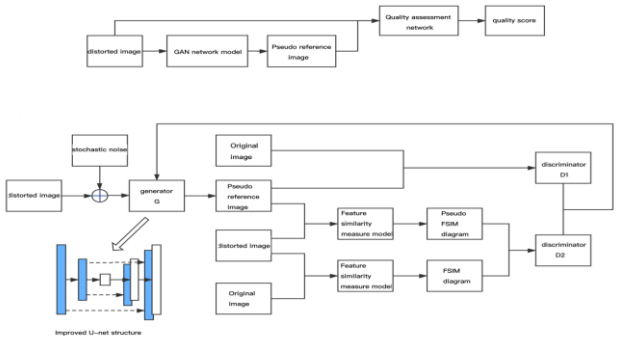

Fig.2 Architecture diagram of the improved generative adversarial network model

Figure 2 shows the architecture of the improved generator network model. The Pix2Pix structure is used as the basis, which is based on the process of getting the desired output image from one input image and can be seen as a mapping between images and images.

The structure of the generator model G is improved with a U-net network structure. According to the FSIM algorithm, the feature similarity map (referred to as FSIM map) is output by learning the feature similarity between the distorted image and the corresponding undistorted original image.

According to the game idea of generative adversarial network, the discriminator model D is needed to distinguish 2 groups of images: the undistorted original image and the pseudo-reference image, and the pseudo-FSIM map and the FSIM map, so a dual discriminator structure is used, and the performance of the discriminator will be degraded if only one discriminator is used to distinguish two groups at the same time. D1 is a convolutional neural network (CNN) to distinguish the pseudo-reference image from the original image to ensure that the generated the structure of D2 is similar to that of D1 and is used to distinguish the pseudo FSIM image from the FSIM image. Adversarial training using this discriminator ensures that the pseudo FSIM map is judged to be as false as possible and the FSIM is judged to be as true as possible.

3.3. Improved U-net network

In this paper, a new generator structure is used, as shown in Figure 3. it utilizes a full-scale jump connection [12] (skip connection) and deep supervisions [13] (deep supervisions) approach. The full-scale skip connection directly sums the feature maps from different levels and uses them as input for the next level. This jump connection is able to retain the detailed information of the low-level feature maps while passing the semantic information of the high-level feature maps, thus effectively improving the model performance. After each decoding layer generates a feature map, the deep supervision mechanism feeds it into a convolutional layer for further processing and performs an upsampling operation on the result using bilinear upsampling. The segmentation result obtained after upsampling is then multiplied with the output of the classification module and processed by sigmoid to obtain the final deeply supervised output. Problems such as gradient disappearance and overfitting can be avoided, and the generalization ability of the model can be improved.

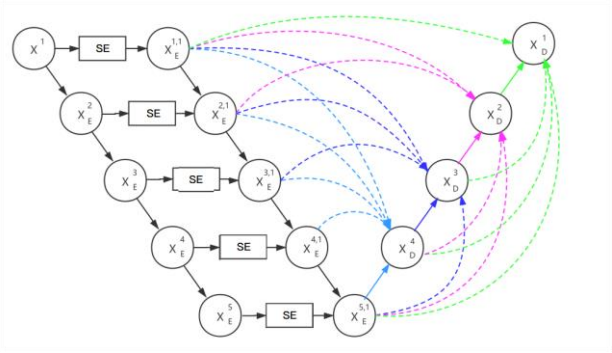

Fig.3 Improved U-Net network

To improve the accuracy of the task, an Attention mechanism is added in this paper. It is a resource allocation strategy by Attention to the result of downsampling in each layer and then upsampling. It allows the neural network to pay more attention to the detailed information related to the features and speeds up the processing time of the information to improve the efficiency of the computer in processing the information. Specifically, SeNet's attention mechanism consists of two steps: the squeezing operation and the excitation operation. In the squeezing operation, the feature maps of each channel are compressed into a vector by global average pooling. This vector represents the global feature information, where each element corresponds to the average of the feature values on the corresponding channel.

In this paper, the SeNet attention mechanism is chosen, and the structure is shown in:

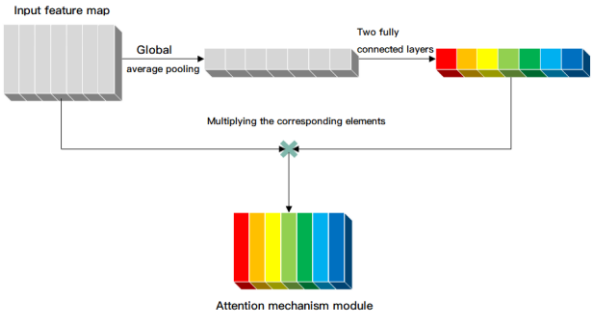

Fig.4 Attention mechanism module

3.4. Image quality assessment network

The IQA network consists of three main components: a feature extraction network, a metric fusion network, and a quality prediction network [14]. The feature extraction network extracts features from the input image, while the metric fusion network combines multiple quality metrics to generate a fused metric vector. Finally, the quality prediction network predicts the final quality score based on the fused metric vector. This is shown in Figure 5.

In the feature extraction network module, two feature extractors share network parameters for extracting the input pseudo-reference image block and distorted image block features, respectively. Specifically, the weights of certain convolutional layers are set to be the same so that multiple convolutional layers share the same set of weights, thus reducing the number of parameters and improving the generalization ability of the model as well as the efficiency

and performance of the model.

In order to solve the problem of distorted images with different local distortions, a multi-scale feature extractor scheme is proposed in this paper. This approach uses an image pyramid to generate images at different scales and applies a feature extractor to extract features at each level. In this way, feature vector information at different scales can be obtained to detect overall and detailed distortions in order to improve model performance and robustness.

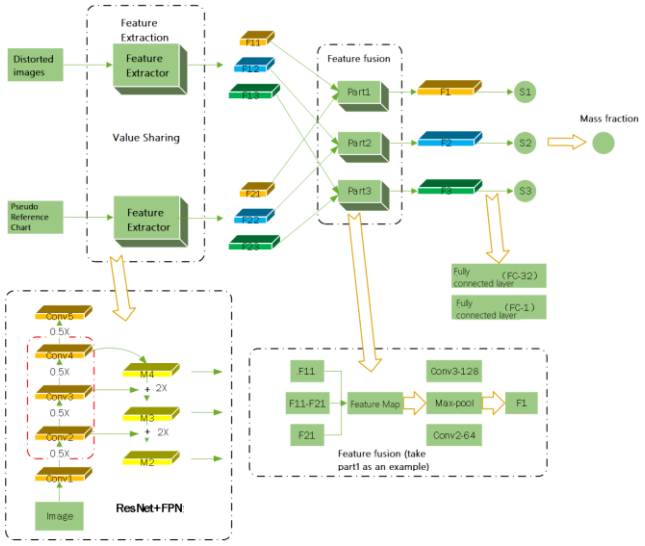

Fig.5 Image Quality Assessment Network

In this paper, the ResNet network is used as the backbone, where different blocks of output feature mappings correspond to different semantic information. To optimize the alignment of features with different scales, the feature extractor incorporates the design of Feature Pyramid Networks (FPN) [42], where the basic idea is to construct a feature pyramid from the input image, with high-level features corresponding to smaller image regions. This allows the network to detect objects at multiple scales and resolutions more efficiently than traditional CNNs. In Figures 5, different colors indicate elemental features with different scales.

The ResNet network is used to extract features from the input image. The feature pyramid network then takes these features and generates a pyramid-shaped feature map with different scales. The feature pyramid network works by combining features from the earlier levels with features from the later levels.

To finally evaluate the quality score of an image, we regress the fused features using two fully connected layers. Compared to the full-reference DeepIQA network, we optimize the number of output channels of the first fully-connected layer by reducing it to 32, thus reducing the computational burden of training the model and also reducing the number of parameters of the algorithm, improving the training efficiency of the model. The scoring regression module consists of three scoring regression blocks and a weighted average operation to calculate the final image block quality scores. The scores of all image blocks are aggregated to obtain the quality prediction score of the whole image.

4. Experimental results

4.1. Data sets and assessment metrics

To verify the effectiveness of the model proposed in this paper, this experiment was conducted on two real distorted image datasets, LIVE Challenge (LIVEC) and KonIQ-10k, respectively.

The LIVE Challenge dataset ((LIVE In the Wild Image Quality Challenge Database) contains images from real scenes, which include various environments and shooting conditions. These images are used to evaluate the performance of the algorithm in real scenarios. There are 1162 images from real scenes.

The KonIQ-10k dataset is a large-scale dataset for evaluating image quality and contains 10,073 images from the Internet and the subjective quality scores associated with them. Developed by the German Institute for Image Processing and Computer Vision (TNT), the dataset contains images from different domains and scenes such as natural landscapes, human portraits, buildings, animals, etc.

In order to compare the prediction accuracy of the image quality assessment methods from an objective point of view during the testing phase, two assessment metrics were used: the Pearson linear correlation coefficient (PLCC) and the Spearman rank-order correlation coefficient (SROCC) correlation coefficient (SROCC). These metrics can provide an objective measure of accuracy to help evaluate and compare different image quality assessment methods. The closer the two metrics are to 1 the better the model performance.

PLCC mainly measures the linear correlation between the prediction results and the manual subjective assessment results. The formula is as follows:

$$PLCC = \frac{\sum_{i=1}^{N}\left(s_i - \bar{s}\right)\left(p_i - \bar{p}\right)}{\sqrt{\sum_{i=1}^{N}\left(s_i - \bar{s}\right)^2 \sum_{i=1}^{N}\left(p_i - \bar{p}\right)^2}}$$

Where N denotes the number of test images, s_i denotes the subjective score of the ith image, p_i denotes the predicted score of the ith image, and denotes their mean.

SROCC is a nonparametric statistic that measures the monotonic relationship between two variables. It calculates the correlation based on the rank order of the variables rather than the values of the variables themselves. The formula is as follows:

$$SROCC = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N\left(N^2 - 1\right)}$$

Where, d_i^2 denotes the rank difference between the subjective and predicted scores of the ith test image.

4.2. Parameter Setting

The parameters were set to achieve better training results. For distorted images, we used a ratio of 8:1:1 to divide the training, validation and test sets. For the choice of optimizer, we used Adam optimizer for training, set the initial learning rate to 10-5, and set the weight decay rate to 0.0005, a value that can try to avoid overfitting during the training process.

4.3. Comparative experiments on a single data set

In order to evaluate the effectiveness of the algorithms proposed in this paper, this experiment was conducted to compare with eight higher performance blind image quality assessment algorithms. These nine algorithms are BRISQUE, HOSA and CORNIA, which are based on manual features,

and WaDIQaM-NR, TS-CNN, DB-CNN, PQR and HyperIQA, which are five classical algorithms based on deep learning similar to the algorithms in this paper. We choose these algorithms to ensure the completeness of the comparison range.

First, this section conducts experimental comparisons on the LIVEC dataset, and averages are taken after conducting multiple experiments. The experimental results are shown in Table 1.

Table 1. Comparison of different methods（LIVEC Dataset）

Algorithm	LIVEC	
	SROCC	PLCC
BRISQUE	0.607	0.630
HOSA	0.638	0.677
CORNIA	0.612	0.653
WaDIQaM-NR	0.669	0.679
TS-CNN	0.657	0.668
DB-CNN	0.848	0.867
PQR	0.856	0.881
HyperIQA	0.859	0.879
Methods of this paper	0.867	0.881

As can be seen from Table 1, the SROCC as well as PLCC indexes of this paper's method outperform the manual feature-based methods on the LIVEC real distortion image dataset, and the improvement effect is more obvious, and the same better performance is achieved compared with the other six deep learning-based methods.

Then, the KonIQ-10k dataset is selected for comparison experiments, and the experimental results are shown in Table 2.

Table 2. Comparison of different methods（KonIQ-10kDataset）

Algorithm	KonIQ-10k	
	SROCC	PLCC
BRISQUE	0.670	0.683
HOSA	0.671	0.693
CORNIA	0.552	0.570
WaDIQaM-NR	0.795	0.806
TS-CNN	0.724	0.731
DB-CNN	0.875	0.883
PQR	0.878	0.883
HyperIQA	0.903	0.916
Methods of this paper	0.905	0.917

As can be seen from Table 2, the method in this paper also has a good performance on the KonIQ-10k dataset and achieves more accurate predictions.

4.4. Cross-dataset comparison experiments

In order to verify the generalization ability of the algorithms proposed in this paper, comparative experiments are conducted in this section, and six advanced algorithms based on deep learning are selected to evaluate the generalization ability of the models. The experimental results are detailed in Table 3.

Table 3. SROCC for cross-dataset testing

Test set	Training set	SROCC					
		WaDIQaM-NR	TS-CNN	DB-CNN	PQR	HyperIQA	Method of this article
KonIQ-10k	LIVEC	0.680	0.652	0.752	0.768	0.782	0.783
LIVEC	KonIQ-10k	0.708	0.661	0.751	0.754	0.769	0.770

According to the data shown in Table 3, the experimental results of the two sets of cross-datasets show that the proposed method in this chapter has higher scores for both SROCC and PLCC metrics in real distortion image testing, indicating that the method has better generalization performance and can be applied to real-life distortion scenes.

5. Concluding remarks

To solve the problem that NR-IQA cannot simulate the HSV process, this paper adopts methods such as improving the U-Net structure and introducing the channel attention mechanism SeNet structure to improve the accuracy of the generated pictures. Also, this paper incorporates a feature similarity measurement system to generate pseudo-FSIM and FSIM maps, and a dual discriminator structure to improve the performance of the discriminator. To improve the prediction accuracy and generalization of the assessment network of image quality. In this paper, we propose a multi-scale feature extractor scheme to generate images at different scales by image pyramids and apply feature extractors at each level to extract features in order to obtain feature vector information at different scales to detect overall and detailed distortions.

The improved method proposed in this paper is effective in improving the prediction accuracy and also has a high degree of similarity with the perceptual consistency of the human visual system. However, we still need to face some problems. Due to the richness of image distortion types, each with different characteristics, but according to the available literature reports and experimental results, there does not exist an image quality assessment algorithm that can show advantages over other algorithms on all distortion types. Therefore, future research should focus on the design of general-purpose and high-performance image quality assessment algorithms. Only in this way can we meet the needs of different application scenarios and provide more comprehensive and accurate image quality assessment services.

References

[1] V Kumar, Bawa V S. No reference image quality assessment metric based on regional mutual information among images[J]. 2019.

[2] Oszust M. Local Feature Descriptor and Derivative Filters for Blind Image Quality Assessment[J]. IEEE Signal Processing Letters, 2019:1-1.

[3] Liu Y, Gu K, Zhang Y, et al. Unsupervised Blind Image Quality Assessment via Statistical Measurements of Structure, Naturalness, and Perception[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(4):929-943.

[4] Bosse S , Maniry D , Muller K R , et al. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment[J]. IEEE Transactions on Image Processing, 2017:1-1.

[5] Kang L, Ye P , Li Y, Doermann D. Convolutional Neural Networks for No-Reference Image Quality Assessment[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014.

[6] Bianco S, Celona L, Napoletano P , et al. On the Use of Deep Learning for Blind Image Quality Assessment[J]. Signal Image & Video Processing, 2016.

[7] Lin K Y, Wang G. Hallucinated-IQA: No-Reference Image Quality Assessment via Adversarial Learning[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[8] Ren H, Chen D , Wang Y . RAN4IQA: Restorative Adversarial Nets for No-Reference Image Quality Assessment[J]. 2017.

[9] Pan D, Shi P, Hou M, et al. Blind predicting similar quality map for image quality assessment[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6373-6382.

[10] Cao Yudong, Cai Xibiao Reference free image quality evaluation algorithm based on enhanced adversarial learning [J] Computer Applications, 2020, v.40; No.363(11):72-77.

[11] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.

[12] Jv A, Yqc A , Jing W B . FaultFace: Deep Convolutional Generative Adversarial Network (DCGAN) based Ball-Bearing failure detection method[J]. Information Sciences, 2021, 542:195-211.

[13] Mao C, Huang L , Xiao Y , et al. Target Recognition of SAR Image Based on CN-GAN and CNN in Complex Environment[J]. IEEE Access, 2021, PP (99):1-1.

[14] Yanding Peng, Jiahua Xu, Ziyuan Luo, Wei Zhou, Zhibo Chen; Multi-Metric Fusion Network for Image Quality Assessment. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2021, pp. 1857-1860.