# Construction and Application of Quality Assessment Model of No Reference Images Two-Stream Convolutional Neural Network

Dong Kang
Department of Information Technology, Hainan College of Economics and Business, Haikou, 571127, China
E-mail: kangdong@hceb.edu.cn

**Keywords:** screen content image, no-reference image, feature fusion, two-stream convolutional neural network, quality assessment

**Received:** June 12, 2024

*With the advancement of information and multimedia technologies, screen content images have been widely applied in multiple fields. However, in image transmission across devices, image distortion may occur due to various reasons. To solve this problem, this paper presents a no-reference image quality assessment (IQA) model for screen content images (SCIs) using areal feature fusion (AFF) and a two-stream convolutional neural network (TSCNN). The model leverages transfer learning for improved performance. Experimental results show that the proposed method achieves a Spearman rank correlation coefficient of 0.9242 and a Pearson linear correlation coefficient of 0.9284, outperforming traditional methods. The results show that the proposed two algorithms are effective in evaluating no-reference images and have high generalization ability.*

*Povzetek: Razvit je nov model za ocenjevanje kakovosti slik zaslonskih vsebin, temelječ na dvosmerni konvolucijski nevronski mreži in arealni fuziji značilk.*

## 1 Introduction

Screen content images (SCIs) are the output content of digital devices such as computer, television and mobile phone screens, occupying an increasingly large proportion in modern life. The types of information processed on the screen are diverse, including text, images, videos, etc. The quality of the output image straightly influences the user's vision and usage experience. Therefore, in the fields of digital media processing, human-computer interaction, and visual applications, the processing and optimization of SCIs are of great significance [1-2]. At present, there are mainly subjective and objective evaluation methods for image quality assessment (IQA). Objective assessment methods are mainly contained full, semi and no references. Image processing algorithms and artificial intelligence technology are used to calculate image quality scores through models [3]. In application, it is often hard to get information about reference images, so the no reference IQA has great guiding role in the assessment of SCIs.

This research is divided into four parts. The first part is the research on the application of two-stream convolutional neural network (TSCNN) in images by experts and scholars at home and abroad, and introduces the existing IQA models. In the second part, the features in different regions of the image are extracted. A no reference IQA model based on areal feature fusion (AFF) is constructed. Based on TSCNN, a no reference IQA model is constructed. The third part tests and analyzes the model, while the fourth part summarizes the article and points out its shortcomings.

## 2 Related works

As a classic deep learning visual model, the TSCNN has achieved good image recognition performance based on the imitation of human visual perception process. Some scholars have conducted relevant research on the application of TSCNN in the field of images. X Liao et al. found that forensic doctors use images to determine whether the operation is correct during technical operations. Based on this problem, a sequential forensics framework of image operator chain with convolutional neural network (CNN) was proposed. The TSCNN architecture was used to capture evidence of tampering artifacts and local noise residuals. The experimental findings denoted that this method had stability in practical applications and achieved significant performance improvement [4]. M Attique Khan et al. proposed a TSCNN information fusion framework to classify multiple skin cancers through images. By using fusion-based contrast enhancement technology, the enhanced images were provided to the pre trained architecture. Secondly, the proposed features were extracted through a pre trained network for deep feature extraction and downsampling operations. Finally, the parallel maximum coefficient correlation method was used to fuse features, and a classifier was used to classify the image. The research outcomes indicated that the model achieved testing accuracy of 96.5%, 98.0%, and 89.0% under three different datasets [5]. Y Sun et al.

proposed a TSCNN model with multi-level feature fusion to realize stable image recognition under light conditions. Firstly, RGB images were obtained through sensors, a database was established, and data augmentation was performed on the database. The experiment findings showed that the raised model could accurately recognize gestures, with an average detection accuracy improvement of 1.08% and an average accuracy improvement of 3.56% compared to the single channel model. Under the conditions of occlusion and different lighting intensities, the average recognition rate was 93.98% [6]. Z Zhang et al. believed that the semantic segmentation of road scene images played an important role in industrial applications, but the requirements for the diversity of target objects and the variability of high light in different scenes were getting higher and higher. To solve this problem, a TSCNN was put forward, including spatial and detail paths. The experiment findings expressed that this method improved speed and recognition accuracy compared to other methods, and obtained good global feature information [7]. X Li et al. found that when processing hyperspectral images, CNN would be limited by the number of samples, and could not satisfy the corresponding test requirements. Based on this problem, a 2D TSCNN architecture was adopted to improve the network's expression ability. This method used used correlation to identify the features with the largest amount of information. The experiment findings on several hyperspectral datasets showed that this method was effective in practical applications [8].

Some scholars have conducted relevant research on the IQA. K Ding et al. believed that traditional IQA typically compared distorted image pixels with the original image pixels. Therefore, a full reference IQA model was proposed. Using CNN, a differentiable function was constructed to transform an image into a multi-scale over complete representation. The proposed measurement parameters were jointly optimized by combining texture similarity and structural similarity. The findings denoted that the improved method could interpret human perception scores on traditional image quality and texture databases [9]. Y Tian et al. raised a new full reference IQA method. Firstly, the brightness components of the reference and distorted images were transformed using radial symmetry to extract symmetry information. The experimental outcomes expressed that the raised method was conformed with the human visual system, and a large number of simulation results on dense light field datasets showed that the proposed method outperformed other algorithms in evaluating the quality of light field images [10]. Z Huang et al. found that traditional natural image processing methods were not suitable for different screen images, so they proposed a hash method for SCI perception. Firstly, it input the screen image through a joint preprocessing operation to obtain the max gradient amplitude and relevant direction from three color channels. Finally, it output the features to construct a hash sequence. The research outcomes denoted that this method was superior to the most advanced algorithms and could provide accurate predictions [11]. Z Pan et al. found that traditional no reference IQA methods were only feasible to particular scenarios and lack practicality. Therefore, a no reference IQA method based on multi-branch CNN was proposed. This method included a spatial feature extractor and a weight mechanism. Besides, a position vector was put forward to construct a weight mechanism to improve performance. The research outcomes denoted that the performance of this method was superior to traditional no reference IQA methods [12]. Z Zhang et al. found that existing IQA models rarely used 3D models to measure color information. Based on this, a no reference IQA model for 3D models was proposed. Firstly, the 3D space was projected into the geometric and color feature domains related to quality, and then quality perception features were extracted using 3D natural scene statistics and entropy. Finally, a support vector machine regression model was used to treat quality perception as a visual quality score. The research findings indicated that this method outperformed traditional no reference IQA models and improved the accuracy of estimation [13].

Through the research of numerous scholars at home and abroad, it is known that TSCNN can simultaneously capture different types of features and have achieved certain results in the field of image recognition. However, existing no reference IQA methods are often based on specific distortion types or assumptions, resulting in poor performance of the model when facing unknown or mixed distortions. The TSCNN can comprehensively understand image quality by processing different feature streams in parallel, thereby improving the robustness and generalization ability of the model. In view of this, a no reference image quality evaluation model based on TSCNN has been developed, which is expected to evaluate the quality of different images more quickly and accurately. The summary table in the related works section is shown in Table 1.

Table 1 Summary of research status of TSCNN s and image quality evaluation methods

| Reference | Model | Result |
|---|---|---|
| **X Liao et al. [4]** | Image operator chain sequential forensics framework based on TSCNN | It can capture evidence of tampering artifacts and local noise residuals, and has good stability in applications. |
| **M Attique Khan et al. [5]** | TSCNN information fusion framework | It can accurately classify images with a |

| | | |
|---|---|---|
| | | classification accuracy of over 89%. |
| **Y Sun et al. [6]** | TSCNN model based on multi level feature fusion | Compared with the single channel model, the average detection accuracy has increased by 1.08%, and the average accuracy has increased by 3.56%. |
| **Z Zhang et al. [7]** | A TSCNN composed of spatial paths and detail paths | Low resolution feature maps learn global information, while high-resolution feature maps extract local details. |
| **X Li et al. [8]** | 2D TSCNN | Can simultaneously extract spectral features and global spatial features |
| **K Ding et al. [9]** | A full reference image quality model that resamples textures with clear tolerances | Better explain human perceptual scores on traditional image quality databases and texture databases. |
| **Y Tian et al. [10]** | A full reference IQA method based on symmetry and depth feature models | This method is consistent with the human visual system and performs better than other algorithms on dense light field datasets |
| **Z Huang et al. [11]** | A hash method for SCI perception | This method provides accurate prediction of SCI quality, which is superior to state-of-the-art algorithms |
| **Z Pan et al. [12]** | No reference IQA method based on TSCNN | Overcoming the limitations of traditional no reference IQA methods that are only applicable to specific scenarios |
| **Z Zhang et al. [13]** | A no reference quality assessment model for 3D models represented by point clouds and networks | Innovatively using 3D models for measuring color information has improved the accuracy of estimation in no reference IQA models |

# 3 Construction of a no reference IQA model

To extract the features of the text and image parts of the SCI and accurately perceive the image quality, this part is composed of two sections to construct a model. The first section extracts the feature information of text and image based on AFF. In the second part, based on the improved TSCNN, the transfer learning model is introduced to build an IQA model.

## 3.1 Construction of no reference IQA model based on AFF

SCI is a new form of image commonly used for transmission between various electronic devices [14]. Due to the varying degrees of distortion in the text and image parts during the dissemination of SCI, it affects the user's visual experience. AFF is a method that integrates the feature information of different regions to obtain global features. Therefore, SCI is divided into text and image regions, and based on the AFF method, a no reference image feature extraction model is constructed. The gradient histogram of multiple derivatives is utilized for feature extraction of text area, and the wavelet subband energy and multi-directional gradient local binary patterns histogram are applied for feature extraction of image area [15]. These methods are utilized to extract different types of feature information, including texture, shape, color, etc. Finally, the feature information is fused to obtain a global feature representation. The schematic diagram of AFF method is denoted in Figure 1.

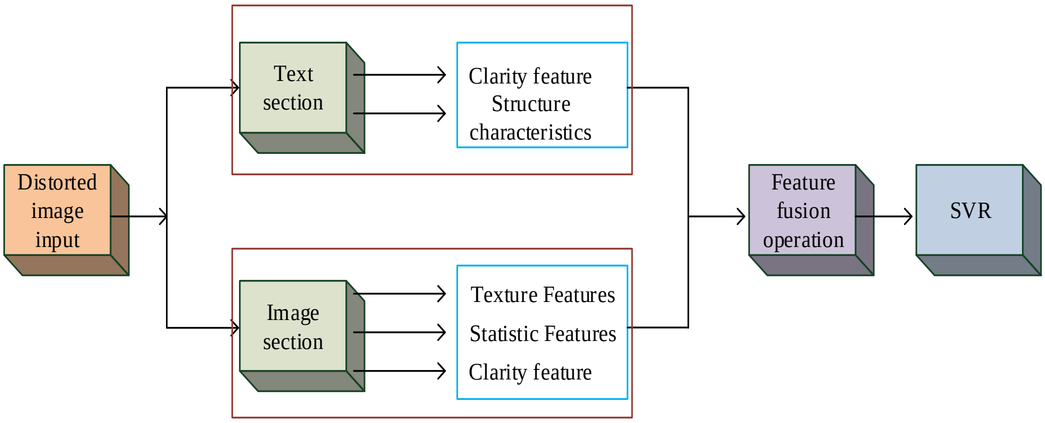


Figure 1: Schematic diagram of AFF method

Before feature extraction of the image, the ultra lightweight optical character recognition method based on the fly propeller depth learning framework is used to recognize the text area of the image [16]. Optical character recognition in this method can use computer technology to recognize, transform printed, handwritten or other forms of text information, and achieve ultra lightweight through model compression and acceleration. For the definition measurement of text area, digital orthophoto measurements (DOM) are introduced for the edge change processing of text image [17]. To identify the edges of the image, one-dimensional filtering is lied to process the text image. The direction of the image is divided into the x and y directions, and the definition equation in the x direction is shown in equation (1).

$$\Delta DOM_x(i,j)=[I_M(i+2,j)-I_M(i,j)]-[I_M(i,j)-I_M(i-2,j)] \quad (1)$$

In equation (1), $(i,j)$ is the pixel coordinate of the image. $I_M(i,j)$ denotes the grayscale value of the median filtered image at pixel $(i,j)$, with a deviation value of 2. The clarity equation in the x-direction is denoted in equation (2).

$$S_x(i,j)=\frac{\sum_{i-w\le i+w}\left|\Delta DoM_x(k,j)\right|}{\sum_{i-w\le i+w}\left|I(k,j)-I(k-1,j)\right|} \quad (2)$$

In equation (2), $I(k,j)$ indicates the pixel values of the image at $(k,j)$. $\sum_{i-w\le i+w}\left|\Delta DoM_x(k,j)\right|$ expresses the sum on a window of size $2w+1$.

$\sum_{i-w\le i+w}\left|I(k,j)-I(k-1,j)\right|$ represents the contrast on a window of size $2w+1$. When the clarity is greater than the preset threshold, it indicates that the image is a clear feature point. The calculation for the proportion of clear feature points in the x and y directions to the total feature points is denoted in equation (3).

$$\begin{cases} R_{\lambda(x,y)}=\dfrac{SP_x}{EP_x} \\ R_{\lambda(x,y)}=\dfrac{SP_y}{EP_y} \end{cases} \quad (3)$$

In equation (3), $\lambda$ indicates the x or y direction. $SP_y$ and $EP_x$ refer to the amount of clear and edge pixels in the x and y directions, respectively. To capture the subtle structural changes in the text image, a multi-derivative histogram of oriented gradients (HOG) is used, and the HOG operator is introduced to calculate the gradient direction and size in the image [18]. Based on the calculated values, it counts the amount of different gradient directions within a local image block, and generates feature vectors that describe the image structure. The gradient increase of the first order differential is indicated in equation (4).

$$X^n(x,y)=\sqrt{G_h^n(x,y)^2+G_v^n(x,y)^2} \quad (4)$$

In equation (4), $X^n(x,y)$ is the amplitude and direction of the n-order differential; $G_h^n(x,y)$ and $G_v^n(x,y)$ mean the gradient amplitudes in the horizontal and vertical directions of n-1 order differentiation, respectively. The calculation expression is shown in equation (5).

$$\begin{cases} G_x^n(x,y)=X^{n-1}(x,y)\otimes p_h \\ G_y^n(x,y)=X^{n-1}(x,y)\otimes p_v \end{cases} \quad (5)$$

In equation (5), $\chi^{n-1}$ refers to n-1 order differentiation. $p_h$ and $p_v$ mean filters in the horizontal and vertical directions, respectively. $\otimes$ expresses convolution operations. Due to changes in lighting and contrast in the image, the same image target may have different gradients at different positions, resulting in significant differences in the magnitude of the gradient vector [19]. To avoid this situation, normalize the gradient histogram of each image block, as shown in equation (6).

$$f'_{m,j} = \frac{f_{m,j}}{\left\|\vec{f}_m\right\|_2 + \varepsilon}, \vec{f}_m = [f_{m,1}, f_{m,2}, \cdots f_{m,36}] \quad (6)$$

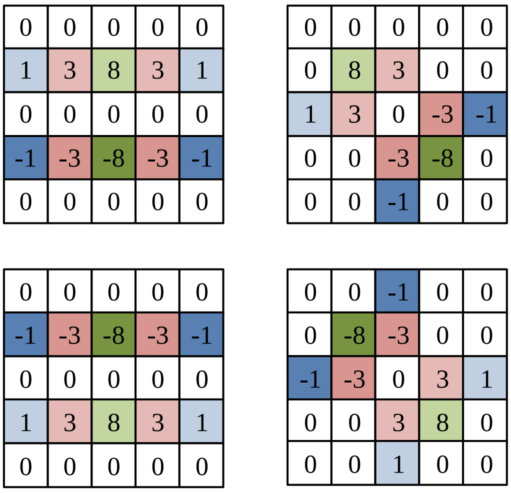
In equation (6), $f_{m,j}$ represents the j-th feature of the m-th image block before normalization. $f'_{m,j}$ represents the j-th feature of the m-th image block after normalization. $\vec{f}_m$ is the 36-dimensional HOG feature vector of the m-th image block. $\left\|\vec{f}_m\right\|_2$ means the norm of $\vec{f}_m$ is 2. $\varepsilon$ is a constant. Finally, it integrates the HOG features of all text images, as shown in equation (7).

$$f_j = \frac{1}{N_B}\sum_{m=1}^{N_B} f'_{m,j}, (j = 1, 2, \cdots 36) \quad (7)$$

In equation (7), $f_j$ means the j-th HOG feature, and $N_B$ denotes the amount of text image blocks. Based on the above calculation, the structural feature information of the text image area can be gained. For texture feature extraction in image regions, the local binary pattern (LOG) operator is used to extract in multi-directional gradient maps. The calculation method for multi-directional gradient maps is shown in equation (8).

$$G_i = g_i \otimes I, (i = 1, 2, \cdots, 8) \quad (8)$$

In equation (8), $G_i$ indicates the i-th gradient graph, and $g_i$ refers to the i-th directional gradient operator. The schematic diagram of the gradient multi-directional operator is shown in Figure 2.



| | | | | |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1 | 3 | 8 | 3 | 1 |
| 0 | 0 | 0 | 0 | 0 |
| -1 | -3 | -8 | -3 | -1 |
| 0 | 0 | 0 | 0 | 0 |

| | | | | |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 8 | 3 | 0 | 0 |
| 1 | 3 | 0 | -3 | -1 |
| 0 | 0 | -3 | -8 | 0 |
| 0 | 0 | -1 | 0 | 0 |

| | | | | |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| -1 | -3 | -8 | -3 | -1 |
| 0 | 0 | 0 | 0 | 0 |
| 1 | 3 | 8 | 3 | 1 |
| 0 | 0 | 0 | 0 | 0 |

| | | | | |
|---|---|---|---|---|
| 0 | 0 | -1 | 0 | 0 |
| 0 | -8 | -3 | 0 | 0 |
| -1 | -3 | 0 | 3 | 1 |
| 0 | 0 | 3 | 8 | 0 |
| 0 | 0 | 1 | 0 | 0 |

Figure 2: Schematic diagram of gradient multi-directional operator

The rotation invariant uniform mode is calculated on each gradient graph, as shown in equation (9).

$$RLBP_P^R = \begin{cases} \sum_{P=0}^{P-1} s(g_p - g_c), & \psi(LBP_p^R) \le 2 \\ P+1, & Otherwise \end{cases} \quad (9)$$

In equation (9), $g_c$ indicates the intensity value of the center pixel value; $g_p$ means the intensity value of the pixel value within the radius R neighborhood of the center pixel; $P$ denotes the amount of adjacent pixels; $\psi$ is used to calculate the amount of bitwise jumps. For the natural image portion in the image region, the no reference image spatial quality evaluator (BRISQUE) is used to statistically assess the features, and the mean subtracted contrast normalized (MSCN) is subtracted from the fitted mean to estimate the parameters of the generalized Gaussian distribution (GD) and the asymmetric generalized GD [20]. The calculation method for the generalized GD is expressed in equation (10).

$$f(x, \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} \exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha}\right) \quad (10)$$

In equation (10), $x$ means the eigenvalues; $\alpha$ refers to the shape parameters; $\sigma^2$ indicates the variance; $\beta$ stands for the coefficients; $\Gamma$ represents the gamma function. The calculation expression for asymmetric generalized GD is shown in equation (11).

$$f(x,v,\sigma_l^2,\sigma_r^2)\begin{cases}\dfrac{v}{(\beta_l+\beta_r)\Gamma(\frac{1}{v})}exp(-(\dfrac{-x}{\beta_l})^v), x<0\\ \dfrac{v}{(\beta_l+\beta_r)\Gamma(\frac{1}{v})}exp(-(\dfrac{-x}{\beta_r})^v), x\geq 0\end{cases} \quad (11)$$

The parameters in equation (11) are the same as those in equation (10). By exporting the eigenvalues through generalized GD and asymmetric generalized GD, the eigenvalues of the natural image in the image region can be obtained. The method of extracting wavelet subbands is utilized to extract the clarity features of the image region. It calculates the total logarithmic energy in each level of wavelet subband based on the logarithmic energy of each decomposed wavelet subband, as shown in equation (12).

$$E_n = \frac{\frac{1}{2}(E_{LH_n}+E_{HL_n})+\alpha E_{HH_n}}{1+\alpha} \quad (12)$$

In equation (12), $\alpha$ expresses the parameter, and $E_{LH_n}$, $E_{HL_n}$, and $E_{HH_n}$ represent the logarithmic energy of different subbands. After calculating the capabilities in different subbands, the statistical features $E_2$ and $E_3$ of the two image regions are ultimately obtained. The flowchart is shown in Figure 3.

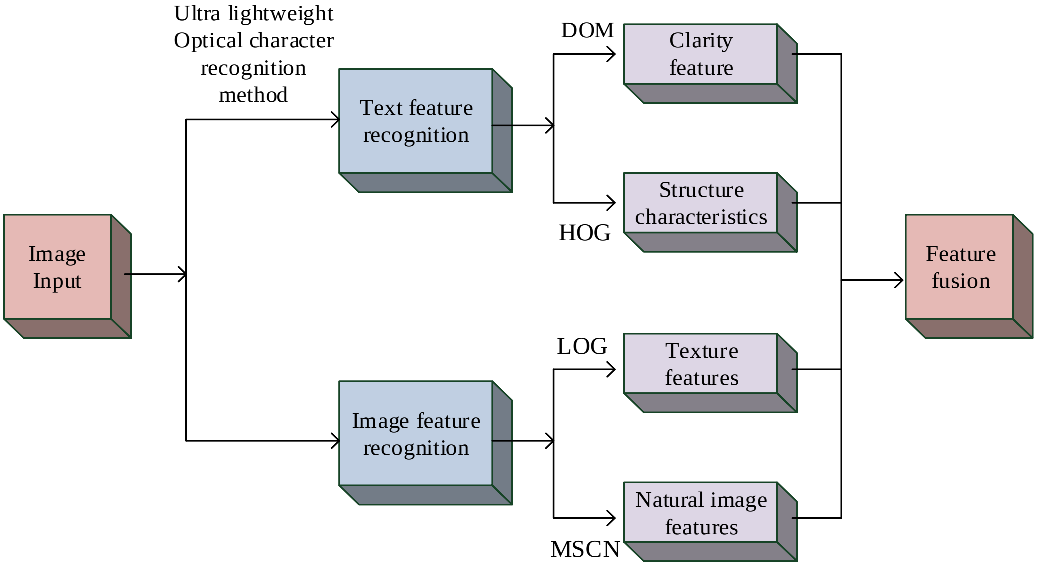


Figure 3: Flow chart of image feature extraction

## 3.2 Construction of IQA model based on TSCNN

Most image learning models input is grayscale, which makes the model insensitive to distortion types such as brightness and contrast. Therefore, TSCNN is used to learn distorted images. The TSCNN can be divided into two sub models, one is the CNN for processing RGB images, and the other is the gradient image branch for processing optical flow information [21]. The two sub-models have similar structures. The RGB branch network in the TSCNN can obtain the visual features in the image by analyzing the static information in the RGB image [22]. To prevent the over fitting phenomenon of the TSCNN, transfer learning is introduced into the RGB branch network to improve the network’s learning ability. The structure of the improved TSCNN is shown in Figure 4.

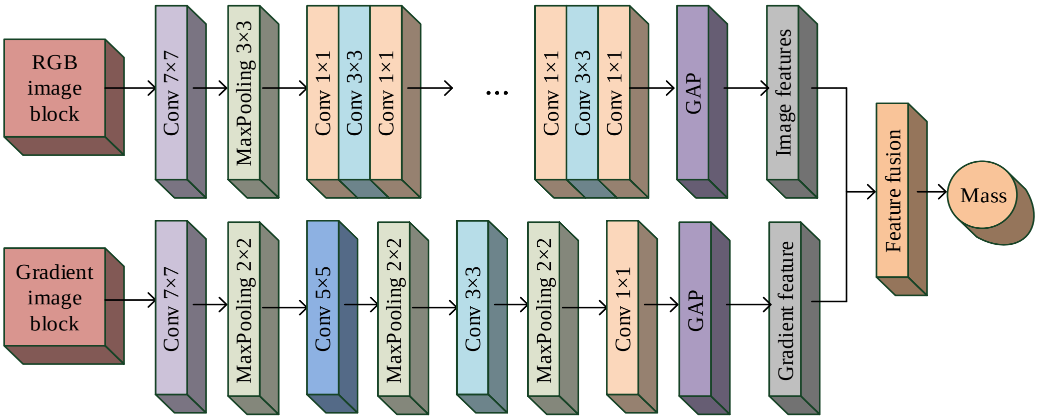


Figure 4: Structure Diagram of Improved TSCNN

The gradient image branch input in the TSCNN is the gradient information of the optical flow image. Each pixel represents the displacement information of the object between two adjacent images. By using convolutional layers in gradient image branches, each gradient information is extracted, and the output of the convolutional layer is transmitted to the next one for higher-level feature extraction and classification [23-24]. After preprocessing the gradient image, the original gradient image is transferred into a standard input format. Usually, the Sobel operator is applied to work out the difference between adjacent frames of images, to calculate the gradient value of each pixel point. The gradient image calculatio expression is shown in equation (13).

$$G=\sqrt{G_x^2+G_y^2} \quad (13)$$

In equation (13), $G_x$ and $G_y$ mean the gradient maps in the horizontal and vertical directions. residual neural network (ResNet) is used in RGB image branch networks to solve gradient vanishing and explosion problems in deep networks. The ResNet has different residual structures according to the number of layers, including ResNet18, 34, 50, 101, 152. Meanwhile, ResNet is divided into conv_1, conv2_x, conv3_x, conv4_x, and conv5_x. It removes the average pooling and fully connected layers from ResNet50 and extracts conv5_x feature map output. The extracted feature maps are used as feature vectors extracted by the RGB image branch network through a global average pooling (GAP) layer. In the gradient image branch network, ReLU function is utilized as the activation function of each convolution layer, and the characteristic expression of gradient flow is shown in equation (14).

$$F_g=f(\theta_g,I_g) \, (14)$$

In equation (14), $F_g$ expresses the quality characteristics of gradient flow; $f(\theta_g,I_g)$ denotes the structure of gradient flow network; $\theta_g$ means the parameters of gradient flow network; $I_g$ represents the gradient image block. The feature maps extracted from conv5_x are subjected to the GAP operation for image features. The comparison between GAP and the original operation is shown in Figure 5.

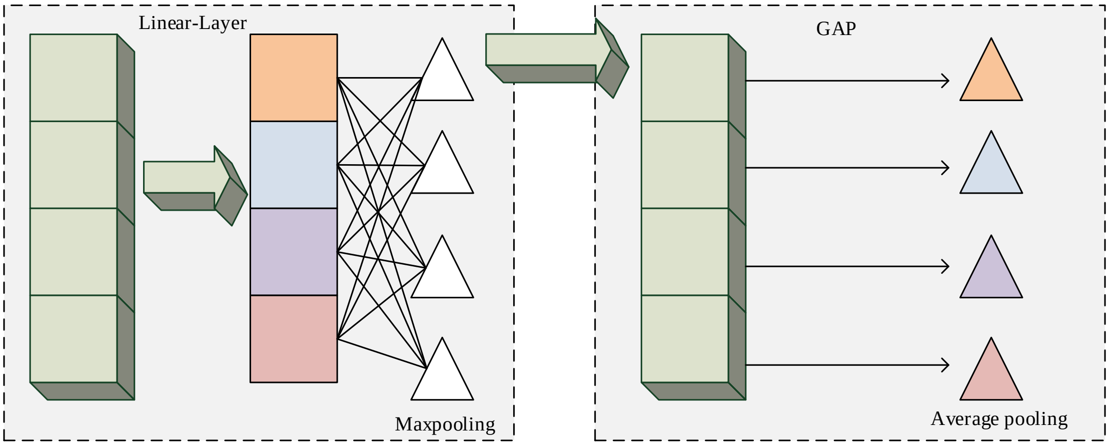


Figure 5: Comparison between GAP and original operation

The GAP operation directly performs GAP on the feature map output from the last convolutional layer, obtaining the global average features of the entire image. The specific operation of GAP is to input a set of feature maps with a size of H×W×C. Average pooling of all pixels is processed in the same channel, outputting a set of average feature vectors containing C elements, with a size of 1×1×C. Compared to fully connected layers, GAP operation reduces the parameters of the model, avoids overfitting problems, and thus improves the generalization performance of the model [25]. Therefore, the image stream feature expression is changed to as shown in equation (15).

$$F_r = f(\theta_r, I_r) \quad (15)$$

In equation (15), $F_r$ indicates the quality characteristics of the image stream; $f(\theta_r, I_r)$ denotes the RGB network structure; $\theta_r$ means the parameters of the image stream branch network; $I_r$ expresses the RGB image block. The output feature vectors of two branches can be fused on different layers. The high-level feature vectors generated by the two branches are concatenated together, and then the concatenated feature vectors are sent to the fully connected layer for classification. After obtaining gradient and RGB image features, the concatenate connection method is used to fuse the features. The fused feature expression is shown in equation (16).

$$F = concat\,(f_g, f_r) \quad (16)$$

In equation (16), $F$ denotes the fused features. It is input into the fully connected layer for linear regression operation, and the one-dimensional feature vector obtained is utilized as the quality score [26]. Afterwards, the image and gradient blocks are input into the corresponding RGB image branch network and gradient image feature network for training. The average output quality score of the two blocks is regarded as the overall quality score [27]. The expression is shown in equation (17).

$$\frac{1}{N_p}\sum_{i=1}^{N_p} f(F,\theta) = q \quad (17)$$

In equation (17), $N_p$ stands for the amount of distorted image blocks; $f$ refers to the network mapping function; $\theta$ expresses the parameters of the network; $q$ represents the distorted image's quality prediction score. The L1 norm is used as the loss function, and the expression is expressed in equation (18).

$$L = \frac{1}{N}\sum_{i=1}^{N} \left\| q_i - y_i \right\|_1 \quad (18)$$

In equation (18), $N$ refers to the amount of images in the training set; $q_i$ denotes the predicted image quality score; $y_i$ indicates the training label of the image block as the distorted image's subjective quality score. The fusion operation of feature vectors generated by two branches can reflect the interactive information between two features to improve the accuracy and robustness of visual tasks [28]. The flowchart is shown in Figure 6.

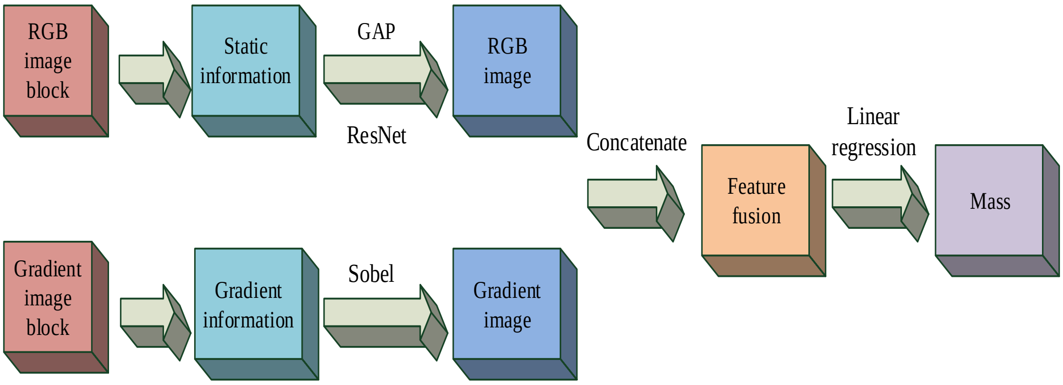


Figure 6: Flow chart of IQA model

# 4 Testing of a no reference IQA model

This part is divided into two sections for testing. The first section tests the no-reference IQA model based on AFF, and analyzes the effectiveness of feature extraction and fusion operations of different regions in quality assessment. The second section tests the IQA model with TSCNN, and analyzes its performance in the actual data set.

## 4.1 Performance test of no reference IQA model based on AFF

To test the no reference IQA model with AFF, the hardware environment of the experiment was set to MATLAB2018Ra, the computer was configured with 8GB memory, i5-6200U processor, and the traditional full and no reference types methods were selected, including peak signal to noise ratio (PSNR), structural similarity index (SSIM), natural image quality evaluator (NIQE), structure-guided IQA (SIQE), and gradient similarity score (GSS). After 10 experiments, the average value of the time required to evaluate an 512×512 size image on the CSIQ dataset is compared with the no reference IQA model based on AFF. The test results are shown in Figure 7.

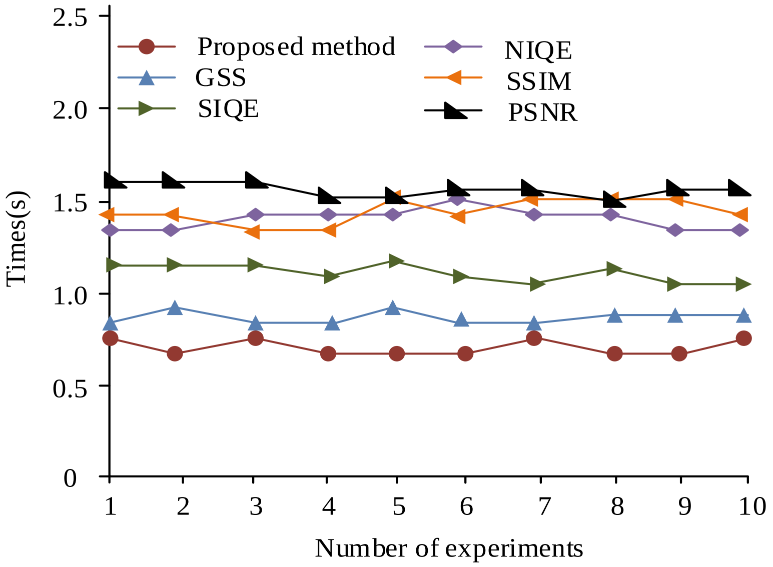
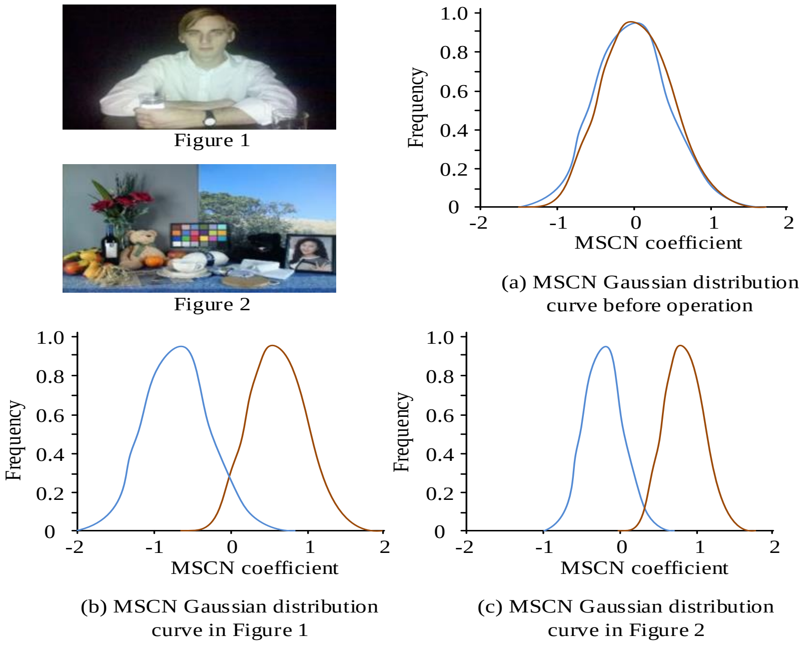


Figure 7: Time required for image assessment using different methods

In Figure 7, the average time consumed by the no reference IQA model based on AFF, GSS, SIQE, NIQE, SSIM and PSNR was 0.65s, 0.82s, 1.22s,1.42s, 1.51s, and 1.65s, respectively. The no reference IQA model based on AFF consumed the least time and had the fastest extraction efficiency. To prove the performance of the method of segmenting images for recognition, two images were randomly selected from the CID2013 database. Using frequency as the vertical axis and MSCN value as the horizontal axis, a GD map was drew, as shown in Figure 8.



(a) MSCN Gaussian distribution curve before operation

(b) MSCN Gaussian distribution curve in Figure 1

(c) MSCN Gaussian distribution curve in Figure 2

Figure 8: MSCN GD curves for different graphs

Figure 8 (a) shows the initial MSCN coefficient curves of the two images. They had similar coefficient values and could not be directly distinguished and evaluated for the features in the images. Figures 8 (b) and 8 (c) show the MSCN coefficient curves of the two images after feature recognition and extraction. The MSCN coefficients of the same area in different images followed a GD, and effectively distinguished between non textured and textured regions of the image, which was conducive to image recognition and assessment. To analyze the accuracy, monotonicity and predictability of the algorithm, Spearman rank order correlation coefficient (SROCC), Pearson linear correlation coefficient (PLCC), Kendal rank order correlation coefficient (KROCC) and root mean square error (RMSE) were selected as assessment indicators. The test results on the CID2013 dataset are displayed in Table 2.

Table 2: Test results of different methods

| Index | PSNR | SSIM | NIQE | SIQE | GSS | Proposed method |
|---|---|---|---|---|---|---|
| SROCC | 0.3998 | 0.4253 | 0.4836 | 0.4982 | 0.4955 | 0.5263 |
| PLCC | 0.1892 | 0.2956 | 0.3016 | 0.3241 | 0.3521 | 0.3852 |
| KROCC | 0.3655 | 0.4026 | 0.4589 | 0.4925 | 0.5219 | 0.5745 |
| RMSE | 1.2652 | 1.1758 | 1.1225 | 1.1104 | 1.0358 | 1.0012 |

From Table 2, the four index values of the five models had little difference. The no reference IQA model based on AFF’s SROCC, PLCC, KROCC and RMSE were 0.5263, 0.3852, 0.5745 and 1.0012, respectively, which were better than the other five methods. Therefore, the no reference IQA model based on AFF divided and extracted different features in the image, which could better meet the needs of human vision.

## 4.2 Test and analysis of IQA model based on TSCNN

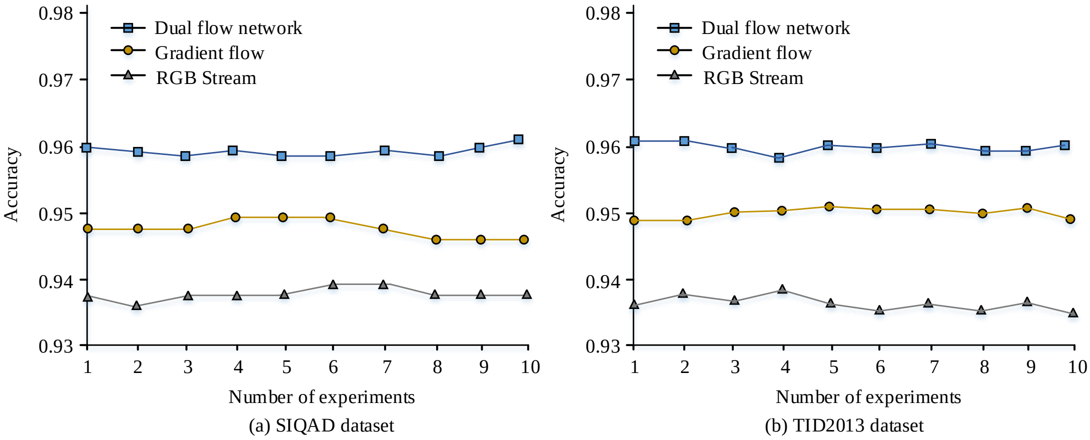
To evaluate the effectiveness of the IQA model using TSCNN, experiments were conducted on the SIQAD dataset and TID2013 dataset, which contained 1000 high-quality raw images and corresponding distorted images of different resolutions, such as 640*480, 1280*720, and 1920*1080. The types of distortion included amplitude distortion, phase distortion, harmonic distortion, intermodulation distortion, transient distortion, and interface distortion, which could comprehensively test and verify the robustness of IQA algorithms. It randomly selected 60% distorted images for training, 10% for image validation, and 30% for testing.

The experiment was conducted in the Window-10 system, and various parameter settings are shown in Table 3.

Table 3: Experimental environment parameter setting table

| Compilation tools and environment | Local settings | Cloud settings |
|---|---|---|
| Equipment system | Window-10 | Linux-5.4.109 |
| CPU | Intel(R)Core (TM)i5-10300H@2.50GHz@RAM:16G | Intel(R)Xeon(R)CPU@2.50GHz@RAM:16G |
| Graphics card settings | GTX1650Ti | Nvidia-K80(CPU)Tesla-T4(TPU) |
| Video storage | 4G | 12G |
| Compilation tools | PyCharm-2020.1.1 | Colaboratory |
| Compiler language | Python | Python |
| Language framework | 3.8 | 3.8 |
| Framework | PyTorch-1.9.0(CPU) | PyTorch-1.10.0(+GPU) |

Due to the large amount of parallel computing required for model training, there is a high demand for CPU processing. Therefore, Intel (R) Xeon (R) CPU @ 2.50GHz was chosen as the CPU. The locally set GTX 1650 Ti was suitable for graphics rendering, while the cloud set Nvidia K80 and Tesla T4 could provide better computing power. PyCharm 2020.1.1 in integrated development environment provided various functions such as code editing control, which was more suitable for local development. Colaboratory could greatly reduce the threshold for deep learning experiments. This configuration allowed for smoother acquisition of experimental results. To verify the effectiveness of the TSCNN in image processing under the joint action of RGB branch and gradient image branch, the dual channel in the TSCNN was compared with the single stream network of RGB branch and gradient image branch. The experiment outcomes are indicated in Figure 9.

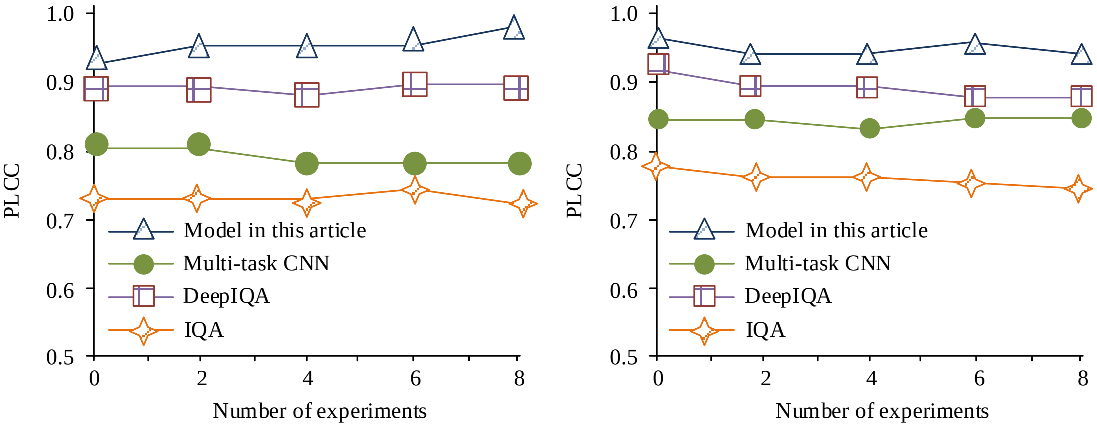


(a) SIQAD dataset

(b) TID2013 dataset

Figure 9: Accuracy Comparison of single stream networks and TSCNN

From Figure 9(a), the accuracies of the individual RGB branch network test, individual gradient branch network test and TSCNN test were 93.7%, 95.2% and 96.1%, respectively. Therefore, the testing accuracy of the TSCNN was the highest, and the gradient network was slightly better than the RGB network. The TSCNN had the best performance. From Figure 9 (b), whether tested on the SIQAD dataset or the TID2013 dataset, the results showed that the dual stream network had the highest testing accuracy, indicating that the method had good robustness. To examine the assessment quality of the model under the distortion type, the traditional IQA model, multi-task convolutional neural network (Multi-task CNN), and deep IQA (DeepIQA) were selected to compare with the PLCC value of 0.7 of the image assessment model based on the TSCNN. The test results under exposure distortion and jitter conditions are shown in Figure 10.



(a) Accuracy curve under exposure distortion

(b) Accuracy curve under jitter distortion

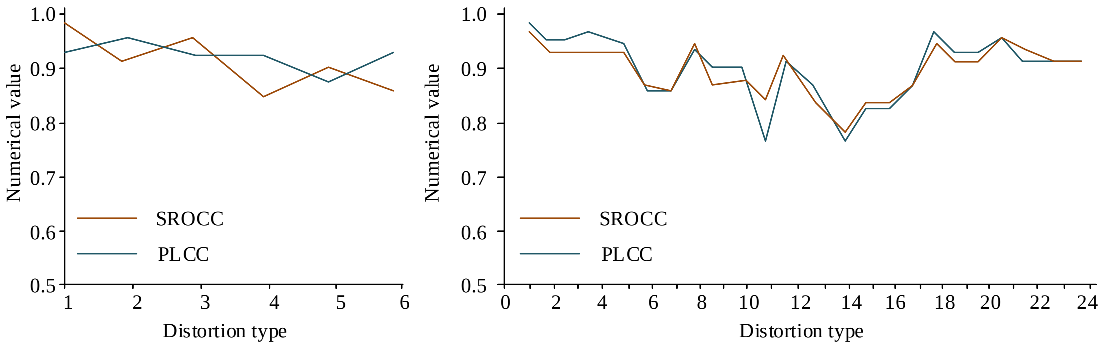
Figure 10: Accuracy curve under exposure and iitter distortion conditions

From Figure 10 (a), under the condition of exposure distortion, the average PLCC value of the image assessment model based on the TSCNN, IQA, Multi-task CNN and DeepIQA was 0.9253, 0.7302, 0.7958 and 0.8952, respectively. From Figure 10 (b), under jitter conditions, the average PLCC of the image assessment model based on the TSCNN, IQA, Multi-task CNN and DeepIQA was 0.9524, 0.7751, 0.8315 and 0.9042, respectively. Compared to the other three models, the image assessment model based on TSCNN had a higher PLCC value and a more linear correlation between the image and the original image. The IQA, Multi-task CNN, DeepIQA, NIQE and SIQE methods were compared with the IQA model with TSCNN on the SIQAD dataset and TID2013 dataset. The experimental outcomes of SROCC, PLCC, KROCC and RMSE values are expressed in Table 4.

Table 4: Test results on the SIQAD dataset and TID2013 dataset

| **Method** | SIQAD dataset | | | | TID2013 dataset | | | |
|---|---|---|---|---|---|---|---|---|
| | **SROCC** | **PLCC** | **KROCC** | **RMSE** | **SROCC** | **PLCC** | **KROCC** | **RMSE** |
| IQA | 0.3568 | 0.6895 | 0.5284 | 20.5412 | 0.3568 | 0.6895 | 0.5284 | 20.5412 |
| Multi-task CNN | 0.9085 | 0.7423 | 0.7893 | 8.5935 | 0.9085 | 0.7423 | 0.7893 | 8.5935 |
| DeepIQA | 0.8952 | 0.8534 | 0.5912 | 12.4950 | 0.8952 | 0.8534 | 0.5912 | 12.4950 |
| NIQE | 0.6815 | 0.6895 | 0.5026 | 11.2595 | 0.6815 | 0.6895 | 0.5026 | 11.2595 |
| SIQE | 0.8836 | 0.7946 | 0.7028 | 9.5348 | 0.8836 | 0.7946 | 0.7028 | 9.5348 |
| Proposed method | 0.9242 | 0.9284 | 0.8059 | 7.8523 | 0.9242 | 0.9284 | 0.8059 | 7.8523 |

From Table 4, in the SIQAD dataset, the SROCC value of the image evaluation model based on the TSCNN was 0.9242, the PLCC value was 0.9284, the KROCC value was 0.8059, and the RMSE value was 7.8523. The results were not significantly different from those tested in the TID2013 dataset, indicating that the method has good robustness. Meanwhile, these four indicators showed that the IQA model based on TSCNN had stronger expression ability for image information in the field of IQA, and was superior to the other five models, with better performance. To verify the model's generalization ability, SIQAD dataset and TID2013 were selected for testing. The number of distortion types was 6 on the CSIQ and 24 on the TID2013, respectively. The test results are shown in Figure 11.



(a) Test results on the CSIQ dataset

(b) Test results on the TID2013 dataset

Figure 11: Test result curves on different datasets

Figures 11 (a) and 11 (b) show the SROCC and PLCC values tested by the model on the SIQAD dataset and TID2013 datasets, respectively. The IQA model based on TSCNN achieved good experimental results for most distortion types on two different datasets, proving its good generalization ability in distortion type testing across datasets. To prove the assessment quality of the model, that is, the relationship between quality score and subjective quality score, six distortion types were selected on the SIQAD dataset. The Manders overall overlap coefficient (MOC) was used as the horizontal axis, and the predicted MOC value was used as the vertical axis. The results of the scatter plot are shown in Figure 12.

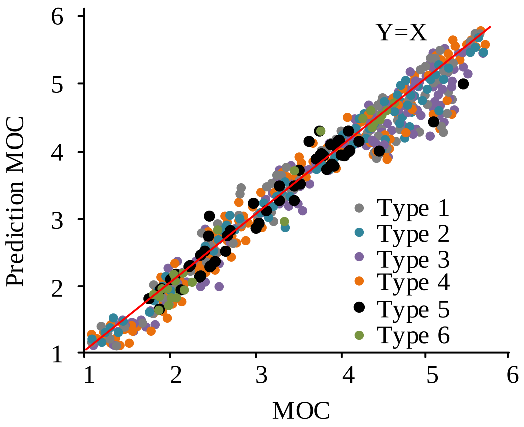


Figure 12: Scatter plot of subjective and objective quality scores

From Figure 12, the scatter points of the six types of distortion had better fitting effects on the Y=X line. The vast majority of distortion type data points were close to the Y=X line, and only a few data points did not fit the curve. This could indicate that the quality score predicted by the IQA model based on TSCNN was close to the subjective quality score, demonstrating the effectiveness of evaluating no reference images.

# 5 Discussion

Testing was conducted on the SIQAD dataset to evaluate the performance of various models on distorted images. The accuracy of the RGB branch network test alone was 93.7%, the accuracy of the gradient branch network test alone was 95.2%, and the accuracy of the TSCNN was 96.1%. This was because the RGB branch mainly focused on the original pixel information of the image, including intuitive visual features such as color and brightness. However, it was not sensitive enough to distortion effects such as blurring and noise. The gradient information branch reflected the changes in pixel values in an image and could capture structural features such as edges and textures of the image. The dual stream network could comprehensively utilize the advantages of both feature information by fusing the outputs of RGB and gradient branches. This fusion not only preserved the basic visual information and color features of the image, but also enhanced the network's ability to recognize image details and structural distortions. Therefore, dual stream networks can demonstrate higher accuracy in evaluating distorted images.

Under the condition of exposure distortion, the average PLCC values of the image evaluation model, IQA model, Multi-task CNN model, and DeepIQA model based on TSCNN were 0.9253, 0.7302, 0.7958, and 0.8952, respectively. Under shaking conditions, the average PLCC values of these four models were 0.9524, 0.7751, 0.8315, and 0.9042, respectively. Compared to the results of Z Zhang et al. [7] and Z Pan et al. [12], the PLCC value of the image evaluation model based on TSCNN was higher, and the image was more linearly correlated with the original image. This was because the dual stream structure allowed the model to simultaneously process different levels of information in the image, such as first stream processing global information (such as overall brightness, contrast, etc.), and second stream processing local details (such as edges, textures, etc.). This design enabled the model to cope with effects such as exposure distortion.

The image evaluation model based on TSCNN could simultaneously process different levels of information in the image and comprehensively capture the features of the image. Compared to traditional single stream convolutional networks or other types of IQA models, the image evaluation models of TSCNN typically exhibited higher accuracy in NR-IQA tasks.

# 6 Conclusion

To extract and evaluate the no reference image’s features and quality, this research constructed an IQA model with AFF and a model with TSCNN. The research findings of the no reference IQA model based on AFF indicated that the average time spent on model feature extraction was 0.65s. From the MSCN coefficient curve, the MSCN coefficients of different images in the same region followed a GD, and effectively distinguished between non textured and textured regions of the image, which was conducive to image recognition and assessment. The no reference IQA model based on AFF had SROCC value of 0.5263, PLCC value of 0.3852, KROCC value of 0.5745, and RMSE value of 1.0012, which were superior to the other five methods. The experimental results of the no reference IQA model with TSCNN showed that the accuracy of the single RGB branch network test was 93.7%, the accuracy of the single gradient branch network test was 95.2%, and the accuracy of the TSCNN test was 96.1%. On the SPAQ dataset, the IQA model based on TSCNN had SROCC values of 0.9242, PLCC values of 0.9284, KROCC values of 0.8059, and RMSE values of 7.8523. The SROCC and PLCC values tested on the CSIQ dataset and TID2013 dataset demonstrated good generalization ability in distortion type testing across datasets. The predicted and subjective quality scores of the model followed the Y=X curve, proving the effectiveness of evaluating no referenced images. There are also shortcomings in this study. Due to the small and single size of the SCI database, the experimental limitations caused by insufficient training data should be addressed in future research.

# References

[1] Das R K, Ahmed N, Pohrmen F H, Maji A K, and Saha G. 6LE-SDN: An edge-based software-defined network for Internet of Things. IEEE Internet of Things Journal, 7(8):7725-7733, 2020. https://doi. 10.1109/JIOT.2020.2990936.

[2] Kou L, Liu C, Cai G, Zhou J N, Yuan Q D, and Pang S M. Fault diagnosis for open-circuit faults in NPC inverter based on knowledge-driven and data-driven approaches. IET Power Electronics, 13(6):1236-1245, 2020. https://doi. 10.1049/iet-pel.2019.0835.

[3] Chen M, and He Y. Multiple open-circuit fault diagnosis method in NPC rectifiers using fault injection strategy. IEEE Transactions on Power Electronics, 37(7):8554-8571, 2022. https://doi. 10.1109/TPEL.2022.3150885.

[4] Liao X, Li K, Zhu X and, Liu K R. Robust detection of image operator chain with two-stream convolutional neural network. IEEE Journal of Selected Topics in Signal Processing, 14(5):955-968, 2020. https://doi. 10.1109/JSTSP.2020.3002391.

[5] Attique Khan M, Sharif M, Akram T, Kadry S, and Hsu C H. A two-stream deep neural network-based

intelligent system for complex skin cancer types classification. International Journal of Intelligent Systems, 37(12):10621-10649, 2022. https://doi. doi.org/10.1002/int.22691.

[6] Sun Y, Weng Y, Luo B, Li G, Tao B, Jiang D, and Chen D. Gesture recognition algorithm based on multi-scale feature fusion in RGB-D images. IET Image Processing, 17(4):1280-1290, 2022. https://doi. 10.1049/ipr2.12712.

[7] Z Zhang T, and Liu P Nie. Real-Time Semantic Segmentation for Road Scene Based on Data Enhancement and Dual-Path Fusion Network, Acta Electronica Sinica, 50(7):1609-1620, 2022. https://doi. 10.12263/DZXB.20210611.

[8] X Li, M Ding, and A Piurica. Deep Feature Fusion via Two-Stream Convolutional Neural Network for Hyperspectral Image Classification, IEEE Trans. Geosci. Remote Sens, 58(4):2615-2629, 2020. https://doi. 10.1109/TGRS.2019.2952758.

[9] Ding K, Ma K, Wang S and Simoncelli E P. Image Quality Assessment: Unifying Structure and Texture Similarity, 44(5):2567-2581, 2020. https://doi. 10.1109/TPAMI.2020.3045810.

[10] Tian Y, Zeng H, Hou J, Chen J, Zhu J, and Ma K K. A Light Field Image Quality Assessment Model Based on Symmetry and Depth Features, 31(5):2046-2050, 2020. https://doi. 10.1109/TCSVT.2020.2971256.

[11] Huang Z, and Liu S. Perceptual hashing with visual content understanding for reduced-reference screen content image quality assessment, IEEE Trans. Circuits Syst. Video Technol, 31(7):2808-2823, 2020. https://doi. 10.1109/TCSVT.2020.3027001.

[12] Pan Z, Yuan F, Wang X, Xu L, Shao X, and Kwong S. No-reference image quality assessment via multibranch convolutional neural networks. IEEE Transactions on Artificial Intelligence, 4(1):148-160, 2023. https://doi. 10.1109/TAI.2022.3146804.

[13] Zhang Z, Sun W, Min X, Wang T, Lu W, and Zhai G. No-reference quality assessment for 3d colored point cloud and mesh models. IEEE Transactions on Circuits and Systems for Video Technology, 32(11):7618-7631, 2022. https://doi. 10.1109/TCSVT.2022.3186894.

[14] Xu M, Li C, Zhang S, and Le Callet P. State-of-the-art in 360 video/image processing: Perception, assessment and compression. IEEE Journal of Selected Topics in Signal Processing, 14(1):5-26, 2020. https://doi. 10.1109/JSTSP.2020.2966864.

[15] Kose K, Bozkurt A, Alessi-Fox C, Brooks D H, and Dy J G. Rajadhyaksha M and Gill M. Utilizing machine learning for image quality assessment for reflectance confocal microscopy. Journal of Investigative Dermatology, 140(6):1214-1222, 2020. https://doi. 10.1016/j.jid.2019.10.018.

[16] Greffier J, Dabli D, Hamard A, Akessoul P, Belaouni A, Beregi J P, and Frandon J. Impact of dose reduction and the use of an advanced model-based iterative reconstruction algorithm on spectral performance of a dual-source CT system: A task-based image quality assessment. Diagnostic and Interventional Imaging, 102(7-8):405-412, 2021. https://doi. /10.1016/j.diii.2021.03.002.

[17] Ueda T, Ohno Y, Yamamoto K, Murayama K, Ikedo M, and Yui M. Deep learning reconstruction of diffusion-weighted MRI improves image quality for prostatic imaging. Radiology, 303(2):373-381, 2022. https://doi. 10.1148/radiol.204097.

[18] Higashigaito K, Euler A, Eberhard M, Flohr T G, Schmidt B, and Alkadhi H. Contrast-enhanced abdominal CT with clinical photon-counting detector CT: assessment of image quality and comparison with energy-integrating detector CT. Academic radiology, 29(5):689-697, 2022. https://doi. 10.1016/j.acra.2021.06.018.

[19] Zan J. Research on robot path perception and optimization technology based on whale optimization algorithm. Journal of Computational and Cognitive Engineering, 1(4):201-208, 2022. https://doi. 10.47852/bonviewJCCE597820205514.

[20] Sartoretti T, Landsmann A, Nakhostin D, Eberhard M, Roeren C, Mergen V, and Euler A. Quantum iterative reconstruction for abdominal photon-counting detector CT improves image quality. Radiology, 303(2):339-348, 2022. https://doi. 10.1148/radiol.211931.

[21] Li B, Zhang W, Tian M, Zhai G, and Wang X. Blindly assess quality of in-the-wild videos via quality-aware pre-training and motion perception. IEEE Transactions on Circuits and Systems for Video Technology, 32(9):5944-5958, 2022. https://doi. 10.1109/TCSVT.2022.3164467.

[22] Oshima K, Onishi T, Kim S J, Ma J, and Hasegawa M. Efficient wireless network selection by using multi-armed bandit algorithm for mobile terminals. Nonlinear Theory and Its Applications, IEICE, 11(1):68-77, 2020. https://doi. 10.1587/nolta.11.68.

[23] Wang S, Wang H, and Huang L. Adaptive algorithms for multi-armed bandit with composite and anonymous feedback//Proceedings of the AAAI Conference on Artificial Intelligence, 35(11):10210-10217, 2021. https://doi. 10.1609/aaai.v35i11.17224.

[24] Chunxia T. Research on the multilevel security authorization method based on image content. Acta Electronica Malaysia, 1(1):18-20, 2017. http//:doi.org/ 10.26480/aim.02.2017.17.19

[25] Huang Y, Xu H, Gao H, Ma X, and Hussain W. SSUR: an approach to optimizing virtual machine allocation strategy based on user requirements for cloud data center. IEEE Transactions on Green Communications and Networking, 5(2):670-681, 2021. https://doi. 10.1109/TGCN.2021.3067374.

[26] Yang Y, and Song X. Research on face intelligent

perception technology integrating deep learning under different illumination intensities. Journal of Computational and Cognitive Engineering, 1(1):32-36, 2022. https://doi. 10.47852/bonviewJCCE19919.

[27] Guo H, Guo S, Xu J, and Tian X. Power switch open-circuit fault diagnosis of six-phase fault tolerant permanent magnet synchronous motor system under normal and fault-tolerant operation conditions using the average current Park's vector approach. IEEE Transactions on Power Electronics, 36(3):2641-2660, 2020. https://doi. 10.1109/TPEL.2020.3017637.

[28] Fath A H, Madanifar F, and Abbasi M. Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems. Petroleum, 6(1):80-91, 2020. https://doi. 10.1016/j.petlm.2018.12.002

.