IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 September 2023, accepted 27 October 2023, date of publication 16 November 2023,
date of current version 1 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3334154

RESEARCH ARTICLE

Stereoscopic Image Quality Evaluation Method for Visual Communication Design

XU HAN[1], QIUYUE SHAN[1], AND TIANSHU CHU[2]
[1]Department of Environmental Arts, Hebei University of Environmental Engineering, Qinhuangdao 066000, China
[2]School of Artificial Intelligence, Xidian University, Xi'an 710000, China

Corresponding author: Xu Han (hanxu_7292@163.com)

ABSTRACT With the development of computer vision technology, the demand for accurate recognition of stereoscopic image quality in the market is increasing. Accurate recognition of stereoscopic image quality is of great significance for providing high-value intelligent image services, which is also the motivation for this study. Accurately recognizing the quality of stereoscopic images is of great significance for image analysis and computer vision applications. In fields such as autonomous driving, medical image analysis, and industrial detection, accurate stereo image quality can provide reliable input for algorithms and improve the accuracy of analysis and recognition. The field or issue of this study is image quality evaluation, which aims to find higher performance stereoscopic image quality evaluation methods. Therefore, this study draws inspiration from the idea of ensemble learning and designs two Convolutional Neural Network (CNN) stereoscopic image quality evaluation methods based on the semantic features of stereoscopic images and the local detail perception module, and fuses them to form a mixed evaluation model. This study aims to solve the problem of image quality assessment, which is to accurately identify the quality of stereoscopic images and provide high-value intelligent image services. With the continuous development of computer vision technology, the demand for accurate recognition of stereoscopic image quality is increasing, which is also the motivation of this study. This study drew inspiration from the idea of ensemble learning and designed two hybrid evaluation models based on stereo image semantic features and local detail perception modules. Convolutional Neural Network (CNN) was used to achieve stereo image quality evaluation. This is one of the main contributions of this article. In order to evaluate the performance of the designed model, the LIVE 3D Phase I dataset was used for testing experiments. The expected results show that when the number of test samples is 500, the overall measurement values of the Spearman rank ordered correlation coefficient (SROCC) and Pearson linear correlation coefficient (PLCC) of the designed ICNN1 and ICNN2 stereo image quality evaluation models are 0.940, 0.949, and 0.940, 0.949, respectively. These results are significantly higher than the selected contrastive deep learning models and machine learning models. In addition, the designed model has relatively low computational time but high computational memory consumption, which is one of the main gaps compared to other studies. In summary, the model designed in this study has great application potential in improving the accuracy of stereo image quality recognition, and is particularly suitable for the Chinese visual design industry. Future research can further explore the market-oriented application of this recognition model.

INDEX TERMS Stereoscopic images, visual communication design, evaluation model, convolutional neural network.

I. INTRODUCTION

Stereoscopic images (SIs) refer to the use of special techniques to convert two-dimensional (2D) images into three-dimensional images, which can enhance the realism and impression of images, and have a significant impact on visual communication design. This is also the motivation behind this study [1], [2]. Three-dimensional (3D) images have been widely used in advertising design, product packaging, poster promotion, film and television production, and other fields.

The associate editor coordinating the review of this manuscript and approving it for publication was Venkateshkumar M.

However, in the process of production and use, image quality issues have become one of the important factors restricting its development.

Because the quality of SIs is the most crucial aspect in SI design, it also directly affects the presentation effect of SIs and the visual perception of the audience. Therefore, the evaluation method of SI quality has become a highly valuable research direction.

Currently, subjective evaluation-based methods are widely used in the study of SI quality. However, due to the significant individual differences in human beings and the reliance on subjective opinions in evaluation results, the subjective rationality of subjective evaluation results is difficult to guarantee, and it is very cumbersome and time-consuming to use. Therefore, it is very important and urgent to study a SI quality evaluation method based on objective evaluation, which is also the gap between previous research and this study [4], [5]. Although objective evaluation methods are standardized and scientific, they often overlook the characteristics and requirements of SIs in visual communication design.

Designing a more suitable method for evaluating the quality of SIs is an urgent problem to be solved in the design field. SI quality evaluation, as an emerging research field, has been rarely involved in industry related research, and only some insights from similar fields can be referenced. For example, currently in the field of advertising design, more attention is paid to the quality evaluation methods of 2D images, while in the field of 3D modeling, 3D models are evaluated by measuring geometric attributes, surface quality, and observing objects from different perspectives.

These methods have different applicability in different fields, but they cannot effectively solve the practical problem of SI quality evaluation. Based on the above considerations, this study draws on the ideas of convolutional neural networks, stereo image semantic features and local detail perception modules, with a view to using convolutional neural networks (CNN) to realize a stereo image quality evaluation system to explore the relationship between visual perception and image features, and effectively improve the image quality of stereo images. In order to meet the game, film, advertising and other fields of stereo image quality requirements. This is the main contribution of this paper. This research has important practical significance for improving the effectiveness of stereo image design and promoting the application of stereo image widely. This study has important practical significance for improving the effectiveness of SI design and widely promoting the application of SIs.

II. RELATED WORKS

Image quality evaluation has been a popular field of image computer processing in recent years, attracting a large number of scholars' attention. Relevant research by industry experts and scholars is shown in Table 1. Table 1 shows the numbering of references in this study, a summary of the main content, and an analysis of the advantages and disadvantages of the design methods for each reference.

TABLE 1. Summary of relevant research.

Reference number	Summary	Advantages and disadvantages
[6]	They designed seven image evaluation models based on commonly used algorithms and calculation rules	The evaluation algorithm of fusion filtering backprojection technology has the closest evaluation quality to manual evaluation labels
[7]	Design an image quality evaluation model based on improved CNN algorithm	The evaluation accuracy of neonatal brain images using this model is higher than that of traditional models
[8]	An improved pulse coupled neural network is proposed. In the calculation process of this network, the role of internal activity terms in convolution calculation is considered, which can better utilize the pixel information of adjacent spaces on sonar images.	This algorithm can further improve the accuracy of underwater target segmentation and evaluation, and has certain adaptability to different optimization frameworks
[9]	Design a pear image evaluation method based on an improved neural network algorithm.	It can greatly improve the taste recognition speed of pears, and the recognition accuracy is high, but the calculation speed is slow
[10]	A novel dual path deep neural network has been designed to measure the quality of sonar images	This model can quickly and accurately evaluate the quality of sonar images
[11]	Design an image quality evaluation method based on improved CNN algorithm	This method effectively improves the evaluation accuracy of traditional evaluation models, and the computational time does not significantly increase
[12]	Design an evaluation method based on high-pass filter and improved neural network algorithm	The evaluation accuracy of this evaluation model has increased by 14.9% compared to the basic algorithm before improvement

In summary, although previous researchers have designed a large number of improved automation models to improve the image quality evaluation, there are still quite few models

TABLE 1. *(Continued.)* **Summary of relevant research.**

[13]	Design an improved artificial neural network and select signal-to-noise ratio and contrast-to-noise ratio as evaluation indicators to conduct SI denoising experiments.	After processing with this improved model, the signal-to-noise ratio of the image has significantly decreased and is lower than that of the comparison model, indicating good denoising performance. But the model designed for research has the drawbacks of slow computation speed and slow training speed
[14]	A hybrid improved CNN and attention module model for breast volume ultrasound imaging and magnetic resonance imaging quality recognition was designed.	This model has higher recognition accuracy for breast volume ultrasound imaging and magnetic resonance imaging quality than traditional CNN models and common image quality recognition models, and the calculation results are more stable.

specifically designed to deal with the problem of SI distortion, and solving this problem is of great significance for fields such as autonomous driving. This is the starting point for conducting this study.

III. CNN SI QUALITY RECOGNITION MODEL FOR VC

A. CNN IMAGE QUALITY RECOGNITION MODEL BASED ON SI SEMANTIC FEATURES

Now, the first step is to design a CNN image quality evaluation model based on SI semantic features, explore the advantages and disadvantages of this model, and prepare for future improvements. Neural network algorithms have powerful nonlinear data fitting capabilities and are increasingly being applied in the data processing, and this ability cannot be improved without hidden layer neurons. The multi-layer perceptron model is the basis of most current neural network algorithms, and its typical computing structure is shown in Figure 1.

Multi-layer perceptron, CNN, backfeed neural network and their improved neural network all calculate the output $h_{w,b}(x)$ of neurons in a similar way. The same is true for the CNN image quality recognition model designed now, as shown in Formula (1).

$$h_{w,b}(x) = f\left(\sum_{i=1}^{n} W_i x_i + b\right) \tag{1}$$

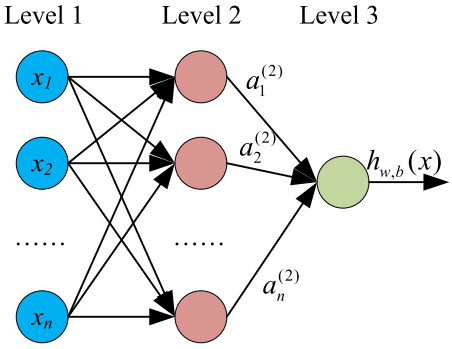

FIGURE 1. Typical computing model of multi-layer perceptron.

In Formula (1), W and b represent the weights and biases of neurons, respectively; $f(\cdot)$ means the activation function of the neuron, and x refers to the input data of the algorithm. If $z_i^{(l)}$ is the sum of the bias term of the ith neuron in the l layer and all input weighted values, then $z_i^{(l+1)}$ can be calculated according to Formula (2).

$$z_i^{(l+1)} = \sum_{j=1}^{n} W_{ij}^{(l)} x_j + b_i^{(l)} \tag{2}$$

At the same time, the output value $a_i^{(l)}$ of the activation function of the i th neuron in the l layer can be calculated according to Formula (3).

$$a_i^{(l)} = f\left(z_i^{(l)}\right) \tag{3}$$

The forward propagation of neural networks is completed through Formulas (2) and (3). Then it selects the type of activation function of the recognition model. There are two types of commonly used neural network activation function, and the sigmoid calculation method is shown in Formula (4).

$$f(z) = sigmoid(z) = \frac{1}{1 + \exp(-z)} \tag{4}$$

The other type is the tanh function, as shown in Formula (5).

$$f(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{5}$$

The function change images of the two functions are shown in Figure 2. As shown in Figure 2, when the input value of the tanh activation function is very large or very small, the output value changes very little, which may cause the neural network to converge in advance or slowly in the training, so it chooses to use sigmoid as the activation function [15], [16].

It redesigns the back propagation in the quality identification model. For a group of input samples (x, y), the loss function value $J(W, b)$ of the network can be calculated according to the forward propagation rules. Therefore, the partial derivative corresponding to the weight value and offset term can be calculated according to the residuals of the output layer. Then, the gradient descent method can be used to obtain

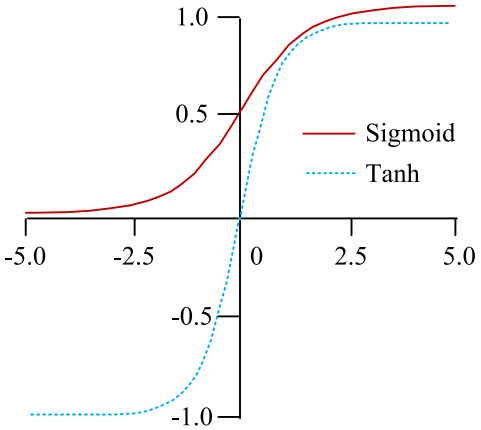

FIGURE 2. Function Images of two typical neural network activation functions.

the function for updating the network parameters, as shown in Formulas (6) and (7).

$$W^l = W^l - \alpha \frac{\partial J(W, b)}{\partial W^l} \tag{6}$$

In Formula (6), α denotes the learning rate during the algorithm training. The expression for updating paranoid items is shown in Formula (7).

$$b^l = b^l - \alpha \frac{\partial J(W, b)}{\partial b^l} \tag{7}$$

The commonly used pooling methods in CNN include maximum pooling and average pooling, and their calculation methods are shown in Figure 3. Since maximum pooling can improve the performance of algorithm feature extraction, it is used to construct the model.

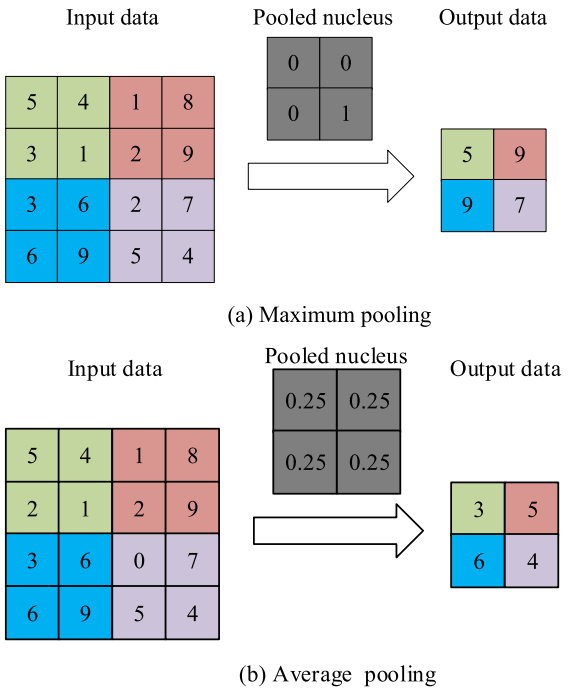

FIGURE 3. Common CNN pooling module calculation methods.

Most existing SI quality evaluation methods tend to segment the entire image into multiple small blocks without overlap for calculation. However, the evaluation results obtained by this method cannot represent the entire image and lack some accuracy. Therefore, based on the above content, a CNN algorithm with multi-channel parallel input structure is now constructed, as shown in Figure 4. As shown in Figure 4, this method can extract the overall semantic features of the entire image, which is beneficial for improving the evaluation quality of distorted images [17], [18]. This algorithm is mainly divided into two parts, which are used for monocular and binocular feature extraction fusion work [19]. When the input of the whole algorithm is a distorted SI, a dual column parallel channel is used in the network to extract semantic information from the left and right views. This module can also simulate the process of image information of human eyes imaging on the retina [20], [21], [22]. For each column of network channel, it is also necessary to use parallel networks to extract receptive field information. Only the size of convolution kernel in each column structure is different. In the binocular feature fusion module, the fully connected layer is used to achieve the fusion convolutional layer for feature extraction. In Figure 4, "MP" represents the maximum pooling module, the values marked on the convolutional module represent the size of the convolutional kernel, and "FC" represents the fully connected layer [23], [24].

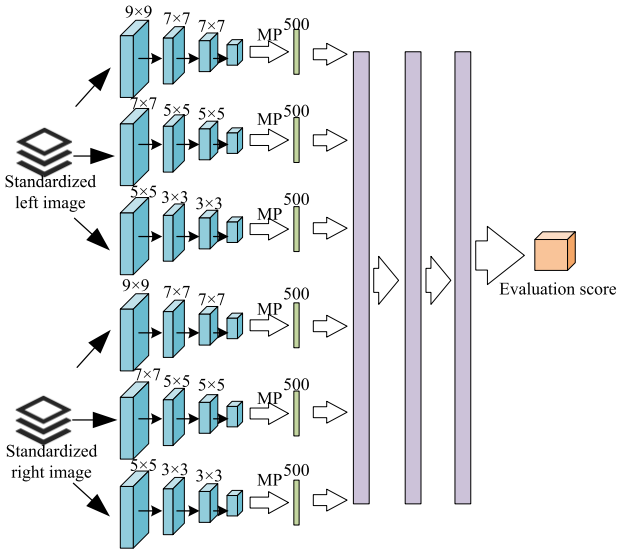

FIGURE 4. CNN with multi-channel parallel input structure.

In the study, Spearman Rank Ordered Correlation Coefficients (SROCC) and Pearson Linear Correlation Coefficients (PLCC) were selected as the evaluation indicators for the model's performance in SI quality evaluation. The calculation method for SROCC indicators is shown in Formula (8).

$$SROCC = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2 - 1)} \tag{8}$$

In Formula (8), N and d_i respectively represent the difference between the output rank of the total number of samples to be tested and the subjective evaluation rank of the ith

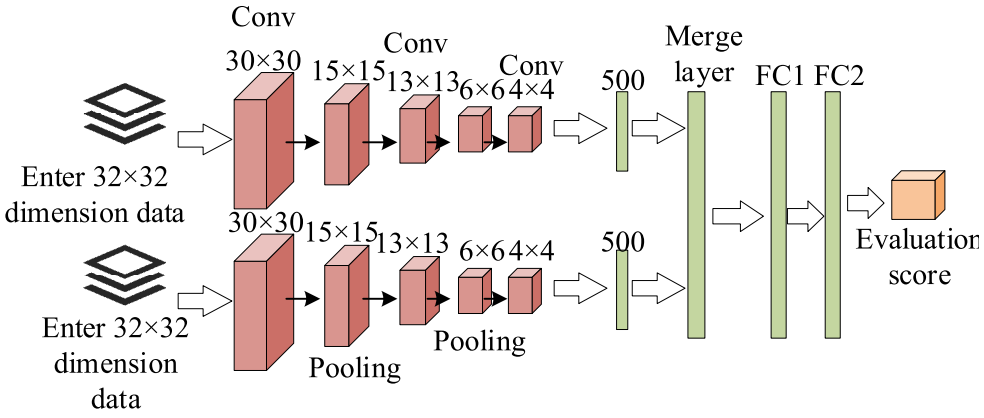

FIGURE 5. Improved CNN image quality recognition model based on hybrid local detail perception.

distorted image after objective quality evaluation ranking. The calculation method of PLCC is shown in Formula (9).

$$PLCC = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{N} (y_i - \bar{y})^2}} \tag{9}$$

In Formula (9), x_i and y_i respectively represent the objective quality evaluation value and subjective quality evaluation value of the i distorted image. Finally, ReLU function is more appropriate for the activation function in the study, and the calculation method is shown in Formula (10).

$$f(z_i^l) = \max\left(0, \sum w_i * a_i\right) \tag{10}$$

In Formula (10), $\sum w_i * a_i$ indicates the visual layer output.

B. CNN IMAGE QUALITY RECOGNITION MODEL BASED ON HYBRID LOCAL DETAIL PERCEPTION

The disadvantage of CNN image quality evaluation model based on SI semantic features is that if the training sample set is small, the model's ability to extract local distorted features will also deteriorate. To address this drawback, the designed model has been further improved. The overall structure of the improved model is shown in Figure 5. The input image size of the network in Figure 5 is all 32 × 32 size and normalized. At the same time, as the distortion level of each region of the image used for training the model in this study is generally consistent, the subjective evaluation score of the entire image can be used to represent the quality score of the image. And the model is also constructed using a symmetrical structure, which is used to extract feature information from left and right views.

Moreover, the improved model has a total of ten layers, among which the first five layers follow the classic CNN structure components: the convolutional layer is repeated with the pooling layer module, and a convolutional layer is placed at the end.

At the same time, to preserve the detailed data of the input image more clearly, the convolutional kernels are all based on 3 × 3 and a step size of 1 for parameter components. The size of the pooling layer is uniformly 2 × 2 and the step size is also 2. Then the seventh layer is designed as a fusion layer, which is used to fuse the left and right view features, and the remaining layers are all fully connected layers. To match the research purpose, only one neuron is retained in the last layer, with the aim of outputting objective quality prediction values of the image. The specific structure of each layer of this neural network is shown in Table 2. Due to table format limitations, Table 2 does not display the complete structure of the neural network. In fact, a fully connected layer consisting of 1000 neurons, 300 neurons, and one neuron is connected sequentially.

In the SI quality model, the image saliency detection step also needs to exist. This module is constructed by imitating the attention mechanism in the human visual system. The role of image saliency detection is to make the algorithm pay more attention to designated or predetermined regions with certain rules. It is of great significance for tasks such as image segmentation, object detection, and image quality evaluation. The mature visual attention mechanisms currently validated by the market are mainly divided into stimulus driven bottom-up types and task driven top-down types.

The bottom-up visual attention mechanism driven by stimuli specifically refers to the way in which image processing attention regions are guided by the characteristics of the image itself, such as color and brightness. For example, in images with a gray background, the black region can be considered a prominent region. The task driven top-down type of visual attention mechanism refers to the way in which the focus area is influenced by set rules, such as the requirement to locate the vehicle position in the image. Under this rule, the vehicle in the image is considered the area that the attention mechanism should focus on. This study adopts a bottom-up type of visual attention mechanism, and below is the design of a visual attention mechanism for SI quality evaluation.

The first step of image saliency testing based on visual attention mechanism is to analyze the pixel value differences in the input image using a locally known kernel function, with the aim of obtaining local structural information of

TABLE 2. Structural parameters of improved CNN image quality recognition model based on mixed local detail perception.

LeftViewInputData		RightViewInputData	
LayerNumber	ParametersandTypes	LayerNumber	ParametersandTypes
Conv1	323×3convolutionkernel	Conv1	503×3convolutionkernel
Pooling1	two×Maximumpoolingof2	Pooling1	two×Maximumpoolingof2
Conv2	643×3convolutionkernel	Conv2	643×3convolutionkernel
Pooling2	two×Maximumpoolingof2	Pooling2	two×Maximumpoolingof2
Conv3	643×3convolutionkernel	Conv3	643×3convolutionkernel
FC1	800neurons	FC1	800neurons

the image and the size and shape of the kernel. The locally known kernel function $K(x_l - x_i)$ is calculated according to Formula (11).

$$K(x_l - x_i) = \frac{\sqrt{\det(C_l)}}{h^2} \exp\left\{\frac{(x_l - x_i)^T C_l (x_l - x_i)}{-2h^2}\right\} \tag{11}$$

In Formula (11), l is the pixels in the image, and $l = 1, \ldots, P^2$ (P^2 is the number of pixels in the local window); C_l is the covariance matrix composed of pixel gradient vectors; $x_l = [x_1, x_2]^T$ indicates the spatial coordinate, and $x_l = [x_1, x_2]^T$ means the global smoothing parameter in significance testing. The second step is to normalize the local control kernel function in the first step and compare the similarity of feature matrices between pixels and neighboring pixels. The next step is to calculate the significance value S_i of the corresponding point based on the similarity of the feature matrix obtained in the previous step. The calculation method is shown in Formula (12).

$$S_i = \frac{1}{\sum_{j=1}^{N} \exp(\frac{-1+\rho(F_i, F_j)}{\sigma^2})} \tag{12}$$

F_i and F_j represent the feature matrices of the current pixel i and adjacent pixel j, respectively; ρ is a function that calculates the cosine similarity of a matrix. After calculating the salient values of each point in the image, a salient map of the image can be drawn for subsequent training purposes. So far, it has been possible to obtain objective quality evaluation prediction values from both local detail information and overall semantic information perception, but the final evaluation value needs to consider both aspects at the same time, and all need to combine the test results of both.

Now it chooses to combine the data obtained from the significance test with the CNN model data. Because the saliency map output from the image saliency test can reflect the degree of interest of the set rules in each region of the image, it can also be considered that the proportion of saliency regions can reflect the total proportion of information that is concerned. The higher this proportion, the more information the algorithm can obtain from it during the calculation process. The quality evaluation results calculated in this region also have a greater impact on the overall image quality evaluation. Based on this judgment, it is necessary to combine the overall and local quality evaluation results in a weighted manner, and the calculation method is shown in Formula (13).

$$Q = P_s Q_g + (1 - P_s) Q_l \tag{13}$$

In Formula (13), Q denotes the final quality evaluation value; Q_g and Q_l respectively refer to the overall quality evaluation score and local quality evaluation score of the image, and P_s represents the proportion of significant areas in the SI to the overall image, calculated according to Formula (14).

$$P_s = \frac{D}{S} \tag{14}$$

In Formula (14), D means the total area of the area in the image where the pixel saliency value is greater than the average saliency value of the overall image, and S represents the area of the overall image, calculated according to Formula (15).

$$S = \Phi(S) = \frac{1}{M} \sum \sum_{(i,j) \in S} S(i, j) \tag{15}$$

In Formula (15), $S(i, j)$ indicates the significant value of (i, j) pixel positions on the image, and M denotes the total number of pixels in the image. At this point, the CNN SI quality recognition model for VC has been designed.

IV. PERFORMANCE TESTING OF IMAGE QUALITY EVALUATION MODEL

A. EXPERIMENTAL PLAN DESIGN AND TRAINING PROCESS ANALYSIS

For the purpose of verifying the performance of the two SI quality evaluation models proposed in this study, a distortion image evaluation experiment based on the LIVE 3D Phase I dataset has been designed. The dataset was divided into a training set and a testing set in a 7:3 ratio.

There were five types of image distortion on this dataset, including WN, JPEG, JP2K, BLUR, and FF. SROCC and PLCC were selected as the criteria for evaluating the quality of the model. To compare the performance of the SI quality evaluation model designed in this study, classic Faster RCNN, GoogLeNet structures, as well as XGBoost and Support Vector Machine (SVM) algorithms in machine learning were selected to construct a comparative evaluation model. The initialization mode of neuron parameters of each neural network algorithm was completely consistent. The learning rate, iteration times, and hyperparameter of a single batch of training sizes were determined through multiple experimental runs.

First, in needs to compare the changes of the loss function values of each comparison model during the training in Figure 6. Because there were many comparison models in Figure 6, the data that were divided into deep learning model and machine learning model according to the algorithm type

of building models were stored in sub graphs (a) and (b) respectively. The horizontal axis of the two sub graphs represents the number of iterations, the vertical axis represents the value of the loss function, and different line styles represent different algorithm models.

It notes that ICNN1 and ICNN2 in Figure 6 represent the CNN image quality recognition model based on SI semantic features and the CNN image quality recognition model mixed with local detail perception designed in this study. From the analysis of Figure 6, at the beginning of training, the loss function of each algorithm model showed a trend of rapid decline. When the number of iterations reached 20, the loss function value of each model declined slowly, and then the loss function value of each model stabilized near a certain value. On the whole, the convergence speed of ICNN1 and ICNN2 models designed in this study was not much different from that of the comparison model, but the loss function after convergence was significantly lower than that of other models, which were 7.62 and 4.85 respectively.

It is the reanalysis of the changes in the SROCC and PLCC total measure of each algorithm during the training. The statistical results of the SROCC total measure data are shown in Figure 7. The same comparative model was divided into deep learning models and machine learning models based on the algorithm type used to construct the model. The data of both models were stored in subgraphs (a) and (b), respectively.

The horizontal axis of the two subgraphs represents the number of iterations, the vertical axis represents the SROCC total measure, and different line styles represent different algorithm models. Analyzing Figure 7, the overall SROCC of each algorithm still exhibited a pattern of rapid growth first, followed by a slowdown in growth rate, and fluctuating repeatedly around a certain value. And the convergence speed of the SROCC total measure numerical values of each algorithm was basically consistent with that in Figure 6.

Faster RCNN, GoogLeNet and XGBoost three models in machine learning had the fastest convergence speed, and the convergence was completed when the number of iterations reached about 20, 28 and 21 respectively. The ICNN1 and ICNN2 models designed in this study had a slow convergence speed, and the training was completed in the 52nd and 46th iterations respectively, which was related to the network hierarchy and internal calculation steps of ICNN1 and ICNN2 models. From the training effectiveness, when the number of iterations was 100, all models had completed training. At this time, the total SROCC measurement data of ICNN1, ICNN2, Faster RCNN, GoogLeNet, XGBoost, and SVM models were 0.925, 0.938, 0.910, 0.873, 0.907, and 0.826, respectively. By analyzing Figure 7 (b), the recognition and evaluation quality of the two more traditional machine learning algorithm models was worse than that of all neural network algorithms.

The statistical results of the PLCC total measure data for each algorithm during the training process are shown in Figure 8. The horizontal axis, subgraph, and line styles in Figure 8 were consistent with those in Figure 7, but the

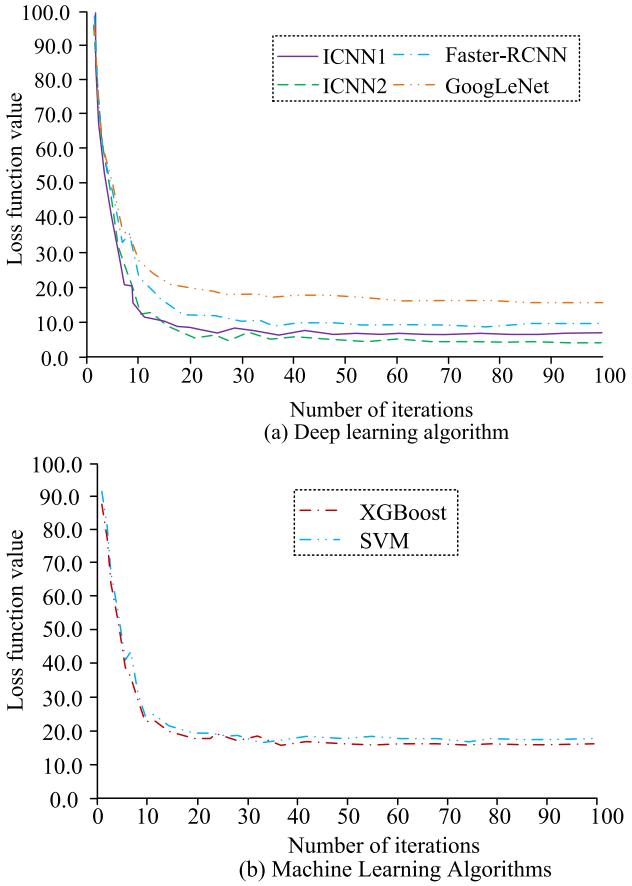

FIGURE 6. Changes in loss function values of various neural network algorithms during training.

vertical axis had different meanings, representing the PLCC total measure values of each algorithm during the training. By analyzing Figure 8, the changes in the total PLCC measure of each model during the training were generally consistent with the changes in the total SROCC measure.

These graphs exhibited a pattern of rapid growth followed by a slowdown in growth rate, and fluctuating repeatedly around a certain value. However, the difference was that the PLCC total measure data of each model after training was slightly higher than the SROCC total measure data under the same conditions. For example, when the number of iterations reached 100, the PLCC total measure data of ICNN1 and ICNN2 models were 0.936 and 0.957, respectively, which were 0.011 and 0.019 higher than the SROCC total measure. Similarly, by analyzing Figure 8 (b), the PLCC recognition evaluation quality of the two more traditional machine learning algorithm models was still worse than that of all neural network algorithms.

B. ANALYSIS OF TEST SET CALCULATION RESULTS FOR THE MODEL

After the training was completed, the SROCC measurement values of each algorithm on the test set would be analyzed again. The statistical results are shown in Table 3. The total SROCC measurement values of the ICNN1 and ICNN2 SI

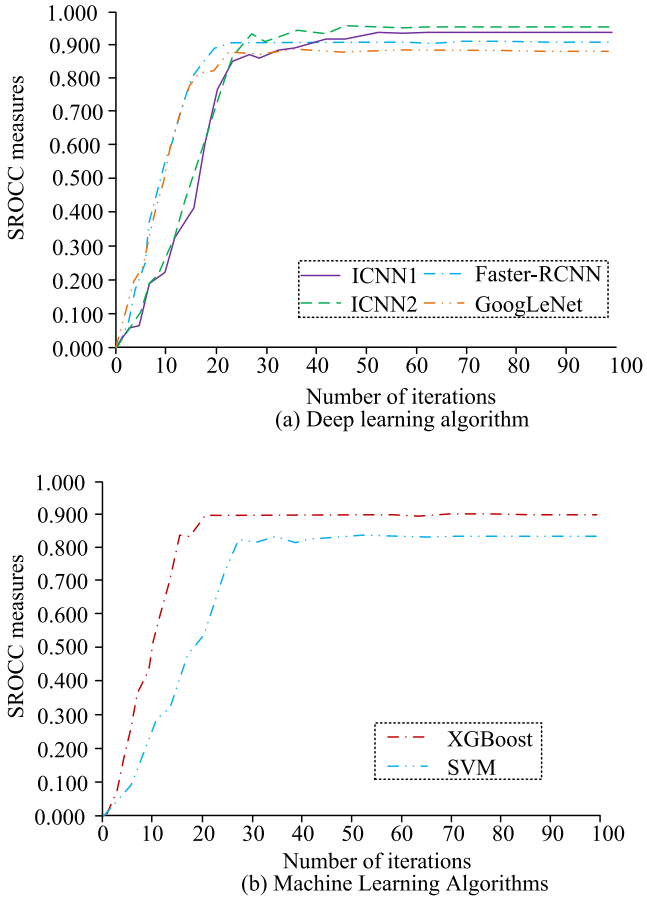

FIGURE 7. Change curve of SROCC total measure for each algorithm during training.

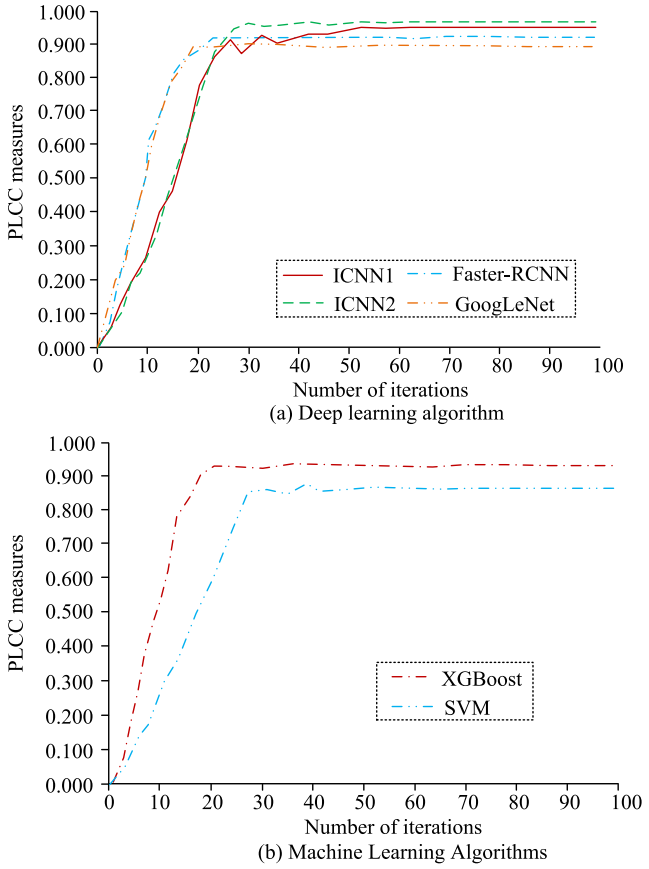

FIGURE 8. PLCC total measure change curve of each algorithm during training.

quality evaluation methods designed in this study were significantly higher than those of other comparative models, with values of 0.940 and 0.949, respectively.

The ICNN2 model with a mixture of local and global image distortion abstract information was the highest. Secondly, the total SROCC measure values of the deep learning model and the SROCC measure values on various distortion problems were generally higher than those of the machine learning algorithm. Only the XGBoost algorithm had slightly higher values on the total SROCC measure values, WN, JP2K, and FF measures than the GoogLeNet algorithm. The total SROCC measure value of the SVM algorithm and the SROCC measure value of various distortion problems were the smallest, and the evaluation ability of stereo distorted images was the worst.

The statistical results of the overall SROCC total measure and PLCC total measure on the test set after each algorithm training are shown in Figure 9. The horizontal axis of the two subgraphs in Figure 9 represents different numbers of test set samples, while the vertical axes of subgraphs (a) and (b) represent the SROCC total measure and PLCC total measure, respectively.

Different line styles represent different evaluation models. Observing Figure 9, when the number of samples participating in the test was small, the total measurement values of SROCC and PLCC of each evaluation model fluctuated significantly. However, as the number of samples to be tested increased, the total measurement values of SROCC and PLCC of each model tended to stabilize. After the number of test samples exceeded 300, the test indicators of each model had completed convergence. At this time, the total measurement values of SROCC and PLCC for the GoogLeNet evaluation model were 0.92, 0.94, 0.89, 0.83, and 0.93, 0.95, 0.90, and 0.84, respectively. The overall values of each model in subgraph (a) in Figure 9 were slightly better than the corresponding data in subgraph (b) under the same conditions.

After the training was completed, the PLCC measurement values of each algorithm on the test set were analyzed again. The statistical results are shown in Table 4. Table 4 showed that the total PLCC measurement values of the ICNN1 and ICNN2 SI quality evaluation methods designed in this study were still higher than the other models, with values of 0.951 and 0.960, respectively. At the same time, the ICNN2 model with a mixture of local and global image distortion abstract information had the highest indicator value. At the same time, the total PLCC measure value of SVM algorithm and the PLCC measure value of various distortion problems were still the smallest, and the evaluation ability of stereo distorted images was the worst.

TABLE 3. Statistics of SROCC measures for each algorithm on the test set.

Model Name	ALL	WN	JPEG	JP2K	BLUR	FF
ICNN1	0.940	0.925	0.663	0.913	0.875	0.806
ICNN2	0.949	0.933	0.724	0.935	0.869	0.794
Faster-RCNN	0.926	0.904	0.618	0.887	0.846	0.783
GoogLeNet	0.891	0.843	0.528	0.840	0.816	0.742
XGBoost	0.902	0.855	0.514	0.842	0.811	0.763
SVM	0.831	0.782	0.503	0.769	0.774	0.724

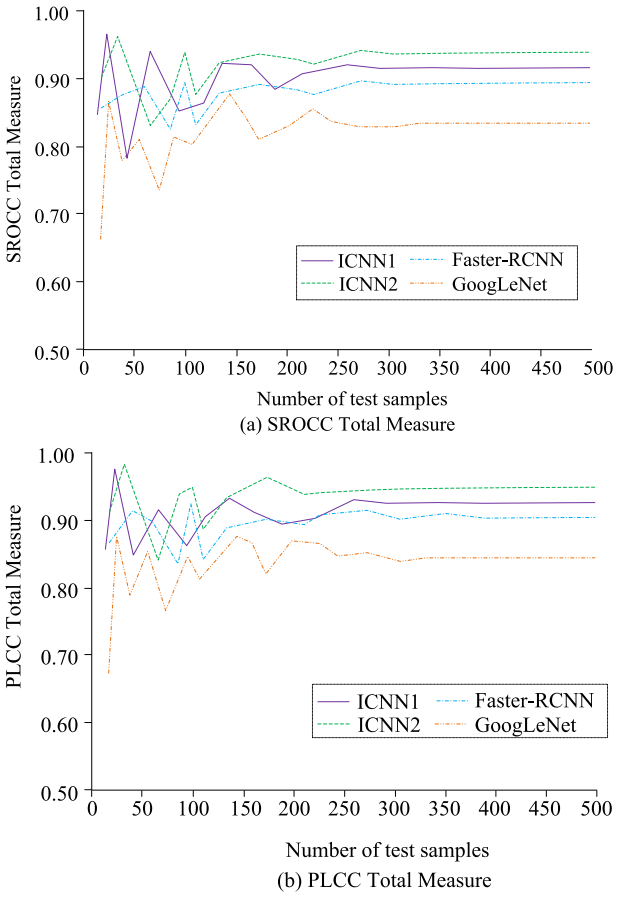

FIGURE 9. SROCC total measure and PLCC total measure of each algorithm on the test set.

Then, it needs to further compare the values and their distributions of each algorithm on the two total measures. The statistical results are shown in Table 4. It noted that the statistical results were obtained under the condition of taking the maximum of 500 test samples. To improve analysis accuracy, each experimental plan was conducted 10 times to calculate the average, standard deviation, minimum, and maximum values of the indicators. According to Table 4, the performance of each model on the mean of SROCC total measure and PLCC total measure was basically consistent with their performance on training results.

Then, the values and their distributions of each algorithm on the SROCC total measure were further compared. The statistical results are shown in Table 5. It noted that the statistical results were obtained under the condition of taking the maximum of 500 test samples. To improve analysis accuracy, each experimental plan was conducted 10 times to calculate

TABLE 4. Statistics of PLCC measures for each algorithm on the test set.

Model Name	ALL	WN	JPEG	JP2K	BLUR	FF
ICNN1	0.951	0.928	0.667	0.916	0.881	0.811
ICNN2	0.960	0.939	0.731	0.939	0.876	0.804
Faster-RCNN	0.932	0.908	0.623	0.895	0.852	0.789
GoogLeNet	0.899	0.848	0.537	0.848	0.825	0.749
XGBoost	0.912	0.862	0.527	0.847	0.817	0.768
SVM	0.843	0.787	0.508	0.774	0.779	0.731

the average, standard deviation, minimum, and maximum values of the indicators. According to Table 5, the mean values of ICNN1 and ICNN2 on this SROCC were higher than those of other deep learning and machine learning models. From the perspective of distribution patterns, the standard deviation of the ICNN1 and ICNN2 models designed in this study was significantly lower than the other comparative models in multiple repeated experiments. Specifically, the total standard deviations of SROCC measures for ICNN1, ICNN2, Faster RCNN, GoogLeNet, XGBoost, and SVM models were 0.017, 0.014, 0.025, 0.028, 0.031, and 0.027, respectively. Moreover, even from the perspective of the maximum and minimum values of the indicators, the two models designed in this study still had better stability than other algorithms.

Next, it analyzed the computational efficiency and memory consumption of each model. Statistical evaluation observation showed that the horizontal axis in Figure 10 was consistent with Figure 9, and the vertical axis represents the computational time in milliseconds. Icons of different colors and styles represent different evaluation models, while the corresponding colored lines represent the fitting lines of the corresponding model's data points.

The fitting lines were fitted using a cubic polynomial equation style. Observing Figure 10, the time consumption of the ICNN1 and ICNN2 neural network SI quality recognition models designed in this study was generally consistent, but the calculation time of the ICNN1 model was significantly lower than that of ICNN2 under the same conditions, which was related to the latter integrating more network structures.

The Faster RCNN evaluation model also had a linear curve between the number of samples to be tested and the calculation time. However, the difference between the former and ICNN1 and ICNN2 models was that the former took less time when calculating fewer samples, but the calculation time started to be higher when calculating more samples than the latter two. The number of samples to be tested - calculation time consuming curve of GoogLeNet model showed exponential growth, which was related to the highest calculation complexity and more internal calculation levels of the model.

Finally, the calculation memory consumption of each model was analyzed. The vertical axis in Figure 11 represents the calculation memory consumption, in MB. Icons of different colors and styles still represent different evaluation models. Observing Figure 11, the changes in computational memory consumption of the ICNN1 and ICNN2 neural networks designed in this study, as well as the comparison model Faster RCNN for SI quality recognition, were

TABLE 5. The numerical values and distribution statistics of each algorithm on the SROCC total measure.

Number	Model Name	SROCC total measure	
		Mean ± standard deviation	maximum-minimum values
#01	ICNN1	0.940±0.017	0.920-0.959
#02	ICNN2	0.949±0.014	0.931-0.968
#03	Faster-RCNN	0.926±0.025	0.893-0.967
#04	GoogLeNet	0.891±0.028	0.861-0.925
#05	XGBoost	0.902±0.031	0.854-0.938
#06	SVM	0.831±0.027	0.802-0.864

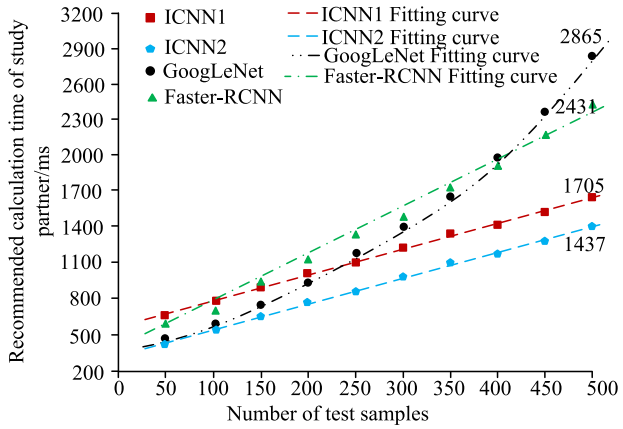

FIGURE 10. Comparison of calculation time consumption of each algorithm on the test set.

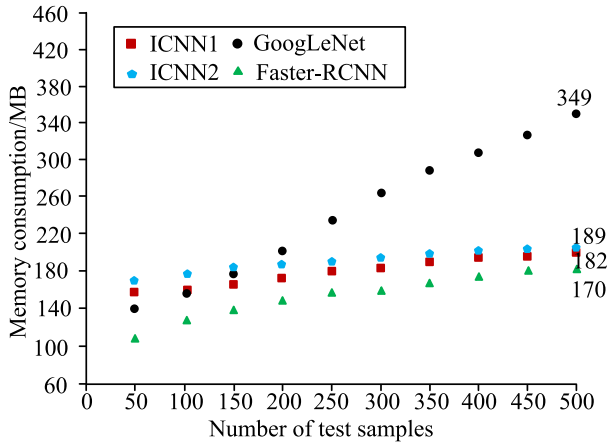

FIGURE 11. Comparison of computational memory consumption of various algorithms on the test set.

still generally consistent. Both of these models had a slow increase in memory consumption as the number of samples to be calculated increased.

The growth rate of the GoogLeNet model's sample size calculation memory consumption curve was significantly faster, indicating that the internal structure of the model was relatively not suitable for calculating a large number of samples and was related to the high level of internal calculation. When the number of contemporary calculation samples was 500, the total computational memory consumption of ICNN1, ICNN2, Faster RCNN, and GoogLeNet evaluation models were 182MB, 189MB, 170MB, and 349MB, respectively.

To further compare the evaluation ability of the model designed this time, a subjective evaluation experiment was designed, in which 40 domestic and foreign SI evaluation experts were selected to manually evaluate the 50 randomly selected images in this dataset. The higher the score, the higher the quality of the images. The designed model and various comparison models were used for the same ten point scale. The scoring results showed that the average absolute difference between the evaluation results of the designed model and the manual evaluation score was 0.36 points, which was significantly lower than the difference in evaluation scores between other comparison models and manual models.

V. CONCLUSION

This study was divided into four parts. The first part introduced the basic concepts and main technical routes of SI quality evaluation and recognition. The second part provided a detailed introduction to the improved ResNet image recognition model for hybrid SI semantic features, local detail perception module, and visual attention mechanism. The third part conducted experimental verification on the model and compared it with other models. Finally, the fourth part summarized the entire article and provided prospects for future research directions.

The test results showed that the convergence speed of the model designed in this study was not significantly different from the comparison model, but the loss function after convergence was significantly lower than other models. The SROCC total measure values of the ICNN1 and ICNN2 SI quality evaluation methods were significantly higher than other comparison models, with values of 0.940 and 0.949, respectively. The ICNN2 model, which mixed local and global image distortion abstract information, had the highest total measure value, and the PLCC total measure values of both models were still higher than the other models, with values of 0.940 and 0.949, respectively.

When the number of samples to be tested was 500, the total SROCC and PLCC measures of ICNN1, ICNN2, Faster RCNN, and GoogLeNet evaluation models were 0.92, 0.94, 0.89, 0.83, and 0.93, 0.95, 0.90, and 0.84, respectively. At this time, their total computational memory consumption was 182MB, 189MB, 170MB, and 349MB, respectively. The experimental data showed that the SI quality evaluation method designed in this study had high evaluation accuracy and stability, and had certain application value. However, the drawback was that it consumed a lot of computational memory. Due to limitations in research conditions, this study was unable to collect more real samples for broader validation experiments, nor did it explore whether deep learning

structures that integrate prior knowledge can further enhance model capabilities. Improving the model based on prior knowledge of previous optimization and design similar to neural networks is another direction for future research.

REFERENCES

[1] Y. Qi, Z. Yang, and L. Kang, "Multi-exposure X-ray image fusion quality evaluation based on CSF and gradient amplitude similarity," *J. X-Ray Sci. Technol.*, vol. 29, no. 4, pp. 697–709, Jul. 2021.

[2] A. Steuwe, C. Rademacher, B. Valentin, M. Koehler, E. Appel, V. Keitel, J. Timm, G. Antoch, and J. Aissa, "Dose optimized chest CT for diagnosis of COVID-19—Evaluation of the image quality and diagnostic impact," *J. Radiol. Protection*, vol. 40, no. 3, pp. 877–891, 2020.

[3] R. A. Helal, R. Jacob, M. A. Elshinnawy, A. I. Othman, I. M. Al-Dhamari, D. W. Paulus, and T. T. Abdelaziz, "Cone-beam CT versus multidetector CT in postoperative cochlear implant imaging: Evaluation of image quality and radiation dose," *Amer. J. Neuroradiol.*, vol. 42, no. 2, pp. 362–367, Feb. 2021.

[4] I. P. Gurov and A. Y. Pimenov, "Evaluation of the influence of scattered radiation on image quality in spectral optical coherence tomography systems with electronic scanning of objects," *J. Opt. Technol.*, vol. 87, no. 7, p. 401, 2020.

[5] Q. Jiang, Z. Peng, G. Yue, H. Li, and F. Shao, "No-reference image contrast evaluation by generating bidirectional pseudoreferences," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6062–6072, Sep. 2021.

[6] K. Jensen, G. Hagemo, A. Tingberg, C. Steinfeldt-Reisse, G. K. Mynarek, R. J. Rivero, E. Fosse, and A. C. Martinsen, "Evaluation of image quality for 7 iterative reconstruction algorithms in chest computed tomography imaging: A phantom study," *J. Comput. Assist. Tomogr.*, vol. 44, no. 5, pp. 673–680, 2020.

[7] T. Vanderhasselt, M. Naeyaert, N. Watté, G.-J. Allemeersch, S. Raeymaeckers, J. Dudink, J. de Mey, and H. Raeymaekers, "Synthetic MRI of preterm infants at term-equivalent age: Evaluation of diagnostic image quality and automated brain volume segmentation," *Amer. J. Neuroradiol.*, vol. 41, no. 5, pp. 882–888, May 2020.

[8] X. Wang, Q. Li, Y. Yu, and Y. Xu, "Evaluation criterion of underwater object clustering segmentation with pulse-coupled neural network," *IET Image Process.*, vol. 14, no. 16, pp. 4076–4085, Dec. 2020.

[9] Y.-S. Xue, S.-Z. Xu, C. Xue, R.-Z. Wang, M.-Y. Zhang, J.-M. Li, S.-L. Zhang, and J. Wu, "Pearprocess: A new phenotypic tool for stone cell trait evaluation in pear fruit," *J. Integrative Agricult.*, vol. 19, no. 6, pp. 1625–1634, Jun. 2020.

[10] H. Zhang, S. Li, D. Li, Z. Wang, Q. Zhou, and Q. You, "Sonar image quality evaluation using deep neural network," *IET Image Process.*, vol. 16, no. 4, pp. 992–999, Mar. 2022.

[11] Y. Yokota, Y. Fushimi, T. Okada, K. Fujimoto, S. Oshima, S. Nakajima, T. Fujii, M. Tanji, N. Inagaki, S. Miyamoto, and K. Togashi, "Evaluation of image quality of pituitary dynamic contrast-enhanced MRI using time-resolved angiography with interleaved stochastic trajectories (TWIST) and iterative reconstruction TWIST (IT-TWIST)," *J. Magn. Reson. Imag.*, vol. 51, no. 5, pp. 1497–1506, May 2020.

[12] C. Chen, L. Wang, Z. Zhang, C. Lu, H. Chen, and J. Chen, "Construction and application of quality evaluation index system for remote-sensing image fusion," *J. Appl. Remote Sens.*, vol. 16, no. 1, pp. 12006.1–12006.15, Aug. 2021.

[13] J. C. Park, K. J. Park, M. Y. Park, M. Kim, and J. K. Kim, "Fast T2-weighted imaging with deep learning-based reconstruction: Evaluation of image quality and diagnostic performance in patients undergoing radical prostatectomy," *J. Magn. Reson. Imag.*, vol. 55, no. 6, pp. 1735–1744, Jun. 2022.

[14] A. V. Nikolaev, L. de Jong, G. Weijers, V. Groenhuis, R. M. Mann, F. J. Siepel, B. M. Maris, S. Stramigioli, H. H. G. Hansen, and C. L. de Korte, "Quantitative evaluation of an automated cone-based breast ultrasound scanner for MRI–3D US image fusion," *IEEE Trans. Med. Imag.*, vol. 40, no. 4, pp. 1229–1239, Apr. 2021.

[15] J.-P. Suuronen, B. Hesse, M. Langer, M. Bohner, and J. Villanova, "Evaluation of imaging setups for quantitative phase contrast nanoCT of mineralized biomaterials," *J. Synchrotron Radiat.*, vol. 29, no. 3, pp. 843–852, May 2022.

[16] W. Wu, J. Qu, J. Cai, and R. Yang, "Multiresolution residual deep neural network for improving pelvic CBCT image quality," *Med. Phys.*, vol. 49, no. 3, pp. 1522–1534, Mar. 2022.

[17] S. Li, "RETRACTED: Deep adversarial model for musculoskeletal quality evaluation," *Inf. Process. Manage.*, vol. 57, no. 1, Jan. 2020, Art. no. 102146.

[18] L. Simani, Y. Oron, O. Handzel, R. A. Eta, A. Warshavsky, G. Horowitz, N. Muhanna, and O. J. Ungar, "Evaluation of the quality of online information on sudden sensorineural hearing loss," *Otol. Neurotol.*, vol. 43, no. 2, pp. 159–164, 2022.

[19] M. T. Löffler, N. Sollmann, S. Mönch, B. Friedrich, C. Zimmer, T. Baum, C. Maegerlein, and J. S. Kirschke, "Improved reliability of automated ASPECTS evaluation using iterative model reconstruction from head CT scans," *J. Neuroimag.*, vol. 31, no. 2, pp. 341–347, Jan. 2021.

[20] K. Yildiz and Z. Yildiz, "Evaluation of nano-filler dispersion quality in polymeric films with binary feature characteristics and fractal analysis," *IET Image Process.*, vol. 186, no. 10, pp. 110173.1–110173.9, 2021.

[21] J. Zhang, Y. Zhao, F. Shone, Z. Li, A. F. Frangi, S. Q. Xie, and Z.-Q. Zhang, "Physics-informed deep learning for musculoskeletal modeling: Predicting muscle forces and joint kinematics from surface EMG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 484–493, 2023, doi: 10.1109/TNSRE.2022.3226860.

[22] J. Zhang, Y. Li, W. Xiao, and Z. Zhang, "Non-iterative and fast deep learning: Multilayer extreme learning machines," *J. Franklin Inst.*, vol. 357, no. 13, pp. 8925–8955, Sep. 2020, doi: 10.1016/j.jfranklin.2020.04.033.

[23] L. Vigoroso, F. Caffaro, and E. Cavallo, "Occupational safety and visual communication: User-centred design of safety training material for migrant farmworkers in Italy," *Saf. Sci.*, vol. 121, pp. 562–572, Jan. 2020, doi: 10.1016/j.ssci.2018.10.029.

[24] A. J. Lazard, "Design cues for tobacco communication: Heuristic interpretations and usability of online health information about harmful chemicals," *Int. J. Med. Informat.*, vol. 141, Sep. 2020, Art. no. 104177, doi: 10.1016/j.ijmedinf.2020.104177.

XU HAN was born in Heilongjiang, China, in 1987. She received the bachelor's degree in visual communication from Yanshan University, in 2010, and the Master of Engineering degree in software from Wuhan University, in 2012.

Since 2015, she has been a Lecturer with the Department of Environmental Arts, Hebei University of Environmental Engineering. She participated in multiple research projects and appearance patents, with research directions, including color and visual aspects.

QIUYUE SHAN was born in Hebei, China, in 1982. She received the Bachelor of Arts and Crafts degree from Tangshan Normal University, in 2005, and the Master of Engineering degree in software from Wuhan University, in 2009.

Since 2011, she has been a Lecturer with the Department of Environmental Arts, Hebei University of Environmental Engineering. She participated in multiple research projects and thesis writing, with research interests, including art and design.

TIANSHU CHU was born in Heilongjiang, China, in 2002. She is currently pursuing the degree in intelligence science and technology with Xidian University.

Since 2020, she has been studying intelligence science and technology with the School of Artificial Intelligence, Xidian University, with the research directions, including computer vision and image understanding.