

Video Quality Assessment based on Deep Neural Network

Zhiming Shi¹*

¹ School of Software&Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330032 China

[e-mail: szmi_2007@126.com]

*Corresponding author: Zhiming Shi

*Received March 15, 2023; revised July 12, 2023; accepted August 5, 2023;
published August 31, 2023*

Abstract

This paper proposes two video quality assessment methods based on deep neural network. (i)The first method uses the IQF-CNN (convolution neural network based on image quality features) to build image quality assessment method. The LIVE image database is used to test this method, the experiment show that it is effective. Therefore, this method is extended to the video quality assessment. At first every image frame of video is predicted, next the relationship between different image frames are analyzed by the hysteresis function and different window function to improve the accuracy of video quality assessment. (ii)The second method proposes a video quality assessment method based on convolution neural network (CNN) and gated circular unit network (GRU). First, the spatial features of video frames are extracted using CNN network, next the temporal features of the video frame using GRU network. Finally the extracted temporal and spatial features are analyzed by full connection layer of CNN network to obtain the video quality assessment score. All the above proposed methods are verified on the video databases, and compared with other methods.

Keywords: Video quality assessment, IQF-CNN, Hysteresis function, Window function, GRU, Accuracy

1. Introduction

With the development of 5th generation communication technology and the continuous growth of internet users, as well as the diversified forms of network access and the wide application of video services, the network video has become the king of internet traffic. The pattern of global new media is constantly changing, especially in the emerging market represented by China. Because there are many factors may impair the network video quality, how to deal with the relationship between video features and video quality and build the corresponding assessment method is a common concern in academic circles in recent years.

1.1 Related work

Now a lot of institutions and research organizations have done research on video quality assessment. The video quality assessment methods are divided into two categories. The first category is subjective assessment method. It can reflect the user experience, but it requires a lot of human resource and time. It can't be widely used in the current video service system. The second category is objective assessment method. It uses related feature information of video to build mathematical model to assessment video quality. Compared with the subjective method, the objective method has better real-time performance. It doesn't need human participation, so people focus on this kind of method.

In additional, the objective method can be modeled by statistics, visual psychology and artificial intelligence. The assessment model based on statistics usually analyzes the correlation between the characteristic features of video and video quality and uses the statistical regression analysis method. Balasubramanyam Appina used the binary generalized Gaussian distribution model to count the motion vector and depth features of 3D video, and proved that the co-variance matrix of the model is proportional to the video quality [1]. Rajiv Soundararajan designed the assessment method by combining statistical model and perception principle. Firstly, the wavelet coefficients and the inter frame difference coefficients of Gaussian scale hybrid model were used to calculate the difference values of spatial and temporal information of reference video and distorted video respectively. Next the difference value of time-space information was combined to obtain the reference entropy difference. The experiment shows that the difference was related to video quality [2]. Deepti Ghadiyaram firstly described the time information, pause information, video content characteristics and perceptible quality information of distorted video, next used Hammerstein Wiener model to build QoE predictors [3]. Wenjuan Shi discussed the application features of video frame, pause information and memory effect of video quality, next used Hammerstein Wiener model to build evaluation model [4]. The above methods mainly use the relationship between video features and video quality to build an assessment model. Due to the specific samples and scenes, the mathematical model is relatively fixed, the scope of application is limited, and the accuracy is not as good as artificial intelligence methods.

The assessment model based on visual psychology usually describes the relationship between the degree of visual stimulation and people's psychological feelings. In many scenes, this relationship usually presents logarithmic characteristics. Fan Zhang proposed a hybrid assessment model based on visual perception features, which comprehensively considered the obvious distortion and fuzzy artifacts of nonlinear models [5]. Mehdi Banitalebi-Dehkordi proposed an assessment model based on visual memory. The model used complete local binary pattern to process saliency map, next established a visual memory model according to statistical information of saliency map. Finally, the visual memory, saliency and frame features were trained by support vector regression machine to

obtain video quality [6]. Chathura Galkandage designed a novel visual perception assessment model based on the motion sensitive response of complex cells in the visual cortex, used to simulate the characteristics of simple and complex cell behavior, and finally determined the weight of each feature through the customized double order multiple stepwise regression algorithm [7]. These models based on visual psychology don't need complex training process and calculation process, but it mainly studies the function model of video quality and some visual features, and can't solve the problem of multiple video features synthesis, which also brings some limitations to the application of this kind of model.

Now the focus is the assessment model based on artificial intelligence, which uses video features for machine learning to build assessment method. Kongfeng Zhu used discrete cosine transform to calculate six frame level features of video image, and converted all frame level features into corresponding video features through time pool, next used multi-layer neural network to train video features and assess video quality [8]. Xiaoming Tao collected a large number of network video data, such as subjective score and network features, and selected features that are highly correlated with subjective assessment, and then used deep neural network to train network features to obtain objective score [9]. Yu Zhang proposed a no reference assessment model based on weakly supervised convolution neural network and resampling strategy. In this method, an eight-layer convolution neural network was constructed to process video features, and then the mapping relationship between the frequency histogram obtained from the training network and the video quality was established. At the same time, resampling strategy was used to improve the mapping relationship [10]. Ali Al-Naji used three fuzzy reasoning systems to build the assessment model, which had nine characteristic indexes, such as peak signal-to-noise ratio, visual signal-to-noise ratio, weighted signal-to-noise ratio, SSIM, multi-layer SSIM and general image quality index, and achieved good results [11]. Mohammed Alreshoodi proposed a cross-layer assessment model to predict the quality of 3D video. The model selected feature indicators from coding and network level, and systematically analyzed the correlation between them and video quality. Then the selected feature index used fuzzy inference system to build the assessment method. The experimental results showed that the system has good effect [12]. Zhiming Shi analyzed the impact of initialization buffer time, number of stuck times, average stuck time, noise ratio, ambiguity, block effect and other characteristics of video quality, next used fuzzy inference system to build assessment model [13]. Domonkos Varga proposed a no-reference video quality assessment method based on CNN combined with long short term memory (LSTM) [14]. In this method, the video sequence is regarded as depth time series. With the help of the features extracted by CNN, the LSTM network can be trained to predict the video quality. Because the process of machine learning is complex and needs a lot of calculation, the efficiency of the assessment model needs to be improved. Jari Korhonen proposed a new approach for learning-based video quality assessment, based on the idea of computing features in two levels so that low complexity features were computed for the full sequence first, and then high complexity features were extracted from a subset of representative video frames, selected by using the low complexity features [15]. Zhengzhong Tu conducted a comprehensive evaluation of leading blind features and models on a fixed evaluation architecture, yielding new empirical insights on both subjective video quality studies and objective model design [16]. Moreover, the accuracy of different training samples is still biased, which needs to be improved and bring great challenges to the assessment model based on artificial intelligence.

But many factors may impair the video quality, the analysis process is complex. As shown in **Table 1**, the existing assessment methods mainly have the following problems. 1. "Weak

comprehensiveness”. Many methods don’t comprehensively consider the factors that impair the video quality. Some focus on the pixel of video frame, some focus on the network performance. They are not comprehensive enough. 2. “Weak robustness”. The relatively fixed mathematical model is constructed through limited experimental scenes and data samples, which easily leads to the weak robustness of the assessment method. 3. “Low accuracy”. Many methods can’t adjust the assessment model according to the actual test situation, resulting in low accuracy. The accuracy of method should be improved.

Table 1. Existing problems and advantage of proposed method

Method	Problem
Existing method	Weak comprehensiveness Weak robustness Low accuracy
Method	Advantage
Proposed method	Good comprehensiveness and robustness Improved accuracy

1.2 The content of paper

In this paper, two video quality assessment methods based on deep neural network are proposed. By comparing the two methods, the application of deep neural network [17] in video quality assessment is studied. (i)The first method uses the IQF-CNN to build image quality assessment method. The LIVE image database is used to test this method, the experiment show that it is effective. Therefore, this method is extended to the video quality assessment. At first every image frame of video is predicted, next the relationship between different frames is analyzed by the hysteresis function and different window function to improve the accuracy of video quality assessment. (ii) The second method proposes a video quality assessment method based on convolution neural network (CNN) and gated circular unit network (GRU). First, the spatial features of video frames are extracted using CNN network, next the temporal features of the video frame using GRU network. Finally, the extracted temporal and spatial features are analyzed by using full connection layer of CNN network to obtain the video quality assessment score. All the above proposed methods are verified on the public video libraries.

The contributions of this paper are as follows. 1.The first objective method uses IQF-CNN to assess video quality. It predicts every image frame of video, and build the relationship between different frames to improve the accuracy of assessment. 2.The second method combines CNN and GRU to assess video quality. It comprehensively considers the spatial and temporal features of video to improve the accuracy of assessment. 3.At last through the analysis of this paper, we can see the research of CNN in the video quality assessment.

The methods proposed in this paper have good applicability, and every step of them has experimental basis. They can be used for actual video quality assessment.

The paper is organized as follows. Firstly, the research status of video quality assessment is introduced. Secondly the two proposed methods are presented. The steps of them and experimental results are given. At last, the paper is concluded, the advantages and disadvantages of them is analyzed.

2. The proposed method

2.1 The first method based on IQF-CNN

As shown in Fig. 1, the process of first method is divided into two steps. Firstly, the input image is pre-processed, next the IQF-CNN is used to train the image and assess them. Secondly the relationships between image frames are analyzed by hysteresis function and different window function to improve the accuracy of this model. At last, the video quality is given, the detail steps of this method will be given as follows.

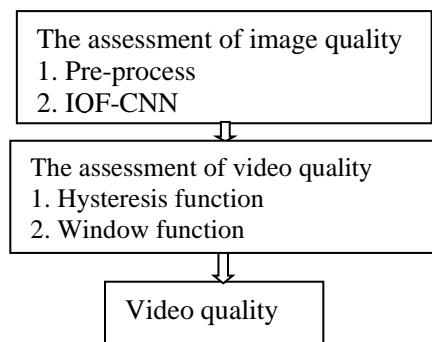


Fig. 1. The process of first method

(1) Step1: The process step of image quality assessment

The process step of image quality assessment is shown in Fig. 2. At first the input image should be normalized, next the processed image is decomposed into many image blocks and trained by IQF-CNN to assess the image quality.

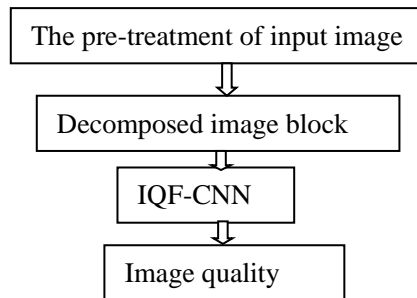


Fig. 2. The process of image quality assessment

$$\tilde{x}(i, j) = \frac{x(i, j) - \mu(i, j)}{\sigma(i, j) + C} \quad (1)$$

$$\mu(i, j) = \sum_{p=-p}^p \sum_{q=-q}^q \omega(p, q) x(i+p, j+q) \quad (2)$$

$$\sigma(i, j) = [\sum_{p=-p}^p \sum_{q=-q}^q \omega(p, q) (x(i+p, j+q) - \mu(i, j))^2]^{1/2} \quad (3)$$

Set $x(i, j)$ is the pixel of image, $\tilde{x}(i, j)$ is the normalized output pixel of image, $\mu(i, j)$ is the average value of image pixel, $\sigma(i, j)$ is the standard deviation of image pixel. ω is the

2-dimensional symmetric Gaussian weight matrix, C is constant. P and Q are the width and height of the image block respectively. In this way, the entire input image will be pre-processed.

Part 1: pseudo code

Begin

Procedure Pre-treatment()

Function Average <input image> ;

Function Standard deviation<input image>;

Function Normalized output<input image>;

End.

Next the input image is divided into non overlapping 32×32 image block. All the image blocks are inputted into the IQF-CNN to give the scores of them. The structure of IQF-CNN is shown in Fig. 3. The first convolution layer has 8 convolution kernels with the size of 3×3 , the convolution step is set 1. So 8 characteristic figures with the size of 30×30 will be obtained. Next the characteristic figures are pooled to maximize. The second convolution layer has 32 convolution kernels with the size of 3×3 . In this way 32 characteristic figures will be obtained. Then two pooling operations are processed, including maximum pooling and minimum pooling to obtain 64 outputs. These outputs pass through two full connection layers. At last, the scores of image block will be calculated.

$$y_k = C_{k1}\alpha_1 + C_{k2}\alpha_2 + \dots + C_{km}\alpha_m + e(n_k) \quad (4)$$

Set C_{km} is the score of image block, y_k is the score of image, α_m is the weight of different image block, $e(n_k)$ is error. m is the number of image block; k is the number of images. The linear least square method is used to calculate α_m .

$$R(\hat{\alpha}) = \sum_{k=1}^n [y_k - \sum_{j=1}^m c_{kj} \hat{\alpha}_j]^2 = \min \quad \frac{\partial R(\hat{\alpha})}{\partial \alpha_m} = 0 \quad (5)$$

So, the score of images will be calculated.

Part 2: pseudo code

Begin

Procedure Image treatment()

Function Divide block <input image>;

Function Set CNN Network<input image block>;

Function Calculate<the score of all image block>;

Function Least Square Method< C_{km} >;

Function Calculate<The score of images>;

End.

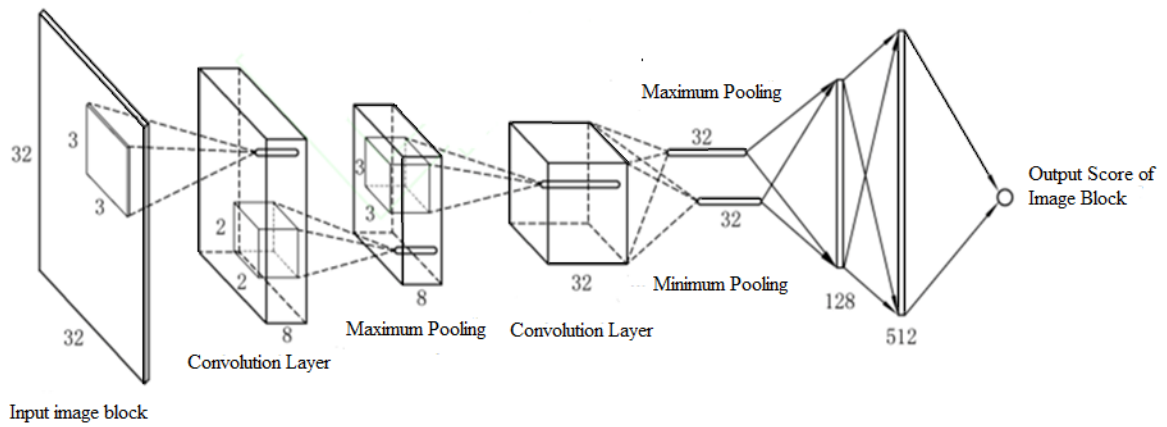


Fig. 3. The structure of IQF-CNN

Table 2. The Spearman correlation coefficient

Method	JPEG2000	JPEG	WN	FF	BLUR
DIIVINE[18]	0.913	0.910	0.984	0.863	0.921
BLLINDS-II[19]	0.929	0.900	0.946	0.889	0.923
BRISQUE[20]	0.914	0.913	0.979	0.877	0.951
CNN-KangLe[21]	0.952	0.977	0.984	0.908	0.956
This method	0.987	0.989	0.985	0.985	0.997

Table 3. The Pearson correlation coefficient

Method	JPEG2000	JPEG	WN	FF	BLUR
DIIVINE	0.922	0.921	0.988	0.888	0.917
BLLINDS-II	0.935	0.968	0.980	0.896	0.930
BRISQUE	0.922	0.973	0.985	0.903	0.942
CNN-KangLe	0.953	0.981	0.984	0.933	0.953
This method	0.952	0.971	0.984	0.932	0.948

Next the image database of LIVE is used to test. LIVE has five distortion types of images, including JPEG200, JPEG, WN, FF, BLUR. Every image has subjective score. The proposed method can give the objective score of images. So, the subjective and objective score of every image can be compared to give the similarity coefficient. The spearman correlation coefficient (SROCC) and pearson linear correlation coefficient (PLCC) of DIIVINE and other assessment methods are compared with this method. **Table 2** and **Table 3** show that this method can improve the similarity coefficient between the subjective and objective scores.

(2) Step2: The video quality assessment

Based on the above derivation, the method is extended to video quality assessment. Because the video has many image frames, the relationship between different image frames should be analyzed, next the frame weights of different windows will be added to the image frames. At first the above model is used to calculate the scores of all the image frames. Due to the retention effect of human vision, the scores of image frames need to be corrected. The hysteresis function is designed. Set i th image frame score is $f(i)$, $a(i)$ and $b(i)$ are the correction scores. $a(i)$ is the maximum scores of the front T image frames, $b(i)$ is the minimum scores of the front T image frames.

$$\begin{cases} a(1) = b(1) = f(1) \\ a(i) = \max[f(t)], t = \{\max(i-T, 1) \dots i-1\} \\ b(i) = \min[f(t)], t = \{\max(i-T, 1) \dots i-1\} \end{cases} \quad (6)$$

$$Q(i) = \alpha a(i) + (1 - \alpha) b(i)$$

The $a(i)$ and $b(i)$ is used to simulate the retention effect of human vision. To balance the scores of them, $Q(i)$ is calculated as the last score of image frame, α is the adjustable weight value. In this way, the score of every image frame will be updated.

The different window functions are designed to improve the accurate of video quality assessment. The Rectangular window, Hanning window and Hamming window are used to give the weight to different image frames. So, the video quality can be calculated by equation (7).

$$\omega(i) = 1, 1 \leq i \leq N$$

$$\omega(i) = 0.5(1 - \cos(2\pi \frac{i}{N+1})), 1 \leq i \leq N$$

$$\omega(i) = 0.54 - 0.46 \cos(2\pi \frac{i}{N-1}), 1 \leq i \leq N$$

$$video\ quality = \sum_{i=1}^N \omega(i) Q(i) \quad (7)$$

Part 3: pseudo code

Video treatment

Hysteresis function (input scores of image frame);

Update (the scores of image frame);

Rectangular window (input image frame);

Calculate (the score of videos).

Table 4. The correlation coefficient of different method

Method	PLCC	SROCC
Rectangle window	0.404	0.395
Hamming window	0.719	0.706
Hanning window	0.869	0.864
PSNR	0.404	0.368
Speed-SSIM ^[22]	0.596	0.585
VQM ^[23]	0.724	0.703
MOVIE ^[24]	0.812	0.789
Method of [15]	0.832	0.831
Method of [16]	0.815	0.817

The video database of LIVE is also tested the effectiveness of this method. As shown in **Table 4**, the similarity coefficient of different window functions and other methods are calculated. The Hanning window has improved the accuracy of assessment more than other methods and got higher similarity coefficient.

The first method has two steps. Firstly, the IQF-CNN is used to train the image and assess them, secondly the relationships between image frames are analyzed by the hysteresis function and different window function to improve the accuracy of this model. All steps of

this method have been validated by related experiments, which show good applicability and can improve the accuracy of objective assessment.

2.2 The second method based on CNN and GRU

As shown in Fig. 4, this method comprehensively considers the spatial and temporal features of video. The CNN is used to extract the spatial features of video and the GRU is used to analyze the relationship between different frames of video and extract the temporal features. At last, the full connection is regressed to predict the video quality. The detail steps of this method are given as follows.

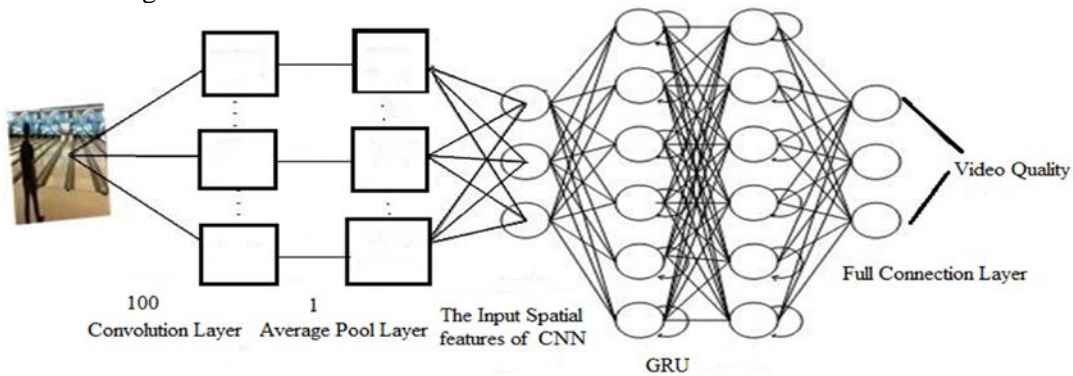


Fig. 4. The model of second method

(1) Step1: The process step of CNN

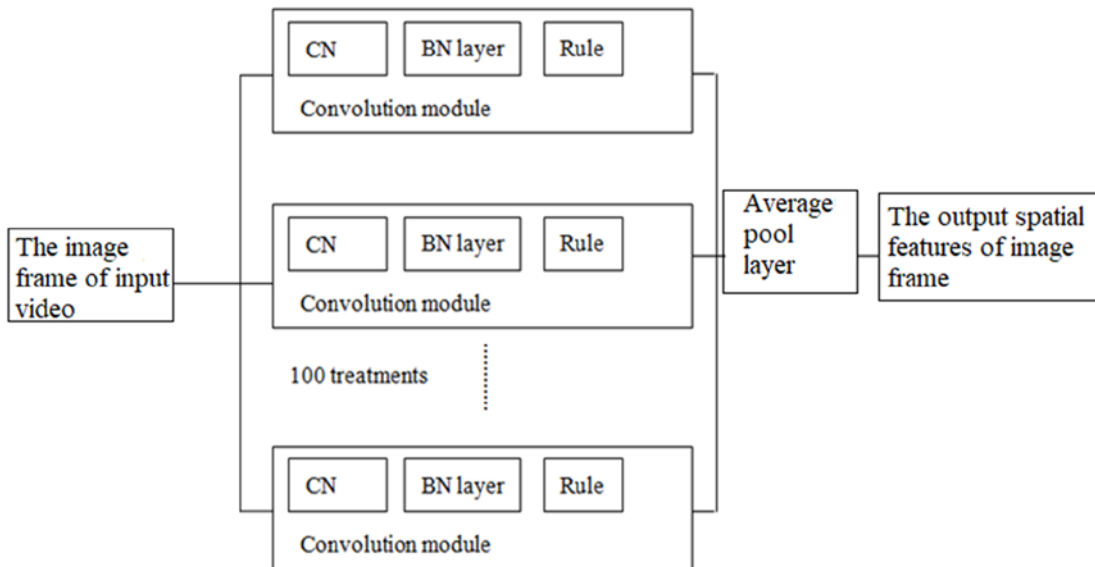


Fig. 5. The process of CNN

As shown in Fig. 5, the first part of this model is the CNN. It designs a CNN with 100 layers of convolution model and 1 layer of average pool. It is used to extract the spatial features of every image frame of test video. Next all the spatial features of different image frame will be processed.

(2) Step2: The process step of GRU

As shown in Fig. 6, GRU is one kind of LSTM. It has two control doors, including update door and reset door. The equation of update and reset door is (8) - (11). R represents the reset door and T represents update door.

$$R_t = \sigma(W_{rx}X_t + W_{rh}h_{t-1} + b_r) \tag{8}$$

$$Z_t = \sigma(W_{zx}X_t + W_{zh}h_{t-1} + b_z) \tag{9}$$

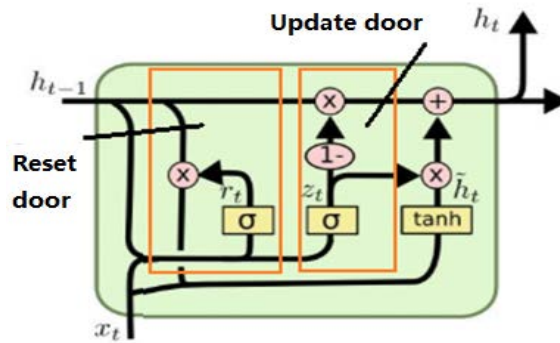


Fig. 6. The model of GRU [25]

$$\tilde{h}_t = \tanh(W_{hx}X_t + W_{hh}(R_t \cdot H_{t-1} + b_h)) \tag{10}$$

$$h_t = Z_t \cdot H_{t-1} + (1 - Z_t) \cdot \tilde{h}_t \tag{11}$$

σ is the activation function, h_{t-1} is the output value of the previous time, \tilde{h}_t is the candidate hidden state, W_{hx} , W_{hh} , b_h is the weight value.

The cyclic kernel of GRU has memory and can extract the temporal features of different image frame by setting the parameters of network. As shown in Fig. 7, the spatial features of different image frame are inputted into the GRU. It designs a GRU with 256 hidden cells and 1 full connection layer. The SGD gradient descent algorithm is used to optimize the structure of GRU, the RMSE (Root Mean Square Error) is used as loss function. The detail parameters set for GRU is shown in Table 5. At last, the objective assessment score of videos is given.

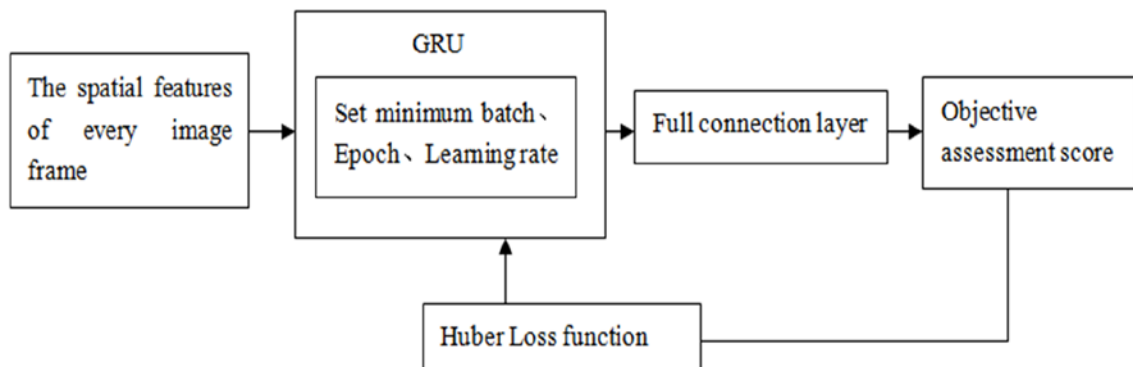


Fig. 7. The process of GRU

Table 5. The parameters set for GRU

Parameter	Value
Max Epoch	50
Decline coefficient of learning rate	0.2
Initial learning rate	0.02
Minimum batch capacity	30

(3) The experimental analysis

The video database of LIVE, IVP and CSIQ are used to test the proposed method. Eighty percent of LIVE videos are used for training and 20 percent of LIVE videos are used for testing. Next this method is applied on IVP and CSIQ. As shown in **Fig. 8**, when the training time is 0 to 100, RMSE decreases sharply, when the training times reach more than 100 times, RMSE is stable. On the other hand, the predicted values of 20 samples are compared with the true values. As shown in **Fig. 9**, when the number of prediction samples exceeds 10, the predicted value and true value are close.

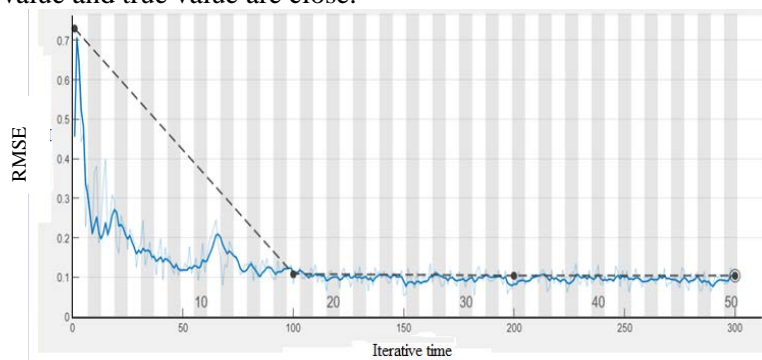


Fig. 8. The RMSE result of different training time

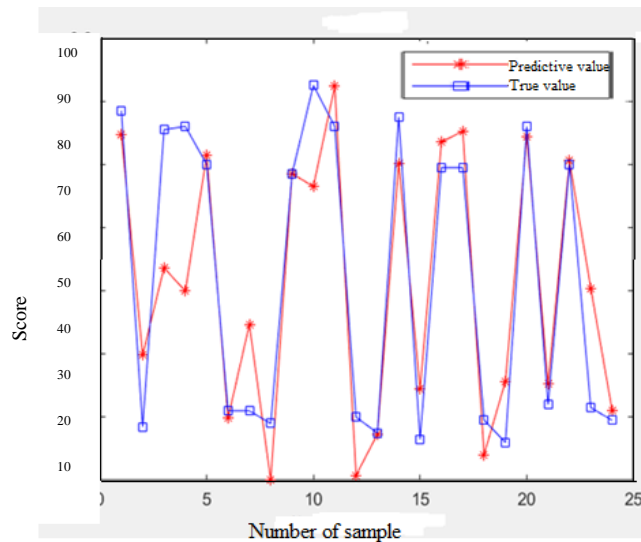


Fig. 9. The value of prediction and true

Table 6. The SROCC similarity of LIVE

Distortion type Method	Wireless	IP	H.264	MPEG-2
SSIM[26]	0.738	0.600	0.762	0.512
STMAD[27]	0.806	0.768	0.904	0.842
BSWQ[28]	0.848	0.802	0.756	0.731
FS-MOVIE	0.859	0.801	0.876	0.872
Ours	0.783	0.742	0.833	0.842
Method of [15]	0.827	0.828	0.831	0.834
Method of [16]	0.814	0.815	0.816	0.819

Table 7. The PLCC similarity of LIVE

Distortion type Method	Wireless	IP	H.264	MPEG-2
SSIM	0.822	0.769	0.705	0.534
STMAD	0.806	0.769	0.904	0.847
BSWQ	0.871	0.835	0.807	0.753
FS-MOVIE	0.813	0.772	0.849	0.860
Ours	0.786	0.745	0.833	0.843
Method of [15]	0.827	0.831	0.829	0.836
Method of [16]	0.813	0.815	0.816	0.818

The experimental results of LIVE are record in **Table 6** and **Table 7**. Other methods are compared with it. The full-reference methods are SSIM, STMAD, the reduced-reference method is BSWQ, the non-reference is FS-MOVIE. This method is superior to the common full reference and reduced reference method, but it is slightly inferior to the non-reference method. This method has a big difference in dealing with IP transmission distortion and wireless network transmission distortion. Because this method does not consider the video distortion caused by network delay, this algorithm in dealing with these two distortion types of videos is not good.

Table 8. The correlation coefficient of IVP

Method	PLCC	SROCC
PSNR	0.566	0.542
SSIM	0.543	0.521
3-SSIM	0.677	0.662
VQM	0.673	0.685
Ours	0.695	0.676
Method of [15]	0.765	0.771
Method of [16]	0.782	0.779

Table 9. The correlation coefficient of CSIQ

Method	PLCC	SROCC
PSNR	0.583	0.578
SSIM	0.718	0.701
STMAD	0.826	0.817
BSWQ	0.826	0.841
Ours	0.756	0.774
Method of [15]	0.812	0.807
Method of [16]	0.816	0.810

In the same way, the video databases of IVP and CSIQ are also tested. **Table 8** and **Table 9** record the correlation coefficient of IVP and CSIQ, this method is close to the effect of many assessment methods.

The second method has two steps. Firstly, the CNN is used to extract the spatial features of video. Secondly the GRU is used to analyze the relationship between different frames of video and extract the temporal features. Thirdly the full connection is regressed to predict the video quality. This method comprehensively considers different features of video, and combines two algorithms to give more accurate objective assessment scores. It has good usability.

3. Conclusion

This paper introduces two video quality assessment methods based on deep neural network. The deep neural network is applied in video quality assessment. The first method uses the IQF-CNN to build image quality assessment method. Next the method is extended to the video quality assessment. So, every image frame of video is predicted. Meanwhile the relationship between different frames is analyzed, and the hysteresis function and different window function are designed to improve the accuracy of video quality assessment. The second method uses CNN and GRU to build video quality assessment method. The spatial features of video frames are extracted using CNN network, next the temporal features of the video frame using GRU network. Finally, the extracted temporal and spatial features are analyzed by using full connection layer of CNN network to obtain the video quality assessment score. All the above proposed methods are verified on different public video databases and compared with other video quality assessment methods.

In the future, we will reconstruct the network structure to improve the accuracy of the proposed method.

Conflicts of interests and Acknowledgments

There is no conflict in this paper. All data generated or analyses during this study are included in this published article.

Acknowledgments

This work is supported by Science and Technology Research Project of Jiangxi Provincial Department of Education(GJJ2200529).

Reference

- [1] Balasubramanyam Appina, Sumohana S. Channappayya, "Full-Reference 3-D Video Quality Assessment Using Scene Component Statistical Dependencies," *IEEE Signal Processing Letters*, vol.25, no.6, pp.823-827, Apr. 2018. [Article \(CrossRef Link\)](#)
- [2] Rajiv Soundararajan, Alan C. Bovik, "Video Quality Assessment by Reduced Reference Spatio-Temporal Entropic Differencing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.23, no.4, pp.684-694, Aug. 2012. [Article \(CrossRef Link\)](#)
- [3] Deepti Ghadiyaram, Janice Pan, Alan C. Bovik, "Learning a Continuous-Time Streaming Video QoE Model," *IEEE Transactions on Image Processing*, vol.27, no.5, pp.2257-2271, Jan. 2018. [Article \(CrossRef Link\)](#)
- [4] Wenjuan Shi, Yanjing Sun, Jinqiu Pan, "Continuous Prediction for Quality of Experience in Wireless Video Streaming," *IEEE Access*, vol.7, pp.70343-70354, May. 2019. [Article \(CrossRef Link\)](#)
- [5] Fan Zhang, David R. Bull, "A Perception-Based Hybrid Model for Video Quality Assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.26, no.6, pp.1017-1028, May. 2015. [Article \(CrossRef Link\)](#)
- [6] Mehdi Banitalebi-Dehkordi, Abbas Ebrahimi-Moghadam, Morteza Khademi, et al, "No-Reference Video Quality Assessment Based on Visual Memory Modeling," *IEEE Transactions on Broadcasting*, vol.66, no.3, pp.676-689, Dec. 2019. [Article \(CrossRef Link\)](#)
- [7] Chathura Galkandage, Janko Calic, Safak Dogan, et al, "Full-Reference Stereoscopic Video Quality Assessment Using a Motion Sensitive HVS Model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.31, no.2, pp.452-466, Mar. 2020. [Article \(CrossRef Link\)](#)
- [8] Kongfeng Zhu, Chengqing Li, Vijayan Asari, et al, "No-Reference Video Quality Assessment Based on Artifact Measurement and Statistical Analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.25, no.4, pp.533-545, Oct. 2014. [Article \(CrossRef Link\)](#)
- [9] Xiaoming Tao, Yiping Duan, Mai Xu, et al, "Learning QoE of Mobile Video Transmission With Deep Neural Network: A Data-Driven Approach," *IEEE Journal on Selected Areas in Communications*, vol.37, no.6, pp.1337-1348, Mar. 2019. [Article \(CrossRef Link\)](#)
- [10] Yu Zhang, Xinbo Gao, Lihuo He, et al, "Blind Video Quality Assessment with Weakly Supervised Learning and Resampling Strategy," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.29, no.8, pp.2244-2255, Aug. 2018. [Article \(CrossRef Link\)](#)
- [11] Ali Al-Naji, Sang-Heon Lee, Javaan Chahl, "Quality Index Evaluation of Videos Based on Fuzzy Interface System," *IET Image Processing*, vol.11, no.5, pp.292-300, Aug. 2017. [Article \(CrossRef Link\)](#)
- [12] Mohammed Alreshoodi, Emad Danish, John Woods, et al, "Prediction of Perceptual Quality for Mobile Video Using Fuzzy Inference Systems," *IEEE Transactions on Consumer Electronics*, vol.61, no.4, pp.546-554, Nov. 2015. [Article \(CrossRef Link\)](#)
- [13] Zhiming Shi, Chengti Huang, "Network Video Quality Assessment Based on Fuzzy Inference System," *The Journal of China Universities of Posts and Telecommunications*, vol.2018, no.1, pp.70-77, Feb. 2018. [Article \(CrossRef Link\)](#)
- [14] Domonkos Vargha, "No-Reference Video Quality Assessment Based on the Temporal Pooling of Deep Features," *Neural processing letters*, vol.50, no.3, pp.2595-2608, Dec. 2019. [Article \(CrossRef Link\)](#)
- [15] Jari Korhonen, "Two Level Approach for No-Reference Consumer Video Quality Assessment," *IEEE Transactions on Image Processing*, vol.28, no.12, pp.5923-5938, Dec. 2019. [Article \(CrossRef Link\)](#)
- [16] Zhengzhong Tu, Yilin Wang, Neil Brikbeck, et al, "UGC-VQA: Benchmarking Blind Video Quality Assessment for User Generated Content," *IEEE Transactions on Image Processing*, vol. 30, pp. 4449 -4464, Apr.2021. [Article \(CrossRef Link\)](#)
- [17] Salah.Khawla Ben, Othmani.Mohamed, Kherallah.Monji, "A Novel Approach for Human Skin Detection using Convolution Neural Network," *Visual Computer*, vol. 38, no. 5, pp. 1833-1843, May. 2022. [Article \(CrossRef Link\)](#)

- [18] Moorthy A.K., Bovik A.C., “Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality,” *IEEE Transactions on Image Processing*, vol.20, no.12, pp.3350-3364, Dec. 2011. [Article \(CrossRef Link\)](#)
- [19] Saad M.A., Bovik A.C., Charrier C., “Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339-3352, Mar. 2012. [Article \(CrossRef Link\)](#)
- [20] Mittal A., Moorthy A.K., Bovik A.C., “No-reference Image Quality Assessment in The Spatial Domain,” *IEEE Transactions on Image Processing*, vol.21, no.12, pp.4695-4708, Dec. 2012. [Article \(CrossRef Link\)](#)
- [21] Kang L., Ye P., Li Y., et al, “Convolutional Neural Networks for No-reference Image Quality Assessment,” in *Proc. of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, United States, pp.1733-1740, June 23-28, 2014. [Article \(CrossRef Link\)](#)
- [22] You J., Ebrahimi T., Perkis A., “Attention Driven Foveated Video Quality Assessment,” *IEEE Transactions on Image Processing*, vol.23, no.1, pp.200-213, Oct. 2013. [Article \(CrossRef Link\)](#)
- [23] Pinson M.H., Wolf S., “A New Standardized Method for Objectively Measuring Video Quality,” *IEEE Transactions on broadcasting*, vol.50, no.3, pp.312-322, Sep.2004. [Article \(CrossRef Link\)](#)
- [24] Seshadrinathan K., Bovik A.C., “Motion Tuned Spatio-temporal Quality Assessment of Natural Videos,” *IEEE transactions on image processing*, vol.19, no.2, pp.335-350, Oct. 2009. [Article \(CrossRef Link\)](#)
- [25] Fugang Liu, Ziwei Zhang, Ruolin Zhou, “Automatic Modulation Recognition based on CNN and GRU,” *Tsinghua Science and Technology*, vol.27, no.2, pp.422-431, Apr. 2022. [Article \(CrossRef Link\)](#)
- [26] Wang Z, Bovik A C, Sheikh H R, et al, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Transactions on Image Processing*, vol.13, no.4, pp.600-612, Apr. 2004. [Article \(CrossRef Link\)](#)
- [27] Vu P V, Vu C T, Chandler D M, “A Spatiotemporal Most-apparent-distortion Model for Video Quality Assessment,” in *Proc. of 2011 IEEE International Conference on Image Processing*. Brussels, Belgium, pp.2505-2508, Sep. 11-14, 2011. [Article \(CrossRef Link\)](#)
- [28] Zhang Shufang, Huang Xiaoqin, “Video Quality Assessment Model in Wavelet Domain based on Background Subtraction,” *Journal of Tianjin University*, vol.50, no.12, pp.1255-1261, Dec. 2017.



Zhiming Shi is a doctor. His major is communication engineering. Now he is a lecturer in Jiangxi University of Finance and Economics. His interests include image processing, video processing and video quality assessment.