Quality Assessment for Networked Video Streaming Based on Deep Learning

Jinkun Guo and Shuai Wan(✉)

Northwestern Polytechnical University, Xi'an, China
guojinkun@mail.nwpu.edu.cn, swan@nwpu.edu.cn

Abstract. There arises the need for quality assessment in networked video streaming since video services have great significance for both users and providers. In this paper, a neural network is proposed to realize networked video streaming quality assessment. Firstly the key parameters of video streaming are extracted, including the bit-rate, the coded bits of each frames, the number of lost packet and so on. Then the neural network is built to study the mapping of these parameters and video quality. The influence on the video quality assessment by different network depth and different layer settings in the neural network is also taken into comparison. The performance of the proposed neural network has been compared with other methods and evaluated by the quality assessment experiment of videos in different resolutions. The results demonstrate the effectiveness and efficiency of video quality assessment based on the neural network.

Keywords: Video quality assessment · Deep learning · Video streaming

1 Introduction

Nowadays, more and more people use online video services. Therefore, quality assessment for networked video streaming is important for both video users and service providers. The traditional video subjective and objective quality assessment methods have been developing for a long time. Each has advantages, but the disadvantages are also prominent. The traditional subjective video quality assessment can achieve the most accurate result but it requires rigorous experiment environments. The traditional objective video quality assessment can avoid the great manpower involved in the experiment but it is generally not very accurate in reflecting the feelings of the viewers.

According to the researches related to the networked video streaming quality assessment, most of the methods and algorithms are based on the no-reference objective video quality assessment. Authors in [1] proposed a reconstruction-based no-reference video objective quality assessment algorithm. It can give out pleasant performances but the features of the testing frames have to be deeply mined first. Wei and Zhang proposed a no-reference video quality assessment method by utilizing the hybrid parameters extracted from compressed video frames [2]. This method could be effective, but the computation burden was high. An objective no-reference video quality assessment method was presented in [3], where the pictures have to be split into small blocks which will introduce much complexity in video decoding. A packet-layer

B. Li et al. (Eds.): IoTaaS 2018, LNICST 271, pp. 90–97, 2019.
https://doi.org/10.1007/978-3-030-14657-3_10

assessment model for video quality was introduced in [4]. The information including bit-rate, frame rate, etc., were provided by packet headers so the payload information was not needed for video quality assessment. Authors in [5] used the packet-layer video quality assessment model with the characteristics of the High Efficiency Video Coding (HEVC) standard which achieved a present correlation between subjective scores and objective scores. Quality of Experience (QoE) reflects the impact of factors that affect the satisfaction of users. Ref. [6] provided a structured way to build an objective QoE model by Principal Component Analysis (PCA) and Analytic Hierarchy Process (AHP) analysis. Ref. [7] was devoted to build a perceptual video quality metric based on features which were used for semantic task and human material perception. In this method, several estimators which could reflect the neuroscientific and psychophysical evidence were gathered together to achieve video quality assessment.

Combined with the deep learning method, which is popular and most suitable for dealing with large data problems at present, the traditional problems encountered in video quality assessment can be effectively avoided and even better results can be achieved. The literature [8–10] put forward different video quality assessment based on back propagation (BP) neural network. In [8], human visual regions of interest were selected and temporal and spatial features were extracted as the input of BP network. Literature [9] presented one kind of BP-based estimate on the network video QoE to construct the mapping model between QoE and quality of service (QoS). Ref. [10] provided an automated and computational video quality assessment method which employed offline deep unsupervised learning processes and inexpensive no-reference measurements at server side and client side, respectively.

The rest of this paper is organized as follows. Section 2 proposes a novel quality assessment method for networked video based on deep learning. Performance evaluation and the comparison of this method with the existed method are provided in Sect. 3 in detail. Conclusions are drawn in Sect. 4.

2 Quality Assessment for Networked Video Streaming Using Deep Learning

The general framework of the proposed method is shown as Fig. 1. Firstly, the useful parameters in a video stream are extracted. Then they are sent to a neural network to get the video quality. The neural network has to be carefully designed and trained for quality assessment. In the experiment, videos in three different resolutions from YUV video sequence are used. For training purposes, video sequences Slide Editing, Johnny in 720P; Crowdrun, Harbour in 4cif; Bus, Carphone in cif are used. For testing purposes, video sequences Shield in 720P, Crew in 4cif and Claire in cif are used.

The subjective video quality can be calculated in terms of the 5-point absolute category rating (ACR) mean opinion score (MOS) scale according to the recommendation P.1201.1. The ITU-T P.1201-series of Recommendations specifies models for monitoring the audio, video and audiovisual quality of IP-based video services based on packet-header information. MOS used in the experiment as the label of videos represents the mean opinion score of video quality by viewers with the consideration of influence brought by video encoding, packet-loss in the transmission, rebuffering and the screen size.

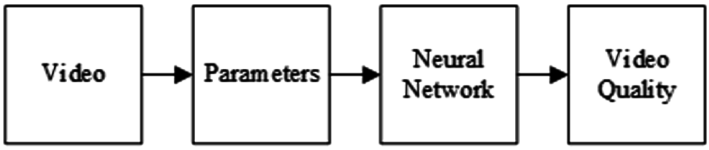

Fig. 1. Framework of proposed video quality assessment method.

2.1 Extract Key Parameters

Firstly, the original reference video is encoded and formatted by ffmpeg. The quantization parameters range from 0 to 51, producing 52 compressed videos which quality are successively reduced. Secondly, the Wireshark software [11] is used to simulate the network transmission of compressed video and capture each video packet during the transmission process.

After the completion of the Wireshark sending and receiving video data packets, each of the original video will generate a pcap file. A pcap file contains all the data packets in the process of video transmission. The detailed information of each video packet can be got, which contains comprehensive information for video streaming transmission on the network.

Since there are so many parameters in every data packet and most of them are not relative to the video quality assessment, the parameters below are chosen as the key values to assess the video quality: Br (the coded bits of each frames), bit-rate, Psize, Isize, Vplef (video packet-loss event frequency), Vir (impairment rate of video stream), Vairf (average impairment rate of video frequency), Vdsize (the screen size), Vdiag (number of pixels on the diagonal of screen) since they represent video impairment from different aspects. For example, Vir reflects the video impairment among frames and Vairf reflects video impairment within each frame.

Because of the video size, there is a difference in the number of packets that can be transmitted. Matlab software is used to obtain the mean value of the key parameters extracted from each video. The data sets of key parameters and MOS of compressed video are divided into train set and test set. To be precise, the key parameters extracted from videos of two different scenes in each resolution and the corresponding quality value MOS are selected as the train set. The key parameters and MOS of one video scene in every resolution are divided as test set.

2.2 Construct the Neural Network

Fully connected neural network minimizes the mean square error between network output and label by back-propagating error, which can adjust the weights and biases of network effectively. It has great performance of non-linear modelling. In order to achieve the optimal network model, the experiments try to build fully connected networks in different depth with the comparison among different activation function types based on the parameters and MOS extracted and calculated in advance.

The fully-connected back-propagation neural network is built based on TensorFlow. The loss function is the mean square error between the output of the network

and the ground truth of MOS calculated by recommendation P.1201.1. In the back forward propagation process, Adam optimization algorithm is chosen to optimize the loss function. The Adam algorithm dynamically adjusts the learning rate of each parameter according to the first and second order moment estimate of the gradient of the loss function. It is also based on the gradient descent method, but in each iteration parameters have a definite range of learning step so that it will not lead to a large learning step size because of a large gradient and the parameter values are relatively stable. In this experiment, the training iterations are 200000 epochs and after every 500 epochs the current error value will be displayed, which is the difference between the network output and the video real quality. After the training process, the network structure and parameters will be saved which can be used directly in the testing process. For all the experiment, we have used the computer with two GeForce GTX 1080Ti Graphics Cards and Intel i7 7700k CPU and it will take around 330 s for the whole training and testing process.

3 Experimental Results and Analysis

The input of the fully connected neural network are key parameters Br, Psize, Isize, Vplef, Vir, Vairf, Vdsize, Vdiag, which are extracted from the videos in YUV video sequence. This dataset is commonly used in the video coding and decoding. Videos in the YUV video sequence are in the 4:2:0 YUV format. And also, MOS is computed as the label of each video. The Pearson correlation coefficient (PCC) is used to test the linear correlation between experimental video quality and real video quality. The range of PCC is [−1,1]. The greater the PCC is, the higher the linear correlation degree will be.

3.1 Comparison Among Different Neural Network Depth

The results from three-layer fully connected network 1 and four-layer fully connected network 2 are compared in this experiment. The number of neurons in each layer of network 1 are 10, 20, 30, respectively. The number of neurons in each layer of network 2 are 10, 20, 25, 30, respectively. The training set used in each training session is the same, and videos in each resolution are tested. Figures 2 and 3 show the results and errors comparison between two networks testing video Shield in 720P. The x-axis represents 52 distorted video sequences. The y-axis indicates the experimental video quality ranging from 0 to 5. It can be seen that increasing depth can enhance the ability of studying the mapping of key parameters and video quality by fully-connected back-propagation neural network.

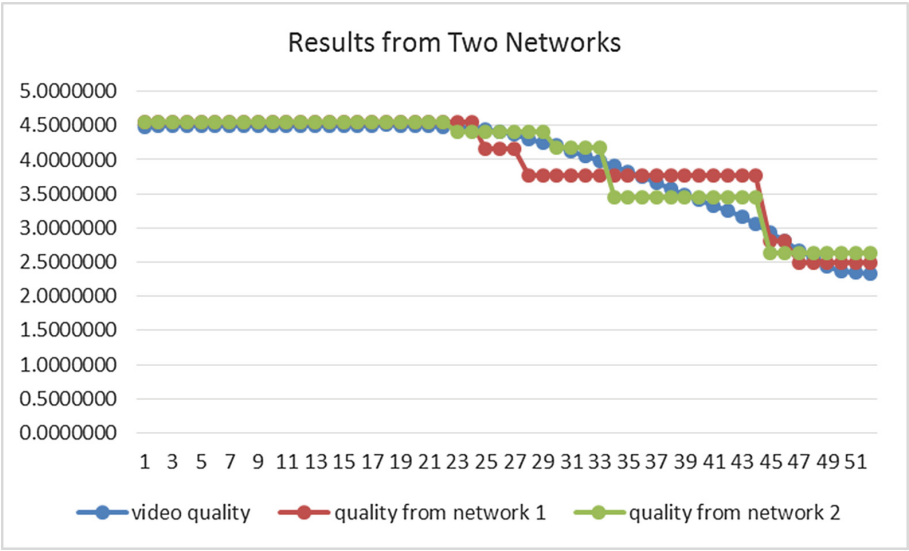

Fig. 2. The results of video Shield in 720P.

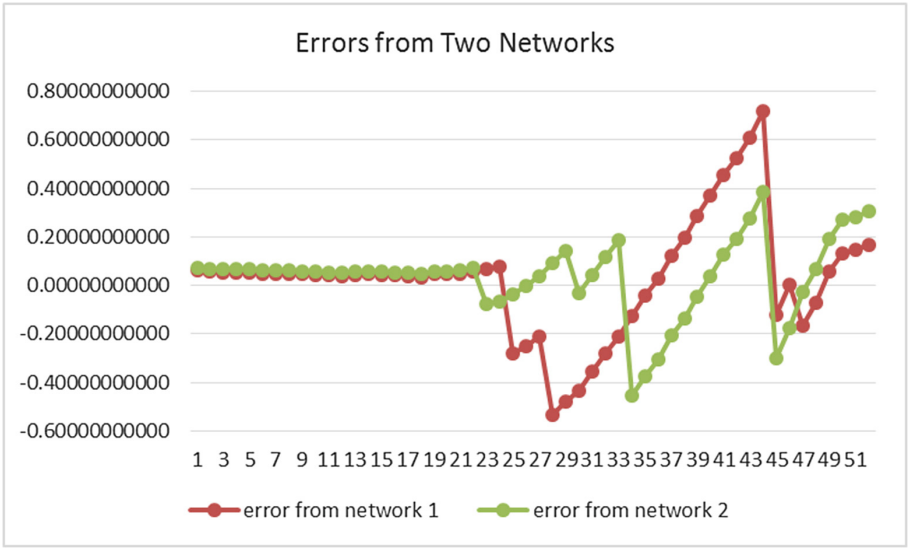

Fig. 3. The errors of video Shield in 720P.

3.2 Comparison Among Different Activation Function

In this experiment a four-layer fully connected neural network is chosen. The number of neurons in each layer are 10, 20, 25, 30, respectively. The second and third layers use ReLU activation function. The fourth layer uses softsign activation function. The

experiment compares the results from two different structures: the first layer in network 1 uses softsign and network 2 first layer does not use any activation function. The training sets are the same in each training session, and videos in each resolution is tested. The x-axis represents 52 distorted video sequences. The y-axis indicates the experimental video quality ranging from 0 to 5. Figures 4 and 5 show the results and errors from two networks testing video Shield in 720P. The results and errors indicate that softsign activation function can improve the fitting capacity of the neural network and ReLU can ensure the convergence of computation results.

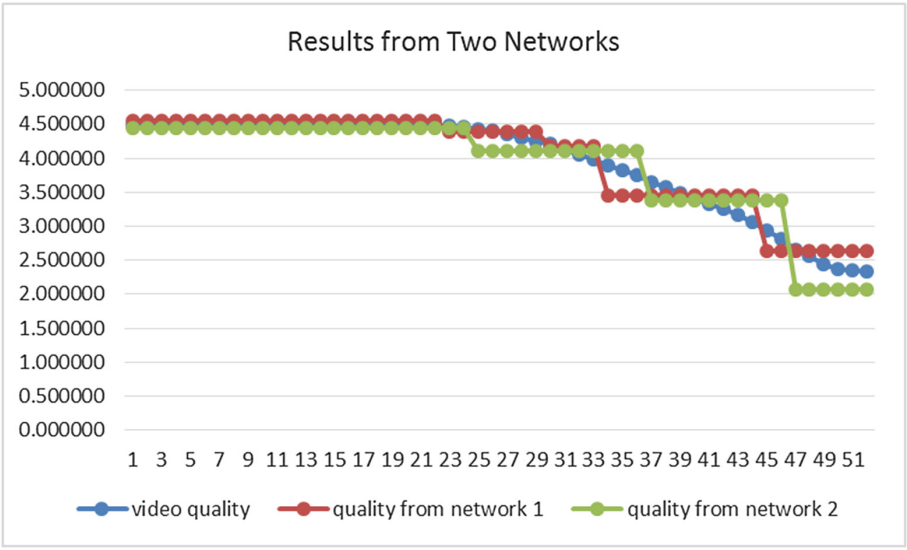

Fig. 4. The results of video Shield in 720P.

3.3 Analysis

After a large number of experiment, the optimal fully-connected back-propagation neural network is in four layers, of which the second and third network layer use softsign activation function, and the first, fourth network layer use ReLU. The PCC of video Crew in 4cif is 98.02% and the average error is 0.1502; the PCC of video Coastguard in cif is 89.47% and the average error is 0.2302; the PCC of video Shield in 720P is 97.57% and the average error is 0.1217. Compared with the performance of the method FRAME-FEBP proposed by Wei and Zhang in [2], which relevance is 75.20% in such video sequence, the results show that the four-layer fully-connected back-propagation neural network can perform the subject video quality assessment quite well.

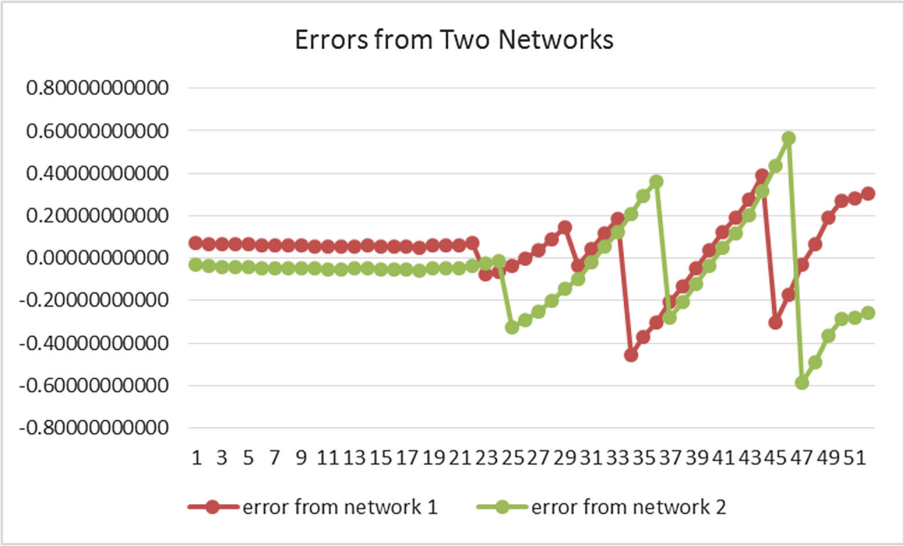

Fig. 5. The errors of video Shield in 720P.

4 Conclusions

This paper proposed a method of quality assessment for networked video streaming based on deep learning, where a fully connected neural network is designed. The network can conclude results which are very close to the traditional ways by learning the mapping between key parameters extracted from the networked video streaming and MOS calculated according to the recommendation P.1201.1. Experimental results have shown that the proposed model provides a high-performance and feasible solution to evaluate the quality of networked video streaming.

References

1. Wu, Z., Hu, H.: Reconstruction-based no-reference video quality assessment. In: Region 10 Conference (TENCON), pp. 3075–3078. IEEE (2016)
2. Wei, B., Zhang, Y.: No-reference video quality assessment with frame-level hybrid parameters for mobile video services. In: 2016 2nd IEEE International Conference on Computer and Communications (ICCC), pp. 490–494. IEEE (2016)
3. Jing, W., Zedong, W., Fei, W., et al.: A no-reference video quality assessment method for VoIP applications. In: 2016 IEEE 13th International Conference on Signal Processing (ICSP), pp. 644–648. IEEE (2016)
4. Yang, F.Z., Wan, S.: Overview of state-of-the-art and future of networked video quality assessment. J. Commun. (2012)
5. Guo, J., Zheng, K., Hu, G., et al.: Packet layer model of HEVC wireless video quality assessment. In: 2016 11th International Conference on Computer Science and Education (ICCSE), pp. 712–717. IEEE (2016)

6. Youssef, Y.B., Mellouk, A., Afif, M., et al.: Video quality assessment based on statistical selection approach for QoE factors dependency. In: 2016 IEEE Global Communications Conference (GLOBECOM), pp. 1–6. IEEE (2016)
7. Deng, B.W., et al.: Video quality assessment based on features for semantic task and human material perception. In: IEEE International Conference on Consumer Electronics-China. IEEE (2017)
8. Jiangbo, X., Xiuhua, J.: No-reference high definition video quality assessment based on BP neural network. In: Proceedings of the 2011 International Conference on Future Computer Science and Application (FCSA 2011 V1), p. 4. Intelligent Information Technology Application Association (2011)
9. Yao, H., Huang, Y.: BP-based estimate on network video QoE. Comput. Eng. Des. **38**(1), 1–6 (2017)
10. Vega, M.T., et al.: Deep learning for quality assessment in live video streaming. IEEE Signal Process. Lett. **PP**(99), 1 (2017)
11. https://www.wireshark.org/