Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Dataset-driven new set of video quality evaluations towards reference-free video quality evaluations

Feiran Yang[1], Guozhong Wang[1,✉], Haiwu Zhao[1]

[1]*Shanghai University Of Engineering Science, School of Electronic and Electrical Engineering, Shanghai, 201600, China*

ABSTRACT

The popularity of network video service, which leads to the specification of the network video service quality is becoming more and more urgent, and human is the ultimate watch users of network video, Evaluation of human observers on the video's perception of the situation is becoming more important, depending on the source video network video know nothing, you will need to refer to the participation of video quality evaluation algorithm. However, when the quality evaluation is done by human rating, it is time-consuming and laborious, so the computer is required to make an objective evaluation of the video. In the objective evaluation, the excellent performance of convolutional neural network based on deep learning in feature extraction contributes to the rapid development of the research field of video quality evaluation. However, the development of deep learning algorithms requires appropriate data sets for training and testing. The existing data sets are relatively small in scale and not comprehensive in terms of video content types and distortion types. Therefore, it is necessary to provide a new data set to evaluate the quality of video without reference, expand the scale of the data set, expand the content and distortion type of the video. At the same time, considering the new development of network video services, the video resolution is positioned as high definition, and the original video sampling ratio is 4:2:2. The dataset is freely available to relevant researchers for scientific research.

Keywords: Network video; Video quality evaluation; Video data set; Subjective quality rating

1. Introduction

With the rapid development of multimedia and network technology, network video service is more and more popular. For example, the number of network video applications such as short video, Digital cinema and shared conference film is growing rapidly, and digital video is entering People's Daily life through the network more and more [1]. Digital videos typically go through several stages of processing before reaching the end user, who tends to be a human observer. The effect of most processing stages will reduce

✉ Corresponding author.
E-mail address: yangfeiran111@163.com (G. Wang).
Received 10 April 2024; Revised 24 August 2024; Accepted 12 October 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-041

the quality of the video, for example, video distortion caused by compression coding and changing the resolution of the source video will affect the user's viewing experience [2]. Because of this, web video needs a quality standard. The most reliable measure of video quality is the subjective evaluation of video by human observers, which is called subjective video quality assessment. However, subjective experiments are not feasible for most applications, because they require a lot of human and material resources [3]. To solve this problem, researchers are committed to the study of objective Video quality evaluation algorithm (VQA) to achieve the ultimate goal of matching human perception [4].

Objective quality evaluation algorithms include reference and No reference [5], and the source video cannot be known for network video. Therefore, an objective No reference video quality evaluation algorithm (NR-VQA) method is needed.

In recent years, when studying no-reference video quality assessment, traditional no-reference video quality assessment is often used [6]. Studies have found that it is difficult to train quality assessment models with traditional machine learning methods and the results are poor. Based on this problem, research on video quality assessment methods based on deep learning has become a hot topic. In order to predict video quality objectively, accurate and effective video quality assessment methods are needed [7]. The most important thing to develop these algorithms is to establish the corresponding video data set for training and testing. The content richness and quality of the data set play a crucial role in the evaluation results. At present, there is no special data set for network video quality research at home and abroad, but there are many public general video quality assessment data sets. For example, LIVE Video Database [8], VQEGHD5 Video Database [9], IVPL Video Quality Database [10], CSIQ Video Quality Database [11], etc.

The LIVE Video Database [8] has 10 source videos, and four distortions are added: wireless transmission, IP transmission, H.264 compression coding and MPEG-2 compression coding. There are 150 distorted videos in total, and the resolution of the video is 768x432. The VQEGHD5 Video Database [9] has nine source videos with four distortions added: H.264 compression coding, MPEG-2 compression coding, slice error transmission and stop error transmission. There are 135 distorted videos in total, and the video resolution is 1920×1080. IVPL Video Quality Database [10] has 10 source videos, and four kinds of distortion are added: correction compression coding, H.264 compression coding, IP transmission, and MPEG-2 compression coding. There are 128 distorted videos in total, and the video resolution is 1920×1088. There are 12 source videos in CSIQ Video Quality Database [11], and six distortions have been added: JPEG compression coding, H.264 compression coding, HEVC compression coding, wavelet compression coding, packet loss transmission, and additive white Gaussian noise transmission distortion. There are 216 distorted videos in total, and the video resolution is 832x480.

Today, studio video has reached a higher quality playback standard, and studio video uses YUV4:2:2 format, while the current public video datasets use YUV4:2:0 format, which makes the quality of the public video datasets can not meet the studio quality standard. At present, these four datasets contain fewer types, the types of distortion are not comprehensive, and the video quality is not high. Aiming at the existing problems, this paper creates a video dataset in YUV4:2:2 format with more categories, richer scenes, more comprehensive distortion types, larger scale and higher quality for no-reference video quality assessment research.

Finally, four publicly available NR-VQA algorithms [12] are evaluated on our dataset. Pearson linear correlation coefficient (PLCC) and Spearman's rank correlation coefficient were used coefficient, SROCC) to evaluate the performance of objective video quality assessment algorithms. Experiments show that the data set in this paper can get better results when used in these classical algorithms, which shows the effectiveness of the data set in this paper.

2. Database Creation process

In the experiment, a new video dataset was constructed, which consisted of 1288 video sequences, including

14 uncoded and compressed source videos and 91 distorted video sequences corresponding to each source video. These video sequences were used for subjective scoring experiments [13].

2.1. *Source Video*

All the source videos provided in this paper are raw data without any post-processing, and do not contain audio information [14].

These 14 categories of videos integrate the characteristics of other datasets and are designed according to the scene, situation, and personnel complexity, and the scene of rainy day is added. The video content covers most scenarios of network video, and compared with the existing data sets, it is more suitable for the research of network video quality assessment. Table 1 shows the content comparison between the data set in this paper and the existing data set. Example frames of 14 source video sequences of different scenes are shown in Figure 1. All videos have a duration of 10s.

At present, the video sequence format is used in 4:2:0 format, the horizontal and vertical direction of the chrominance will be halved, and the 4:2:2 video format video, only in the horizontal direction of the chrominance component halved, so 4:2:2 playback video quality is higher than 4:2:0, and 4:2:2 is in line with the quality requirements of the studio. The 4:2:2 format video is more suitable for the research of high-quality network video. Therefore, this paper applies the 4:2:2 format to the self-created data set, and the comparison between the data set and the source video format of the existing data set is shown in Table 2.

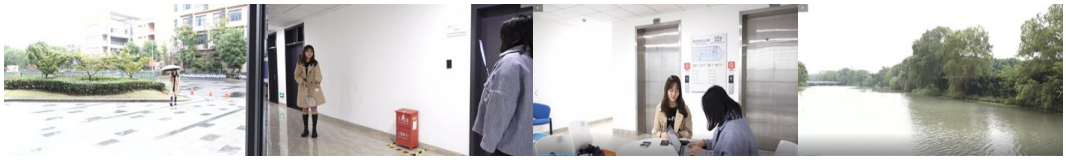
Table 1. Data set content comparison

	Indoor							Outdoor		
	daytime					night		daytime		
	Rain day	Clear day				Clear day		Rain day	Clear day	
	Rain, people, umbrella sports	Road traffic street lights alternate non-motor sky leaves	College building two people meet to talk flower bed	Bridge, lake, people, leaves, sky duck	Traffic flow	Intersection people, street lights alternate non-motor vehicles	Bridge, tree branches, lake people, sports	A man wandered in from the rain	People, badminton, fire hydrants, sports	People, textbooks, desks, chairs, computers, AC elevators, flashing lights
LIVE [8]	/	√	/	√	√	/	/	/	/	/
IVPL [9]	/	/	√	√	√	√	/	/	/	/
CSIQ [10]	/	/	/	√	√	/	/	/	/	√
VQEG [11]	/	/	/	√	√	/	/	/	√	√
Article	√	√	√	√	√	√	√	√	√	√

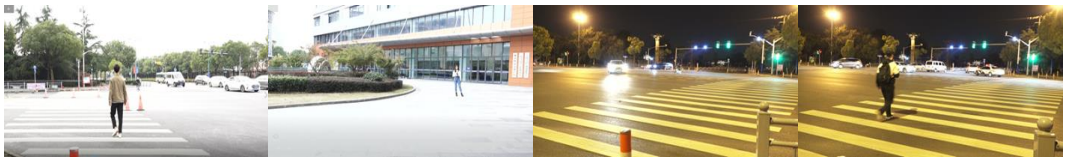
Table 2. Comparison of self-generated data sets with existing video data sets

	LIVE [8]	VQEGHD5 [9]	IVPL [10]	CSIQ [11]	Article
Original video	10	9	10	12	14
Distorted video	150	135	128	216	1274

Type of distortion	4	4	4	6	6
Resolution ratio	768x432	1920x1080	1920x1088	832x480	1920x1080
Source Video	YUV420	YUV420	YUV420	YUV420	YUV422

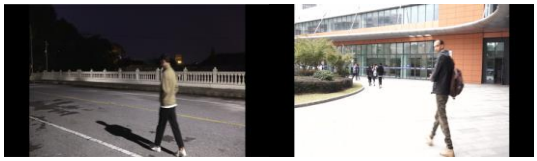


(a) (b) (c) (d)



(e) (f) (g) (h)

(i) (j) (k) (l)



(m)(n)

Fig. 1. sample image of source video sequence content((a) girl walking on a rainy day; (b) Playing badminton; (c) Two people sit down and communicate; (d) During the day, leaves and water move with the wind; (e) crossing zebra crossings at traffic lights; (f) Face-to-face communication between the two; (g) Night traffic; (h) Alternating outdoor night traffic lights crossing zebra crossings; (I) the night leaves and the lake move with the wind; (j) daytime traffic at traffic light junctions; (k) Walking alone in the daytime with the leaves moving in the wind; (l) Running outdoors at night; (m) Walking alone at night with the leaves moving in the wind; (n) a group of people hurried walking in front of the college)

2.2. Distortion operation

The experimental design created video sequences with six categories of distortion from each source video sequence using six different distortion processes. These include downsampling, demosaicing, basic feature distortion, noise, filtering, and X264 compression. Since the main goal of the dataset designed in this paper is to target the online video viewing experience, the selected distortion is considered to be the most common and prominent distortion encountered in online videos [15].

1) Downsampling: The source video sequence collected in this paper is in 4:2:2 format. The color differences in these source video sequences are downsampled to the 4:2:0 format. The color difference in the first row is added to the corresponding color difference in the second row and averaged.

2) Mosaic: using Adobe Premiere Pro 2020 version software, in the current network video playback, the most common is the rectangular Mosaic, so this paper designs the use of rectangular Mosaic, and the placement of different shape areas. Mosaic affects the recognition of faces, the recognition of traffic lights at intersections, the cognition of building annotations, and the overall feeling of video Mosaic in bright places at night. Each frame of the video is coded locally by taking the average of pixels. According to the different content of the video, the coding position, coding range and the size of the Mosaic are selected.

3) Basic features. The Fengyun video Converter 2021 version is used to modify the saturation, brightness and contrast in the basic characteristics of the video sequence. Ten distortion sequences in units of ten are created for saturation, ten distortion sequences in units of ten are created for brightness, and nine distortion sequences in units of ten are created for contrast. Change the frame rate and resolution of the source video sequence to 1920x1080p@25fps 4:2:2 format video. The frame rate was changed by fixed resolution, and 10 distortion sequences were designed. The resolution was changed at a fixed frame rate and 15 distortion sequences were designed.

4) Noise: salt and pepper noise is a common noise in digital images, which is shown as random black and white pixels on the image. It is a noise caused by signal pulse strength. The most common noise in network video is salt and pepper noise, and the video sequence is composed of frame by frame digital images. The YUV video is read by frame, and salt and pepper noise is added to each frame, and finally saved into YUV format video by frame. Specify the value range of SNR as [0,1], fix its value as 0.5, calculate the total number of pixels, randomly obtain each pixel position with noise, specify the pixel value as 0 or 255, repeat the operation until all pixel operations are completed, and finally output the video sequence with noise.

5) Blur: Median filtering has a good filtering effect on impulse noise, especially when filtering noise, it can protect the edge of the signal from being blurred [16]. Because the shape and size of the window have great influence on the filtering effect, different image contents often use different window shapes and sizes. According to the characteristics of the video sequence in this paper, the video content contains many images of slowly changing objects with long contours, so it is most appropriate to use a square window, and the size of the window should be no

6) encoding and decoding: Currently, H.264 is the most widely accepted and used video compression standard. The video is compressed and encoded using X264 to generate MP4 format file. The codec distortion is achieved using the criteria described previously, by varying CRF and QP fixed values. Because the distortion degree will be different when the CRF and QP values are different, firstly, the QP value is fixed as 18, and the CRF value is changed, according to the CRF value range of [0,51], the 16 levels of X264 compression are fixed. Secondly, the CRF value is fixed as the default value of 23, and the QP is changed. However, since CRF takes motion into account, the setting of QP will also be different according to the motion situation in the video sequence. For severe motion, the QP will be set a little higher, and for less severe motion, the QP will be reduced accordingly.

2.3. *Subjective test experiment*

2.3.2. Scoring design. According to the BT.500technical standard issued by the International Telecommunication Union (ITU), this paper designs a scoring bar [17], as shown in Figure 2. The single stimulus method was selected, and the five-point scale of the single stimulus method was changed to a hundred-point scale according to the problem of the large number of data sets in this paper, thus reducing the computational difficulty, as shown in Table 3. After the observers had watched all the videos, the authors read the data and recorded it.

Table 3. Scoring standard of data set in this paper

Fractional segmentation	marking standard
0-20	very bad
20-40	bad
40-60	ordinary
60-80	goog
80-100	Very goog

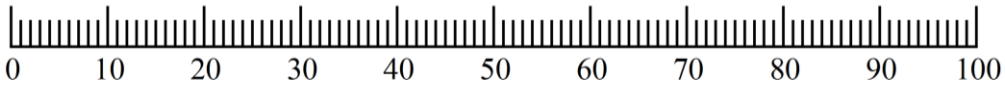


Fig. 2. scoring strip

2.3.1. Experimental Environment Design. The evaluation lighting conditions of the subjective test experiment meet the requirements of the laboratory environment of the ITU-BT.500 standard [18]. The experiment uses Dell Lingyue display screen, the aspect ratio of the display is 16:9; The maximum deviation Angle of the observer from the normal of the screen center was 31°. The viewing distance is 2m-4m. The test sequences are played and stored in original YUV sequence and MP4 format using a high-performance workstation.

Twenty-nine observers participated in the subjective testing experiment, with 13 females and 16 males. Twenty-five of the observers had not participated in any video processing related work and were called "non-experts", while the remaining four were engaged in research in image and video processing and were regarded as "experts". All the observers had normal vision.

When watching videos, there was a 5s interval between the two videos, 45min for a cycle, 180 videos were watched and 30min rested. Every two cycles are one stage, and in each stage, five videos are arranged as training videos to stabilize the viewpoint of observers [18]. After the video demonstration, the observer drew the score in the corresponding position of the scoring bar, and proceeded to play the next video after scoring a score [19].

3. Subjective score processing

3.1. *Data processing*

Data processing after subjective scoring includes three steps. First, the scoring bar score is read, recorded, and the average value of the most original data is calculated $\bar{u}_{ij}$.

$$\bar{u}_{ij} = \frac{1}{N}\sum_{i=1}^{N} u_{ij} \tag{1}$$

Where, u_{ij} is the scoring of the i observer for j video sequences, and N is the total number of observers.

Second, when presenting test results, all averages should have an associated confidence interval, which is derived from the standard deviation and size of each sample, using a 95% confidence interval [17]. Formula:

$$[\bar{u}_{ij} - 1.96\frac{S_{ij}}{\sqrt{N}}, \bar{u}_{ij} + 1.96\frac{S_{ij}}{\sqrt{N}}] \tag{2}$$

The standard deviation S_{ij} is given by:

$$S_{ij} = \sqrt{\sum_{i=1}^{N}\frac{(\bar{u}_{ij} - u_{ij})^2}{(N-1)}} \tag{3}$$

Abnormal observers were excluded according to ITU-R BR 500.11 [20]. The scores for each video were tested with $\beta 2$ for normal distribution.

$$\beta_{2j} = \frac{m_4}{(m_2)^2} \tag{4}$$

$$m_x = \frac{\sum_{i=1}^{N} (z_{ij} - \bar{z}_{ij})^x}{N_j} \tag{5}$$

$$z_{ij} = \frac{(u_{ij} - \bar{u}_{ij})}{S_{ij}} \tag{6}$$

Where, m_x is x the order matrix, z_{ij} is the standard score, and $\bar{z}_{ij}$ is the average of the standard scores.

If β_{2j} belongs to [2,4], then the score belongs to the normal distribution. If video j is normally distributed, then:

$$if: z_{ij} \geq \bar{u}_{ij} + 2S_{ij}, then: P_i = P_i + 1 \tag{7}$$

$$if: z_{ij} \leq \bar{u}_{ij} - 2S_{ij}, then: Q_i = Q_i + 1 \tag{8}$$

Where P_i, Q_i are the counters associated with each observer.

If videoj is not normally distributed, then:

$$f: z_{ij} \geq \bar{u}_{ij} + \sqrt{20} S_{ij}, then: P_i = P_i + 1 \tag{9}$$

$$if: z_{ij} \leq \bar{u}_{ij} - \sqrt{20} S_{ij}, then: Q_i = Q_i + 1 \tag{10}$$

The observer's score needs to be rejected if the following two conditions hold [20].

$$\frac{P_i + Q_i}{N} > 0.05, \left| \frac{P_i - Q_i}{P_i + Q_i} \right| < 0.3 \tag{11}$$

After the experiment rejected the scores of six observers, the average $\bar{u}'_{ij}$ of the remaining values was calculated according to Equation (1).

Thirdly, after removing the scores of 6 observers, the Laida criterion was used to eliminate the outliers of the overall data [21]. When a score value exceeds this interval, it is not a random error but a gross error, and it will be eliminated.

$$| u_{ij} - \bar{u}'_{ij} | > 3S_{ij} \tag{12}$$

Finally, according to formula (1), the mean value of the score $\bar{u}''_{ij}$ after excluding the abnormal values was calculated. Of the 29 observers, the data of 1 expert and 5 non-experts were eliminated, and the data values of 3 experts and 20 non-experts were finally retained.

Figure 3 is the MOS case distribution for all the converted videos, and Figure 4 shows the MOS distribution for each individual video distortion category. The abscissa represents the score range, and the ordinate represents the number of videos in the score range. The MOS distribution of the source videos is somewhat like a Gaussian distribution, however, the distorted videos exhibit different distribution shapes as they reflect different classes and levels of distortion.

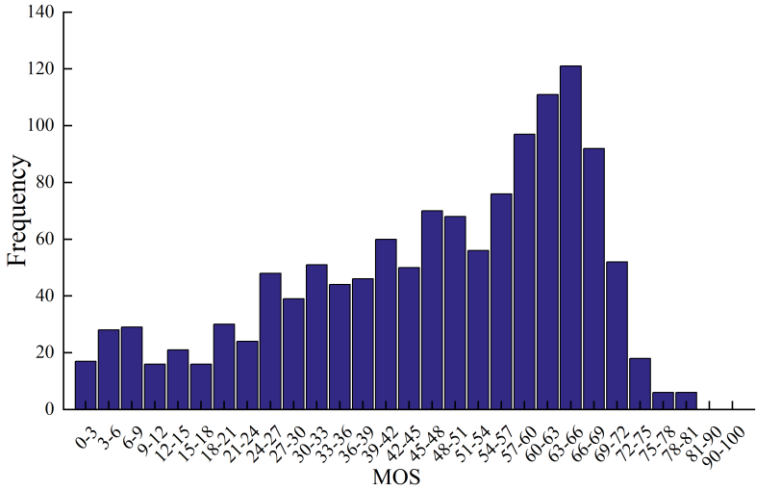


Fig. 3. MOS distribution of all videos

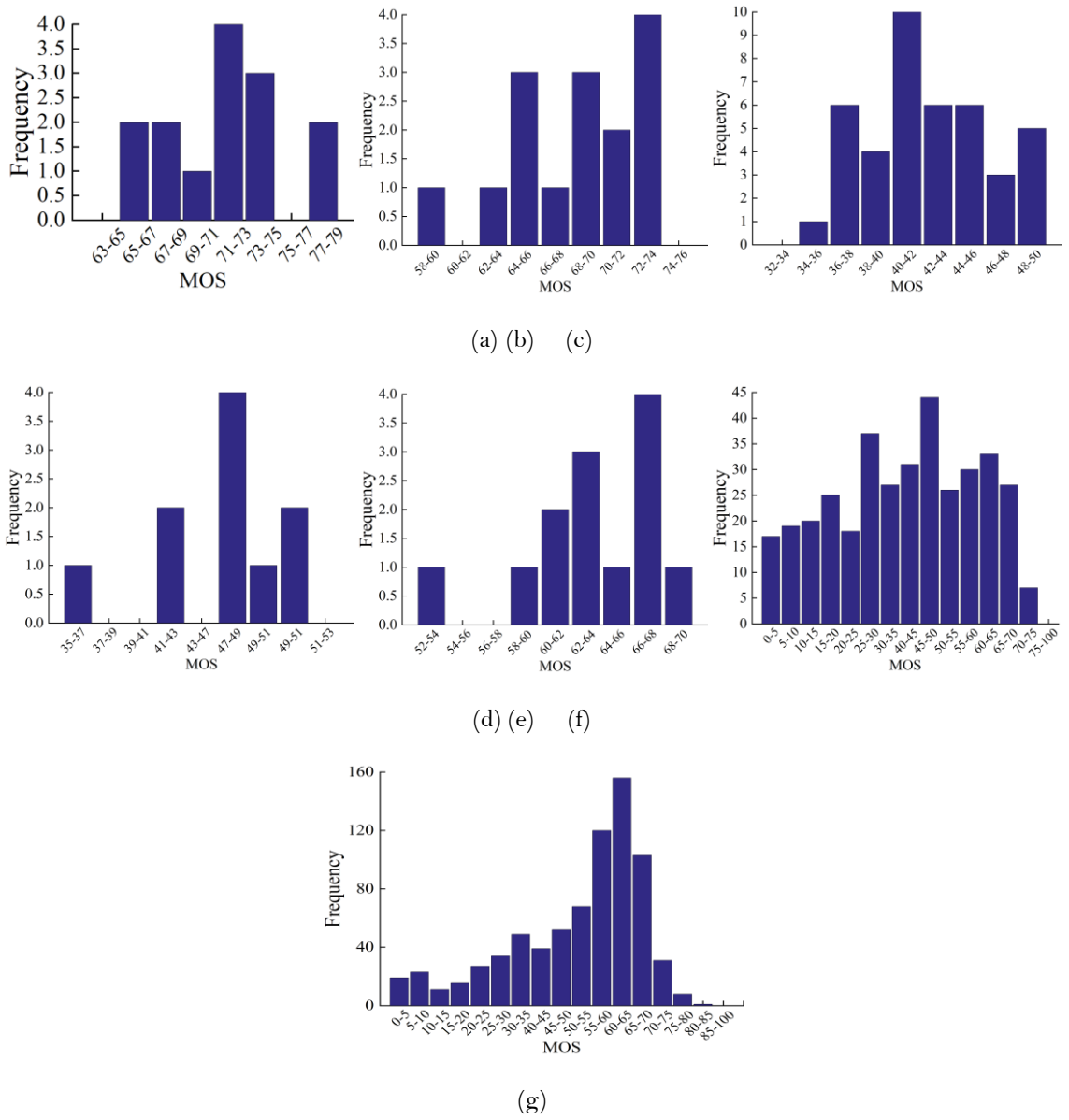


Fig. 4. MOS distribution of source video and distorted video ((a) Source video; (b) Down-sampling; (c) Mosaics; (d) noise; (e) fuzzy; (f) Basic characteristics; (g) codec)

3.2. Subjective score data performance

Table 4 and Table 5 show the measure of observer rating agreement for each of the different distortion types. The observers were again divided into two groups, resulting in two sets of data. These tables list the PLCC and SROCC computed for the entire dataset and for each distortion type. It can be observed that SROCC is slightly lower than PLCC, probably because observers have difficulty providing correctly ranked ratings for videos of very similar quality. However, predictions are generally made in a linear fashion, and the results show very high inter-observer agreement across all distorted video types. Figure 5 shows the distribution of time and space complexity of the data set in this paper, and it can be seen that the source video sequences are widely distributed, and different test sequences can be selected according to requirements.

Table 4. PLCC of maximum, median and minimum values of two groups of data

	ALL	Downsampling	Mosaic	Noise	Dim	Essential feature	Compress
max	0..977	0.978	0.981	0.998	0.967	0.927	0.987
mid-value	0.964	0.976	0.964	0.987	0.956	0.911	0.978
min	0.931	0.968	0.933	0.979	0.921	0.899	0.954

Table 5. SROCC of maximum, median and minimum values of two groups of data

	ALL	Downsampling	Mosaic	Noise	Dim	Essential feature	Compress
max	0.969	0.925	0.971	0.966	0.924	0.911	0.947
mid-value	0.955	0.896	0.932	0.932	0.918	0.897	0.955
min	0.932	0.828	0.901	0.922	0.869	0.833	0.932

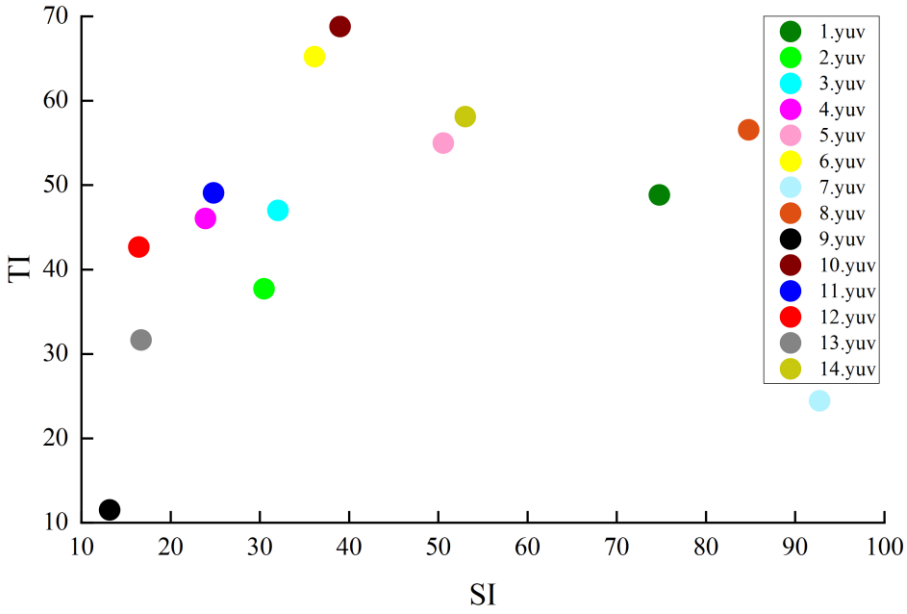


Fig. 5. Source video complexity distribution

4. Experimental results and analysis

4.1. Benchmark experiments

Several publicly available deep learning-based no-reference objective assessment algorithms are evaluated on the newly created dataset to show the effectiveness of the new dataset. Given the MOS value, the no-reference video quality assessment algorithms are compared.

The scores predicted by various NR models were compared with the subjective scores, and four mainstream NR VQA methods based on deep learning were used for training and testing. They are VSFA [22], RAPIQUE [23], CNN-TLVQM [24], and MLSP [25] algorithms. PLCC and SROCC are used to evaluate the performance of NR VQA algorithm.

In the VSFA [22] model, researchers apply a pre-trained image classification neural network as a deep feature extractor, and then use a gated recurrent unit and subjectively inspired temporal pooling layer to integrate the deep features. In CNN-TLVQM [23], researchers use convolutional neural networks to extract spatial features of video frames and combine hand-crafted statistical temporal features with spatial features acquired by convolutional neural networks trained for image quality assessment by universal transfer learning. The RAPIQUE [24] model combines the advantages of the quality-aware scene statistics function and the semantic-aware deep convolution function, and designs the first general and efficient spatio-temporal band-pass statistical model for video quality prediction. Finally, MLSP [25] relies

on multi-level spatial pooling for deep feature extraction, and the MLSP model is suitable for large-scale training.

4.2. Experimental results and analysis

In the experiments, each of these algorithms is tested in 1000 random training sessions, during which the video sequence is split into two non-overlapping parts. Of these, 80% of the data is used for training and 20% for testing. The values of PLCC and SROCC are also calculated. Table 6 and Table 7 show the performance indicators of the four VQA models on the source video sequences and video sequences with different distortions. It can be seen that the SROCC of all video sequences in the proposed dataset is above 0.7. In addition, the PLCC results among different models also remain highly consistent. To further visualize the properties of this video dataset, in Figure 6 a scatter plot between MOS values and predicted scores for subjective experiments is depicted. Through FIG. 6, it can be observed that most of the points are located near the reference line, which indicates that the four VQA models achieve good performance on the video dataset designed in this paper.

In order to analyze the characteristics of the data set in this paper more comprehensively, the generalization of the data set in this paper is verified by cross-data sets. The proposed dataset and VQEGHD5 dataset are used for cross-dataset testing in the experiments. The results of the four VAQ models evaluated in the cross-dataset experiments are shown in Table 8 and Table 9, and it can be seen from the tables that the generalization ability of the proposed dataset is good.

Table 6. PLCC of the compared NR VQA models (The scores of the top performing algorithm are boldfaced)

VQA model	all	Downsampling	Mosaic	Noise	Dim	Essential feature	Compress
VSFA [22]	0.733	0.767	0.666	0.639	0.797	0.753	0.752
RAPIQUE [23]	0.794	0.691	0.551	0.745	0.619	0.853	0.819
CNN-TLVQM [24]	0.803	0.807	0.698	0.829	0.715	0.829	0.877
MLSP [25]	0.819	0.814	0.752	0.801	0.821	0.844	0.855

Table 7. SROCC of the compared NR VQA models (The scores of the top performing algorithm are boldfaced)

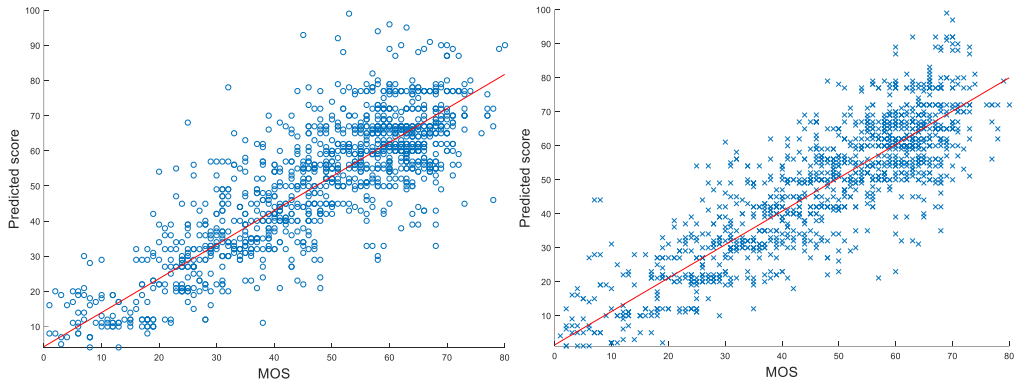
VQA model	all	Downsampling	Mosaic	Noise	Dim	Essential feature	Compress
VSFA [22]	0.701	0.701	0.636	0.712	0.768	0.701	0.667
RAPIQUE [23]	0.703	0.635	0.640	0.689	0.699	0.815	0.804
CNN-TLVQM [24]	0.774	0.761	0.602	0.801	0.691	0.812	0.829
MLSP [25]	0.799	0.793	0.716	0.783	0.759	0.809	0.811

Table 8. Evaluation on cross-datasets

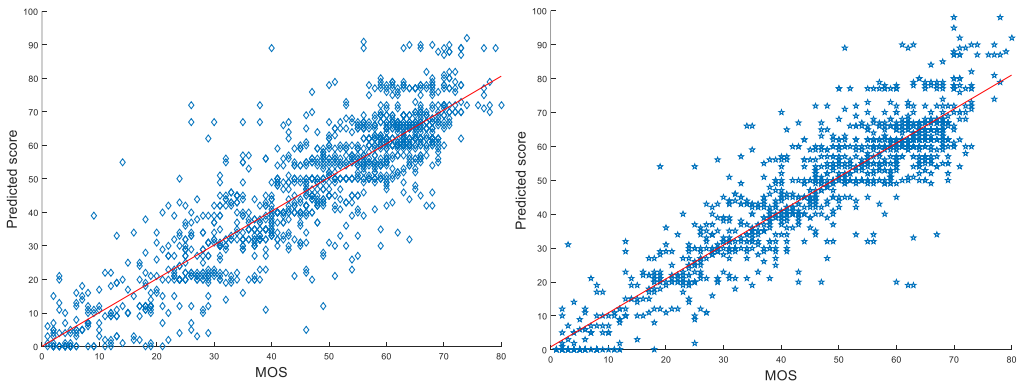
training set	Article	
test set	VQEGHD5	
MODEL	PLCC	SROCC
VSFA [22]	0.705	0.675
RAPIQUE [23]	0.711	0.687
CNN-TLVQM [24]	0.735	0.709
MLSP [25]	0.799	0.794

Table 9. Evaluation on cross-datasets

training set	VQEGHD5	
test set	Article	
MODEL	PLCC	SROCC
VSFA [22]	0.695	0.651
RAPIQUE [23]	0.671	0.657
CNN-TLVQM [24]	0.719	0.699
MLSP [25]	0.782	0.752



(a) (b)



(c) (d)

Fig. 6. Scatter plot of predicted scores and MOS scores of output from different models (a) VSFA; (b) RAPIQUE; (c) CNN-TLVQM ; (d) MLSP

5. Conclusion

This paper creates a new high-definition video quality dataset, which is the first dataset in 4:2:2 format for web video quality research. The dataset has the advantages of rich video content is diverse distortion

complexity, and wide distribution of spatio-temporal information, Including 1288 videos and 6 distortion types. In this paper, four mainstream VQA algorithms are used to verify the effectiveness of this dataset. In the future work, the laboratory will focus on deep learning network models to improve the performance of this dataset.

The dataset created in this paper can be a convenient platform for researchers and developers to build algorithms and a common test set for performance comparison. At the same time, a subset of the information in this dataset can be used, for example, for daytime-only scenes, and researchers can choose different video sequences according to their needs. Data gathering continuously updated, synchronous update in: https://pan.baidu.com/s/1Yo4HfUMp1IOYJr4hwf6Uuw.

Funding

Foundation items: National Key R&D Program of China: Media convergence network intelligent collaborative computing and cross-network collaborative transmission (2019YFB1802700).

References

[1] ZHU Ze,SANG Qingbing,ZHANG Hao. No Reference Video Quality Assessment Based on Spatio-Temporal Features and Attention Mechanism[J].Laser &Optoelectronics Progress, 2020,57(18):351-359.

[2] KORSHUNOV P and EBRAHIMI T. UHD video dataset for evaluation of privacy[C]// The Institute of Electrical and Electronic Engineers(IEEE). Proceedings of Quality of Multimedia Experience. 2014:232-237[2022-07-21].

[3] XING F C, WANG Y G, WANG H P, et al. DVL2021: An ultra high definition video dataset for perceptual quality study[J]. Journal of visual communication and image representation 2022,28(4):351-363

[4] YE Jihua, LIU Yan, LI Hanxi, et al. Construction of a Facial Expression Dataset in Natural Scenes[J].Journal of Data Acquisition and Processing, 2019,34(01):58-67.

[5] LEYVA R,SANCHEZ V, LI C J, The LV dataset: A realistic surveillance video dataset for abnormal event detection[C]// The Institute of Electrical and Electronic Engineers(IEEE). Proceedings of Biometrics and Forensics , 2017:1-6[2022-07-22].

[6] GUO Zhipeng. Research on universal no reference video quality assessment method[D].Tianjin: Tianjin University,2017.

[7] ZRRMAN E,AKAR G B,KONUK B, et al. A comparative study on no-reference Video Quality Assessment metrics[C]// The Institute of Electrical and Electronic Engineers(IEEE). Proceedings of Signal Processing and Communications Applications Conference ,2014: 1774-1777[2022-07-23]

[8] SESHADRINATHAN K, SOUNDARARAJAN R, BOVIK A C, et al. Study of subjective and objective quality assessment of video[J]. IEEE transactions on Image Processing 2010,19(6):1427-1441.

[9] Video Quality Experts Group. Report on the validation of video quality models for high definition video content [EB/OL].(2010-06-04)[2022-07-06]. http://www.its.bldrdoc.gov/vqeg/projects/hdtv/hdtv.aspx,2010.

[10] ZHANG F, LI S, MA L, et al. IVP subjective quality video database[EB/OL].(2011-12-15)[2022-07-06]. http://ivp.ee.cuhk.edu.hk/research/database/subjective, 2011.

[11] VU P V, CHANDLER D M. ViS3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices[J]. Journal of Electronic Imaging, 2014, 23(1): 013016-013016.

[12] TU Z Z, WANG Y L, N. Birkbeck, BIRKBECK N, et al. UGC-VQA: Benchmarking Blind Video Quality Assessment for User Generated Content[J]. IEEE Transactions on Image Processing,2021 15(4): 4449-4464.

[13] MA J,LI S Y,YOU Z X, et al. A New 3D Video Database for Stereoscopic Video Comfortable Prediction[C]// The Institute of Electrical and Electronic Engineers(IEEE). Proceedings of Signal and Image Processing, 2020: 463-467[2022-07-22]

[14] CHEN M Y,ALREGIB,JUANG B H, A new 6D motion gesture database and the benchmark results of feature-based statistical recognition[C]// The Institute of Electrical and Electronic Engineers(IEEE).Proceedings of Emerging Signal Processing Applications. Palms Springs:IEEE,2012:131-134[2022-07-05].

[15] GU Yuse, JIANG Qiuping, SHAO Feng, et al. A real-world quality evaluation dataset for enhanced underwater images[J].

[16] SUN Han,LIU Zeshan, LIN Yuhan. A Survey of Salient Object Detection Based on Deep Learning[J]. Journal of Data Acquisition and Processing, 2023,38(01):21-50.

[17] International Telecommunication Union. ITU-R Rec.BT.500.13-2012 Methodology for the subjective assessment of the quality of television pictures[S]. Geneva, 2012.

[18] SHANG Z X, EBENEZER J P, WU Y J, et al. Study of the subjective and objective quality of high motion live streaming videos[J]. IEEE Transactions on Image Processing, 2022,31:1027-1041.

[19] HUANG Ruiyu. Research on subjective video quality evaluation method.[D].Xian: Xidian University,2010.

[20] International Telecommunication Union. ITU-R BT.500.11-2002 Methodology for the subjective assessment of the quality of television pictures[S]. Geneva 2002.

[21] STREIJL R C, WINKLER S, HANDS D S. Perceptual Quality Measurement-Towards a More Efficient Process for Validating Objective Models[J].IEEE Signal Processing Magazine, 2010,27(4):136-140.

[22] LI D Q, JIANG T T, JIANG M, Quality assessment of in-the-wild videos[C]// Proceedings of 27th ACM International Conference on Multimedia, 2019:2351-2359[2022-07-22].

[23] TU Z Z, YU X X, WANG Y L, et al.RAPIQUE: Rapid and accurate video quality prediction of user generated content[J] IEEE Open Journal of Signal Processing. Signal Process 2021,23(6): 425–440.

[24] KORHONEN J, SU Y C, YOU J Y, Blind natural video quality prediction via statistical temporal features and deep spatial features[C]// Proceedings of 28th ACM International Conference on Multimedia, 2020:3311-3319[2022-07-23].

[25] GöTZ-HAHN F, HOSU V,LIN H H,et al. KonVid-150k: A Dataset for No-Reference Video Quality Assessment of Videos in-the-Wild[J] IEEE Access 2021,5(5):72139-72160.