

# VDPVE: VQA Dataset for Perceptual Video Enhancement

Yixuan Gao<sup>1\*</sup>    Yuqin Cao<sup>1\*</sup>    Tengchuan Kou<sup>1\*</sup>    Wei Sun<sup>1</sup>    Yunlong Dong<sup>2</sup>  
 Xiaohong Liu<sup>2†</sup>    Xiongkuo Min<sup>1†</sup>  
 Guangtao Zhai<sup>1†</sup>

Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University<sup>1</sup>  
 John Hopcroft Center, Shanghai Jiao Tong University<sup>2</sup>

{gaoyixuan, caoyuqin, 2213889087, sunguwei, dongyunlong, xiaohongliu,  
 minxiongkuo, zhaiguangtao}@sjtu.edu.cn <sup>\*†</sup>

## Abstract

Recently, many video enhancement methods have been proposed to improve video quality from different aspects such as color, brightness, contrast, and stability. Therefore, how to evaluate the quality of the enhanced video in a way consistent with human visual perception is an important research topic. However, most video quality assessment methods mainly calculate video quality by estimating the distortion degrees of videos from an overall perspective. Few researchers have specifically proposed a video quality assessment method for video enhancement, and there is also no comprehensive video quality assessment dataset available in public. Therefore, we construct a Video quality assessment dataset for Perceptual Video Enhancement (VDPVE) in this paper. The VDPVE has 1211 videos with different enhancements, which can be divided into three sub-datasets: the first sub-dataset has 600 videos with color, brightness, and contrast enhancements; the second sub-dataset has 310 videos with deblurring; and the third sub-dataset has 301 deshaked videos. We invited 21 subjects (20 valid subjects) to rate all enhanced videos in the VDPVE. After normalizing and averaging the subjective opinion scores, the mean opinion score of each video can be obtained. Furthermore, we split the VDPVE into a training set, a validation set, and a test set, and verify the performance of several state-of-the-art video quality assessment methods on the test set of the VDPVE.

## 1. Introduction

Due to the rapid development of social media applications, the demand for higher video quality is also increas-

ing. Therefore, many video enhancement methods have emerged recently. These video enhancement methods aim to improve video quality from different aspects, such as color, contrast, brightness, and stability, to give people a more comfortable viewing experience. However, the quality of videos enhanced by different methods is distinct. Therefore, it is non-trivial to propose a promising method that enables the quality assessment of enhanced videos.

Traditional video quality assessment (VQA) methods assess the video quality mainly dependent on the distortions in video. Subjective VQA refers to rating the quality of videos by inviting many subjects to participate in a subjective experiment. Objective VQA refers to calculating video quality by approximating the perceptions of people about video quality through various computational models. For subjective VQA, researchers first process the obtained subjective opinion scores, such as screening invalid subjects and normalizing scores, and then average all the subjective opinion scores for each video to get the mean opinion score (MOS). For example, the LIVE Video Quality Database [31], proposed by Seshadrinanathan *et al.* in 2010, is the most famous VQA dataset. This dataset contains 10 original videos with different contents and their corresponding distorted videos that are compressed and transmitted. A total of 38 subjects were invited to participate in the subjective experiment. In addition, other well-known VQA datasets include CSIQ [42], MCL-V [19], and MCL-JCV [44]. The distorted videos in these datasets are synthetically distorted and cannot represent videos with authentic distortions. Therefore, many user-generated content (UGC) VQA datasets were proposed. For example, the first UGC VQA dataset is the Camera Video Database (CVD2014) [28]. This dataset includes 234 authentically distorted videos captured by 78 different video capture devices. In addition, the authors in [11] proposed the KoNViD-1k dataset. This dataset consists of 1200 videos, and 642 subjects were invited to rate these

\*Equal contribution.

†Corresponding authors.

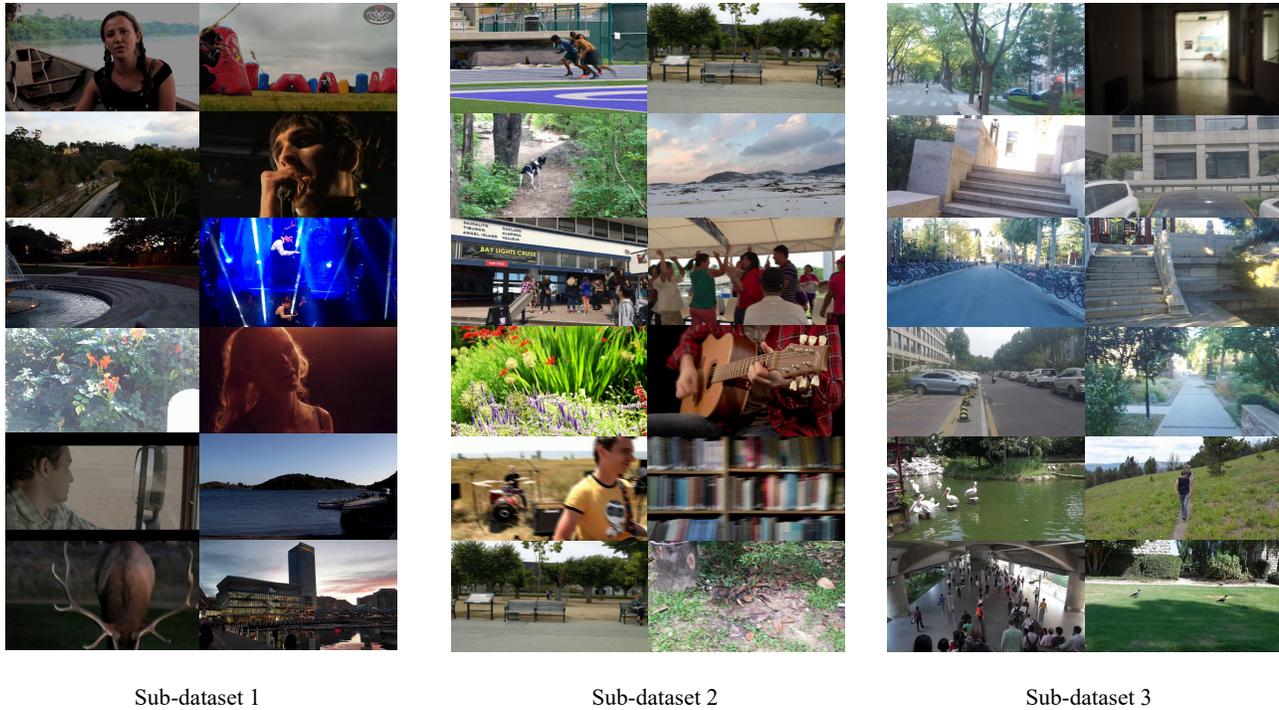


Figure 1. Sample frames of representative enhanced videos in the three sub-datasets of VDPVE.

videos. The video quality obtained by subjective VQA is usually considered to be consistent with people’s perceptions.

However, subjective VQA has some disadvantages, such as complex preparation, high cost, and time-consuming, which makes it difficult to apply in practice [3, 6, 7, 51]. Objective VQA can avoid the disadvantages of subjective VQA, which can be specifically divided into full-reference (FR) VQA [1, 38], reduced-reference (RR) VQA [36], and no-reference (NR) VQA [24, 30, 37, 49, 52]. The FR VQA method requires all the information about the reference video to calculate the quality of the distorted video, for example, SSIM [47] and ST-VMAF [1]. And Kim *et al.* [13] proposed a Deep Video Quality Assessor (DeepVQA), which predicts video quality by learning a spatiotemporal visual sensitivity map. The RR VQA method only requires partial information about the reference video to calculate the quality of the distorted video [36]. The NR VQA method, on the other hand, does not require any information about the reference video to calculate the quality of the distorted video. For example, the well-known V-BLIINDS [30] used a natural bandpass spatio-temporal video statistics model to calculate video quality. Recently, Sun *et al.* [37] proposed a VQA method developed specifically for UGC videos. This method predicts the quality of UGC videos by training an end-to-end spatial feature extraction network to learn quality-aware spatial feature representations directly

from the original pixels of a video frame. Compared with subjective VQA, objective VQA is more convenient, faster, and easier to apply in the industry. However, objective VQA methods often require VQA datasets constructed by subjective VQA methods to verify their effectiveness and feasibility. Therefore, a successful VQA dataset can promote the development of objective VQA methods.

At present, few researchers have proposed objective VQA methods for video enhancement. The reason might be that there is no successful and effective VQA dataset for video enhancement. To break through this limitation, this paper constructs a VQA Dataset for Perceptual Video Enhancement (VDPVE) with a large number of videos and a wide range of enhancements. Specifically, the VDPVE includes 1211 enhanced videos, which can be divided into three sub-datasets: the first sub-dataset consists of 600 videos with color, brightness, and contrast enhancements; the second sub-dataset consists of 310 videos with deblurring; and the third sub-dataset consists of 301 deshaked videos. We also carried out a subjective experiment and obtained 21 subjective opinion scores for each enhanced video. During data processing, we reject an invalid subject and normalize the subjective opinion scores. Finally, we can obtain the mean opinion score (MOS) for each enhanced video. The proposed VDPVE successfully fills in the blank of such kind of dataset in the field of video processing. It can not only promote the development of VQA

methods for video enhancement, but also improve the performance of video enhancement methods to a certain extent.

The rest of the paper is organized as follows. In Section 2, details of the proposed VDPVE are provided. Section 3 describes the subjective experiment in detail. In Section 4, we test the performance of some VQA methods on the VDPVE. Finally, we conclude this paper in Section 5.

## 2. Dataset Construction

The VDPVE includes 1211 enhanced videos in total. Specifically, they can be divided into three sub-datasets. The first sub-dataset consists of 600 videos with color, brightness, and contrast enhancements. The second sub-dataset consists of 310 videos with deblurring. The third sub-dataset consists of 301 deshaked videos. Figure 1 shows sample frames of representative videos in these three sub-datasets. Next, we would like to introduce the construction process of these three sub-datasets in detail.

### 2.1. Sub-dataset 1:

The 600 videos with color, brightness, and contrast enhancements are obtained by using six enhancement methods on 79 original videos selected from four datasets.

#### 2.1.1 Original Video

These 79 original videos are selected from the following four datasets:

- LIVE-Qualcomm [9]: the dataset contains 208 videos in total. These videos have a duration of 15s and are subject to one of the following six distortions: artifacts, color, exposure, focus, sharpness, and stabilization. These videos were captured by the following eight mobile devices: Samsung Galaxy S5, Samsung Galaxy S6, HTC One VX, Apple iPhone 5S, Nokia Lumia 1020, LG G2, Samsung Galaxy Note4, and Oppo Find 7, which are widely used to capture videos. All videos have a resolution of  $1920 \times 1080$ .
- V3C1 [2]: the dataset is the first partition of the Vimeo Creative Commons Collection (V3C) [29], which is designed to represent real network videos with diverse video contents. There are 7475 videos in the V3C1 dataset, which were selected from videos uploaded to Vimeo from 2006 to 2018. The resolutions of most videos are  $1280 \times 720$  and  $1920 \times 1080$ .
- KoNViD-1k [11]: the dataset is composed of 1200 videos, which were all obtained from the large public video dataset YFCC100m [39] through “fair sampling”. The video length in KoNViD-1k is 8s, and the following three frame rates are used for encoding: 24 frames per second (FPS), 25 FPS, and 30 FPS, which

correspond to 27%, 5%, and 68% of the videos, respectively. There are 12 resolutions in total, and the resolution of most videos is  $1280 \times 720$ , followed by  $1920 \times 1080$ .

- LIVE-VQC [35]: the dataset contains 585 videos with unique content that were captured by 101 different devices (most of them are smart phones) and has extensive authentic distortions. The main resolutions are  $404 \times 720$ ,  $1024 \times 720$ , and  $1920 \times 1080$ . The average video duration is 10s.

We select 7, 26, 13, and 33 original videos from the LIVE-Qualcomm, V3C1, KoNViD-1k, and LIVE-VQC datasets, respectively. We set the resolution of all original videos to  $1280 \times 720$ , and the duration is 8s or 10s.

#### 2.1.2 Enhancement Method

We use eight enhancement methods to enhance the color, brightness, and contrast of these original videos, including ACE [8], AGCCPF [10], BPHEME [43], MBLLEN [25], SGZSL [55], DCC-Net [53], and two commercial softwares: CapCut and Adobe Premiere Pro. Specifically, ACE [8] is an effective color correction and enhancement method based on a simple model of the human visual system. AGCCPF [10] uses the gamma correction of brightness pixels and weighted probability distribution to enhance the contrast and brightness of the image. BPHEME [43], namely brightness preserving histogram equalization with maximum entropy, is a brightness enhancement method. MBLLEN [25], SGZSL [55], and DCC-Net [53] are low-light enhancement methods. CapCut and Adobe Premiere Pro are popular video editing softwares. Users can enhance the color, contrast, and brightness of videos through these two softwares.

For ACE [8], AGCCPF [10], BPHEME [43], MBLLEN [25], SGZSL [55], and DCC-Net [53], we use the default parameters to enhance videos. When using Adobe Premiere Pro to process videos, the brightness and contrast parameters are set to 50.

#### 2.1.3 Enhanced Video

We enhance the color, brightness, and contrast of 79 original videos through the above eight enhancement methods, and obtain a total of 632 enhanced videos. We delete 32 videos with extremely serious distortions and finally obtain 600 videos with color, brightness, and contrast enhancements. Figure 2 shows the sample frames of two original videos and their corresponding enhanced videos in the sub-dataset 1.



Figure 2. Sample frames of two original videos and their corresponding enhanced videos in the sub-dataset 1.

## 2.2. Sub-dataset 2:

The 310 videos with deblurring constitute the second sub-dataset of the VDPVE. These enhanced videos are obtained by using five deblurring methods on 62 original videos selected from four datasets.

### 2.2.1 Original Video

We first selected 22, 24, 10, and 6 videos from the LIVE-VQC [35], Youtube-UGC [46], LIVE-YT-HFR [27], and BVI-HFR [26] datasets, respectively.

- Youtube-UGC [46]: this dataset consists of 1098 20-second videos, which were sampled from YouTube. These videos are not always professionally curated and often suffer from diverse authentic distortions, like blockiness, blur, banding, noise, and so on. The frame rates of videos vary from 15 to 60 FPS.
- LIVE-YT-HFR [27]: this dataset is comprised of 480 10-second videos with 6 different frame rates, obtained from 16 diverse contents. The videos are processed at five compression levels at each frame rate. The frame rates of videos vary from 24 to 120 FPS.
- BVI-HFR [26]: this dataset contains 88 videos with four different frame rates (from 15 to 120 FPS), including 22 120 FPS source sequences that were captured natively using a RED Epic-X video camera.

We select 22 and 24 original videos with a frame rate of 30 FPS from the LIVE-VQC and Youtube-UGC datasets, and select 10 and 6 original videos with a frame rate of 120 FPS from the LIVE-YT-HFR and BVI-HFR datasets. We

set the resolution of all original videos to  $1280 \times 720$ , and the duration is 8s.

Post-processing is then performed on these original videos to obtain the blurred videos. On the one hand, we manually add motion blur to the original videos selected from the LIVE-VQC and Youtube-UGC datasets to obtain the blurred videos through the OpenCV toolbox. On the other hand, for the original videos selected from the LIVE-YT-HFR and BVI-HFR datasets, we take every 8 frames of each video as a group and average them as the current frame, where the adjacent groups overlap 4 frames, to convert the original videos with a frame rate of 120 FPS into the blurred videos with a frame rate of 30 FPS.

### 2.2.2 Enhancement Method

The following five enhancement methods are used to deblur the 62 blurred videos: ESTRNN [56], DeblurGANv2 [15], FGST [18], BasicVSR++ [4], and Adobe Premiere Pro. Specifically, ESTRNN [56] realizes the deblurring of the current frame by fusing the effective layered features of the past and future frames. DeblurGANv2 [15] firstly introduces the feature pyramid network into deblurring, which can work flexibly with a variety of backbones to find a balance between performance and efficiency of deblurring. FGST [18] proposes a flow-guided sparse window-based multi-head self-attention module to deblur videos. BasicVSR++ [4] proposes a second-order grid propagation and flow-guided deformable alignment model to effectively utilize the spatiotemporal information in unaligned video frames to achieve video deblurring.

For ESTRNN [56], DeblurGANv2 [15], FGST [18], and BasicVSR++ [4], we use the default parameters to enhance



Figure 3. Process of enhancing videos in the sub-dataset 2.

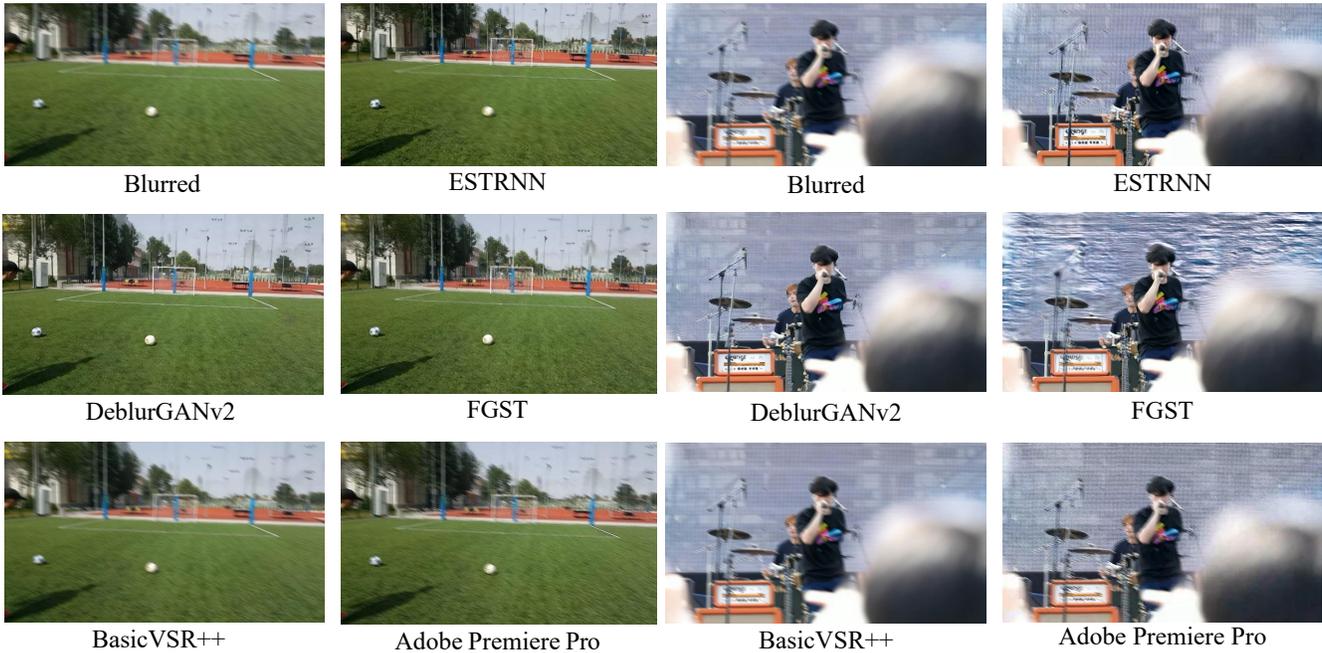


Figure 4. Sample frames of two videos and their corresponding deblurred videos in the sub-dataset 2.

videos. When using Adobe Premiere Pro to process videos, we set the sharpness to 100.

### 2.2.3 Enhanced Video

Figure 3 shows the process of enhancing videos in the sub-dataset 2. We deblur the 62 blurred videos through the above five enhancement methods, and obtain a total of 310 enhanced videos. Figure 4 shows the sample frames of two blurred videos and their corresponding enhanced videos in the sub-dataset 2.

### 2.3. Sub-dataset 3:

The 301 deshaked videos constitute the third sub-dataset of the VDPVE. These enhanced videos are obtained by using seven enhancement methods on 43 original videos selected from two datasets.

#### 2.3.1 Original Video

We select 31 and 12 original videos from the DeepStab [45] and NUS [20] datasets, respectively.

- DeepStab [45]: this dataset contains 60 synchronized steady and unsteady video pairs captured by specially designed hand-held hardware. The length of these videos is within 30s with a frame rate of 30 FPS.
- NUS [20]: this dataset contains 174 videos with a length of 10s to 60s. These videos were selected from previous publications, Internet, and the authors' own captures.

We set the resolution of all original videos to  $1280 \times 720$  and the duration to 10s.

Table 1. Details of enhanced videos of all sub-datasets in the VDPVE. # means the number.

Dataset	# Method	# Original	# Enhanced	Resolution	Duration
Sub-dataset 1	8	79	600	1280 × 720	8s, 10s
Sub-dataset 2	5	62	310	1280 × 720	8s
Sub-dataset 3	7	43	301	1280 × 720	10s

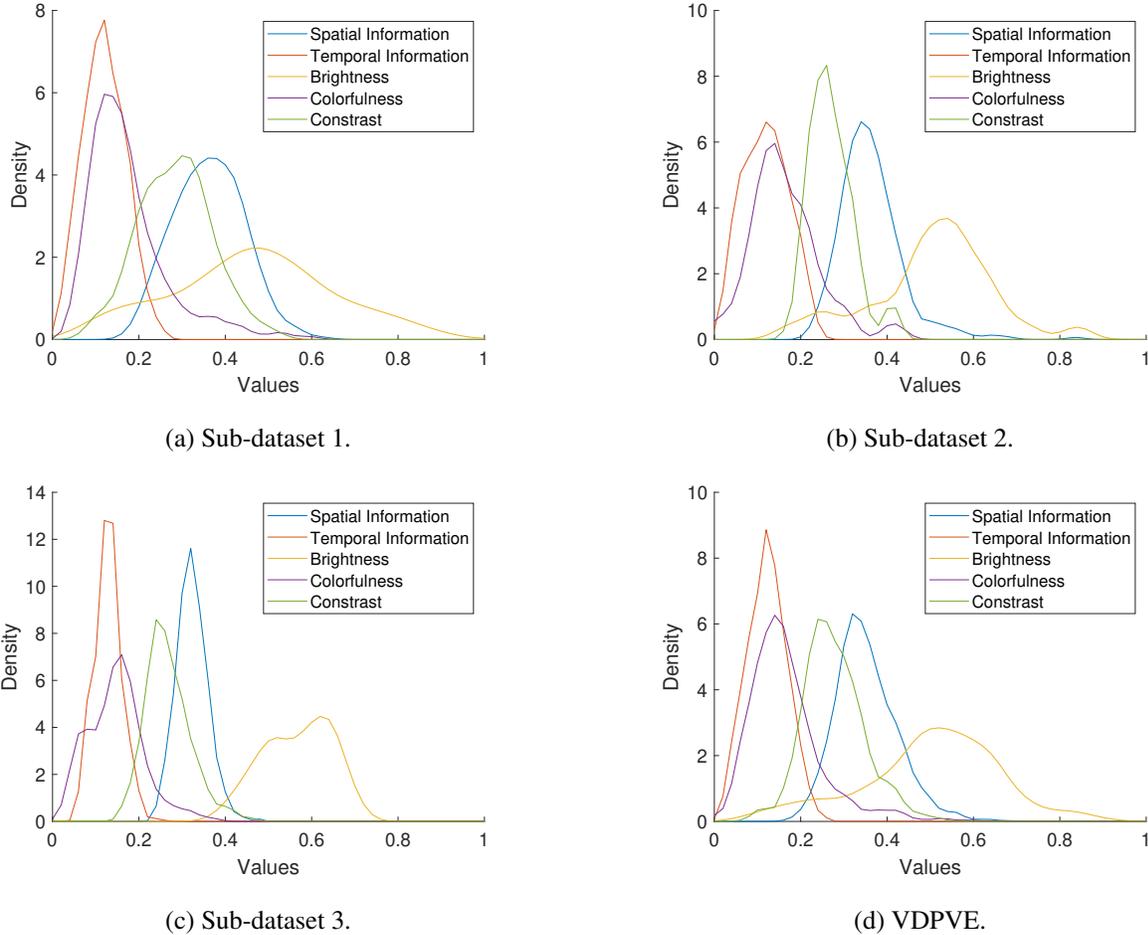


Figure 5. Distribution of attribute values over the three sub-datasets and the VDPVE.

### 2.3.2 Enhancement Method

We use seven enhancement methods to stabilize original videos, including GlobalFlowNet [12], DIFRINT [5], PWStableNet [54], Yu [50], CapCut (most stable mode), CapCut (minimum cropping mode), and Adobe Premiere Pro. Specifically, GlobalFlowNet [12] uses a two-stage process to stabilize the video. DIFRINT [5] uses frame interpolation technology to generate in-between frames, thus reducing the shake between frames. Zhao *et al.* believed that different pixels might have different distortions, thus they

proposed a pixel-level video stabilization method based on deep learning, PWStableNet [54]. The proposed method is based on a multistage cascaded encoder-decoder architecture and learns per-pixel distortion mapping from continuous unstable frames. Yu [50] uses optical flow to analyse motion and learn stability directly.

For GlobalFlowNet [12], DIFRINT [5], PWStableNet [54], and Yu [50], we use the default parameters to enhance videos. When using CapCut to process videos, we use the most stable mode and the minimum cropping mode. When

using Adobe Premiere Pro to process videos, we use the default parameters for stabilization.

### 2.3.3 Enhanced Video

For sub-dataset 3, we deshake 43 original videos through the above seven methods, and obtain a total of 301 enhanced videos.

## 2.4. Video Analysis

Table 1 summarizes the details of enhanced videos in all sub-datasets in the VDPVE. In order to measure the content diversity of the dataset, we calculate five video attributes of the VDPVE and three sub-datasets, including spatial information, temporary information, brightness, colorfulness, and contrast. Specifically, Figure 5(a) shows the distribution of attribute values over the sub-dataset 1. Figure 5(b) shows the distribution of attribute values over the sub-dataset 2. Figure 5(c) shows the distribution of attribute values over the sub-dataset 3. Figure 5(d) shows the distribution of attribute values over the VDPVE.

## 3. Subjective Experiment

In this section, we introduce the subjective experiment in detail.

### 3.1. Subjects

We invited 21 subjects from 20 to 30 years old to participate in this subjective experiment. Before the experiment, we tested the visual conditions of all subjects. All 21 subjects had normal (corrected) vision and color vision.

### 3.2. Experimental Procedures

The enhanced videos were displayed on the monitor at their original resolution. Subjects were asked to score the enhanced videos within the range of [0, 100]. The scoring criteria are as follows: a score in the range between 0 and 20 means that the quality of the enhanced video is poor; a score in the range between 20 and 40 means that the quality of the enhanced video is bad; a score in the range between 40 and 60 means that the quality of the enhanced video is fair; a score in the range between 60 and 80 means that the quality of the enhanced video is good; a score in the range between 80 and 100 means that the quality of the enhanced video is excellent.

### 3.3. Data Processing

After the subjective experiment, each enhanced video has 21 subjective opinion scores. Then, we reject one invalid subject and obtain 20 subjective opinion scores for each video. Finally, we normalize the scores and calculate the MOS for each enhanced video. Figure 6 shows the distribution of MOSs for all enhanced videos in the VDPVE.

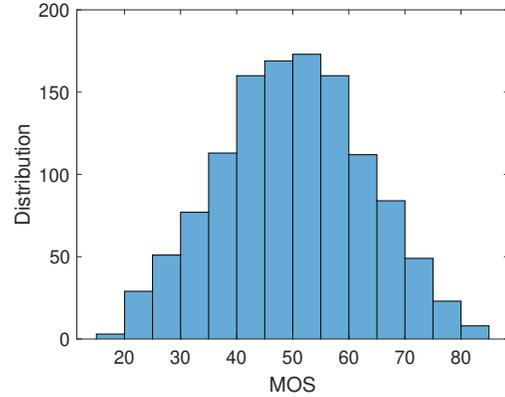


Figure 6. Distribution of MOSs for all enhanced videos in the VDPVE.

Table 2. Mean confidence intervals for different MOS ranges. “CI” means the confidence interval.

MOS	Mean CI	MOS	Mean CI
[15, 20]	[12.6, 19.8]	[20, 25]	[19.1, 26.7]
[25, 30]	[23.8, 32.4]	[30, 35]	[28.3, 36.8]
[35, 40]	[32.9, 42.5]	[40, 45]	[37.6, 47.6]
[45, 50]	[42.5, 52.0]	[50, 55]	[47.5, 57.4]
[55, 60]	[52.5, 62.2]	[60, 65]	[57.6, 67.2]
[65, 70]	[62.9, 71.7]	[70, 75]	[67.3, 77.0]
[75, 80]	[72.9, 81.2]	[80, 85]	[78.0, 84.5]

It can be seen from the figure that the quality of enhanced videos in VDPVE is between 15 and 85. We also calculate the mean 95% confidence intervals for different MOS ranges in Table 2. It can be seen from the table that the length of the mean confidence interval of each MOS range will not exceed 10.

## 4. Experiment

In this section, some experiments are carried out on the proposed VDPVE.

### 4.1. Set Splitting

We first randomly split the enhanced videos in the VDPVE into a training set, a validation set, and a test set according to the ratio of 7:1:2. The enhanced videos generated from the same original video are divided into the same set, and the ratio of the training set, the validation set, and the test set of each divided sub-dataset is also about 7:1:2. The numbers of enhanced videos in the training set, validation set, and test set of the VDPVE are 839, 119, and 253, respectively.

Table 3. Prediction performance of the state-of-the-art VQA methods on the test set. “All” means the entire test set of the VDPVE. The best performances are in bold.

Test Set	Sub-dataset 1		Sub-dataset 2	
Method	SRCC	PLCC	SRCC	PLCC
V-BLIINDS [30]	0.4561	0.4959	0.5048	0.5408
TLVQM [14]	0.5203	0.5721	0.2743	0.3130
VIDEVAL [40]	0.3991	0.3972	0.2887	0.4348
RAPIQUE [41]	0.5161	0.5826	0.4711	0.4789
FastVQA [48]	<b>0.7083</b>	<b>0.7571</b>	<b>0.7479</b>	<b>0.7429</b>
VSFA [17]	0.5240	0.4900	0.6196	0.6060
BVQA [16]	0.5742	0.5632	0.6025	0.6621
SimpleVQA [37]	0.4999	0.5722	0.7007	0.6934
Test Set	Sub-dataset 3		All	
Method	SRCC	PLCC	SRCC	PLCC
V-BLIINDS [30]	0.6895	0.6822	0.5652	0.5503
TLVQM [14]	<b>0.8074</b>	<b>0.8203</b>	0.5474	0.5509
VIDEVAL [40]	0.7111	0.6903	0.5005	0.4724
RAPIQUE [41]	0.5554	0.5736	0.5434	0.5393
FastVQA [48]	0.6368	0.6433	<b>0.7350</b>	<b>0.7310</b>
VSFA [17]	0.5874	0.6281	0.5871	0.5424
BVQA [16]	0.7486	0.7671	0.6995	0.6674
SimpleVQA [37]	0.7420	0.7324	0.6340	0.6354

## 4.2. Performance

Then, we test the performance of several state-of-the-art VQA methods on the test set, including V-BLIINDS [30], TLVQM [14], VIDEVAL [40], RAPIQUE [41], FastVQA [48], VSFA [17], BVQA [16], and SimpleVQA [37]. Specifically, all methods are trained on the entire training set, verified on the entire validation set, and then tested on the three test sets of the three sub-datasets and the entire test set of the VDPVE, respectively. Pearson linear correlation coefficient (PLCC) and Spearman rank correlation coefficient (SRCC) are used to measure the prediction performance of these state-of-the-art VQA methods. Specifically, SRCC measures the prediction monotonicity, while PLCC measures the prediction accuracy. Table 3 shows the prediction performance of these state-of-the-art VQA methods. From the table, we can see that FastVQA [48] gets the best prediction performance on the sub-dataset 1, sub-dataset 2, and the entire VDPVE. The prediction performance of TLVQM [14] in the sub-dataset 3 is the best.

## 5. Conclusion

To make up for the lack of the VQA dataset for video enhancement in the field of video processing, this paper con-

structs the VDPVE. The VDPVE includes a total of 1211 enhanced videos, involving a wide range of enhancement methods. Among them, there are 600 videos with color, contrast, and brightness enhancements; 310 videos with deblurring; and 301 deshaked videos. At the same time, we invited 21 subjects (20 valid subjects) to rate all enhanced videos in the VDPVE to obtain the MOS for each enhanced video. Finally, we split the VDPVE and verify the performance of several popular VQA methods on the test sets. The proposed VDPVE can provide the basis for the development of VQA methods for video enhancement, and further promote the performance of video enhancement methods.

Recently, in addition to enhancing the color, contrast, and brightness of videos, as well as deblurring and deshaking videos, other video enhancement methods such as video super-resolution [21–23, 32] and video frame interpolation [33, 34] have also attracted the attention of a large number of researchers. Therefore, in future work, we will specifically build a VQA dataset for video super-resolution and video frame interpolation, to propose more efficient VQA methods for video enhancement and improve the performance of video enhancement methods.

## References

- [1] Christos G Bampis, Zhi Li, and Alan C Bovik. Spatiotemporal feature integration and model fusion for full reference video quality assessment. *IEEE Trans. Circuits Syst. Video Technol.*, 29(8):2256–2270, 2018. 2
- [2] Fabian Berns, Luca Rossetto, Klaus Schoeffmann, Christian Beecks, and George Awad. V3c1 dataset: an evaluation of content characteristics. In *Proc. ACM Int. Conf. Multimed. Retr.*, pages 334–338, 2019. 3
- [3] Yuqin Cao, Xiongkuo Min, Wei Sun, and Guangtao Zhai. Attention-guided neural networks for full-reference and no-reference audio-visual quality assessment. *IEEE Trans. Image Process.*, 2023. 2
- [4] Kelvin CK Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. Basicvsr++: Improving video super-resolution with enhanced propagation and alignment. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, pages 5972–5981, 2022. 4
- [5] Jinsoo Choi and In So Kweon. Deep iterative frame interpolation for full-frame video stabilization. *ACM Trans. Graph.*, 39(1):1–9, 2020. 6
- [6] Yixuan Gao, Xiongkuo Min, Wenhan Zhu, Xiao-Ping Zhang, and Guangtao Zhai. Image quality score distribution prediction via alpha stable model. *IEEE Trans. Circuits Syst. Video Technol.*, 2022. 2
- [7] Yixuan Gao, Xiongkuo Min, Yucheng Zhu, Jing Li, Xiaoping Zhang, and Guangtao Zhai. Image quality assessment: From mean opinion score to opinion score distribution. In *Proc. ACM Int. Conf. Multimedia*, pages 997–1005, 2022. 2
- [8] Pascal Getreuer. Automatic color enhancement (ace) and its fast implementation. *Image Process. On Line*, 2:266–277, 2012. 3
- [9] Deepti Ghadiyaram, Janice Pan, Alan C Bovik, Anush Krishna Moorthy, Prasanjit Panda, and Kai-Chieh Yang. In-capture mobile video distortions: A study of subjective behavior and objective algorithms. *IEEE Trans. Circuits Syst. Video Technol.*, 28(9):2061–2077, 2017. 3
- [10] Bhupendra Gupta and Mayank Tiwari. Minimum mean brightness error contrast enhancement of color images using adaptive gamma correction with color preserving framework. *Optik*, 127(4):1671–1676, 2016. 3
- [11] V Hosu, F Hahn, M Jenadeleh, H Lin, H Men, T Szirányi, S Li, and D Saupe. The konstanz natural video database, 2020. 1, 3
- [12] Jerin Geo James, Devansh Jain, and Ajit Rajwade. Globalflownet: Video stabilization using deep distilled global motion estimates. In *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, pages 5078–5087, 2023. 6
- [13] Woojae Kim, Jongyoo Kim, Sewoong Ahn, Jinwoo Kim, and Sanghoon Lee. Deep video quality assessor: From spatiotemporal visual sensitivity to a convolutional neural aggregation network. In *Proc. Eur. Conf. Comput. Vis.*, pages 219–234, 2018. 2
- [14] Jari Korhonen. Two-level approach for no-reference consumer video quality assessment. *IEEE Trans. Image Process.*, 28(12):5923–5938, 2019. 8
- [15] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proc. IEEE Int. Conf. Comput. Vision*, pages 8878–8887, 2019. 4
- [16] Bowen Li, Weixia Zhang, Meng Tian, Guangtao Zhai, and Xianpei Wang. Blindly assess quality of in-the-wild videos via quality-aware pre-training and motion perception. *IEEE Trans. Circuits Syst. Video Technol.*, 32(9):5944–5958, 2022. 8
- [17] Dingquan Li, Tingting Jiang, and Ming Jiang. Quality assessment of in-the-wild videos. In *Proc. ACM Int. Conf. Multimedia*, pages 2351–2359, 2019. 8
- [18] Jing Lin, Yuanhao Cai, Xiaowan Hu, Haoqian Wang, Youliang Yan, Xueyi Zou, Henghui Ding, Yulun Zhang, Radu Timofte, and Luc Van Gool. Flow-guided sparse transformer for video deblurring. *arXiv preprint arXiv:2201.01893*, 2022. 4
- [19] Joe Yuchieh Lin, Rui Song, Chi-Hao Wu, TsungJung Liu, Haiqiang Wang, and C-C Jay Kuo. MCL-V: A streaming video quality assessment database. *J. Vis. Commun. Image Represent.*, 30:1–9, 2015. 1
- [20] Shuaicheng Liu, Lu Yuan, Ping Tan, and Jian Sun. Bundled camera paths for video stabilization. *ACM Trans. Graph.*, 32(4):1–10, 2013. 5
- [21] Xiaohong Liu, Lei Chen, Wenyi Wang, and Jiying Zhao. Robust multi-frame super-resolution based on spatially weighted half-quadratic estimation and adaptive btv regularization. *IEEE Trans. Image Process.*, 27(10):4971–4986, 2018. 8
- [22] Xiaohong Liu, Lingshi Kong, Yang Zhou, Jiying Zhao, and Jun Chen. End-to-end trainable video super-resolution based on a new mechanism for implicit motion estimation and compensation. In *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, pages 2416–2425, 2020. 8
- [23] Xiaohong Liu, Kangdi Shi, Zhe Wang, and Jun Chen. Exploit camera raw data for video super-resolution via hidden markov model inference. *IEEE Trans. Image Process.*, 30:2127–2140, 2021. 8
- [24] Wei Lu, Wei Sun, Xiongkuo Min, Wenhan Zhu, Quan Zhou, Jun He, Qiyuan Wang, Zicheng Zhang, Tao Wang, and Guangtao Zhai. Deep neural network for blind visual quality assessment of 4k content. *IEEE Trans. Broadcast.*, 2022. 2
- [25] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mblen: Low-light image/video enhancement using cnns. In *BMVC*, volume 220, page 4, 2018. 3
- [26] Alex Mackin, Fan Zhang, and David R Bull. A study of high frame rate video formats. *IEEE Trans. Multimedia*, 21(6):1499–1512, 2018. 4
- [27] Pavan C Madhusudana, Xiangxu Yu, Neil Birkbeck, Yilin Wang, Balu Adsumilli, and Alan C Bovik. Subjective and objective quality assessment of high frame rate videos. *IEEE Access*, 9:108069–108082, 2021. 4
- [28] Mikko Nuutinen, Toni Virtanen, Mikko Vaahteranoksa, Tero Vuori, Pirkko Oittinen, and Jukka Häkkinen. CVD2014—A database for evaluating no-reference video quality assessment algorithms. *IEEE Trans. Image Process.*, 25(7):3073–3086, 2016. 1

- [29] Luca Rossetto, Heiko Schuldt, George Awad, and Asad A Butt. V3c—a research video collection. In *Proc. MultiMedia Model.*, pages 349–360. Springer, 2019. 3
- [30] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind prediction of natural video quality. *IEEE Trans. Image Process.*, 23(3):1352–1365, 2014. 2, 8
- [31] Kalpana Seshadrinathan, Rajiv Soundararajan, Alan Conrad Bovik, and Lawrence K Cormack. Study of subjective and objective quality assessment of video. *IEEE Trans. Image Process.*, 19(6):1427–1441, 2010. 1
- [32] Zhihao Shi, Xiaohong Liu, Chengqi Li, Linhui Dai, Jun Chen, Timothy N Davidson, and Jiying Zhao. Learning for unconstrained space-time video super-resolution. *IEEE Trans. Broadcast.*, 68(2):345–358, 2021. 8
- [33] Zhihao Shi, Xiaohong Liu, Kangdi Shi, Linhui Dai, and Jun Chen. Video frame interpolation via generalized deformable convolution. *IEEE Trans. Multimedia*, 24:426–439, 2021. 8
- [34] Zhihao Shi, Xiangyu Xu, Xiaohong Liu, Jun Chen, and Ming-Hsuan Yang. Video frame interpolation transformer. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 17482–17491, 2022. 8
- [35] Zeina Sinno and Alan Conrad Bovik. Large-scale study of perceptual video quality. *IEEE Trans. Image Process.*, 28(2):612–627, 2018. 3, 4
- [36] Rajiv Soundararajan and Alan C Bovik. Video quality assessment by reduced reference spatio-temporal entropic differencing. *IEEE Trans. Circuits Syst. Video Technol.*, 23(4):684–694, 2012. 2
- [37] Wei Sun, Xiongkuo Min, Wei Lu, and Guangtao Zhai. A deep learning based no-reference quality assessment model for ugc videos. In *Proc. ACM Int. Conf. Multimed.*, pages 856–865, 2022. 2, 8
- [38] Wei Sun, Tao Wang, Xiongkuo Min, Fuwang Yi, and Guangtao Zhai. Deep learning based full-reference and no-reference quality assessment models for compressed ugc videos. In *Proc. Int. Conf. Multimedia and Expo Worksh.*, pages 1–6. IEEE, 2021. 2
- [39] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. Yfcc100m: The new data in multimedia research. *Commun. ACM*, 59(2):64–73, 2016. 3
- [40] Zhengzhong Tu, Yilin Wang, Neil Birkbeck, Balu Adsumilli, and Alan C Bovik. Ugc-vqa: Benchmarking blind video quality assessment for user generated content. *IEEE Trans. Image Process.*, 30:4449–4464, 2021. 8
- [41] Zhengzhong Tu, Xiangxu Yu, Yilin Wang, Neil Birkbeck, Balu Adsumilli, and Alan C Bovik. Rapique: Rapid and accurate video quality prediction of user generated content. *IEEE Open J. Signal Process.*, 2:425–440, 2021. 8
- [42] Phong V Vu and Damon M Chandler. ViS3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices. *J. Electron. Imaging*, 23(1):013016–013016, 2014. 1
- [43] Chao Wang and Zhongfu Ye. Brightness preserving histogram equalization with maximum entropy: a variational perspective. *IEEE Trans. Consum. Electron.*, 51(4):1326–1334, 2005. 3
- [44] Haiqiang Wang, Weihao Gan, Sudeng Hu, Joe Yuchieh Lin, Lina Jin, Longguang Song, Ping Wang, Ioannis Katsavounidis, Anne Aaron, and C-C Jay Kuo. MCL-JCV: a JND-based H. 264/AVC video quality assessment dataset. In *Proc. IEEE Int. Conf. Image Process.*, pages 1509–1513, 2016. 1
- [45] Miao Wang, Guo-Ye Yang, Jin-Kun Lin, Song-Hai Zhang, Ariel Shamir, Shao-Ping Lu, and Shi-Min Hu. Deep online video stabilization with multi-grid warping transformation learning. *IEEE Trans. Image Process.*, 28(5):2283–2292, 2018. 5
- [46] Yilin Wang, Sasi Inguva, and Balu Adsumilli. Youtube ugc dataset for video compression research. In *Proc. IEEE Int. Workshop Multimed. Signal Process.*, pages 1–5, 2019. 4
- [47] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004. 2
- [48] Haoning Wu, Chaofeng Chen, Jingwen Hou, Liang Liao, Annan Wang, Wenxiu Sun, Qiong Yan, and Weisi Lin. Fast-vqa: Efficient end-to-end video quality assessment with fragment sampling. In *Proc. Eur. Conf. Comput. Vis.*, pages 538–554, 2022. 8
- [49] Fuwang Yi, Mianyi Chen, Wei Sun, Xiongkuo Min, Yuan Tian, and Guangtao Zhai. Attention based network for no-reference ugc video quality assessment. In *Proc. IEEE Int. Conf. Image Process.*, pages 1414–1418. IEEE, 2021. 2
- [50] Jiyang Yu and Ravi Ramamoorthi. Learning video stabilization using optical flow. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, pages 8159–8167, 2020. 6
- [51] Guangtao Zhai, Wei Sun, Xiongkuo Min, and Jiantao Zhou. Perceptual quality assessment of low-light image enhancement. *ACM Trans. Multimed. Comput. Commun. Appl.*, 17(4):1–24, 2021. 2
- [52] Zicheng Zhang, Wei Lu, Wei Sun, Xiongkuo Min, Tao Wang, and Guangtao Zhai. Surveillance video quality assessment based on quality related retraining. In *Proc. IEEE Int. Conf. Image Process.*, pages 4278–4282. IEEE, 2022. 2
- [53] Zhao Zhang, Huan Zheng, Richang Hong, Mingliang Xu, Shuicheng Yan, and Meng Wang. Deep color consistent network for low-light image enhancement. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, pages 1899–1908, 2022. 3
- [54] Minda Zhao and Qiang Ling. Pwstabenet: Learning pixel-wise warping maps for video stabilization. *IEEE Trans. Image Process.*, 29:3582–3595, 2020. 6
- [55] Shen Zheng and Gaurav Gupta. Semantic-guided zero-shot learning for low-light image/video enhancement. In *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, pages 581–590, 2022. 3
- [56] Zhihang Zhong, Ye Gao, Yinqiang Zheng, and Bo Zheng. Efficient spatio-temporal recurrent neural network for video deblurring. In *Proc. Eur. Conf. Comput. Vis.*, pages 191–207, 2020. 4