Image/Video Quality Assessment Challenges: A Comprehensive Review

Muhammad Uzair

Faculty of Engineering, Department of Electrical Engineering, Islamic University of Madinah, Madinah, Saudi Arabia
Email: muzair@iu.edu.sa

***Abstract*—Lossy coding schemes which generates distortions in the images/videos are used by video encoders to meet the bandwidth requirements. Similarly, preprocessing, transmission, and post-processing steps generates additional artifacts. Measuring these artifacts is crucial for the development of digital video systems. However, there are many issues/challenges while measuring the quality of an image/video efficiently and effectively. Many existing research papers present the issues/challenges while doing quality assessment. However, existing work does not present all the issues and challenges comprehensively and discusses either one or two kinds of issues during quality assessment. However, this work presents a comprehensive overview of all kinds of issues & challenges while measuring image/video quality, i.e., ranging from subjective to objective, using 3D video and new coding tools, and the effect of one type of distortion on other types of distortion, issues due to subjective databases, etc. The paper also describes issues/challenges related to Quality of Experience (QoE) by the end users, issues in real time streaming (e.g., edge & cloud based streaming, etc.), issues in deep learning, machine learning, artificial intelligence generated & enhanced contents, etc. The work also discusses other miscellaneous issues/challenges and recommends for future directions based on the current work understanding. Overall, this work will guide the research community in taking into account these issues and developing more efficient ways of measuring image/video quality, along with developing efficient encoders and 3D encoding methods for the further development of digital video systems.**

***Keywords*—issues/challenges, quality, image/video, artifacts**

I. Introduction

Capturing, storing, receiving, viewing, utilizing, and sharing images/videos have been revolutionized by advanced digital image and video processing technologies. The general user's attraction has also been enhanced by many folds due to the growth of multimedia applications, i.e., High-definition Television, e-commerce, etc. Hence, image/video quality assessment is not only an important & integral part of image and video processing, and plays a key role in developing digital video systems.

Digital video systems also provide a standard for other systems, but they also introduce new kinds of artifacts/distortions compared to analog systems. As a result, conventional analog techniques are inadequate for assessing the quality of digital videos. Digital content undergoes various distortions during processes such as capture, compression, transmission, decoding, and playback. This is because the compression techniques (to produce a low bit rate for low bandwidth requirements) used by the video codecs during the quantization process are lossy. The perceptibility of these distortions is also heavily influenced by the specific content of the image or video [1–2].

Despite the digital imaging revolution, the biological hardware of humans (Human Visual System—HVS) is the same. Along with HVS properties, distortions also happen due to the spatial and temporal characteristics of the images/videos, i.e., blocking, ringing, motion compensation, etc. Similarly, different kinds of losses, i.e., packet losses, jitter, flickering, etc., occurs during video transmission through wireless networks [3]. Therefore, the whole video processing, for example, pre-processing, compression, transmission and post- processing generates artifacts on the decoder side, i.e., reconstructed video. Moreover, artifacts are also introduced due to the effect of one type of distortion on the other artifacts, new kinds of artifacts because of latest coders; and artifacts due to the 3D coding, etc. Similarly, end-user satisfaction is also critical aspect of digital video applications. Moreover, there are also challenges in real time streaming even due to edge & cloud based streaming, and also challenges due to the new quality measurement approaches, i.e., deep learning, machine learning, etc. There are also challenges in measuring quality for artificial intelligence generated & enhanced contents, etc. Therefore, Quality of Experience (QoE) by the end user is another issue/challenge in the case of digital video applications [1, 4].

The research community has proposed different kinds of algorithms to measure image/video quality. The main goal of these algorithms is that their assessment should correlate well with the subjective evaluation. However, there are many limitations in these proposed algorithms such as existing algorithms do not take into account all kinds of distortions and are distortion specific; current Full Reference (FR) metrics are not feasible in real time because of the bandwidth constraint; current HVS algorithms are very complex, time-consuming and

Manuscript received February 21, 2025; revised April 1, 2025; accepted June 6, 2025; published December 19, 2025.

doi: 10.18178/joig.13.6.651-672

expensive; many existing Reduced Reference (RR) and No Reference (NR) metrics correlate well with HVS, but are unable to measure network & temporal distortions; existing spatial measuring distortion algorithms do not estimate HVS effects particularly, etc. For a sound digital video quality system, it is very imperative that the algorithm be able to detect different kinds of distortions effectively and efficiently and should address all of the shortcomings of existing metrics and address all existing issues/challenges [4–5].

As new challenges and issues that arise in image and video quality assessment, an effective algorithm must be developed to detect all distortions accurately. It should maintain a low rate of false detections and missed artifacts, while ensuring consistency and accuracy in identifying distortions. Additionally, the algorithm should demonstrate strong overall performance. These aspects become important as the nature of video artifacts is complex and involves multiple algorithms and techniques.

Numerous works have been done in the past that describe different kinds of issues/challenges while estimating image/video quality [6–9]. However, the published papers either address issues and challenges related to the subjective assessment of image/video quality [6–9], or focus on objective assessment challenges [10]. Some discuss challenges related to spatial and temporal artifacts [4–5], while others examine the interactions between different types of artifacts [11]. Additionally, certain papers explore challenges introduced by modern coding techniques [11–12], 3D coding [13], or miscellaneous issues such as geometric distortion, supra-threshold distortion, and the HVS [14]. Furthermore, some papers focus solely on image quality evaluation without considering video quality assessment or the quality of graphics images [14]. Similarly, some recent works discuss only challenges in edge & cloud based streaming, and issues while measuring quality by using deep learning, machine learning, and issues/challenges in measuring quality for Artificial Intelligence Generated & Enhanced Contents (AIGC & AIEC), etc. [15–17]. However, the existing work does not describe all the issues/challenges collectively and comprehensively while assessing image/video quality assessments. This work, however, offers a thorough overview and analysis of the various challenges involved in measuring video quality, while also addressing the limitations of previous studies.

The rest of the paper is structured as follows: Section II describes a comprehensive study of different kinds of subjective quality assessment methods and their issues/challenges. Section III describes a comprehensive study of objective quality assessment methods and their challenges. Section IV provides challenges/issues related to the spatial & temporal artifacts and their effect on other kinds of artifacts. Section V presents challenges due to the new types of coding standards. Section VI describes challenges due to the 3D coding. Section VII describes issues related to the subjective databases. Section VIII describes issues related to the QoE by the end users. Section IX presents issues related to real time processing of video. Section X describes issues related to deep learning, Machine learning, and AIGC & AIEC. Section XI describes other major/miscellaneous issues. Section XII presents the recommendations. Section XIII presents conclusion.

The following section describes subjective quality assessment challenges/issues while assessing image/video quality.

II. Subjective Uality Measurement Issues/Challenges

Subjective quality measurement is one approach to estimate image/video quality. In subjective quality assessment, a viewer assigns a score to an image/video based on perception. This quality assessment approach is very accurate, but it is complex, time-consuming, risky (Error risks during test design), and expensive (dedicated rooms and human resources needed). Generally, blur, noise, and compression-related artifacts are considered by human observers in this process, and other kinds of artifacts are ignored. Similarly, many other aspects are ignored in this process due to the lack of understanding of supra threshold distortions, natural images, images having nontraditional distortions, manifold & heightened distortions, etc. [6–7].

The contents of video sequences significantly influence the activation of various cognitive, emotional, and behavioral responses in viewers. Additionally, emotions such as disgust, fear, anger, and sadness should be considered, as they affect conative processes differently than emotions like happiness and surprise. Hence, further research is needed to explore the connection between video contents, emotional responses, and the overall perceived video quality. Additionally, factors such as gender, culture, and demographics should be considered when conducting subjective quality evaluations. Table I shows the most famous subjective quality assessment methods with their issues/challenges [6–9, 18, 19].

TABLE I. Subjective Quality Assessments Issues/Challenges

Approach	Description	Result evaluation	Issues/Quality level
Single Stimulus Continuous Quality Evaluation (SSCQE)	Viewers see an image for a short duration of time; Images are displayed in random order & include reference images which is not known to the observer.	Observer evaluates the image by using these five classifications: exceptional, decent, reasonable, reduced, or bad; MOS enhances as many decisions can be obtained in limited time; Process is efficient as n+1 judgments are required to measure n situations (reference image requires one extra judgement).	Inconsistent: A continuous method will be better to avoid quantization effects as compared to categorical scales; Takes into account changes in scene complexity, which may produce substantial-quality variations; Presentation time is variable for all samples affect overall length and efficiency; Well-trained observers are required to obtain a persistent quality scores.

Double stimulus categorical rating (DSCQS)	Reference & test images are presented in random order one after another, i.e., analogous to SSCQE; Subjects are unaware of reference & test images settings.	A continuous scale ranging from 0 to 100 (bad to excellent) is used to evaluate images; SSCQE & DSCQS methods dominate video quality assessment.	More time-consuming as compared to single stimulus quality categorical rating; A preferred choice when no significant change in quality exists between successive video clips.
Double Stimulus Impairment Scale (DSIS)	Viewers know the locations of reference & test images; Reference sequence is shown first, followed by the test sequence & shown only once.	A five-level scale (discrete) ranging from very bad to imperceptible is used to rate the sequence; it is better suited for visible distortions, i.e., transmission errors.	Both (DSCQS & DSIS) approaches are not feasible with scenes having changes in complexity due to a single rating, i.e., may produce reasonable quality changes that are not evenly distributed over time.
Absolute Category Rating (ACR)	Similar to SSCQE; Only tested videos are shown to viewers from bad to excellent.	Viewers give a single score for the overall quality of the video using a five-level discrete scale.	Due to the lack of reference video, an observer may not provide a good quality score.
Absolute Category Rating-Hidden Reference (ACR-HR)	Like ACR, except the reference image is also exposed to the observer.	The evaluators' rate using Mean Opinion Score (MOS) and a final quality assessment are calculated using a differential quality score.	A better quality score provided as a reference image is available, but more time consumption.
Forced-choice pair- wise comparison	A pair of images depicting the same scene under changed situations is presented to viewers; Viewers are forced to specify an image, even if there is no quality difference.	No specific amount of time to choose one image by the viewers; More reliable than the rating method; Popular in computer Graphics.	Consistency is low; it is very tedious & time-consuming for a large number of conditions; For n videos, its time complexity is $O(n)^2$, as compared to other methods, which have only O(n).
Similarity judgment	Similar to the forced choice pair method where viewers choose not only an image of higher quality, but also assess quality using continuous scale, i.e., the forced method does not tell how much difference among images.	An average quality score is obtained by the video quality ratings specified by the viewer; Viewers have the choice to use a marker at the'0' position if there is no quality difference among pair; Used in functional measurement approaches that depends on comparative decisions.	The same issues apply to force choice pair methods as more tedious work.
Quality ruler (QS)	Both reference & tested images are known to viewers and are shown in series.	The difference in quality is detected among the reference and tested images by viewers.	Consistent as compared to SSCQE; Tedious.
Simultaneous Double Stimulus for Continuous Evaluation (SDSCE)	Two parallel scenes are viewed by the observer to specify quality by comparing reference & impaired video.		The drawback of this method is that the viewer must shift attention between the right and left presentations; Exhaustive & time-consuming.
Subjective Assessment Methodology for Video Quality (SAMVIQ)	Similar to ACR; Processed video is shown to the viewers according to their need & pace, and the score is assigned immediately.	Difference between SAMVIQ and ACR is that the observers can rerun videos to enhance their evaluations in SAMVIQ whereas in ACR viewers do not have this option.	SAMVIQ provides accurate measurement but takes more time than ACR.

Next section describes objective quality assessment metrics and their challenges/issues.

III. Issues/Challenges with Objective Quality Evaluation

Objective quality assessment is another way of estimating image/video quality. In this approach, any mathematical and/or statistical method is used to automatically assessing video quality without human involvement. Many objective quality metrics have been proposed by the research community to evaluate image/video quality based on three criteria's, i.e., Full Reference (FR—original information is accessible at the receiver side), Reduced Reference (RR—little original information is accessible at the receiver side), and No Reference (NR—no original information is accessible at the receiver side) [20].

However, all of the proposed algorithms have different kinds of issues while assessing image/video quality, i.e., existing metrics are impairment type and are not generalized to assess all kinds of artifacts; presented FR algorithms are not feasible in real-time due to bandwidth constraint; presented HVS based metrics are very complex and not able to take into account HVS effects efficiently; difficult to select some information for RR metrics and this problem increases with video contents having temporal variations; many existing RR & NR algorithms correlate well with subjective rating, but their performance deteriorates in case of network & temporal losses; most existing algorithms which assess compression impairments generally do not estimate network & temporal distortions neither HVS effects, etc. [20–21].

In Ref. [22], the effectiveness of various objective video quality metrics is evaluated based on several performance criteria, including accuracy, monotonicity, stability, and complexity. The findings reveal that factors such as video resolution and types of distortion significantly affect the performance of these metrics. Additionally, the results indicate that metrics designed to account for temporal motion tend to perform better than others. Metrics like Foveated Mean Squared Error (FMSE) and Motion-based Video Integrity Evaluation (MOVIE) are more flexible with resolution variations compared to Video Quality Metric (VQM) and Temporal Motion-based Video

Integrity Evaluation (TMOVIE), although the computational complexity of MOVIE is high. Table II provides feasibility/issues, i.e., correlation & feasibility, of different objective metrics [20–22].

TABLE II. Objective Quality Assessments Issues/Challenges

Approach	Metrics	Performance	Feasibility/Issues
FR (Pixel-based)	PSNR/MSE, etc.	Low	Low (bandwidth constraint); Not feasible in real time
FR (Structural information & Similarity based)	SSIM, Radon, PIQ, NQM, UQI, IFC, VIF, etc.	High	Low (bandwidth constraint); High computational complexity; Not feasible in real-time
HVS based	PSNR-HVS-M, VSNR, JNBM, MAD, etc.	Medium	Low; Complex & High computational complexity; No network artifacts measurement
RR (Spatial, Temporal & Network artifacts based)	Application-oriented (MGA-based IQA, RRIQA, etc.)	High	Medium/High (less feasible than NR)
NR	BIQI, BLIND, BLIND II, BRISQUE, DIVINE, MREBN, etc.	High	Generally able to measure specific distortion
Data Hiding		Medium	Low/Medium; Extra Overheads
Network distortions		Medium	Medium; No compression distortions estimation
Learning oriented metrics	MMF, etc.	Medium/High	Low due to FR

TABLE III. Correlation Coefficients for FR-IQA Metrics

Database	FR Metrics	FSIMc	MDSI	HPSI	VQCS	SPSIM	GMSD
TID2008	PLCC	0.87	**0.91**	0.90	0.87	0.89	0.87
	SROCC	0.88	**0.92**	0.91	0.89	0.91	0.89
	KROCC	0.69	**0.75**	0.73	0.71	0.73	0.70
	RMSE	0.64	**0.53**	0.56	0.64	0.60	0.64
TID2013	PLCC	0.87	**0.90**	0.89	0.90	**0.90**	0.85
	SROCC	0.85	0.89	0.87	0.89	**0.90**	0.80
	KROCC	0.66	0.71	0.69	0.71	**0.72**	0.63
	RMSE	0.59	0.51	0.55	0.54	**0.51**	0.64
KADID-10k	PLCC	0.85	0.86	**0.88**	0.86	0.87	0.80
	SROCC	0.85	**0.88**	**0.88**	0.87	0.87	0.84
	KROCC	0.65	**0.70**	0.69	0.68	0.68	0.66
	RMSE	0.56	0.54	**0.50**	0.53	0.52	0.64
PIPAL	PLCC	0.61	0.59	**0.64**	0.55	0.57	0.62
	SROCC	**0.58**	0.58	**0.58**	0.53	0.56	0.58
	KROCC	0.41	0.40	0.41	0.37	0.39	**0.41**
	RMSE	0.104	0.106	**0.101**	0.11	0.108	0.103

TABLE IV. Correlation Coefficients for NR-IQA Metrics

NR Methods	CLIVE			KonIQ-10k		
	PLCC	SROCC	KROCC	PLCC	SROCC	KROCC
BIQI	0.51	0.48	0.32	0.68	0.66	0.47
BLIINDS-II	0.47	0.44	0.29	0.57	0.57	0.41
BMPRI	0.54	0.48	0.33	0.63	0.61	0.42
BRISQUE	0.52	0.49	0.34	0.70	0.67	0.49
CurveletQA	0.63	0.62	0.42	0.73	0.71	0.49
DIIVINE	0.61	0.58	0.40	0.70	0.69	0.47
ENIQA	0.59	0.56	0.37	0.76	0.74	0.54
GM-LOG-BIQA	0.60	0.60	0.38	0.70	0.69	0.50
GWH-GLBP	0.58	0.55	0.39	0.72	0.69	0.50
IL-NIQE	0.48	0.41	0.28	0.46	0.44	0.30
NBIQA	0.62	0.60	0.42	0.77	0.74	0.51
NIQE	0.32	0.29	0.20	0.31	0.40	0.27
OG-IQA	0.54	0.50	0.36	0.65	0.63	0.44
PIQE	0.17	0.10	0.08	0.20	0.24	0.17
Robust BRISQUE	0.52	0.48	0.33	0.71	0.66	0.47
SSEQ	0.48	0.43	0.30	0.58	0.57	0.42
SGL-IQA	**0.70**	**0.66**	**0.47**	**0.81**	**0.79**	**0.59**

Table III provides a quantitative performance comparison of six Full-Reference (FR) quality assessment metrics on four different databases. The metrics have been evaluated with respect to Pearson Linear Correlation Coefficient (PLCC), Spearman Rank Order Correlation Coefficient (SROCC), Kendall Rank Correlation Coefficient (KROCC), and Root Mean Square Error (RMSE). The best correlation is shown in bold [23].

Similarly, Table IV presents the performance comparison of several state-of-the-art NR quality assessment metrics, evaluated using PLCC, SROCC, and KROCC on two different databases. The best correlation is shown in bold, and the second best correlation is shown in green [24].

Next section describes issues due to spatial & temporal distortions and their effect on each other.

IV. Challenges due to the Effect of Spatial and Temporal Distortions on other Artifacts

Block-based Discrete Cosine Transform (DCT) coding with motion compensation and quantization is a common compression technique in current and emerging video codecs. This process generates different kinds of compression artifacts in the transform domain. Similarly, when successive frames in a video sequence are coded inconsistently, temporal artifacts are produced. This occurs due to variations in prediction methods, quantization settings, motion compensation, or a combination of these elements [11, 23].

Most image/video quality measurement algorithms can only measure one kind of distortion (spatial or temporal) efficiently instead of multiple distortions. However, multiple distortions interact with each other and may cross-mask and effect each other. Also, the occurrence of one type of distortion and/or trying to reduce/remove one type of distortion may also introduce/enhance other kinds of distortions. Therefore, further research is needed for efficient algorithms that can assess multiple distortions simultaneously or at least measure those distortions that are related/affecting each other more accurately [25]. Table V shows the spatial distortions (not all spatial distortions, but those that generally affect other distortions) and their relation with other kinds of distortions, highlighting the issues/challenges during image/video quality assessment [11, 25–27].

Similarly, temporal distortions also have impact/relation with other artifacts, which makes it very difficult for existing algorithms to efficiently assess image/video quality. This is because assessing one type of distortion and ignoring others (which may be introduced due to other types of distortions) does not provide an efficient estimation of image/video quality. Table VI shows temporal distortions (not showing all temporal distortions) and their relation with other distortions, highlighting the issues/challenges during image/video quality assessment [11, 25–27].

TABLE V. Spatial Artifacts Effect/Relation with other Artifacts

Artifact	Occurrence reason; Spatial extension	Coexisting artifacts	Relation/effect on other artifacts
Blocking	Happens due to the independent quantization of adjacent blocks; Spatial extension goes up to 64×64 macro block in HEVC staring from 4×4 blocks		Motion compensation mismatch increases due to macro block partitioning if no de-blocking filter; Blur is produced due to de-blocking filtering;
Blurring	Lack of high-frequency components; 4×4 block in H.264 and H.265	Ringing at sharp edges; color bleeding in chroma	Generate sharpening artifacts or noise by introducing high-frequency components during decoding; It may occur with mosaic pattern where high spatial activity exists
Ringing	Insufficient approximation of steep edges; 4×4 block in H.264 and H.265	Blurring	Ringing effect also introduces Mosquito effect which is related to the high frequency distortions.
Staircase effect	Diagonal edges are not estimated sufficiently; Global spatial extent	Basis image effect may exist	Blocking & mosaic pattern effect may exist; Associated with ringing & increases when step size is equal to the size of a macro block
False edges	MC transfers blockiness distortion from reference to predicted frame		Associated with blocking distortion
Mosaic Patterns	Removal of AC coefficient cause apparent mismatch between adjacent blocks; Global spatial extent		Associated with blocking effect
Color bleeding	Coarse quantization of high frequency chroma components; 64×64 macro block (HEVC)		Color bleeding in chroma sub sampled images also exist.
Basis Image effect	Elimination of all DCT coefficients except one; 4×4 block in H.264 and H.265		Introduce mosaic patterns; In low spatial activity regions may introduce blocking & blurring

TABLE VI. TEMPORAL ARTIFACTS EFFECT/RELATION WITH OTHER ARTIFACTS

Artifact	Occurrence reason	Relation/effect on other artifacts
Flickering	Frame to frame coarse quantization; Video quality degrades considerably	Blocking & blurring reduce in H.264, but flickering may increase
Fine granularity flickering	Due to blocking distortion and slow motion in regions with low to mid-level energy	
Mosquito noise	Due to the quantization of high frequency components	Exist with MC mismatch
Floating	Produced as skip mode is used by the encoder to copy a block from one frame to another	Produces illusive motion in certain ranges
Texture floating	Skip mode & texture region with no motion is used by the encoder to copy a block from one frame to another	Masks other kinds of distortions in high energy texture and edge regions
Edge neighborhood floating	Happens as a result of using skip mode; Stationary areas may appear as moving with object boundaries	Looks like mosquito effect distortion in stationary areas
Motion compensation (MC)	Occurs due to incorrect motion estimation	Produces mosquito noise
Smearing	Occurs when the recorder is not able to change beam intensity rapidly with resolution	Causes loss of spatial resolution & blurring
Down/up sampling	Occurs due to the discarding of even fields during down sampling & making vertical resolutions as half	Creates jitter & spatial variations
Temporal pumping artifact	Variations in quantization between adjacent pictures	Causes significant quality fluctuations between adjacent frames in a GOP

Next section describes issues/challenges due to new video coding tools.

V. ISSUES/CHALLENGES DUE TO NEW VIDEO CODECS

Most spatial and temporal distortions described in previous sections are present when encoding video using codecs like MPEG-4 Part 2, H.264, and others. However, the introduction of newer coding technologies, such as Scalable Video Coding (SVC), Multi-View Coding (MVC), and Next Generation Video Coding (NVC), may give rise to additional artifacts. Similarly, after HEVC, the Joint Video Experts Team (JVC) introduced a new video compression standard known as Versatile Video Coding (VVC). Additionally, many new video codecs, such as VP9, VP10, and AV1, have also been developed by different companies. Following Fig. 1 shows future video coding [26, 28].

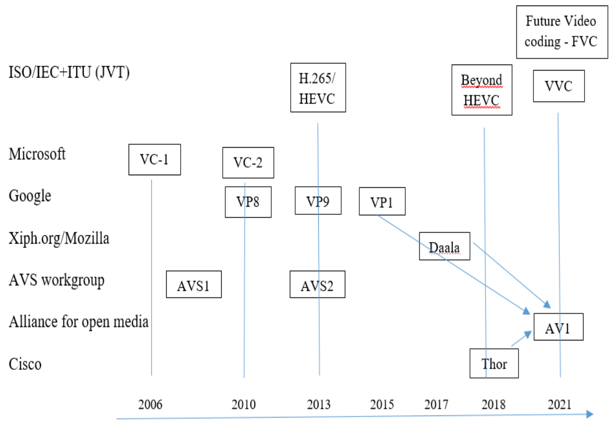

Fig. 1. Future video coding.

The main task of designing a new video codec is to provide high-throughput & low-energy solutions for highly complex videos. However, new video codecs have a lot of issues/challenges as discussed below.

A. *Inter-frame Prediction*

Due to high complexity and memory bandwidth, inter-frame prediction for HEVC, AV1 and VVC is extremely difficult and demands a continuous research effort. Especially, AV1 and VVC codecs needs more attention as they provide support for more block sizes & reference frames with increasing resolutions & frame rates, which introduces new artifacts and makes it difficult to assess video quality efficiently [26].

1) Macro blocking effects

In inter-predictive coding, macroblocks are typically divided into sizes such as 16×16, 16×8, and 8×16. Newer codecs, however, allow for finer partitioning down to 4×4 macroblocks. Although, it provides better matches, but enhances other artifacts [23].

2) Effect of different coding modes

New artifacts are also produced due to different coding modes in inter-frame prediction. The flickering or pumping artifact occurs due to changes in the coding mode of a specific section in consecutive frames.

B. *Intra-frame Prediction*

The intra-frame prediction also has many challenges due to the support of high number of supported block sizes and the complexity of the novel intra-coding modes. In H.265/HEVC/VVC, the macro blocks can be 16×16, 32×32, or up to 64×64 luma sample compared to 16×16 pixels in H.264. Although coding efficiency improves, it also amplifies ringing artifacts due to the higher number of coefficients and samples, which provide more opportunities for overshooting and undershooting. Similarly, the minimum transform size in H.264, H.265, and VC-1 is 4×4, which can lead to increased blockiness due to more prominent block borders [11, 27].

C. Transforms & Quantization

The transform & quantization steps include new combinations of different transforms, an increased number of supported sizes, and high throughput requirements in new codecs. The minimum DCT transform size in H.264, H.265, and VC-1 is 4×4, as opposed to the 8x8 used in MPEG-2 and MPEG-4 Part 2. While this reduction in transform size may lead to increased prominence of block borders, it also helps reduce ringing within each transformed block by limiting the space for overshooting and undershooting. Additionally, the use of integer transforms in newer codecs introduces new artifacts that have not yet been studied [11, 27–28].

D. Multi-View Coding (MVC)

MVC is a 3D view of a scene or part of it obtained by performing multiple views of a scene. However, MVC introduces or enhances many artifacts making quality assessment more difficult as discussed below [11, 27–28].

1) Depth map quantization

- Depth ringing (depth bleeding) artifacts: A depth map, which is compressed similarly to textures, represents the distance between each pixel and the camera. MPEG-4 Part 2 clearly defines the coding of depth maps. Depth ringing (depth bleeding) is produced by the quantization of depth maps, and it is more prominent at steep edges [11, 27].
- Card board (puppet theater) effect: These effects are produced by depth estimation error and harsh quantization, where light and dark colors have different depths. This artifact manifests as 2D layers, rather than smooth depth transitions. Additionally, the coding of the depth map and texture may interfere with each other due to their superposition.

2) Frame packing artifacts

A second type of MVC (available in H.264 & H.265) is a kind of stereoscopic video coding. Interleave coding (a form of frame packing) can lead to artifact crosstalk enhancing many kinds of artifacts. Additionally, side-by-side, top-bottom frame packing, column and row alternation, and checkerboard arrangements can also contribute to crosstalk. In checkerboard setups and frame alternation configurations, color bleeding occurs, and MC mismatches are amplified as a result [11, 28].

3) Artifacts in H.264/H.265 due to MVC

A backwards-compatible method of coding, known as the third type of Multi-View Coding (MVC) in H.264, is available. However, this approach exacerbates Motion Compensation (MC) mismatch artifacts, which require further investigation.

E. Scalable Video Coding (SVC)

It discusses to the process of decoding portions of a bit stream to achieve a lower frame rate, spatial resolution, or video quality.

1) Temporal scalability produced artifacts

Frame rate defines temporal scalability and produces different kinds of artifacts as described in Table V.

2) Spatial scalability produced artifacts

Spatial resolution defines spatial scalability where data is decoded at lower resolutions to reduce the bit rate, with the possibility of achieving higher resolutions. The implementation of spatial scalability differs across MPEG-4 Part 2, and H.264 SVC [27–28].

3) Quality scalability produced artifacts

A special case of spatial scalability, which relies on multiple layers, involves using generated video streams to predict and decode the video at different quality levels. Coarse-grain quality scalability and inter-layer inter-prediction use similar techniques, leading to comparable artifacts. Further research is required to determine if any new artifacts arise when the enhancement layer is removed during the quality scalability process, and whether these artifacts differ from the codec drift seen in MPEG-2 [11, 28].

F. Angular Intra-prediction Produced Artifacts in H.265

This new intra-prediction mode may produce pumping artifacts in H.265, and more research is needed to understand this artifact.

G. Interpolation Filter Produced Artifacts in H.265/HEVC

In H.265/HEVC, a 6-tap directional or a 12-tap DCT-based interpolation filter is used for subsampling, in contrast to the Wiener and bilinear filters employed in H.264. This difference in filtering alters the signal characteristics, potentially introducing new artifacts that require further investigation [27–28].

H. In-loop Filters

In-loop filtering is the last step in encoding which improves the video quality by reducing/eliminating artifacts generated during the encoding process. The in-loop filters complexity has also increased in advanced codecs. Also, the challenges, i.e., connecting in-loop filters in line, usage of different filters, filtering large number of samples, storing samples & intermediate results, etc., are still preventing them from generating the required results up to standards, and artifacts are still present in final coded video [26].

I. Memory Issues

Memory issues really hurts inter prediction process in an encoder, which ultimately is responsible of many other kinds of distortions.

Table VII briefly presents the artifacts generated due to use of new codecs [11, 26, 28, 29].

Next section describes issues/challenges due to 3D coding.

TABLE VII. CHALLENGES DUE THE USE OF NEW CODECS

New coding tools effects/Artifact	Occurrence reason	Relation/effect with other artifacts
Transform sizes & new transforms effects	Fewer transform coefficients are produced in H.264 & H.265 in comparison to DCT; less loss of signal energy	Ringing reduces due to the transform size (4×4) in H.264/H.265, i.e., limited space for over & undershooting; Blockiness may increase due to smaller transform sizes (4×4) in H.264 and H.265 and VC-1; No information available whether integer transform used in H.264 and H.265 produces any kinds of new artifacts or not
Macro block partitioning effects	Precise matching can be achieved by conducting a distinct search for each section of a macroblock	Macro block partitioning increases blocking, MC, & mosquito noise
Large size macro block (32×32 & 64×64) effects in H.265	An increase in macro block size increases coding efficiency	Increase ringing artifacts
Different coding modes effects	Due to different coding modes in subsequent frames	Produces flickering or pumping artifact
Depth ringing or depth bleeding (Multi View Coding-MVC)	Depth maps coding exist in MPEG-4 part 2 (kind of MVC), i.e., 3D representation of a scene obtained by coding of multiple views of a scene	Depth map quantization produces depth ringing or depth bleeding
Card board effect (MVC)	Depth estimation errors and excessive quantization can lead to a "cardboard effect," where lighter colors appear to have more depth than darker colors	Superposition may be produced among depth map & texture coding with combined existence & masking, which may also happen
Frame packing effect (MVC)	Frame packing is the second type of Multi-View Coding (MVC), where both left & right views are encoded used by a single view, utilizing Supplemental Enhancement Information (SEI)	Using checkerboards arrangements (i.e., frame packing, etc.), color bleeding propagates across views; Interleaving of views enhances mosquito noise, pumping and MC mismatch; MC mismatches increases by using frame alternation arrangements; Cross talk artifacts may happen
MVC artifacts in H.264	Backward compatible way of coding creates new artifacts, i.e., MVC uses similarity of many views at any moment	Backwards compatible way of coding may enhance MC mismatch
Artifacts due to temporal scalability	A layered coding (scalable video coding) created by dropping packets to minimize bandwidth; produce smaller frame rate & spatial resolution	Mosquito noise & MC mismatch are introduced; To avoid pumping artifacts, quantization parameter is increased in higher temporal layers; Jerkiness may be introduced due to the dropping of fine temporal resolution layers
Artifacts due to spatial scalability	Spatial resolution defines spatial scalability, where data is decoded at lower resolutions to decrease the bit rate, while also allowing for the prediction of higher resolutions	Blocking artifacts may increase by up-sampling in H.264; MC mismatches, blurring & mosquito noise enhances due to up-sampling
Artifacts due to quality scalability	Special case of spatial scalability, i.e., generated video streams is used to forecast and decode video with different qualities	Produces same artifacts as coarse grain quality scalability; Enhancement layer removal produces drift between encoder & decoder introducing new artifacts
Angular intra prediction mode artifacts (H.265)	New coding mode used in H.265	Pumping artifacts may increase
Artifacts due to interpolation filter (H.265/HEVC)	6-tap directional or a 12-tap DCT based interpolation filter usage in H.265/HEVC generate these artifacts in comparison to Wiener & bilinear filter used in H.264	New filters change features of signals & may introduce new artifacts

VI. 3D CODING ISSUES/CHALLENGES

Three-Dimensional (3D) video is rapidly gaining popularity in both broadcasting and cinema, as it offers viewers a more immersive and captivating visual experience. By adding depth and realism, 3D technology enhances the sense of presence, drawing audiences deeper into the content and making the viewing experience more engaging. Different types of 3D video formats & coding strategies have been currently created and co-exist. However, the 3D video transmission chain has new/different processes as compared to 2D, creating new issues/challenges as shown in Fig. 2 [13, 30].

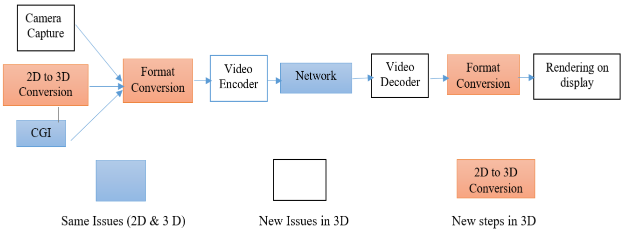

Fig. 2. Block diagram of 3D transmission.

The most detailed and realistic 3D representations are created using Computer-Generated Imagery (CGI), which demands significant computational power. However, very little information can be achieved about a 3D structure from a 2D still image. Few existing algorithms use monocular cues, i.e., focus information, texture or shades, etc., to extract information about 3D from 2D to provide a depth map (2D plus depth). However, information derived from 2D images only reveals the first object in the line of sight (in a monocular view), which creates challenges when a second perspective is introduced or when hidden portions of an object need to be recovered. This requires techniques like in-painting algorithms to restore the occluded areas of the object [13].

A. Signal Acquisition & Format Conversion in 3D

When capturing stereo content, two cameras are used to generate a stereo pair. However, this process often introduces various artifacts, such as vertical and color misalignments, differing focus points, temporal offsets, and geometric distortions. Similarly, converting 3D content between formats can be challenging. For a lossless conversion, the output must be a subset of the input, such as when converting from MVD (Multiple Video plus Depth) to a single view plus depth format. However, in many cases, conversion introduces noticeable distortions. For example, when assessing a depth map from stereoscopic video, the ambiguity of features between the two images can lead to significant errors, such as background objects being incorrectly placed in the foreground. Similarly, while reconstructing the original stereoscopic view, the depth perception often contains numerous inaccuracies [13, 30].

B. 3D Signals Transmission

For compression, transmission, and storage of 3D video, Multi-View Coding (MVC) is commonly used, such as in the standard format for Blu-ray Discs. Binocular rivalry can occur when 3D video signals are transmitted independently or in combination. Additionally, information loss may occur during format conversions in 3D video transmission. Various stages in the 3D transmission chain—such as camera capture, conversion, or rendering—can introduce different types of artifacts, including geometric distortions or the "puppet theater" effect. These artifacts make it more challenging to accurately assess video quality in 3D content [13, 30].

C. 3D Signals Display

Format conversions also occur on the display side, particularly when a 2D plus depth representation is transmitted or a different viewpoint is used. Typically, Depth Image-Based Rendering (DIBR) is employed to generate separate images for left & right eyes. However, the displayed viewpoint often differs from the transmitted view, requiring in painting algorithms to fill in the previously occluded regions. This process can introduce artifacts, especially around the edges of the reconstructed areas [30].

D. 3D vs 2D

3D video quality subjective assessment is less advanced than 2D, primarily due to the lack of uniform procedures for calculating display features and the challenge of characterizing the various dimensions of perceived quality in 3D. Similarly, objective quality measurement for 3D video is more complex than for 2D, due to additional transmission steps, artifacts, and other factors. 3D video content has a greater effect on perceived visual quality as compared to 2D, with key issues including object clipping by the frustum, which causes visual discomfort, camera movements that destabilize the sense of orientation and fast-moving objects in the foreground. Furthermore, limited research has been conducted on understanding content and visual attention in 3D videos. Studies suggest that observers generally pay attention to certain sections of interest within an image, leading to increased interest in using visual attention models, such as saliency maps, for more accurate quality assessments in 3D [13, 30].

Moreover, 2D objective quality measurement techniques are not applicable directly to 3D as there is no trustworthy subjective datasets for 3D video. This is due to the fact that in 2D videos, an objective score is directly correlated with subjective value, but in 3D it is the rendered version and not the signal itself which needs to be analyzed. Similarly, a multidimensional observer's opinion (visual fatigue & depth perception, etc.) along with more HVS aspects (e.g. binocular rivalry, binocular suppression, etc.) should be taken into account [13, 30].

Next section describes Issues/challenges due to subjective databases.

VII. Issues/Challenges Related to Subjective Databases

The research community has developed numerous subjective databases to establish correlations with objective quality metrics. In 2005, the LIVE image quality database was created, comprising twenty-nine original images and seven hundred & seventy-nine distorted images representing five different distortion kinds. The SSCQ method was used to develop this database. To ensure consistency, Mean Opinion Score (MOS) scaling was applied to account for variations in rating scales across different subjects, and realignment was performed to prevent significant bias in MOS values due to specific distortion types or levels [31].

The TID2008 was developed using 24 originals and 1 computer, and it used 24 originals and 1 computer-generated image. In this database, 18 images were taken from LIVE, and only a difference of sizes was created via cropping. Seventeen hundred distorted images were created in this way, incorporating seventeen different kinds of distortions, each with four distortion levels. The paired comparison method was used to develop this database without the application of explicit MOS scaling or realignment. The largest public database (TID2013) to date was created by extending TID2008 with 7 new distortion types and 1 extra distortion level. Similarly, the CSIQ database was developed with thirty original images and eight hundred & sixty-six distorted images, covering 6 distortion kinds and 4 to 5 distortion levels. This database was created using the multi-stimulus absolute category rating method [32].

The LIVE Multiply Distorted (MD) database and the LIVE Wild Image Quality Challenge database (LIVE Challenge) were created by a mix of image distortions. LIVE MD contains fifteen original images and four hundred & five distorted types. The LIVE Challenge database (total 1162 images) expands this by including distorted images taken from mobile devices. MOS scores for this database were crowdsourced using the Amazon Mechanical Turk platform. In addition to these, several smaller databases, such as IVC, Toyama-MICT, Cornell A57, WIQ, ECVQ, and EVVQ, have also been developed by the research community [33].

A major problem with all these databases is the inclusion of a limited number of source images, i.e., none of the databases has more than 30 source images. These databases do not cover all types of contents, distortions, or the full variety of real-world images (except for cartoon videos in EVVQ database), raising concerns about the reliability of existing objective Image Quality Assessment (IQA) models. For instance, test results have shown that the efficiency of some of the most accurate No-Reference IQA (NR-IQA) models considerably declines when evaluated using the LIVE Challenge database. The limited scope of available subjective testing is primarily due to the high costs and time requirements associated with conducting such tests. For example, evaluating 1700 distorted images in the TID2008 database is expensive & time-consuming, even when using just twenty-five original images [31–33].

Moreover, viewing conditions, i.e., screen category & resolution, room arrangement, lighting, screening distance, and other experiments-related conditions, etc., create a major challenge for developing a standard subjective database. Similarly, subject age, gender, and qualifications (generally university students are taken as viewers) are also problems, and different results are produced while developing a database.

The LIVE database primarily contains videos where Temporal Information (TI) increases linearly with Spatial Information (SI), whereas the ECVQ & EVVQ databases feature videos that include both low TI with high SI and high TI with low SI. Additionally, ECVQ and EVVQ databases are generated in different resolutions compared to LIVE. As a result, there is a significant need for a new database containing a sufficient amount of source videos with a broad diversity of contents and varying levels of spatial and temporal activity. However, testing such a database using traditional subjective methods is extremely challenging, if not impossible. Therefore, innovative and reliable approaches must be developed to evaluate and create these subjective databases, as seen in [31–33].

Next section describes Issues/challenges due to Quality of Experience (QoE).

VIII. Video Quality Measurement Challenges & Quality of Experience (QoE)

It refers to the level of satisfaction or dissatisfaction a user experiences when using an application or service. QoE reflects viewer's overall contentment with the quality of a streamed video. QoE is strongly linked to subjective and objective quality measurement methods as shown in the Fig. 3 [34]. The assessment of QoE provides an overall gauge of the subjective/objective quality metrics performances and associated issues/challenges [35–36].

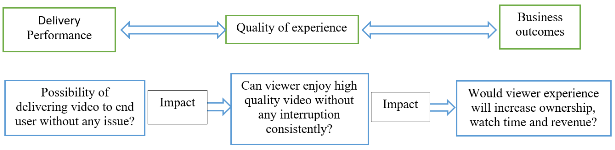
Delivery Performance
Quality of experience
Business outcomes
Possibility of delivering video to end user without any issue?
Impact
Can viewer enjoy high quality video without any interruption consistently?
Impact
Would viewer experience will increase ownership, watch time and revenue?

Fig. 3. Relation between quality measurement & QoE.

Research has shown a 10.4% increase in viewer's engagement compared to a lower-quality resolution video. If a video quality is not assessed accurately, there is an enormous chance that the video provider will lose business as viewers will switch to some other source provider. While measuring QoE, it is not only testing the performance of subjective & objective video quality metrics, but many other factors are tested which contribute towards QoE, i.e., content providers, service & network providers, delivering, measuring, and controlling QoE, etc. [14, 36–38]. There are many QoE Influencing Factors (IFs) that can interrelate and should not be understood as an isolated entity, as shown in Table VIII [36–39]. These factors play a vital role and present a great challenge when measuring video quality accurately. Therefore, understanding these IF (challenges/issues) also greatly helps to estimate the video quality more accurately.

TABLE VIII. QoE Influencing Factors

Type	Explanation	Examples
Influencing Factor related to human	Any characteristic of a human user.	Demographic, socioeconomic background, physical & mental composition, viewer's emotional conditions such as attractiveness, enjoyment, expectations, etc.
Influencing Factors related to system	Properties & characteristics defining quality of a service technically.	Capturing media, communication, storing, reproduction, end-to-end transfer of information, Network quality, jitter, latency, consistency/ obtainability of server, content associated (e.g., color, 2D/3D), media associated (e.g., synchronization, sampling rate), & device associated (e.g., personalization, security, privacy, effectiveness/efficiency), etc.
Influencing Factors related to context	Factors describing user's environment, i.e., physical, temporal, social, economic, etc.	Temporal (e.g., usage frequency, time), physical (e.g., location), economic, technical (e.g., networks availability), tasks & social (e.g., groups) etc.
Influencing Factors related to contents	Information related to the contents provided by the service provider.	Video codec, video format/resolution, duration, motion patterns, type & contents, etc.

There are many other issues/challenges related to QoE due to its multidisciplinary and subjective nature, i.e., difficulty in getting access to operator's network data & traces in realistic environments; no open source video database availability, improve & control QoE, etc. These issues/challenges also influence a lot while assessing image/video quality accurately as described.

A. QoE-Aware Encoder & Decoder Optimization

Optimizing encoders and decoders remains a key challenge, particularly in the context of QoE-aware image and video compression. This includes effective rate control, compression-introduced artifacts, complexity, transcoding between different standards & time complexity in line with emerging standards while maintaining the efficiency and performance for both QoE & while assessing video quality [36–38].

B. Robust, Realistic & Continuous Time-Varying QoE Model

All QoE influencing factors are not taken into account by existing QoE models to provide a robust, realistic & continuous time model. A comprehensive CTVQ model should be developed to consider spatial and temporal artifacts, perceptual saturation and smoothing, as well as hysteresis effects. Human response should be observed at each interval to build a comprehensive dataset and to evaluate such models. More work is needed to develop a cross-layer QoE approach for deployment in next-generation networks. This is particularly important for time-varying QoE modeling and prediction [36–38].

C. Human Centralization

Future applications such as Virtual & Augmented Reality (VR & AR), Mission Critical (MC), Tactile Internet (TI), and teleportation will demand significant resources to ensure a high-quality QoE for end users. These technologies may lead to issues like motion sickness, addiction, discomfort, and eyestrain, all of which need careful consideration [37, 39].

D. Standardization Efforts

Different QoE models show different performance under different realistic conditions, i.e., correlation with different MOS, which shows all real-time constraints and aspects are not taken into account. Therefore, standardization is required so that QoE models can be used efficiently, keeping in view the user's expectations.

Next section describes challenges in real time processing.

IX. Real-Time Processing for Video Quality Metrics

Real-time streaming refers to the continuous transmission and immediate playback of audio, video, or multimedia content over the internet with minimal latency. To ensure smooth delivery and playback across diverse network conditions and devices, real-time streaming technologies employ protocols such as Real-Time Messaging Protocol (RTMP), HTTP Live Streaming (HLS), and Dynamic Adaptive Streaming over HTTP (DASH). However, real-time streaming introduces several challenges, including latency, network congestion, and scalability requirements. One major issue arises from real-time transmission over protocols like User Datagram Protocol (UDP), which, while low-latency, can lead to quality degradation such as frame drops, flickering, and packet loss. Addressing these problems demands robust error recovery mechanisms and real-time Video Quality Assessment (VQA) metrics capable of monitoring and adapting to user experience in dynamic environments [40–41].

In live streaming, videos must be encoded, transmitted, and decoded almost instantly, leaving minimal time for quality evaluation or optimization. This constraint is particularly critical in high-motion content, such as sports broadcasts, where distortions like stuttering, motion blur, and de-interlacing artifacts can occur under constrained bandwidth. Moreover, the complex and abrupt temporal variations in such content challenge conventional VQA models, which often fail to capture the degree of real-time distortion. As a result, VoD-optimized VQA techniques frequently fall short when applied to live-streaming scenarios. Therefore, content providers deploy advanced streaming infrastructures, leverage Content Delivery Networks (CDNs), and apply real-time optimization strategies to address these issues [17, 40].

Even, if a quality metric demonstrates high accuracy while assessing video quality, but requires significant processing time to evaluate the quality of the received video, it may not be suitable for real-time applications. Therefore, highly optimized video quality algorithms with low computational complexity are essential to meet the real-time streaming demands while maintaining high accuracy. These algorithms can also provide feedback to enhance quality when received video is degraded. One such real-time quality monitoring tool is the Real-Time Monitor (RTM), which enables continuous quality measurements in live streaming scenarios, and offers real-time feedback to the content source to enable corrective actions. Additionally, the Haar Wavelet-based Perceptual Similarity Index (HPSI) is a novel and computationally efficient full-reference IQA metric. The fast computation time of HPSI makes it highly beneficial for real-time quality prediction, allowing it to evaluate video quality effectively while minimizing processing delays [40, 42].

Similarly, Training and deploying large AI models require substantial computational resources in real-time processing. Beyond the inherent complexity of the models themselves, the efficiency of the hardware used is pivotal in determining the overall speed and accuracy of quality predictions. Additionally, the relationship between power consumption and video quality in streaming contexts is important. It is crucial to understand how power-saving methods impact visual quality, particularly in the context of emerging technologies that aim to optimize both energy use and performance [41].

Table IX [23] presents the computational efficiency of eight Full-Reference (FR) quality metrics. The three fastest metrics are highlighted in bold, indicating their ability to assess video quality with high efficiency as mentioned in Table III in real-time streaming scenarios. These metrics demonstrate that achieving high-quality predictions does not necessarily require sacrificing computational speed, making them suitable for practical deployment in time-sensitive applications.

TABLE IX. COMPUTATION TIME COMPARISON OF IQA METRICS

Database	FSIMc	MDSI	HPSI	VCGS	SPSIM	LGV	SWLGV	GMSD
TID2008	0.08	**0.014**	0.02	0.21	0.12	0.43	5.98	0.019
TID2013	0.10	**0.016**	0.03	0.21	0.11	0.33	6.13	0.019
KADID-10k	0.13	**0.026**	0.04	0.29	0.14	0.45	5.89	0.028
PIPAL	0.17	0.01	0.01	0.09	0.17	0.61	2.62	**0.009**

A. Edge-Based Live Streaming

As demand for high-definition content across a wide range of devices continues to grow, traditional approaches to live video streaming face significant challenges, as mentioned earlier, particularly with bandwidth limitations and network congestion. To address these issues, edge and cloud computing integration has become increasingly critical for ensuring seamless user experiences with high quality and accuracy. Edge computing plays a pivotal role by decentralizing computational resources and strategically placing them closer to the end user, at the network edge, rather than relying solely on centralized data centers, i.e., reducing latency and enhancing responsiveness during video playback, etc. This approach not only mitigates the effects of network congestion but also improves the overall quality of service, making it particularly valuable for real-time streaming [43].

Moreover, edge nodes can dynamically adapt to fluctuating network conditions and viewer preferences by integrating machine learning algorithms and predictive analytics, ensuring optimal content delivery in real-time. This capability allows edge computing to significantly enhance Quality of Service (QoS) by optimizing content distribution, reducing network congestion, and intelligently adjusting streaming parameters based on network performance and device capabilities. Therefore, edge computing is set to redefine the landscape of streaming services by improving efficiency, scalability, reliability, and accuracy. Furthermore, edge computing can alleviate the strain on centralized data centers by handling much of the data processing locally, leading to reduced energy consumption, costs, and environmental impact while simultaneously enhancing the user experience through faster and more responsive streaming [43, 44].

Issues/Challenges in edge-based streaming: Edge computing also has several issues/challenges, as mentioned [41, 43, 44]. One of the primary concerns is the constrained computational, storage, and network resources available at the edge. 2) Unlike the high-speed, low-latency networks typically found in cloud data centers, edge computing operates across various wired and wireless networks with differing bandwidth, reliability, and latency characteristics. These network variations can pose significant obstacles in delivering consistent and high-quality video streaming, as well as complicating the implementation of streaming video analytics that requires real-time processing. 3) The diverse range of edge devices introduces significant hardware heterogeneity, making it challenging to develop a universal solution 4) Resource-intensive deep learning models that perform efficiently in cloud environments may encounter significant performance issues at the edge due to the limited processing power, memory, and storage available on edge devices. 5) Shared ownership of devices: Another challenge unique to edge computing is the issue of shared ownership of edge devices, leading to conflicts and inefficiencies. 6) Securing edge computing systems poses a greater challenge than cloud-based environments, as devices are more vulnerable to external attacks. 7) Budgetary constraints: From a financial perspective, edge infrastructures often operate under tight budgetary constraints compared to cloud infrastructures maintained by well-resourced companies. 8) Edge computing spans a wide range of devices, from small edge devices to micro data centers (EDCs). However, EDCs, which host significant IT resources, are particularly vulnerable to higher temperatures due to the density of computational hardware. 9) Devising strategies for environmentally and economically sustainable deployment in edge computing is a considerable challenge. 10) Energy-aware, fault-tolerant operation is crucial in edge computing, as it allows automatic operation even under varying conditions while maintaining performance. 11) For edge computing to be effective in real-time video streaming and data analytics, the edge nodes must be capable of achieving high throughput and reliability under diverse workloads, i.e., able to accommodate additional workloads, etc. 12) Scalability is another significant issue, especially when operating in environments with a large number of edge devices. 13) Edge computing heavily relies on network connectivity to transfer data between edge devices and centralized servers, but it can be unreliable, often subject to latency and bandwidth limitations due to the heterogeneous nature of the underlying infrastructure. 14) Edge devices typically have limited computational and storage resources, complicating resource allocation and allocating full computational power required for demanding tasks like real-time video processing, machine learning, or complex analytics [41, 43, 44].

B. Cloud-Based Live Streaming

Cloud computing offers a robust and scalable solution for real-time video streaming and live streaming sports events like FIFA, YouTube TV, and more. Cloud companies have large-scale data centers equipped with thousands of servers, massive storage capacities, AI, machine learning framework, high-speed networks, and dedicated operational and maintenance teams. By leveraging the vast resources of cloud infrastructure, video streaming platforms can efficiently deliver high-quality content to a global audience. The key benefits of using cloud computing for video streaming include Scalability, Content Archiving, Content Delivery, Encoding and Transcoding, Cost-Effective Broadcasting, Operational Excellence, etc. [15, 45].

Cloud-based live streaming performance monitoring ensures high-quality, uninterrupted viewing experiences. Providers can quickly identify and resolve issues that may affect the stream by monitoring quality metrics in real-time. Several tools are available to assess the quality of received cloud-based live streaming as described [15, 45]. 1) Video player platforms with built-in analytics offer valuable insights for monitoring user engagement, video completion rates, and other essential quality metrics, ensuring an optimal viewing experience for users. 2) Network Performance Monitoring (NPM) tools like Pingdom and PRTG are crucial for identifying issues or bottlenecks that could impact the quality of live streaming. 3) Content Delivery Networks (CDNs) are critical for ensuring video content's efficient and smooth delivery, especially for real-time streaming applications. 4) Software-oriented video quality monitoring programs such as Structured Similarity Indexed Metric (SSIM) Wave, StreamEye, and Video Multimethod Assessment Function (VMAF) evaluate video stream quality and spot any problems that can degrade the viewing experience. 5) Tools like Sentry and Rollbar track faults and find errors during live streaming service. 6) Real-time Monitoring Dashboards are essential for ensuring that live streaming services run smoothly by providing immediate insights into Key Performance Indicators (KPIs) [15, 45].

TABLE X. COMPUTATION OF DEEP LEARNING MODELS FOR CLOUD BASED VIDEOS STREAMING

Algorithm	Model Architecture	Performance
DeepConvLSTM	CNN-LSTM	99.2% accuracy
DeepMIML	CNN	96.2% accuracy
C3D	3D CNN	94.9% accuracy
DeepVS	CNN	97.4% accuracy
ST-ResNet	CNN-ResNet	96.3% accuracy
MS-TCN	TCN	96.8% accuracy
ResLSTM-NN	CNN-LSTM	93.5% accuracy
Efficient Net-FC	CNN	97.8% accuracy

TABLE XI. EDGE VS CLOUD COMPUTING

Parameter	Edge Computing	Cloud Computing
Processing location	Edges of the network	At a central location
Bandwidth Requirements	Low	High
Cost	More expensive	Less expensive
Scalability	Challenging	Easily
Applications	Low latency & real-time decision-making (e.g., IoT devices, medical robotics, AVs, & AR/VR systems, etc.)	No strict latency requirements (e.g., web applications, email, and file storage, etc.)
Security	Reliable	Challenging

Issues/Challenges in cloud-based streaming: Cloud-based live streaming has its own set of challenges such as latency, bandwidth, scalability, security, price, management of multiple cloud service providers, increased migration of Cloud, adaptation of cloud platform, locking with one vendor, etc. By leveraging advanced solutions like CDNs, adaptive bitrate streaming, video compression, and machine learning algorithms, content providers can effectively manage issues related to latency, bandwidth, scalability, security, and costs. Table X presents a comparison of various deep learning model architectures for cloud-based video streaming in terms of their performance [15].

Table XI presents a comparison between edge & cloud computing [15, 43].

Next section describes issues/challenges related to deep learning, machine learning, Artificial Intelligence Generated Contents (AIGC), and Artificial Intelligence Enhanced Contents (AIEC).

X. ISSUES/CHALLENGES RELATED TO DEEP LEARNING, ML, AIGC & AIEC

In recent years, deep learning and machine learning based methods have significantly advanced the Video Quality Assessment (VQA) field. These modern techniques generally outperform traditional visual computing metrics by learning complex representations directly from data, which makes them highly effective across diverse content types and distortion patterns [46–47].

Deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are widely used for evaluating video quality. CNNs excel at interpreting spatial information by preserving pixel correlations, making them effective in analyzing individual video frames. In contrast, RNNs are skilled at modeling temporal dependencies, capturing how video content evolves over time. To simultaneously leverage both spatial and temporal features, 3D Convolutional Neural Networks (3D CNNs) are employed. These networks perform convolutions across both spatial dimensions and time, allowing for the extraction of spatio-temporal features, which leads to more accurate video quality predictions. Temporal Convolutional Networks (TCNs) are employed in video prediction tasks due to their ability to model long-term temporal dependencies through causal and dilated convolutions [42]. Long Short-Term Memory Networks (LSTMs), a specialized form of RNNs, are also widely used in video prediction as they effectively retain and recall information over extended sequences, enabling accurate modeling of temporal dynamics. Similarly, Generative Adversarial Networks (GANs) are increasingly applied in video prediction and reconstruction, leveraging their generative capabilities to produce realistic and temporally coherent video frames. In the context of real-time applications, Reinforcement Learning (RL) has shown promise in dynamic video streaming and quality control, where it enables algorithms to optimize decisions through trial and error, learning effective policies for adapting to network conditions and user preferences [42, 46].

Table XII shows the performance comparison of full reference IQA/VQA, deep learning, and temporal pooling models on FR video quality assessment databases. These metrics have been compared with respect to SROCC/PLCC, and best result is shown in bold and italic [42].

TABLE XII. COMPARISON OF FR IQA/VQA MODELS

Type	**Model**	**User Generated Content**			**AI Generated Content**	
		LIVE-VQC	**MCL-V**	**LIVE-YT-HFR**	**BVI-HFR**	**BVI-VFI**
Knowledge-driven (IQA/VQA)	PSNR	0.69/0.74	0.54/0.54	0.78/0.74	0.25/0.31	0.52/0.47
	SSIM	0.72/0.78	0.71/0.70	0.55/0.54	0.19/0.35	0.60/0.54
	VIF	0.68/0.76	0.74/0.74	0.68/0.70	0.25/0.26	0.53/0.48
	VMAF	0.81/0.81	0.82/0.83	0.77/0.74	0.18/0.37	0.59/0.56
	ST-GREED	0.68/0.70	0.78/0.79	***0.88/0.88***	***0.80/0.83***	0.11/0.21
Deep learning (IQA)	DeepQA	0.86/0.86	0.63/0.63	0.08/0.31	0.02/0.30	0.44 /0.46
	LPIPS	0.52/0.56	0.76/0.75	0.69/0.70	0.23/0.37	0.59/0.59
	DISTS	0.45/0.48	0.79/0.78	0.72/0.72	0.40/0.67	0.50/0.55
	TOPIQ	0.70/0.73	0.73/0.72	0.12/0.30	0.00/0.28	0.23/0.39
Temporal pooling /NN modile (VQA)	FloLPIPS	0.54/0.56	0.54/0.58	0.07/0.16	0.10/0.24	***0.68/0.70***
	DeepVQA	0.91/0.89	-	0.43/0.39	0.14/0.20	0.50/0.35
	C3DVQA	***0.92/0.91***	0.78/0.79	0.73/0.74	-	-
	STRA-QA	-	***0.85/0.86***	0.79/0.80	-	-

The table indicates that VMAF achieves the highest SROCC and PLCC scores on general distortion databases. However, its performance significantly drops on temporal distortion datasets. In contrast, ST-GREED demonstrates a better correlation with subjective scores on databases that include temporal distortions. Deep learning-based Image Quality Assessment (IQA) models, such as LPIPS, generally perform well on standard distortion datasets but struggle with temporal distortions, particularly on High Frame Rate (HFR) and Video Frame Interpolation (VFI) datasets. Incorporating temporal pooling strategies into deep learning-based VQA models such as in FloLPIPS improves their correlation with human perception compared to both knowledge-driven and basic deep learning IQA models. For instance, STRA-VQA achieves strong performance on temporally complex databases like MCL-V and LIVE-YT-HFR [42].

TABLE XIII. COMPARISON OF NR IQA/VQA MODELS

Type	**Model**	**User Generated Content**			**AI Generated Content**	
		LIVE-VQC	**KoNViD-1k**	**YouTube-UGC**	**T2VQA-DB**	**GAIA**
Knowledge-driven (IQA/VQA)	BRISQUE	0.59/0.62	0.65/0.65	0.38/0.39	0.18/0.25	0.09/0.19
	NIQE	0.58/0.62	0.54/0.55	0.23/0.27	0.01/0.20	0.06/0.21
	TLVQM	0.79/0.79	0.77/0.76	0.66/0.65	0.48/0.49	0.46/0.47
	VIDEVAL	0.71/0.72	0.78/0.78	0.77/0.77	0.52/0.53	0.46/0.48
	FAVER	0.78/0.79	0.79/0.79	0.73/0.73	0.51/0.53	0.20/0.26
Deep learning (IQA)	PaQ-2-PiQ	0.64/0.66	0.61/0.60	0.26/0.23	0.15/0.14	0.22/0.23
	DB-CNN	0.63/0.71	0.71/0.73	0.47/0.52	0.01/0.05	0.17/0.17
	MUSIQ	0.62/0.71	0.72/0.74	0.52/0.56	0.07/0.06	0.15/0.16
	CLIP-IQA+	0.72/0.77	0.78/0.78	0.53/0.58	0.07/0.13	0.15/0.16
2D CNNs with simple score/feature averaging (VQA)	RAPIQUE	0.72/0.75	0.80/0.81	0.75/0.76	0.31/0.45	0.27/0.32
	SION	0.73/0.78	0.81/0.81	0.36/0.39	0.24/0.25	0.12/0.15
2D CNN with temporal aggregation networks (VQA)	VSFA	0.69/0.74	0.77/0.77	0.72/0.74	0.10/0.11	0.50/0.52
	BVQA	0.76/0.83	0.81/0.83	0.77/0.79	-	-
3D CNN/Transformation (VQA)	BVQA-T	0.83/0.84	0.83/0.83	0.81/0.82	0.73/0.74	0.52/0.52
	Shen *et al.*	0.76/0.76	0.79/0.78	0.77/0.76	0.27/0.31	0.08/0.14
Transformer (VQA)	FAST-VQA	0.82/0.83	0.85/0.85	0.86/0.86	0.71/0.72	0.52/0.54
	DOVER	0.79/0.83	0.87/0.88	0.87/0.87	0.76/0.76	***0.53/0.55***
	SAMA	***0.86/0.87***	0.89/0.89	0.88/0.88	0.01/0.04	0.23/0.24
Large Multimodality Models (LMMs) (VQA)	COVER	0.80/0.84	0.89/0.89	***0.91/0.91***	0.12/0.24	0.22/0.23
	MaxViTQA	0.85/0.87	***0.89/0.89***	0.89/0.89	0.22/0.25	0.25/0.25
	q-Align	0.77/0.83	0.86/0.87	0.83/0.84	***0.76/0.77***	-
	LMM-VQA	0.83/0.86	0.87/0.87	0.85/0.87	-	-

Similarly, Table XIII presents a performance comparison of No-Reference (NR) IQA/VQA models, including both traditional and deep learning-based approaches. These metrics are evaluated using SROCC and PLCC scores across five different NR video quality assessment datasets focused on User-Generated Content (UGC) and AI-Generated Content (AIGC). The results indicate that traditional models such as BRISQUE and NIQE struggle in dynamic video environments, showing limited ability to capture temporal and content complexities. In contrast, deep learning-based IQA models, which incorporate deep feature representations and end-to-end training, exhibit improved performance on UGC datasets. However, knowledge-driven VQA models such as Two Level Video Quality Model (TLVQM), Video Quality Evaluator (VIDEVAL**)**, and Frame Rate Aware

Video Quality Evaluator w/o Reference (FAVER) often outperform both handcrafted and deep learning-based IQA models across both UGC and AIGC datasets. Notably, the evaluation shows that in NR scenarios, traditional IQA methods and even standard deep learning-based metrics face challenges when applied directly to videos, especially in cases involving complex motion or AI-generated content. In this context, CLIP-IQA+ demonstrates superior performance on UGC datasets, benefiting from the vision-language priors embedded in the CLIP model, which help capture high-level semantics relevant to human perception [42].

Models such as RAPIQUE, SIONR, VSFA, and 2BiVQA, which rely on simple temporal pooling strategies and basic temporal aggregation networks, demonstrate improved performance on UGC datasets. However, their performance deteriorates significantly when applied to AI-generated content, largely due to their limited ability to capture complex visual-text alignment and temporal coherence. In contrast, BVQA models utilizing 3D CNNs show substantial improvements on UGC datasets and maintain reasonable performance on AIGC, owing to their spatiotemporal feature learning capabilities. Notably, Transformer-based and Large Multimodal Models (LMMs) exhibit strong performance across both UGC and AIGC datasets, benefiting from their capacity to model long-range temporal dependencies, high-level semantics, and cross-modal relationships [42].

Tables XIV presents a comparative analysis of various FR VQA algorithms across different kinds of distortions combined together. The results indicate that general-purpose FR Image Quality Assessment (IQA) metrics, such as SSIM, perform poorly in the context of video when compared to video-specific metrics like MOVIE, which better account for temporal features. This underscores the importance of temporal feature extraction in effective video quality assessment. Moreover, deep learning-based VQA models, such as Deep VQA, outperform traditional metrics because they can learn spatio-temporal representations using deep neural networks [46].

TABLE XIV. COMPARISON OF FR VQA ALGORITHMS ON LIVE VQA DATABASE

FR Metrics	SRCC	PLCC
	All Distortions (Wireless, IP, H.264, MPEG-2)	All Distortions (Wireless, IP, H.264, MPEG-2)
PSNR	0.69	0.74
SSIM	0.72	0.76
VIF	0.6861	0.7543
STMAD	0.8301	0.8714
ViS3	0.8651	0.9023
MOVIE	0.7895	0.8123
V-BLINDS	0.8569	0.8716
SACONVA	0.8569	0.8716
DeepQA	0.8687	0.8904
DeepVQA	0.9152	0.8592

Table XV presents the performance of various NR VQA algorithms across three combined databases, KoNViD-1k, LIVE-VQC, & YouTube-UGC. The results reveal that traditional NR IQA models, such as NIQE, consistently exhibit relatively poor performance across all datasets. In contrast, more advanced models like HOSA demonstrate moderate improvements, indicating some advancement within traditional approaches. However, deep learning-based NR IQA models, particularly KonCept512, achieve significantly higher performance, outperforming both traditional IQA and early VQA-specific models. Notably, the most recent architectures such as Simple VQA and Fast VQA show remarkable performance, further validating the shift towards deep learning in video quality assessment [42].

TABLE XV. COMPARISON OF NR VQA ALGORITHMS ON THREE DATABASES

NR Metrics	All-Combined (KoNViD-1k, LIVE-VQC, and YouTube-UGC)		
	SRCC	PLCC	RMSE
NIQE	0.46	0.47	0.61
BRISQUE	0.59	0.59	0.54
GM-LOG	0.58	0.59	0.53
HIGRADE	0.73	0.74	0.46
FRIQUEE	0.73	0.74	0.46
CORNIA	0.69	0.69	0.49
HOSA	0.69	0.68	0.50
KonCept512	0.73	0.73	0.48
PaQ-2-PiQ	0.68	0.69	0.50
V-BLINDS	0.61	0.70	0.51
TLVQM	0.72	0.72	0.47
VMEON	0.25	0.25	0.66
VIDEVAL	0.79	0.73	0.42
RAPIQUE	0.80	0.82	0.39

Based on the above discussion, the quality metrics performance is represented in Fig. 4 [17].

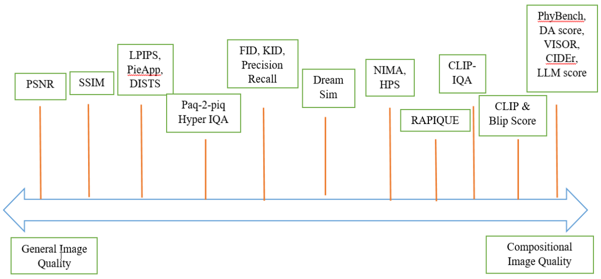

Fig. 4. IQA/VQA models accuracy of assessing image/video quality.

A. Issues/Challenges Related to Deep Learning

1) Modern deep VQA models often follow a two-stage training strategy: pre-training on auxiliary tasks with abundant data, followed by fine-tuning on specialized video quality databases. Although these databases are becoming more comprehensive, their limited scale continues to hinder the training of large deep learning models, primarily due to the high cost and complexity of conducting large-scale subjective quality assessments [47, 48]. This is further constrained by the scarcity of large, annotated psychometric datasets and the incomplete understanding of human visual quality perception. 2) Deep learning-based metrics are resource-intensive, limiting real-time applications. 3) CNN-based methods have reliability issues, i.e., vulnerability to adversarial attacks, etc. In such cases, a Processed Video Sequence (PVS) with a low Video Quality Metric (VQM) value could be slightly modified, resulting in the model incorrectly assigning a

high VQM value (e.g., 5.0), despite no perceptual improvement. 4) Deep learning-based NR methods may be biased toward the specific characteristics of their training data. This bias can undermine the generalizability of the model, leading to unreliable performance on unseen content. 5) Understanding which specific spatial/temporal features or elements of the images or videos have the greatest influence on the final quality score produced by quality models remains a significant challenge. This lack of transparency makes it difficult to improve the performance of the deep learning models. 6) There is a growing interest in using Large Multimodal Models (LMMs) to assess image quality by incorporating detailed descriptive information. This information has the potential to serve as the foundation for more sophisticated scoring systems by capturing the complex nature of images and videos. However, the challenge remains in acquiring those detailed descriptions to create scoring mechanisms that more accurately align with human subjective impressions of quality [47, 48].

B. Issues/Challenges Related to Machine Learning (ML)

Despite the widespread adoption and proven effectiveness of Machine Learning (ML) in Video Quality Assessment (VQA), there are several notable drawbacks in Machine learning approach also. Firstly, it remains unclear how many features should be included in the model, leading to challenges in determining the most effective inputs. Additionally, ML introduces a level of complexity, and risk of overfitting, where the model becomes too tailored to the training data and loses its ability to generalize to unseen content. These challenges become especially apparent in complex mapping functions. This creates a trade-off between performance, where complex models may offer higher accuracy and interpretability, where simpler models, although potentially less effective, may be easier to understand and explain [49, 50].

C. Issues/Challenges Related to AIGC & AIEC [17]

The rapid advancements in machine learning have transformed the landscape of AI-Generated Content (AIGC). Similarly, methods deploying deep learning networks, such as CNNs, GANs, and Auto-encoders have significantly advanced image reconstruction (AIEC) and other related tasks. However, unlike traditional human-created media, AIGC & AIEC also introduce new novel distortions as described:

AI-Generated Content (AIGC) introduces new distortions such as Uncanny Valley, unrealistic object placements, and in the case of video, a lack of temporal coherence, where frame sequences may fail to maintain logical or perceptual continuity. Traditional IQA/VQA algorithms are unable to deal with these new kinds of distortions [17, 42].

New types of artifacts introduced by AIGC also affect various aspects of the perceptual, technical, and aesthetic quality of images/videos. These artifacts are often distinct from those seen in human-created content, making it difficult for existing IQA/VQA methods to effectively capture and evaluate them effectively [17, 42].

AI-generated content may also produce artifacts, such as extra limbs or impossible animals, that do not fall into the usual understanding of "distortions". Generative AI (GenAI) models introduce "generative" distortions, the appearances of which do not comport with commonly encountered video degradations. Evaluating AI-generated content is challenging, as it involves not only assessing technical quality, but also judging common sense, semantic accuracy, and composition. Therefore, need for GenAI content evaluation models is necessary which are able to measure not only technical quality, but also other aspects such as alignment with the input prompts, semantic correctness, biases, and aesthetics, among others [17, 42].

One particularly useful metric that has gained attention is the CLIP score, which measures the alignment between image content and textual descriptions. However, recent research has uncovered discrepancies between CLIP scores, FID (Fréchet Inception Distance) scores, and actual human preferences, highlighting the need for further refinement in these metrics to better align with subjective human judgment [17, 42].

Existing AIGC IQA and VQA subjective datasets lack standardized methodologies for the design and execution of subjective studies specifically focused on the quality of AI-generated content. However, the quality of AI-generated visual content often diverges significantly from that of natural content, necessitating the development of specialized datasets and quality prediction models tailored to AIGC [17, 42].

Existing AIGC models primarily focus on assessing attributes like aesthetics, statistical naturalness, and fidelity, but these metrics are inadequate for thoroughly analyzing Generative Artificial intelligence (GenAI) contents. They fail to capture other relevant attributes and often perform poorly when applied to AI-generated videos and images. However, recent advancements in Vision-Language Models (VLMs) offer promising solutions by providing more holistic approaches to content evaluation. It is clear that, to properly assess AIGC, geometric, structural, and biological consistency, along with visual realism, must be carefully considered in the quality assessment models [17, 42].

IQA and VQA models are also vulnerable to adversarial attacks involving the addition of imperceptible adversarial perturbations, which can significantly affect quality predictions. Attackers deliberately craft adversarial perturbations to exploit model weaknesses. For FR models, these perturbations can often be transferred across different models. This means that adversarial perturbations designed for one publicly available IQA model can also be used to attack a different model, even if the internal details of that second model are unknown to the attacker. As emerging IQA/VQA models are developed for evaluating AI-generated content, these same vulnerabilities are likely to also affect GenAI quality prediction models [17, 51].

Capturing the complexity of physical laws and geometric consistency in AIGC models and datasets remains a significant challenge. As a result, these models often produce unrealistic scenes that deviate from natural visual expectations, making it difficult to accurately predict quality in such cases. The lack of physical fidelity in AI-generated contents complicates quality assessments, as traditional IQA/VQA models may struggle to evaluate content that does not conform to the usual principles of real-world physics and spatial relationships [17, 51].

Next section describes other major issues/challenges.

XI. Other Major Issues/Challenges

A. HVS Models & Natural Images

HVS understanding & its usage in IQA is insufficient. New digital image processing technologies have entirely changed capturing, storing, receiving, viewing, utilizing, and sharing images. Most HVS models are based on the primary visual cortex (VI), and these models are insufficient for image quality assessment. A comprehensive computational HVS model requires a deep understanding of visual neurons, which are highly nonlinear & respond distinctly to naturally occurring stimuli compared to simple, controlled ones. Also, visual perception is another hurdle as psychophysical data for natural scenes is different than the data of simple & controlled stimuli scenes [14, 19].

Understanding the original signals, i.e., natural scenes, is very important and critical to understanding the response properties of the visual cortex along with improved VI models. Therefore, our HVS models must be able to handle such stimuli along with improved VI models. However, developed models have not been tested extensively with natural images as masks and are trained with other unnatural masks, i.e., sine waves, Gabor patterns, etc.

B. Compound & Supra Threshold Distortions

When, a visual target stimulates more than one channel in the HVS, it is called compound distortion. Supra threshold distortion is created by clearly visible targets that have contrast beyond the detection's threshold. Finding compound and supra-threshold distortions is a great challenge while dealing with psychophysical things in IQA algorithms [14].

1) Simple versus compound targets

In HVS, visual targets are typically simple, such as sine-wave gratings, Gabor functions, or extremely managed spatial arrangements confined in space, frequency, and/or orientation. However, compound targets contain many simple elements (e.g., multiple sine waves, or wavelets, etc.). As HVS are nonlinear, the principles derived from simple targets cannot be directly applied to compound targets [14, 19].

2) Perceived dissimilarity of supra-threshold distortions

Visual detection of the distortions provides us with a current understanding of visual perception. However, many images containing supra-threshold distortions have contrasts that are beyond threshold detection. Therefore, current models should be adjusted according to the supra-threshold distortions.

C. Impact of Distortions on Image Perception

Distortion contrast is used to assess quality, assuming that the user is looking for the distortion in an image. However, with severe and spatially interrelated distortions, users try to assess quality based on the interaction between distortions & image's objects. Accurately demonstrating the perceptual properties of such interactions presents an additional challenge [19].

1) Capturing & transparency in IQA

Another challenge lies in understanding (capturing & transparency) whether an image & its background are perceived as a unified stimulus or as two separate elements.

2) Visual strategy role in IQA

The HVS encompasses multiple levels, stretching from individual neurons to whole cognitive processes. However, leveraging the adaptive nature of the HVS for Image Quality Assessment (IQA) remains a significant challenge, particularly for high-quality images with near-threshold distortions and for low-quality images with supra-threshold distortions [19].

3) Effects of higher-level distortion on image appearance

Understanding the interaction between higher-order distortions and images is another issue. For example, if JPEG and JPEG 2000 compress two images, the user will declare JPEG 2000 more clear than JPEG image due to blockiness. Facial expressions and object boundaries in JPEG2000 images are better preserved and less degraded. Similarly, water, skin, and hair are more smooth, as blurring in JPEG 2000 is acceptable. However, current IQA algorithms do not take into account these higher-level perceptual aspects [14].

D. Geometric Changes

Geometric distortions, such as translation, rotation, or viewpoint changes are not currently addressed by any Image Quality Assessment (IQA) methods. Some front-end measurement can be taken to address these geometric changes, but multiple, unknown and compounding geometrical changes with traditional distortions are difficult to assess accurately. Similarly, when there is temporal non-synchronization between frames or changes in viewpoint, such as panning, zooming, or other adjustments, assessing quality on a frame-by-frame basis becomes highly challenging [14, 52].

More radical geometric changes: Other than simple geometric changes, images composed of geometric and photometric changes can also be generated. Similarly, evaluating textures is another challenge. Although humans can distinguish between original and synthesized textures, computationally, it is challenging. Therefore, more research is needed to evaluate geometrical changes.

E. Enhanced Images and Aesthetic Quality

With the latest digital technology, photo editing software, etc., it is possible that the quality of the image can be made better than the original. Existing IQA methods assume that a high-quality image is visually similar to the original. However, with the advent of digital technology, this notion of similarity becomes less applicable, highlighting the need for a different approach to quality assessment—one that remains an open challenge [53].

Enhanced Image database: There is no database for enhanced images other than DRIQ. Enhanced images can severely degrade aesthetic quality. However, an IQA algorithm designed to assess standard enhancement methods may struggle to predict aesthetic quality accurately. Therefore, it is an open research challenge to develop such IQA, which takes into account enhanced images with artistic intent and aesthetic quality.

F. Video Contents

The occurrence probability of distortions and their human tolerance threshold vary across different video content types. Research has shown that viewers tend to rate videos shot in sunlight higher than those taken at night. Additionally, humans are more sensitive to distortions in videos featuring people compared to landscape videos. Even when two scenes have an equivalent compression ratio, viewers may assign different scores as the content of each scene can influence their perception. Furthermore, assessing the quality of "in-the-wild" images or videos requires comparisons across cross-content pairs (i.e., pairs from different reference images or videos), as different content types may be uniquely affected by distortions. However, most of the existing objective quality methods do not fully consider the video content information, which affects the assessment performance, as shown in Fig. 5 [9]. Moreover, most quality methods are tested on cross-content videos and do not perform well in the case of real-world video applications that have abundant video content information. Therefore, more research is required to consider the video contents while assessing video quality [9, 31].

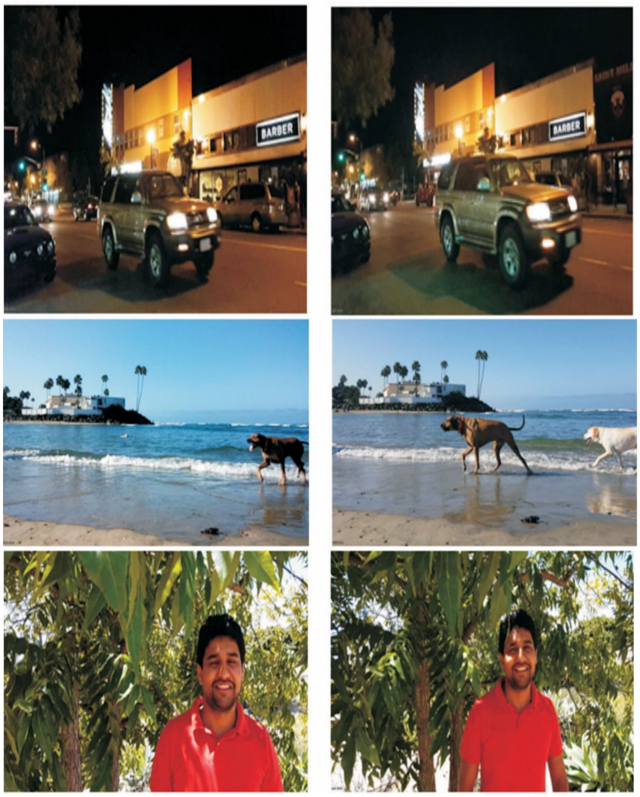

Fig. 5. Six images from six videos: Although distortion level is same but MOS is different in all images showing dependency on contents.

G. Personalized Video Quality Prediction

During subjective testing, MOS score is provided by the average user. But, in reality there is no average user. Therefore, the score obtained has standard deviation variation due to different user's preferences, i.e., perceived quality varies among different viewers. It is very difficult to obtain a personalized video quality which is an important phenomenon for the next generation video quality. Therefore, for subjective studies, physiological, socio cultural, demographics, and psychological factors of each user must be taken into account [6]. Similarly, cognitive biases have different impacts on our interpretation and assessment of video quality which must be taken into account such as attention biasing, choice-supportive, negative & recentness biasing, curse of knowledge, halo effect, and selective perception, etc. [35].

H. Memory Effects

Research has shown that during subjective testing, the viewer retains the poor quality in memory and consequently provides a lower quality score for subsequent frames, even when the frame quality has reverted to an adequate level. All of the existing quality assessment algorithms consider adjacent frames relationship and cannot handle long-term dependencies very well. RR-VQA approaches can be used to improve memory efficiency and reduce bandwidth [53].

I. Newly Emerged Videos Applications

Video quality assessment is also facing new challenges due to the emerging new features and characteristics of videos. For an example, depth perception in 3D videos; different dynamic ranges in HDR videos; Panoramic/first person or free view point videos requiring virtual reality technology, 360° videos, light fields, point clouds, etc. [52].

J. Efficiency

Although many existing QA approaches are effective and show high performance, they are not efficient enough to be deployed in real-world applications. For example, MOVIE takes more than 3 h to assess video quality, which is not feasible in real-time applications.

K. New Applications/Tools Have New Challenges as Described below [36].

1) Image in painting

Image editing operations (covering image distortions, merging or removing two image scene without any effect, etc.) are part of the image imprinting. However, assessing quality for in painted images is very nontrivial. Although, some information is available by edges, contours, gradients, but more research is needed [53].

2) Advanced image coding

Rate-distortion approach should be incorporated during coding for optimization along with traditional spatio-temporal JND, shape/texture & fovea coding, etc. As existing metrics do not assess quality accurately in textured region which needs further attention.

3) Image watermarking & covert communication

Covert messages are embedded as a watermark for covert communication. More HVS properties are required to hide more payload of a watermark. Good HVS metrics are required to efficiently assess the quality of watermarked images [53].

4) Image adaptation

Universal Multimedia Access (UMA) uses image adaptation. However, geometrical distortion happens during the image adoption process. Therefore, better techniques are required to assess the quality degraded by image adaptation [53].

5) Effect of different factors

Quality metrics should be designed to measure quality by taking into account the influence of distortion, aesthetics (lighting, color, contrast, and composition), and visual comfort on each other, rather than just assessing video quality from a single perspective, i.e., aesthetics, or visual comfort, etc. Deep learning techniques should be used for aesthetic quality assessment.

L. Correlation among Objective & Subjective Databases

Existing subjective databases are generally obtained in labs, and existing objective algorithms assess quality from one or two perspectives. Therefore, the correlation between objective and subjective is questionable for real-time applications, and more accurate mapping/fusion is required between subjective & objective quality metrics.

M. Restoration & Enhancement Challenges

Only distortions such as blur, noise, and JPEG compression can be restored. However, restoration and enhancement are a great challenge if the image is distorted by other kinds of distortions.

N. Assessing Quality for Fast Motion Videos

Accurate assessment of the quality of videos with fast motion, rotation, or shear motion is very difficult. Although HVS can track such changes, an automated system shows less efficiency in such cases. Therefore, assessing video quality for graphic videos is inaccurate and exhibits non-linear and complex temporal & spatial characteristics compared to natural videos [54].

O. Metric Selection

FR, RR, and NR can be used for quality assessment. However, the accuracy of the RR & NR assessment is not very good, and FR cannot be used in real-time. Accuracy also depends on application-dependent information and correct metric selection. Different metrics show different levels of quality assessment accuracy for different services, i.e., Video on Demand, Live Streaming, etc.

P. Mean Opinion Score Issue

Mean opinion score is used to verify the existing IQA models, which is a very time-consuming and costly process. MOS is often used without fully considering its scope and limitations, such as neglecting content variations and specific constraints. The fitting process in MOS, particularly the correlation, does not clearly understand how well the MOS values represent quality in precise terms. Additionally, MOS results from different studies and subjective tests cannot be directly compared due to their discrete nature. Furthermore, since the digital image space is defined by the number of pixels in the image, collecting sufficient subjective ratings to cover this entire space becomes a significant challenge [55].

Q. Data Sets & Tools

Comprehensive public databases for continuous time video quality is very limited. Very few publically available databases contain instantaneous subjective image quality assessment, as most contain overall image quality assessment for different video contents. Similarly, tools and techniques for steering, determining, imagining, investigating, and storing subjective scores at different time instances during subjective assessments are rarely available [56].

R. Saliency-Based QoE

Humans generally focus on a particular region of image/video, but existing QoE approaches analyze all spatial & temporal regions equally which leads to a poor subjectively perceived QoE. Therefore, more research is required for the audiovisual Focus of Attention (FOA) for perceived-QoE [55].

S. Video Coding Using Machine Learning

Machine learning should be integrated into new video coding standards to improve learning-based techniques for intra & inter-prediction, sub-pixel interpolation, transformation and quantization, as well as entropy coding, among others. However, training a neural network for Motion Compensation (MC) remains a particularly challenging task. This will generate an encoded video with minimum artifacts and make it much easier to assess video quality more accurately [42].

T. Light Field Imaging

This approach offers the potential for rendering 3D content more comprehensively. While traditional photography captures a 2D projection of light, this technology captures the luminosity of light rays from multiple directions. The applications of light field technology are vast, spanning areas such as gaming, video conferencing, and medical imaging. However, the apprehension, compression, editing, broadcasting, and presentation of vast amounts of light field data present significant challenges that require further research. Additionally, comprehensive light field databases—incorporating subjective and objective scores, various content types, distortions, and details on angular and spatial resolutions—should be developed and publicly released [55].

U. Quality for Computer Graphics

Measuring quality for computer graphics image is another challenge as most IQA metrics are designed for natural images and authenticated accordingly and not suitable for measuring quality for computer graphics images [56].

V. *Standardization Efforts*

Standardization is required for subjective quality evaluation for real-time video quality application for efficient integration of quality assessment & QoE models [57].

In the previous sections, we have presented different kinds of issues/challenges. Building on the discussion above, recommendations are presented in next section to address these issues.

XII. Recommendations

Research indicates that distortions are often compensated at the decoder side, which can introduce additional artifacts, making quality assessment more complex. Therefore, quality-aware codecs should be developed, which will compensate the artifacts on the encoder side and not the decoder side. For example, reducing MC mismatch at the encoder side is much better rather than compensating at the decoder side [7].

Developing artifact-aware encoder while taking into account HVS will also greatly reduce post-processing issues (eliminating any chance of originating new artifacts) due to information provided by the encoder. It will also provide the possibility of developing new metrics & will make artifact detection easier [11].

The HVS effect must be taken into account while doing Rate-Distortion Optimization (RDO) for encoder optimization to further minimize the artifacts originating.

New video coding tools and consequently, newly generated artifacts must be analyzed in detail such as what kinds of new artifacts are generating and how they are affecting other artifacts.

Quality algorithms should be developed with the capability of assessing all kinds of artifacts (existing & new artifacts generated from new codec's & also the effect of one on the other) with great efficiency.

Quality evaluation should be enhanced by improving the pre- and post-processing techniques.

More research is required to find out the effect of one artifact on others such as masking/superposition, creation, decrease, etc. [26].

The HVS effect on other artifacts and vice versa must be analyzed in detail.

More work is required to assess the quality of temporal & new artifacts instead of just focusing on spatial.

Generally, decoders are defined by video coding standards, and encoders have different configurations. However, a uniform encoder/decoder must be designed to reduce artifacts and enhance quality assessment.

Quality metrics should be used for the restoration and enhancement of quality assessment rather than just assessing certain specific distortions.

Detection accuracy of the quality metric should be high enough, i.e., fewer false positives. At the same time, it should avoid missing actual errors (false negatives) [39].

Computational time by quality metrics should not be high in order to optimize the throughput of overall media workflow, i.e., real-time assessment [58].

The accuracy of quality metrics should be content—independent. For example, it should be able to handle natural video, cartoon/animated content, text/graphics within content and so on, efficiently.

Advanced Machine Learning (ML), deep learning, and Artificial Intelligence (AI) methods must be used in developing new quality algorithms to enhance video quality assessment [59].

We have no knowledge beyond HVS model VI which also fails to predict masking in natural images. Therefore, more advanced & accurate HVS models should be devolved to assess quality efficiently [14].

Designing a computational neural model is impossible, as a database for natural images does not exist. An efficient database is essential for training and testing to develop a good HVS model.

A publically available database (consisting of all possible different kinds of videos) must be developed with large subjective ratings to include global & diverse pool of subjects. This can be achieved through crowd-sourcing approach, i.e., Amazon's Mechanical Turk, Facebook, Microworkers. In this way, a lab-oriented approach will be made public and more accurate quality assessment will be possible [38].

Energy-aware video coding standards: Due to the increasing complexity of encoders and energy consumption concerns in real-time video communications, there is a need for more energy-efficient video coding standards. Existing standards, such as H.264, are no longer adequate for emerging applications. Additionally, many devices, especially mobile phones, have energy constraints, which can lead to the generation of additional artifacts. Balancing these conflicting demands is crucial [60].

Efficient resource utilization and fine-tuning of video coding parameters are essential for any application. A deep understanding of these parameters is necessary to strike a balance, ensuring that improving one aspect of video quality doesn't lead to significant loss in another. Otherwise, additional artifacts will be generated, which will make the quality assessment more difficult.

Traditional video assessment approaches are still used for 360° videos which do not measure quality accurately & it is a challenging issue [52].

New algorithms should be designed to achieve higher efficiency using the latest hardware acceleration.

New quality metrics should be developed to assess new kinds of artifacts due to AIGC & AIEC.

A video quality algorithm should have the following features, i.e., it must be comprehensive (be able to measure existing & new artifacts) rather than specific to a particular kind of artifact; Should be NR to work in real time; should correlate strongly with subjective scores, etc.

XIII. Conclusion

Image/video quality estimation is an integral part of developing digital video systems. However, there are many issues/challenges associated with assessing video quality. Moreover, due to the rising demand & growth of digital video applications due to emerging technologies (e.g., new coding tools, 3D coding, etc.), new issues/challenges have also emerged while measuring video quality along with existing challenges. There is also a lack of understanding about these challenges with respect to the end-user satisfaction (QoE). Many existing works describe these issues/challenges while measuring video quality, however they do not collectively and comprehensively describe all the issues concerning video quality assessment. This work presents a comprehensive overview of all kinds of issues & challenges while measuring video quality, i.e., ranging from subjective to objective, using 3D video & new coding tools, the effect of one type of distortion to another type of distortion, etc.

The first contribution of the work is a comprehensive description of the issues/challenges due to the subjective & objective quality assessment methods. Another contribution is the comprehensive description of the issues due to the effect of spatial & temporal artifacts on other artifacts. Another contribution is a detailed presentation of the fresh challenges (new spatial & temporal artifacts) because of the latest coders. Another input of the work is the description of challenges due to 3D coding and issues due to the subjective databases. The paper also discusses challenges/requirements for the end-user satisfaction. Another input of the work is description of challenges in real time streaming including edge & cloud based streaming. Paper also describes issues related to quality measurement by deep learning & machine learning approaches along with challenges in image/video contents generated and enhanced by artificial intelligence comprehensively. Paper also presents a comprehensive description of miscellaneous issues/challenges.

Numerous recommendations have also been suggested as input in this work obtained by a comprehensive survey of different issues/challenges. These recommendations highlight future work and will guide the research community to consider these issues while measuring image/video quality and satisfying the end users.

The challenges, mainly, due to the latest coders, 3D coding, the influence of one artifact on other artifacts, issues in deep learning approaches & contents generated due to artificial intelligence, and issues related to the subjective databases, needs more attention. The work also highlights the need to understand these challenges' source for better quality assessment. More efficient encoders, 3D encoding, and enhanced deep learning approaches can be done with this knowledge. This work can be used in any digital image/video system where the received image/video does not follow prescribed standards and requires advanced analysis.

Conflict of Interest

The author declares no conflict of interest.

References

[1] R. Chaminda, A. Arslan, M. Thanuja *et al.*, "Measuring, modeling and integrating time-varying video quality in end-to-end multimedia service delivery: A review and open challenges," *IEEE Access*, vol. 10, pp. 60267–60293, June 2022.

[2] The Broadcast Bridge Interra Systems, Inc. (June 2015). Video dropouts and the challenges they pose to video quality assessment. [Online]. Available: https://www.thebroadcastbridge.com/content/entry/2863/video-dropouts-and-the-challenges-they-pose-to-video-quality-assessment

[3] N. Cranley, "Dynamic content-based adaptation of streamed multimedia," *Journal of Network and Computer Applications*, vol. 30, no. 3, pp. 983–1006, June 2007.

[4] R. Mantiuk and A. Tomaszewska, "Comparison of four subjective methods for image quality assessment," *Computer Graphics Forum*, vol. 31, no. 8, pp. 2478–2491, April 2012.

[5] M. Uzair and D. Dony, "No-reference transmission distortion modelling for H.264/AVC-Coded video," *IEEE Transactions on Signal and Information Sciences*, vol. 1, pp. 209–221, September 2015.

[6] A. Renuka and M. Azath, "A survey on image/video quality assessment some challenges and limitations," *International Journal of Computer Sciences and Engineering*, vol. 3, no. 4, pp. 42–45, March 2015.

[7] M. Mirkovic, P. Vrgovic, D. Culibrk *et al.*, "Evaluating the role of content in subjective video quality assessment," *Scientific World Journal*, vol. 2014, pp. 1–9, November 2014.

[8] AccepTV, "Subjective video quality assessment: The problems of subjective video quality assessment," *Perceived Video Quality Metrics*, 2018.

[9] L. Dingquan, J. Tingting, and J. Ming, "Recent advances and challenges in video quality assessment," *ZTE Communications*, vol. 13, no. 1, pp. 830–843, March 2019.

[10] V. Mario, S. Rimac, and K. Grgic, "Review of objective video quality metrics and performance comparison using different databases," *Signal Processing: Image Communication*, vol. 28, no. 1, pp. 1–19, August 2013.

[11] M. Uzair, "A comprehensive overview of classical and new perceivable spatial and temporal artifacts in compressed video streams," *The Journal of Engineering, Science and Computing (JESC)*, vol. 2, no. 2, December 2020.

[12] R. Palau, B. Silveira, R. Domanski *et al.*, "Modern video coding: Methods, challenges and systems," *Journal of Integrated Circuits and Systems*, vol. 16, no. 2, February 2021.

[13] Q. Huynh, P. Callet, and M. Barkowsky, "Video quality assessment: From 2D to 3D-challenges and future trends," in *Proc. International Conference on Image Processing*, 2010, pp. 1–5.

[14] M. Damon, "Seven challenges in image quality assessment: Past, present, and future research," *Signal Processing*, vol. 2013, pp. 1–53, February 2013.

[15] T. Kumar, P. Sharma, J. Tanwar *et al.*, "Cloud-based video streaming services: Trends, challenges, and opportunities," *CAAI Transactions on Intelligence Technology*, vol. 9, no. 2, pp. 265–285, March 2024.

[16] K. Lamachine, P. Mazumdar, and M. Carli, "A no reference deep learning based model for quality assessment of UGC videos," in *Proc. IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Shenzhen, China, July 2021.

[17] A. Ghildyal, Y. Chen, S. Zadtootaghaj *et al.*, "Quality prediction of AI generated images and videos: Emerging trends and opportunities," *ACM Digital Library*, vol. 3, pp. 20–36, March 2025.

[18] K. Bouraqia, E. Sabir, M. Sadik *et al.*, "Quality of experience for streaming services: Measurements, challenges and insights," *IEEE Access*, vol. 8, pp. 13341–13361, December 2019.

[19] A. Rogowitz, E. Bernice, N. Thrasyvoulos *et al.*, "Seven challenges for image quality research," *Human Vision and Electronic Imaging*, vol. 19, 2014.

[20] H. Yogita, P. Sapat, and Y. Hemprasad, "A survey on image quality assessment techniques, challenges and databases," *International Journal of Computer Applications*, vol. 7, pp. 34–38, April 2015.

[21] M. Vranješ and A. Snježana, "Review of objective video quality metrics and performance comparison using different databases," *Signal Processing: Image Communication*, vol. 28, no. 1, pp. 1–19, January 2013.

[22] C. Shyamprasad, S. Vijay, and K. Lina, "Objective video quality assessment methods: A classification, review and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, June 2011.

[23] M. Frackiewicz, Ł. Machalica, and H. Palus, "New combined metric for full-reference image quality assessment," *Symmetry*, vol. 1, pp. 1–21, December 2024.

[24] D. Varga, "No-reference image quality assessment using the statistics of global and local image features," *Electronics*, vol. 4, pp. 40–62, March 2023.

[25] A. Kumar and S. Chandramathi, "Video quality assessment methods: A bird's-eye view," *International Journal of Computer and Information Engineering*, vol. 8, no. 5, 2014.

[26] A. Unterweger, "Compression artifacts in modern video coding and state-of-the-art means of compensation," *IGI Global*, May 2013.

[27] L. Callet, V. Gguadin, and D. Barba, "A convolutional neural network approach for objective video quality assessment," *IEEE Transactions on Neural Networks*, vol. 17, no. 5, pp. 1316–1327, 2006.

[28] T. Zhang and S. Mao, "An overview of emerging video coding standards," *Mobile Computing and Communications*, May 2019.

[29] Akamai, "Measuring video quality and performance: Best practice," *White Paper*, April 2021.

[30] N. Yuzhen, Z. Yini, and G. Wenzhong, "2D and 3D image quality assessment: A survey of metrics and challenges," *IEEE Access*, vol. 7, pp. 782–801, December 2018.

[31] T. Liu, Y. Chieh, and W. Lin, "Visual quality assessment: Recent developments, coding applications and future trends," *SIP*, vol. 2, no. 4, pp. 1–20, November 2013.

[32] M. Kede, Z. Duanmu, Z. Wang *et al.*, "New challenges for image quality assessment models," *IEEE Transactions on Image Processing*, vol. 26, no. 2, February 2017.

[33] S. Winkler, "Analysis of public image and video databases for quality assessment," *IEEE Journal on Selected Topics in Signal Processing*, vol. 6, no. 6, October 2012.

[34] T. Zhou, Q. Liu, and C. Chen, "QoE in video transmission: A user experience-driven strategy," *IEEE Communications Surveys & Tutorials*, vol. 9, no. 1, pp. 285–302, September 2017.

[35] M. Abdelhamid, H. Said, and A. Hai, "Quality of experience for multimedia," *Wiley*, October 2013.

[36] K. Zhu, C. Li, and V. Asari, "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. 533–546, June 2015.

[37] Z. Akhtar, K. Siddique, A. Rattani *et al.*, "Why is multimedia quality of experience assessment a challenging problem?" *IEEE Access*, vol. 7, pp. 117897–117915, 2019.

[38] S. Felici and J. Garcia, "Adaptive QoE-based architecture on cloud mobile media for live streaming," *Cluster Computing*, vol. 5, no. 3, pp. 355–369, May 2018.

[39] K. Bouraqia, E. Sabir, M. Sadik *et al.*, "Quality of experience for streaming services: Measurements, challenges and insights," *IEEE Access*, vol. 8, pp. 13341–13368, April 2020.

[40] Video Clarity, "Achieving maximum accuracy with real time video quality measurement," *Tools for Video Analysis*, June 2022.

[41] G. Proietti and R. Beraldi, "A study on real-time image processing applications with edge computing support for mobile devices," in *Proc. IEEE/ACM 25th International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*, Valencia, Spain, September 2021.

[42] Q. Zheng, Y. Fan, L. Huang *et al.*, "Video quality assessment: A comprehensive survey," *Electrical Engineering and Systems Science*, vol. 14, pp. 1–36, December 2024.

[43] A. Ravindran, "Internet-of-things edge computing systems for streaming video analytics: Trails behind and the paths ahead," *IoT*, vol. 4, pp. 486–513, October 2023.

[44] S. Dost, F. Saud, M. Shabbir *et al.*, "Reduced reference image and video quality assessments: Review of methods," *EURASIP Journal on Image and Video Processing*, vol. 2, pp. 61–93, June 2022.

[45] W. Rivera, M. Pineda, and M. Calero, "Cloud media video encoding: Review and challenges," *Multimedia Tools and Applications*, vol. 83, pp. 81231–81278, March 2024.

[46] X. Min, H. Duan, and G. Zhai, "Perceptual video quality assessment: A survey," *Science China Information Sciences*, vol. 67, no. 1, pp. 1–57, November 2024.

[47] O. Izima, R. Fréin, and A. Malik, "A survey of machine learning techniques for video quality prediction from quality of delivery metrics," *Electronics*, vol. 10, pp. 1–44, November 2021.

[48] C. Lee, D. Seok, D. Kim *et al.*, "Reliability of CNN-based no-reference video quality metrics," in *Proc. 15th International Conference on Information, Intelligence, Systems & Applications (IISA)*, Chania Crete, Greece, July 2024.

[49] J. Klink, M. Łuczyński, and T. Uhl, "The use of machine learning in modeling video quality assessment," in *Proc. International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Split, Croatia, September 2024.

[50] J. Søgaard, S. Forchhammer, and J. Korhonen, "Video quality assessment and machine learning: Performance and interpretability," in *Proc. IEEE QoMEX*, vol. 2, December 2015, pp. 31–38.

[51] D. Mohanty, G. Thippeswamy, G. Erappa *et al.*, "Quality of video rendering techniques using artificial intelligence," *ICTACT Journal on Image and Video Processing*, vol. 13, no. 3, February 2023.

[52] S. Webster and G. John, "Methods for image quality assessment," *Wiley Encyclopedia of Electrical and Electronics Engineering*, August 2015.

[53] F. Dufx, "Grand challenges in image processing," *Frontiers in Signal Processing*, vol. 1, April 2021.

[54] O. Miguel, P. Pablo, M. Otoniel *et al.*, "On the performance of video quality assessment metrics under different compression and packet loss scenarios," *Scientific World Journal*, vol. 2014, pp. 1–18, 2014.

[55] R. Serral, E. Cerqueira, M. Curado *et al.*, "An overview of quality of experience measurement challenges for video applications in IP networks?" in *Proc. 8th Wired/Wireless Internet Communications*, pp. 12–21, December 2010.

[56] Z. Icheng and Z. Zhzhao, "Subjective & objective quality assessment for in-the-wild computer graphics images," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 1, no. 1, September 2023.

[57] R. Jinjia and X. Dongliang, "A survey on QoE-oriented VR video streaming: Some research issues and challenges," *Electronics*, vol. 10, no. 17, January 2021.

[58] A. Lewandowska, "Scene reduction for subjective image quality assessment," *Journal of Electronic Imaging*, vol. 25, no. 1, May 2016.

[59] S. Ramakrishnan, *Cryptographic and Information Security—Approaches for Images and Videos*, CRC Press, June 2018.

[60] W. Lin and C. C. J. Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, August 2011.