INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2025, VOL. 71, NO. 3, PP. 1–14
Manuscript received June 5, 2025; revised June 2025. doi: 10.24425/ijet.2025.155450

Unveiling the art of video enhancement: a comprehensive examination of content selection and sequencing for optimal quality in conventional and AR/VR environments

Mehrunnisa, Mikolaj Leszczuk, and Dawid Juszka

***Abstract*—The demand for high-quality video content has grown along with the rise of new technologies. The quality of visual content directly impacts user engagement and satisfaction, highlighting a clear correlation between user expectations and content delivery. Recent studies stress how important it is to pick the right content, especially in fields such as signal processing and multimedia communication. But there are challenges, such as inconsistent content selection, a lack of standards, and not enough data. Using generative AI and machine learning can help address these issues. By embracing technology-driven, inclusive and teamwork-based methods, this review paper reviews better content and sequence choices in both traditional and AR/VR setups for video enhancement. The need for high-quality content has increased with time due to the emergence of new technologies. User engagement and satisfaction are directly proportional to the quality of visual, revealing a direct proportionality between user expectations and content delivery. Recent research in digital media has emphasized the importance of selecting a particular type of content, leading to an optimized user experience. Signal processing, multimedia communication, and image processing have been significant areas of interest for researchers in which content selection is of great importance. Factors such as motion characteristics and visual complexity must be considered for precise results. The main consequence emphasizes the focus on dynamic content, diversity, and UGC as a significant area of interest. Compared to the current literature, challenges such as content selection variability, no standardized criteria, and limited data sets that serve as benchmarks must be considered. Integrating machine learning algorithms into data sets alongside scenario-based criteria can be an essential solution to such problems. Adherence to technology-driven, inclusive, and collaborative approaches leads to a better outcome that ensures productivity.**

***Keywords*—Quality of experience; video quality evaluation; video content selection; traditional vs AR/VR media**

I. INTRODUCTION

UNDERSTANDING the impact of video quality on user experience is paramount in today's digital landscape, where various resolutions, ranging from 8K to 3D, cater to diverse consumer needs. The evolution of technology has brought about a multitude of challenges that directly influence how users perceive and engage with video content. Both objective and subjective methods play an essential role in assessing Quality of Experience (QOE), providing valuable information to improve consumer satisfaction and diversify viewing experiences. In different sectors such as medicine, web design, broadcast television, and content delivery, human factors are increasingly recognized as central to addressing video quality. This underscores the importance of considering aspects such as content creation, delivery mechanisms, and the overall viewing experience. As such, content providers are faced with increased responsibilities in ensuring the effective utilization of resources to meet evolving consumer demands.

Research methodologies employed in this domain often involve the analysis of various content formats, including short/long videos on demand and live streams. Quality metrics such as average bitrate, join time, rendering quality, and buffering ratio are meticulously measured using client-side instrumentation. Of particular importance is the proportion of buffering, which has been identified as a critical factor that influences viewer perception across different types of content [1]. From a commercial perspective, the implementation of tailored policies can significantly enhance user engagement by prioritizing qualitative aspects. In contrast, addressing technical considerations, such as buffer size, presents opportunities to drive innovation and stay informed about emerging trends [2, 3].

In the realm of video editing, significant strides have been made from the cumbersome analog setups of the past to today's user-friendly digital platforms, which have been further enhanced by AI-driven tools. However, it should be noted that these advancements have focused predominantly on traditional video formats, leaving emerging technologies like 3D and AR in the periphery. Editing 360-degree videos, for instance, poses unique challenges due to the inherent distortion when viewed on flat screens. However, with familiarity and specialized tools, editors can navigate these challenges to create immersive experiences for viewers. Similarly, in AR/VR editing, the ability to manipulate viewpoints in all directions requires a nuanced approach distinct from traditional methods. Despite the convergence of techniques such as color grading and

Authors are with AGH University of Krakow, Poland (e-mail: mehrunnisa@agh.edu.pl, mikolaj.leszczuk@agh.edu.pl, juszka@agh.edu.pl).

soundtrack addition, editors may find themselves grappling with new tools and workflows, particularly in stereoscopic editing scenarios[4]. However, these challenges underscore the dynamic nature of the field and the ongoing quest for innovation in delivering compelling visual experiences to audiences worldwide.

II. Significance of Work

In today's digital era, video content has become an indispensable aspect of our daily lives. Whether it is the captivating visuals in professional film productions, the engaging posts on social media platforms, the immersive experiences in virtual reality environments, the dynamic gameplay in video games, or the seamless streaming services we enjoy online, videos have undeniably become ubiquitous across the Internet. Consequently, there has been an unprecedented surge in the demand for proficient video editing tools, skilled professionals, and innovative techniques. However, as technology continues to advance at a rapid pace, the landscape of video editing is in a constant state of evolution. While much of the existing literature on video editing primarily focuses on enhancing 2D formats, there remains a significant gap in addressing the revolutionary concepts associated with 3D and dual 180-degree or 360-degree videos. These cutting-edge formats often require specialized "non-linear" editing tools to effectively manipulate and enhance the content.
Therefore, this research endeavor aims to bridge this gap by exploring the applicability of traditional video enhancement techniques to today's evolving visual content, whether it be in 2D or 3D formats. In doing so, this study seeks to identify and overcome the limitations highlighted in the current literature, thereby contributing to a more comprehensive understanding of video editing practices in contemporary multimedia environments.

III. Research Methodology

This study uses a systematic review of the literature (SLR) as its main research method, following the established protocol outlined by [5]. This method allows for the identification, selection, and analysis of the most relevant research directly related to the study's specific research questions. The choice of SLR was driven by its commitment to transparency, clarity, focus, accessibility, and comprehensiveness, all essential principles for conducting a thorough and reliable review. Before starting the actual review, a detailed review protocol was established, as suggested by [6]. This protocol encompassed several key elements:
Rationale: The justification and purpose of the study were clearly defined.
Search Strategy: The methods used to identify relevant literature (including databases and keywords) were described.
Selection Criteria: The specific criteria for including or excluding studies were established.
Quality Assessment: Defined criteria were established to evaluate the quality of the included studies.
Data Extraction: The process for extracting relevant data from the chosen studies.
Dissemination Strategy: The plan was outlined to communicate the review findings.
The implementation of these protocol steps ensured a structured approach to the SLR for this study. It began with defining the research questions, followed by identifying relevant data sources, establishing inclusion and exclusion criteria, and finally, defining quality assessment criteria.

A. RESEARCH QUESTIONS

This research aims to improve the usability of unstructured data, specifically video content. This focus on usability holds promise for a better and more efficient enhancement of such content. To achieve this goal, the study has formulated the following research questions and objectives:

RQ1: What are the latest techniques for enhancing video content?
Objective: This question aims to identify the recent advances and current methods used in video enhancement.

RQ2: Can these techniques be applied to emerging technologies like Augmented Reality (AR) and Virtual Reality (VR)?
Objective: This question compares existing video enhancement practices with the capabilities of these new technologies, evaluating their potential integration.

RQ3: Is there a need for entirely new video enhancement methods to handle upcoming digital visual content?
Objective: This question explores the potential need for novel video enhancement techniques that differ from traditional approaches, specifically designed for future digital visual content.[7]

B. SELECTION OF ONLINE LIBRARIES

This systematic review used online libraries known for their high-quality publications such as: ACM Digital Library, IEEE Explore, Springer, Elsevier, Scopus, Google Scholar, and Wiley Online Library. These libraries were chosen because of their extensive collection of relevant research.
Further keywords were carefully selected on the basis of the study's objectives. These keywords included synonyms and related terms for "video enhancement," "video editing," and "video quality." In addition, terms related to the application of editing techniques in AR, VR, and traditional video were included. The final search string combined these keywords using the AND operator to ensure that all results included both sets of terms. The search focused on specific sections of the articles, including the title, abstract, and keywords, using the TITLE-ABS-KEY field.

C. INCLUSION AND EXCLUSION CRITERIA FOR LITERATURE SELECTION

Selecting the most relevant studies is crucial to effectively address research questions. This review followed specific criteria to ensure that the chosen studies directly contributed to the investigation. The review focused on recent advances in the field. This ensured that the findings incorporated the latest developments and knowledge. Furthermore, the review only considered studies written in English to maintain consistency and avoid possible language-related ambiguities. Furthermore, the review prioritized reliable sources of information. Only journal articles, conference papers, and book chapters were included. These sources are typically subject to rigorous peer review and editorial processes, ensuring the quality and credibility of the research presented. To further guarantee relevance, the review focused on studies that contained specific keywords directly related to the research questions. This ensured that the chosen studies addressed the intended topics of video enhancement and its potential application in emerging technologies. Finally, the review excluded studies that fell outside the defined criteria. This included studies published before the specified time frame, those not written in English, and those published in sources such as reports, magazines, or formats other than journal articles, conference proceedings, or book chapters. In addition, studies that did not meet the quality assessment criteria established for the review were also excluded. This ensured that the final selection comprised only high-quality and highly relevant studies that contributed significantly to the research questions and objectives.

D. CRITERIA FOR THE EVALUATION OF RESEARCH QUALITY

The integrity and credibility of a systematic review of the literature (SLR) are fundamentally dependent on the methodological quality of the primary studies included. Therefore, a rigorous quality assessment process is essential to ensure that only studies exhibiting robust research design, methodological rigor, and relevance to research questions are incorporated into the review. By systematically identifying and excluding studies with methodological flaws or potential biases, the SLR process improves the reliability and validity of its synthesized findings.

The structured nature of the SLR methodology facilitates the consistent application of quality assessment criteria, thus promoting transparency and reproducibility in the selection process. This approach allows researchers to objectively evaluate the suitability of individual studies, ensuring that the final corpus of literature includes high-quality contributions that offer substantive insight into the research topic.

In the present review, the quality of the selected studies was evaluated based on three carefully defined criteria, each aligned with established standards in the field to ensure methodological soundness, relevance and contribution to the overall research objectives.

QA1-Relevance: Does the study directly answer the research questions of this review?

QA2-Objectives: Does the study clearly define its research goals?

QA3-Contribution: Does the study offer valuable new insights into the relevant field of research of this review?

All chosen studies were evaluated according to these criteria, as shown in Table I. A scoring system was used with ”Yes” = 1 point, ”No” = 0 points, and ”Partial” = 0.5 points. Studies scoring at least 75% (2.5 points or higher) were considered high quality and included in the final review.

TABLE I
QUALITY ASSESSMENT SCORE.

Ref. no	QA1	QA2	QA3	Total
[8]	1	0.5	1	2.5
[9]	0	1	1	2
[10]	1	0.5	1	2.5
[11]	0	1	1	2
[12]	1	1	1	3
[13]	0	1	0.5	1.5
[14]	1	0.5	1	2.5
[15]	0	1	1	2
[16]	1	1	1	3
[17]	0	1	0.5	1.5
[18]	1	1	1	3

IV. REVIEW OF RECENT LITERATURE AND RESEARCH FINDINGS

It is important to realize the effect of video quality on the engagement of a particular audience. Dobrian et al. (2011) have reflected the importance of video quality in multiple dimensions. The advent of the latest technologies has led to the distribution of main-stream video distribution through the Internet, which has also led to increased user expectations for better quality[19].

Merikivi et al. (2019) emphasize binge watching as an essential viewing practice. Viewing practice differs from other techniques, as it involves serialized content. However, the review has two important aspects of binge watching: viewer autonomy and continuity. These play an essential role in maintaining the difference from single-episode sessions [20]. To develop a better understanding, a quadrant framework has been used that includes four essential viewing practices. Binge watching, SEA ”Single Episode Appointment,” MA ”Marathon Appointment”, and CV ”Casual Viewing.” CV has been classified as highly self-scheduled television, whereas SEA is a low-scheduled television. The significant difference between the two is that CV is independent of speed and time, whereas SEA is the opposite. Similarly, binge watching depends on time and speed, whereas MA is the opposite. Insightful future work and research practices can show that the focus on binge watching can increase if accurate behavior data is utilized from transaction-based log files, including platforms like Vimeo and YouTube Red. The correlation between content creation and log data can pave the way for more opportunities, allowing researchers to analyze their influence on binge watching.

A. Using Different Impairments for Subjective Quality Evaluation

L. Pozueco et al. (2015) have utilised different impairments for the evaluation of content types. The main goal was to inform consumers about the level of distortion in video quality [21]. One hundred participants participated in the study regardless of biased aspects, such as the same age range. It comprised different video experiences, profiles, and age ranges. Age ranges were between 20 and 68 years, and content types included sports, news, cartoons, and action-packed films. Different impairments were used to gain better insight into the perceived quality of experience, including buffering events, audio/video loss, desynchronisation, and artefacts. The impairment models included IM-A "asynchronous," IM-P "pixelation", IM-CA "audio cuts," IM-CV "video cuts," and IM-S "stalling." The findings were rearranged into statistical analyses in which the mean differences between the variables were determined. According to the normality of the data, it was analysed through Shapiro-Wilk tests, whereas Bartlett tests were utilised for homoscedasticity. After conducting these tests, the data were compared by the ANOVA test. Therefore, it was concluded that the bit rate of transmission could be highly decreased as it significantly impacts the video quality. Caused a disruption of the video experience. The type of content is of immense importance as the impairment effect is different for each content type, e.g., disorders did not significantly impact QOE in the case of cartoons.

B. Role of affective images in IQA 'Image Quality Assessment'

The extent of image distortion can easily be measured and evaluated through image quality assessment. There has been an extensive demand for the development of lossy compression algorithms due to the storage and transmission of high-quality images and videos. Thus, subjective and objective IQAs and VQAs are utilised effectively for evaluating algorithms for applications like watermarking, image enhancement, restoration, broadcast, transmission, and display systems [22, 23]. Ian et al. (2012) have explored the influence of attractive images on subjective image quality assessment [24]. Images such as close-up shots, animals, people, natural scenes, wide-angle shots, distinct background images, and man-made objects were included in the study. These images were part of the LIVE database and Kodak database. Thus, the utilisation of diversified and larger databases resulted in accurate empirical evidence-based results. A total of 25 participants participated in the evaluation and were randomly chosen as part of the university staff or students. Therefore, the research findings enlightened the reader on the aspect that semantic information had an impact on the perception of scenes, significantly affecting top-to-down visual processing.

C. User Factors and Social Context in VQOE 'Video Quality of Experience'

QOE is vital in determining user satisfaction with a multimedia system or content. Most scenarios reflect the importance of media technical aspects to know the quality of experience. More research has suggested that more than technical elements in the media are needed to determine the QOE, but many other factors are also dependent. Y. Zhu et al. (2015) have enlightened user factors and social extent as the most critical aspects to evaluate the quality of experience [25]. Additionally, the correlation between technical media aspects, user factors, and social context has also been realised, providing us with a better understanding of the genre and level of video bit rate. A total of 60 participants participated in the research, with a mean age of 26.5. Of the 15 genres in total, three genres were used for the study. Sixty-one per cent of the participants were in favor of watching education videos alone, while 51% were in favor of watching comedy videos with friends. Therefore, the results indicated that the endurability and enjoyment level of the users suffered an elevation due to the co-viewers. Co-viewers play an essential role in interest development for a particular content type. As a result of the research, participants were also aware of the levels of video quality and did not significantly impact QOE. In addition to this, cultural background, and gender-based aspects were also of massive importance, as these enlightened different people's perceptions.

D. Impact of Content Desirability on Subjective Ratings of Video Quality

It is essential to realise that there is a direct proportionality between audio and video quality ratings. Redundancy in videos is based on the content being played. 90% redundancy is possible if there is a static shot. Slight deviations in test sequences do not affect user perceptions as the decrease in bit rate remains unnoticed. Rochelle (2011), in a study, reflected on the aspect that the liking developed among users for a particular product led to a higher qualitative rating [26]. In total, four experiments were conducted that had different findings. The first experiment comprised 26 participants who seemed to like and enjoy film clips more than normal ones. Similarly, 30 participants participated in the second experiment, where a positive correlation was observed between video quality and content immersion/liking/effect. Experiment 3 enlightened the moderation of the relationship between video quality and affect. The halo effect was relatively reduced because of the training of the participants. Training the participants also led to an accurate rating of the video quality. Most importantly, the last experiment showed that a qualitative focus on the part of the participants could lead to an improved video quality rating. The relationship between video quality and content immersion was also not disrupted because of that. The accuracy of these subjective ratings was the result of focused attention and training processes. In short, proper training and focused attention could improve feedback quality.

The effect of desirability of content has also been realized by Philip et al. (2010), who examined user ratings for subjective video quality [27]. In most scenarios, the encoding algorithms are implemented by short clips that neglect the desirability of the content. One hundred participants participated in the investigation that evaluated 180 clips from low- to high-quality. However, the research results suggested that there was a strong relationship between subjective video quality ratings and movie content. The connection is strong across various encoding levels when the content is viewed under realistic conditions. The viewer's choices are mainly influenced by factors such as psychological state and demographics, making it difficult to choose a desirable clip for a larger audience. So, the choice of content on the part of the participants can result in a better and more productive outcome, along with being expensive and time-consuming. However, these can result in better precision and accuracy of the results according to the research.

E. Exploiting emotions by viewing films

Robert et al. (1995) have enlightened the effect of films on individual emotions according to the research. The research summarizes five years of research that has led to the revelation of a set of films that trigger emotions such as anger, amusement, contentment, neutrality, disgust, sadness, surprise, and fear. A total of 250 films were evaluated, of which 494 English-speaking clips were selected. According to the discreteness and integrity of the subject, two films were chosen in the best way for each emotion. For research, 494 students, a total of 265 women and 229 men, were included. These were made a part of the viewing session for the psychology course. The participant's age ranged from 17 to 43 years. A level ANOVA factor was used as a statistical technique, revealing that the intensity of the intensity of the target emotion in women was greater than in men. The mean target emotion for men was 4.44 which was comparatively lower than that of women 4.98. There were a total of four ethnicities that were taken into account to realize the effect of ethnicity. These included Caucasian, Asian American, African American, and Hispanic. So, the ANOVA factor did not prove to be a practical approach to the effect of ethnicity. Lastly, another important aspect was that participants who had previously watched these films had higher target emotions than those who had watched those films for the first time. Lastly, the research paved the way for exploring other emotions such as pride, embarrassment, guilt, and contempt[28].

F. Multi-Instrumental Approach for Exploring QOE Effectiveness

Katrien et al. (2015) have focused their research perspective on optimizing user quality and experience [29]. The main goal was to realize how humans perceived multiple features, revealing a qualitative perspective. Therefore, the focus areas included behavioral, cognitive, and affective processes because there has been a lack of research in these areas. Educating the multi-instrument approach is essential to realize the gap between traditional and holistic quality of experience. To evaluate the video quality experience, 27 participants participated in the study, where the QOE assessment paradigm was implemented. In addition, the relationship between overall quality and actual quality of experience was also discovered. Data collection consisted of four types of data. These included behavioral, physiological, gaze tracking, and traditional data. The test procedures were conducted so that three five-minute movies had an error profile different in each scenario. It consisted of approximately one hour of providing an overview of the test instrumentation. The research findings showed that 70% of the clips were acceptable, while 30% were not. Overall quality was a combined blend of higher reported pleasure and overall quality scores. Ratings were higher when there was a match between reality and overall quality expectations. Furthermore, there was an inverse proportionality between overall quality and the annoyance of cutting errors. There also existed a slight relationship between pleasure/joy and desirability of content. Similarly, focused attention and interest also had a similar correlation. Most importantly, there was no correlation between self-reported surprise/joy and overall quality. Therefore, the research findings suggested that traditional QOE measures had to be revised and synchronized with the frustration and delight measures. Further analyzes indicated that the correlation between physiological / behavior measures and self-reported measures had to be realized to better understand the implications.

G. Selecting Video Sequences for VQA Video Quality Assessment - Task-Orientated Testing

MH Pinson (2013) explains the techniques for the choice of video sequences that serve the purpose of subjective experimentation for VQA [30]. The availability of content serves as a major setback for scene selection. For better video quality research, one must realize that personal choices should never be standard for selecting a particular sequence. It should always be based on mandatory video sequences. For a more complete view of subjective testing, there are two main types: entertainment-oriented testing and task-oriented testing. The scene selection approach in both is entirely different. For research, there are three various task-based experiments. These include the effect of quality on sign language comprehension through a video link, public safety performance requirements for public safety, and the overall impact of video on oral comprehension through an audiovisual link. The test design has been constructed to have five essential elements: non-manual markers, vocabulary, finger-spelling, gestures, and spatial/role shifting. To prioritize the difference between entertainment and task-orientated selection, camera work and content editing are essential traits of entertainment-based selection, whereas task-based selection prioritizes the selection of a system serving the same purpose. In conclusion, criteria have been described that inform subjective experiments for VQA. Essential aspects include analysis of response strategies for network error, optimization of coding parameters, codec comparison, and testing/training of the VQ model.

H. Selecting Video Sequences for VQA "Video Quality Assessment" - Entertainment-Orientated Testing

M.H. Pinson et al. (2014) enlightened subjective tests geared toward entertainment for video sequences in this article [31]. Such tests improve product development, leading to increased research opportunities. There are multiple aspects to the quality of experience because different users have different levels of perception and, based on their perceptions, all have different experiences. Therefore, there are many different approaches to measuring QOE, such as customer satisfaction surveys, operations research, sociology, etc., by the variables involved. There can be a difference in user experience if applications are used in the same network conditions and context. To conduct an authentic assessment, it is essential to include significant topics regardless of the aspect that it will be resource-consuming and time-consuming. Increasing recognition among researchers, network operators, and service providers will contribute towards better user services. The article also emphasizes that utilizing LTM – Latent Trait Models like item Rasch Theory can be a helpful approach where latent traits can be used to test service features. However, the design of the questionnaire must also be authentic.

I. Scene Selection for subjective VQA for two '2D' dimension and three '3D' dimension

Marcus et al. (2013) enlightened two-and three-dimensional aspects of scene selection for subjective assessment of video quality [32]. Content availability and convenience serve as the basis for scene selection and have a direct impact on it. However, there have been many innovations in 3d subjective testing due to its subsequent intervention in the market. If the modern-day consumer market is taken into consideration, then these include higher resolutions (8K, 4K, and Full HD) and reconstruction precision, including a wide color gamut and a high dynamic range. Associated media includes emotive devices and audio wavefield synthesis. Augmented reality, social networks, and mobile phones are some of the latest modes of interaction among users. The use of these technologies has paved the way for further development through objective and subjective measurement methods. Similarly, developments in audiovisual technologies have produced several challenges for QOE. To address these challenges, subjective experiments are of immense importance in establishing authentic data applicable for validation, verification, and training procedures. The lack of availability of content has also served as a significant setback in correct scene selection for subjective experimentation. Therefore, it is essential to develop scene pool selection guidelines for subjective assessments such as audiovisual quality assurance.

J. Video Processing and Playback Workflow

Figure 1 presents a comprehensive workflow for video processing and playback, detailing each stage from raw footage acquisition to final display output. The process begins with the ingestion of source footage, which is subjected to a restoration phase aimed at enhancing visual quality by correcting artifacts and improving clarity. Concurrently, a scenecut file provides structural information to guide the cutting process, enabling segmentation of the video content based on scene boundaries. Following segmentation, the video undergoes resolution enhancement, which increases spatial fidelity, often through upscaling or super-resolution techniques. The enhanced video is then passed to the encoding module, where it is compressed into a media bitstream and accompanied by metadata files, with the configuration governed by an external settings file. This encoded output is subsequently processed by the decoder, which reconstructs the video stream for playback. Both the encoder and decoder utilize the same settings to ensure consistency in compression and reconstruction. After decoding, the content enters the display processing stage, which applies final adjustments—such as color grading, frame adaptation, or display-specific optimizations—before rendering. The processed output is directed to the CPVS (Consumer Playback Video Stream) for end-user display, and optionally to the AVPVS (Advanced Video Processing Visualization Stream) for specialized visualization or quality assessment tasks. This modular pipeline supports efficient, scalable, and high-fidelity video delivery in advanced multimedia systems.

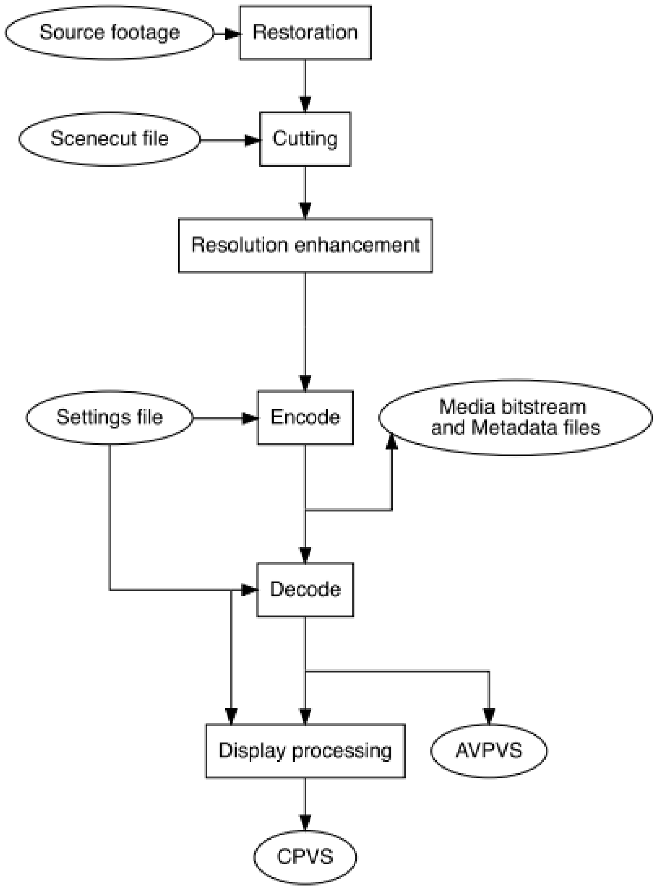

Fig. 1. Flow chart: Processing chain [33]

K. Influence of AI and machine learning on Video Content Selection and QOE/QOS Assessment

The revolution in QOE/QOS and video content selection has been subject to many emerging technologies such as machine learning and artificial intelligence. The main driving forces for maximum productivity are a combined mix of user preferences and behavior. These are analyzed to improve user satisfaction and engagement with fully immersed and personalized content suggestions. Georgios et al. (2022) in their study have enlightened the evaluation and prediction of multimedia-based QOE based on machine learning [34]. Their review-based research reveals that network operating efficiency is easily determined using AI-based techniques, allowing a better end-user experience to provide better services. Furthermore, research reveals that to optimize the delivery of streaming services for mobile multimedia, it is necessary to evaluate and predict the QOE of the end user. Thus, a better understanding of the technical aspects of a network can be gained. ZhiGou et al. (2019) have evaluated the quality of experience through VOIP networks through machine learning algorithms and QOS mapping [35]. They have analyzed the impact of the QOS parameters on QoE through association tests, methods, and experiments. There exists a strong nonlinear relationship between QoE and jitter.

L. Enhancement on Compressed Videos

The introduction of the Versatile Video Coding (VVC) standard in July 2020 marked a significant advance in video compression technology. VVC aimed to achieve comparable perceptual quality to the existing High Efficiency Video Coding (HEVC) standard while halving the bitrate and offering enhanced flexibility in resolution adaptivity, scalability, and multi-view capabilities. However, despite its compression efficiency, VVC still exhibited distortions and quality degradation due to its hybrid block- and transform-based coding approach. To address these issues, [36] proposes a novel quality enhancement method for VVC-compressed videos using advanced deep learning-based multiframe quality enhancement models (MFQE). This approach is founded on the idea of segmenting the decoded VVC video into peak quality and non-peak quality pictures, followed by employing long-short term memory (LSTM) and two subnetworks to achieve superior video quality. The experimental results demonstrate significant quality improvements compared to the original VVC-decoded video. The paper emphasizes the ongoing challenges associated with compression artifacts in VVC-compressed videos, which can diminish the QoE for viewers. As such, there is a pressing need to enhance the QoE of VVC-compressed videos on the decoder side. Previous studies have shown promising results in quality enhancement using deep learning approaches, such as MFQE 2.0, which successfully improved the perceptual quality of HEVC-compressed videos. Experimental validation of the proposed approach demonstrates its effectiveness in enhancing the visual quality of VVC-compressed videos across various test sequences and quantization parameter settings. The evaluation metric, Peak Signal-to-Noise Ratio (PSNR), quantifies the quality improvement achieved by MFQE VVC compared to the original compressed sequences.
The authors of [37] focus on the use of CNN methods for enhancing video quality, extending the enhancement of image quality to the temporal dimension. They address challenges such as understanding spatial context, motion information across frames, and fluctuating quality across frames. To address these challenges, they propose an end-to-end deep learning architecture comprising a temporal structure fusion subnet and a spatial detail enhancement subnet. The temporal structure fusion subnet estimates and compensates for temporal motion across frames, while the spatial detail enhancement subnet reduces compression artifacts and improves reconstruction quality. Quantitative evaluation of the SDTS method demonstrates its effectiveness compared to state-of-the-art algorithms, including single-frame and multiframe-based methods. The SDTS method achieves higher PSNR gains on average, indicating the importance of leveraging both spatial and temporal information for quality enhancement in VVC compressed videos.

M. Enhancement on Real-time Videos

There is a need to emphasize the importance of inclusivity in the development and research efforts of video technology. The paper of [38] investigates the impact of real-time video enhancement techniques on QoE for users with low vision. Traditionally, people with visual impairments, such as low vision or color blindness, have been excluded from subjective video evaluation studies. Therefore, there is a lack of research on video QoE specifically tailored to this user group. Although some studies have considered the needs of users with other disabilities, such as deaf users or elderly users, those with low vision remain underrepresented. The study fills this gap by presenting the results of a lab-based user study conducted with participants with low vision. The participants used a web-based media player to configure individual video quality enhancement filters, such as increased contrast, strong edge enhancement, and decreased brightness, depending on the types of content. Visual performance measurements were also performed for each participant to assess how low vision impacts video QoE. One of the primary research questions addressed in the study is the impact of basic image manipulation techniques on video QoE for users with low vision. Another question relates to how different types of content influence video enhancement. The results indicate that individual video quality filters positively impact the experience of various quality dimensions, including detail, text, and content. However, the study also reveals that the type of content does not significantly affect the rating results, suggesting that further research is needed in this area. The paper describes the methodology used to assess visual impairment among participants, including measurements of visual acuity, contrast threshold, color vision, and need for light. These measurements helped categorize the visual impairments of users, highlighting the highly individual nature of low vision. Furthermore, the study discusses the evaluation of the QoE of adapted video and the challenges in evaluating the improvement of individually adapted videos. Although basic image manipulation techniques showed positive effects on the experience of streaming media of users with low vision, significant results were only found for certain quality dimensions. Participants expressed concerns about the long-term effects of image manipulation and suggested additional features to enhance video QoE, such as selective color manipulation and automatic text/object detection.

Complementing this focus on user-specific needs, recent technological advances shown in [39] introduce efficient frameworks that support real-time video enhancement. ReBotNet, or Recurrent Bottleneck Mixer Network, addresses the challenge of high computational load and latency in conventional video restoration models. It features a dual-branch architecture that combines spatio-temporal feature extraction with temporal consistency enhancements through token-based processing and recurrent training. By leveraging a ConvNext-based encoder and a bottleneck mixer, ReBotNet effectively enhances each video frame in real time, making it well-suited for practical use cases like live video calls and streaming. The model not only outperforms existing methods in terms of computational efficiency and speed, but also introduces curated datasets that reflect real-world video scenarios. The integration of such advanced techniques into accessibility-driven research could significantly improve the quality and inclusivity of media experiences, particularly for visually impaired users. As the field progresses, the combination of user-centric design with cutting-edge real-time technologies offers a promising pathway to more equitable and responsive video systems.

N. Enhancement on Gaming Videos

The rise of streaming platforms such as Twitch.tv and Facebook Gaming has led to increased attention towards streaming gameplay scenes. However, these services face numerous challenges, including low quality source materials due to client devices, network limitations such as bandwidth and packet loss, and the need for low delay requirements. Spatial video artifacts, such as blockiness and blurriness resulting from video compression or up-scaling algorithms, significantly affect the QoE of end-users of passive gaming video streaming applications.

To address these challenges, [40] investigates solutions to enhance the video quality of compressed gaming content. It focuses on super-resolution enhancement techniques, particularly those utilizing Generative Adversarial Networks (GANs), such as SRGAN. The authors propose improvements to SRGAN, including modifications to the loss function and adjustments to the generator network's architecture, to enhance the perceived quality significantly.Given the substantial portion of Internet traffic attributed to video streaming, especially in the gaming sector, optimizing video quality is crucial for user satisfaction.

The study utilizes data sets consisting of both high-quality reference frames and lower-quality compressed or upscaled frames to train the proposed model. Different datasets are constructed to address specific research questions, such as evaluating game-specific versus generic quality enhancement models and examining the impact of content diversity on model performance. The paper details the development of the proposed GAN architecture, emphasizing the importance of loss functions in achieving high-quality enhancement. The authors introduce a modified loss function, incorporating both content loss and adversarial loss, to improve texture recovery and overall perceptual quality. Additionally, they integrate distortion loss based on retrained VGG19 networks for quality prediction, further enhancing the model's performance.

In addition, using objective and subjective measurement techniques, the study evaluates the enhancement power of the proposed model, considering factors such as content diversity, blurred versus blocky artifacts and perceptual quality improvement. The results demonstrate the effectiveness of the proposed approach in improving the quality of the gaming content, paving the way for future research on game-specific enhancement models.

The paper [41] emphasizes the importance of high-quality, inclusive, and immersive experiences in educational games, noting that technical limitations can hinder learning. ReBotNet, with its real-time video enhancement capabilities, addresses these concerns directly by improving the quality of gaming videos, enhancing frame clarity, reducing lag, and stabilizing visuals. This not only ensures a more immersive and engaging experience, but also makes game-based learning more accessible to users with limited hardware or bandwidth. By elevating the visual fidelity of educational games and simulations, ReBotNet plays a crucial role in advancing the integration of video-enhanced, game-based learning in higher education.

O. Video Editing through AI - Diffusion Models

The paper [8] discusses a new method for editing videos based on text prompts using a type of AI called diffusion models. These models aim to improve the visual quality of generated videos while maintaining the original video's layout and motion. The method focuses on ensuring consistency in the edited video by manipulating features within the diffusion model. Unlike other approaches, this method does not require extensive training and can be used with existing text-to-image editing methods. The paper compares the proposed method with various existing techniques and demonstrates superior results in terms of adhering to the editing prompt and maintaining temporal consistency. The evaluation includes measures of edit fidelity and temporal consistency, showing promising outcomes. However, the method has limitations in handling structural changes and may produce visual artifacts in some cases. The paper addresses the challenge of using diffusion-based methods for video editing while maintaining consistency in appearance over time [12]. Existing diffusion models struggle with this task, hindering their practical application in natural video editing. To address this limitation, the paper introduces temporal dependency to text-driven diffusion models, enabling them to generate consistent appearance for edited objects throughout the video. Specifically, a novel interframe propagation mechanism is developed to propagate appearance information across frames, leading to the creation of StableVideo, a text-driven video editing framework capable of consistency-aware editing.

Compared to other methods, such as Tune-a-Video and Dreamix, which focus on geometric or temporal consistency, the proposed approach aims to achieve appearance editing with both geometric and temporal consistency. By leveraging a pre-trained NLA model to propagate edited contents and decompose videos into foreground and background atlases, the method ensures homogeneous appearances and motions across the entire video. The paper presents experiments that demonstrate the effectiveness of the approach in various editing scenarios, including compositing, background replacement, and style transfer. However, limitations are acknowledged, particularly in handling nonrigid objects and specific scenarios involving humans or animals. Future improvements could involve optimizing diffusion models and addressing challenges related to structural deformation.

The authors of [14] introduce vid2vid-zero, a novel method for zero-shot video editing using standard image diffusion models, without the need for training on specific video data. Addresses the challenge of extending the success of text-to-image diffusion models to video editing, which typically requires significant computational resources and training data. The proposed approach leverages the dynamic nature of attention mechanisms to enable effective temporal modeling at test time. At the core of vid2vid-zero are several modules: a null text inversion module for text-to-video alignment, a cross-frame modeling module for temporal consistency, and a spatial regularization module for fidelity to the original video. Using a spatial-temporal attention module, the method achieves bidirectional temporal modeling for video editing without the need for extensive training. The paper discusses existing methods for text-driven image editing, highlighting recent advances in leveraging text-to-image diffusion models for fine-grained control over spatial layout and appearance. However, these methods may face limitations in handling motion changes and maintaining temporal consistency.

Comparisons with existing methods such as Tune-A-Video (TAV) and Plug-and-Play (PnP) demonstrate the effectiveness of vid2vid-zero in achieving both temporal consistency and fidelity to the original video. User preference studies and quantitative evaluations show promising results in terms of quality, text-to-video alignment, and temporal consistency. Furthermore, the paper extends vid2vid-zero to customized video editing, demonstrating its ability to replace objects in a video with customized objects of interest. The comparison results with TAV indicate superior performance in preserving object identity and avoiding artifacts.

[10] paper explores the utilization of pre-trained image diffusion models for text-guided video editing without the need for additional training. The method, termed Pix2Video, operates in two main steps: Initial edits are applied to an anchor frame using a pre-trained structure-guided image diffusion model, and then these edits are progressively propagated to future frames using self-attention feature injection, ensuring temporal coherence. The injection of features from previously edited frames into the self-attention layer of the current frame enables cross-frame attention, resulting in coherent appearance characteristics throughout the video. The evaluation of Pix2Video is conducted on various real video clips, showcasing both localized and global edits. Comparative analysis against state-of-the-art methods demonstrates the effectiveness of Pix2Video in achieving faithful edits while maintaining temporal coherence, without the need for compute-intensive preprocessing or video-specific fine-tuning. Quantitative metrics such as the CLIP score and Pixel-MSE are used to assess fidelity and temporal coherence, respectively. The results indicate that Pix2Video performs competitively or outperforms the baseline methods, showcasing its potential for practical video editing applications.

P. Video Editing in 360/ VR

A recent paper [16] introduces an innovative tool designed to create interactive 360-degree videos infused with storytelling elements, allowing seamless transitions between flat and 360-degree content within a single timeline. By integrating various media types, such as images, audio, and web resources, creators can develop branched stories through non-linear video technology. This approach opens opportunities for broadcasters to blend traditional footage with immersive 360-degree segments, offering audiences an engaging and immersive viewing experience. An essential aspect of 360-degree storytelling discussed in the paper is the choice of point of view (POV), which greatly impacts audience immersion. Whether adopting an objective (third-person) or subjective (first-person) POV, maintaining audience presence is crucial for effective storytelling in VR environments. Sound also plays a vital role in enhancing immersion and directing audience attention within immersive experiences, mirroring real-life scenarios where audio cues guide viewers to critical points of interest.

The paper reviews recent developments in immersive storytelling, including Google's Spotlight Stories for YouTube, which leverages smartphone sensors to enable interactive 360-degree animated videos. It highlights the importance of flexible and scalable systems for interactive 360-degree video design, emphasizing model-based event-driven approaches that utilize web technologies such as HTML, CSS, and JavaScript. The presented web-based editor offers a user-friendly interface for creating interactive 360-degree content, supporting various devices such as desktops, tablets, and head-mounted displays (HMDs). The editor allows for the incorporation of adaptive video formats and interactive elements using WebGL API, enabling seamless playback and interaction across different platforms. The paper also discusses user interaction methods for different devices, including touch inputs for mobile devices, remote control navigation for TVs, and head movement for HMDs. It emphasizes the need for intuitive controls to enhance user engagement and provide a consistent experience across devices. Infrastructure components such as AWS services are utilized for hosting, persistence, and content delivery, ensuring the scalability and security of the platform. Performance tests carried out during the implementation phase demonstrate the efficiency of the tool presented compared to projects ot360 ° 0 ° 0 °, with up to 30% faster performance observed. Subjective tests indicate smooth display performance, enhancing the quality of the viewer experience.

[18] paper presents EditAR, a tool designed to simplify the creation of augmented reality (AR), virtual reality (VR) and video content for digital learning purposes. Unlike traditional methods that require expertise in multiple fields, EditAR allows novices to create content by capturing their interactions within an environment and generating a digital model based on these interactions. Through consultations with experts and user studies, the effectiveness of EditAR was confirmed, showing considerable time savings and improved usability compared to conventional approaches. EditAR aims to make the creation of immersive learning experiences accessible to a wider audience, thereby democratizing the process of creating educational content.

Q. Different Tools for Video Sequence Selection

It is important to realize that the selection sequence is, respectively, of the study requirements and goals. Some of the sequences and approaches are as follows.

1) Public databases: These play an essential role in improving the assessment of video quality by providing diversified and standardized sequences. Such types of databases contribute to an extraordinary resource for research, offering viable solutions that include different types of scenarios, content types, and resolutions [42]. Common examples include

- **UGC-VQA** User Generated Content Video Quality Assessment' (utilised for video quality evaluation for content generated by users)
- **LIVE Video Database** (utilised for video quality evaluation of a diversified content-based scenario)
- **KonIQ – 10k Database** (utilized for incorporating high-quality videos and images for quality assessment)
- **VQEG Databases** 'Video Quality Experts Group' (serves as a reference for VQ evaluation metrics and algorithms)

The public database enlightens comparability and consistency between different studies. Through these databases, researchers can draw valid conclusions and gather information about advances in video quality optimization.

2) Subjective Tools for Evaluation: There are two main aspects of quality assessment according to research [35]; These include subjective and objective evaluation. The subjective assessment is, respectively, of human accessors, whereas the objective assessment is, respectively, of the measurement of subjective quality. In most scenarios, automated metrics can produce inaccurate results. Subjective evaluation tools play a key role in minimizing such errors through a human-centric approach. There are two types of methodology that are essential for subjective evaluation tools.

- **SSCQE** "Single Stimulus Continuous Quality Evaluation" (Real Time Feedback through Continuous Rating for video quality)[43]
- **DSCQE** 'Double Stimulus Continuous Quality Evaluation' (Comparing videos by referencing and distorting)

Subjective scores are effectively compiled using these methodologies to improve human perception. In addition, the evaluation includes aspects such as the overall viewing experience, visual artifacts, and color accuracy. Evaluation through subjective tools can also play an important role in defining the knowledge gap between human perception and technical metrics.

3) Content Aggregators: Content aggregators such as the YouTube Data API and Vimeo API serve as essential tools to access and use large-scale video datasets. These platforms provide researchers and developers with programmatic access to a wide array of multimedia content, encompassing diverse resolutions, encoding formats, and content categories. This access enables the efficient collection of video materials required for tasks such as quality assessment, content analysis, and the development of video processing algorithms. The availability of heterogeneous video sources through these APIs allows the evaluation of algorithms under varying conditions, thereby enhancing their robustness and generalizability. In addition, by facilitating the exploration of different types of content and technical parameters, these aggregators contribute significantly to the research on video quality assessment, where the influence of source variability is a critical factor. As highlighted in previous work [44], the integration of content aggregators into research workflows supports the design of more effective and scalable multimedia systems.

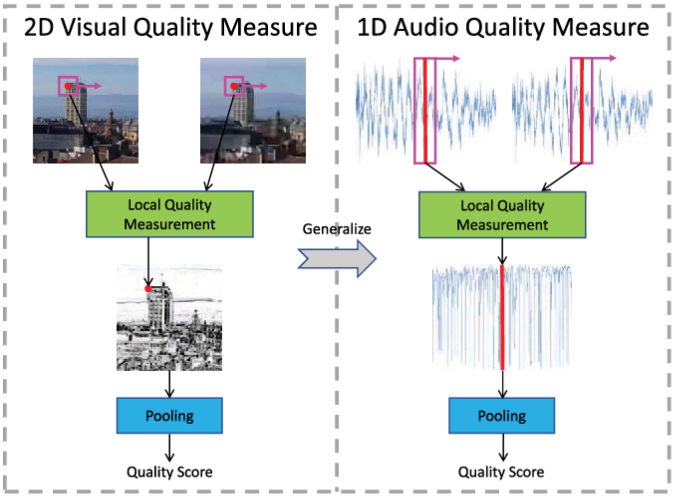

Fig. 2. Work flow: 2D VQA and 1D AQA [45]

4) Objective metrics for quality assessment: The perceptual quality and fidelity of a specific video sequence are quantified through automated algorithms that provide objective metrics [46]. Common examples include

- **PSNR** 'peak signal to noise ratio' (ratio of maximum possible signal power to corrupting noise power)
- **SSIM** structural similarity index (similarity between distorted and reference images)
- **VMAF** 'Video Multimethod Assessment Fusion" (utilization of multiple perceptual features to predict subjective quality through a machine learning model – initially developed by Netflix)

The performance of the processing algorithms, video encoding, and streaming is effectively evaluated through quantitative measures that serve as objective metrics. In short, these metrics provide technical traits for a specific video quality.

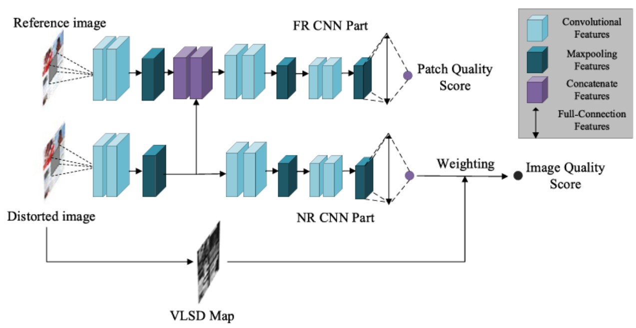

Fig. 3. NR-IQA (No-Reference Image Quality Assessment) architecture [47]

5) Machine learning algorithms: Complex multimedia patterns are effectively predicted by using computational models that serve to optimize and assess video quality [48]. Common types of machine learning algorithms include.

- **CNN's** 'convolutional neural networks' (utilised for video and image classification-related tasks enlightening significant features related to human visual processing)
- **RNNs** (recurrent neural networks) (effective for analysis and prediction of video sequences along with capture of temporal dependencies)

In scenarios where a sequence needs to be selected based on criteria such as complexity, user engagement, or diversity, machine learning algorithms can be used effectively to select and recognize video sequences [49]. The importance of these algorithms lies in aspects like adaptability and automation serving as an effective solution to enlighten maximum productivity.

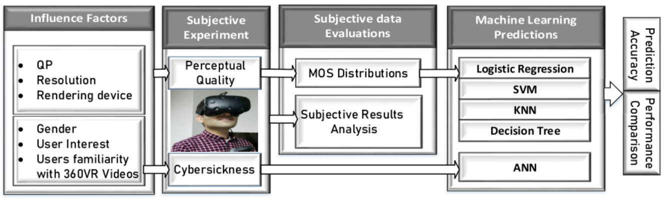

Fig. 4. Machine Learning Prediction and Subjective QoE Evaluation [50]

V. Comparative Analysis

The diversified nature of video quality evaluation and subjective quality evaluation is better analyzed through comparative analysis. It also includes diversified testing approaches and the role of emotions and content. The main objective of the analysis is to provide a comparison of the methodologies and findings that have been previously accessed. Therefore, the main aspects of the analysis are as follows.

- It is important to ensure a detailed understanding of the specified content, as it is important to align with user perspectives. The type of content is of immense importance, as it highlights the variations in video quality impairments. Subjective evaluations are better reflected by prioritizing content characteristics. Furthermore, there is also a further need for a well-defined assessment criterion that should be in accordance with a specific content type. In short, these shall reflect a well-tailored approach that enlightens VQA across different genres of content.
- Most importantly, research studies revealed that there is a correlation between emotional responses and VQA. Most studies suggested that the target emotions in women were higher than in men. While accessing quality, it is important to consider emotional responses to have a better understanding of the video content during the assessment.
- Multi-instrumental approaches are of immense importance as these are mainly focused on cognitive, affective, and behavioral processes. There existed an inverse proportionality between disturbances due to errors and the overall quality that highlighted the minimization of disruptive factors. In addition to this, there also exists a direct proportionality between content desirability and pleasure. This reflected that quality perceptions were the result of cognitive and affective processes. Evaluation methodologies in the perspective of video quality assessment are well refined because of the multi-instrumental approach.
- There exists a significant difference between task-oriented and entertainment-oriented tests. The first important aspect is scene selection, which is different in both. Scenario-based scene selection occurs in task-oriented testing, whereas entertainment testing is irrespective of any scenario. It was only meant for entertainment purposes. Some of the essential elements of task-oriented testing include gestures, fingerwriting, role shifting, and nonmanual markers.

VI. Discussion

A. Review of Primary Literature

The primary literature highlights several key considerations regarding content selection for video streaming and broadcast scenarios. One recurring theme is the importance of including various scenes and complexities in the chosen content. This approach aims to accurately depict the variability encountered in real-world video streaming and broadcast situations. Dynamic content, characterized by scenes that feature rapid motion or intricate visual details, is frequently favored for evaluating the performance of video codecs and streaming algorithms under challenging conditions.
In addition, there is a growing advocacy for content selection that closely mirrors real-world usage patterns, taking into account factors such as resolution, frame rate, and the balance between static and dynamic scenes. Additionally, User-Generated Content (UGC) is gaining prominence as a valuable resource for content selection due to its diversity and authenticity. Researchers are actively exploring methods to incorporate UGC into assessments, despite the challenges associated with its variability. These insights underscore the importance of thoughtful content selection in advancing video streaming and broadcast technologies to meet the evolving needs of users in diverse multimedia environments.

B. Challenges in Current Literature

1) Subjectivity in Content Selection Criteria: The subjective nature of defining criteria, such as determining what qualifies as a "representative" scene or "real-world" content, introduces variability in content selection. This subjectivity poses challenges in ensuring the reproducibility of experiments.

2) Lack of Standardized Criteria: One of the main challenges stems from the absence of universally accepted and standardized criteria for content selection. The diversity in assessment goals and methodologies has resulted in a fragmented landscape, making it challenging to compare the results between studies.

3) Adaptation to Evolving Video Technologies: The rapid evolution of video technologies, including AR and VR, complicates content selection. Current practices may struggle to keep up with emerging formats and usage scenarios.

4) Limited Availability of Benchmark Datasets: The availability of comprehensive and diverse benchmark data sets is limited. This scarcity hampers the ability to perform consistent and comparative evaluations in different studies.

C. RESEARCH CONTRIBUTION

1) Involve various stakeholders, including researchers, industry experts, and end users, in the definition of content selection criteria. This approach ensures a more comprehensive and representative set of guidelines.
2) Encourage the creation and maintenance of open databases that encompass a wide range of content types and characteristics. These databases can serve as reference data sets for researchers and ensure consistency between studies.
3) Develop scenario-based content selection criteria that take into account emerging technologies like AR and VR. This approach ensures that the assessments consider the unique requirements of these technologies.
4) Explore the use of machine learning algorithms to automate aspects of content selection. These algorithms can analyze large data sets and identify content that is in accordance with predefined criteria.

VII. CONCLUSIONS

This review comprehensively examined the evolving landscape of video enhancement, with a focus on content selection and sequencing in both conventional and immersive environments (AR/VR). By synthesizing recent literature, we highlight the critical role of content characteristics, user perception, and contextual factors in shaping Quality of Experience (QoE). The integration of machine learning and AI-driven techniques has emerged as a promising direction for automating and optimizing video quality assessments. However, persistent challenges, such as the lack of standardized criteria, limited benchmark datasets, and the subjective nature of the desirability of the content, continue to constrain progress. To address these gaps, we advocate for the development of inclusive, scenario-based evaluation frameworks and collaborative datasets that reflect real-world diversity. Ultimately, advancing video enhancement requires a multidisciplinary approach that bridges technical innovation with human-centric design, ensuring robust, scalable solutions for next-generation media experiences.

ABBREVIATIONS

The following abbreviations are used in this manuscript:

Abbreviation	Definition
AI	Artificial Intelligence
AR	Augmented Reality
GAN	Generative Adversarial Networks
HEVC	High Efficiency Video Coding
HMDs	Head-Mounted Displays
LSTM	Long-Short Term Memory
MFQE	Multi-Frame Quality Enhancement models
ML	Machine Learning
PSNR	Peak Signal-to-Noise Ratio
QoE	Quality of Experience
UGC	User-Generated Content
VMAF	Video Multimethod Assessment Fusion
VQA	Video Quality Assessment
VR	Virtual Reality
VVC	Versatile Video Coding

REFERENCES

[1] G. Cermak, M. Pinson, and S. Wolf, "The Relationship Among Video Quality, Screen Resolution, and Bit Rate," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 258–262, 2011.

[2] J. Allard, A. Roskuski, and M. Claypool, "Measuring and modeling the impact of buffering and interrupts on streaming video quality of experience," in *Proceedings of the 18th International Conference on Advances in Mobile Computing & Multimedia*, ser. MoMM '20. New York, NY, USA: Association for Computing Machinery, 2021, p. 153–160. [Online]. Available: https://doi.org/10.1145/3428690.3429173

[3] B. Spang, B. Walsh, T.-Y. Huang, T. Rusnock, J. Lawrence, and N. McKeown, "Buffer sizing and Video QoE Measurements at Netflix," in *Proceedings of the 2019 Workshop on Buffer Sizing*, ser. BS '19. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3375235.3375241

[4] J. Leirpoll, D. Osborn, P. Murphy, and A. Edwards, *The Cool Stuff in Premiere Pro*. Berkeley, CA: Apress, 2017. [Online]. Available: http://link.springer.com/10.1007/978-1-4842-2890-6

[5] H. Zhang, M. A. Babar, and P. Tell, "Identifying relevant studies in software engineering," *Information and Software Technology*, vol. 53, no. 6, p. 625–637, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950584910002260

[6] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *Journal of Business Research*, vol. 104, p. 333–339, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0148296319304564

[7] K. Adnan, R. Akbar, and K. S. Wang, "Development of usability enhancement model for unstructured big data using SLR," *IEEE Access*, vol. 9, pp. 87 391–87 409, 2021.

[8] M. Geyer, O. Bar-Tal, S. Bagon, and T. Dekel, "TokenFlow: Consistent Diffusion Features for Consistent Video Editing," no. arXiv:2307.10373, 2023, arXiv:2307.10373 [cs]. [Online]. Available: http://arxiv.org/abs/2307.10373

[9] B. Peng, X. Zhang, J. Lei, Z. Zhang, N. Ling, and Q. Huang, "LVE-S2D: Low-Light Video Enhancement From Static to Dynamic," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 12, pp. 8342–8352, 2022.

[10] D. Ceylan, C.-H. P. Huang, and N. J. Mitra, "Pix2Video: Video Editing using Image Diffusion," no. arXiv:2303.12688, 2023, arXiv:2303.12688 [cs]. [Online]. Available: http://arxiv.org/abs/2303.12688

[11] Y. Ye, J. M. Boyce, and P. Hanhart, "Omnidirectional 360° Video Coding Technology in Responses to the Joint Call for Proposals on Video Compression With Capability Beyond HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 5, pp. 1241–1252, 2020.

[12] W. Chai, X. Guo, G. Wang, and Y. Lu, "StableVideo: Text-driven Consistency-aware Diffusion Video Editing," no. arXiv:2308.09592, 2023, arXiv:2308.09592 [cs]. [Online]. Available: http://arxiv.org/abs/2308.09592

[13] S. D. Roy and M. K. Bhowmik, "Annotation and Benchmarking of a Video Dataset under Degraded Complex Atmospheric Conditions and Its Visibility Enhancement Analysis for Moving Object Detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 3, pp. 844–862, 2021.

[14] W. Wang, Y. Jiang, K. Xie, Z. Liu, H. Chen, Y. Cao, X. Wang, and C. Shen, "Zero-Shot Video Editing Using Off-The-Shelf Image Diffusion Models," no. arXiv:2303.17599, 2024, arXiv:2303.17599 [cs]. [Online]. Available: http://arxiv.org/abs/2303.17599

[15] X. Meng, X. Deng, S. Zhu, X. Zhang, and B. Zeng, "A Robust Quality Enhancement Method Based on Joint Spatial-Temporal Priors for Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 6, pp. 2401–2414, 2021.

[16] L. Bassbouss, S. Steglich, and I. Fritzsch, "Interactive 360° Video and Storytelling Tool," in *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)*, 2019, pp. 113–117.

[17] C. Cao, H. Yue, X. Liu, and J. Yang, "Unsupervised HDR Image and Video Tone Mapping via Contrastive Learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 2, pp. 786–798, 2024.

[18] S. Chidambaram, S. S. Reddy, M. Rumple, A. Ipsita, A. Villanueva, T. Redick, W. Stuerzlinger, and K. Ramani, "EditAR: A Digital Twin Authoring Environment for Creation of AR/VR and Video Instructions from a Single Demonstration," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2022, pp. 326–335.

[19] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the impact of video quality on user engagement," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 4, p. 362–373, 2011. [Online]. Available: https://doi.org/10.1145/2043164.2018478

[20] J. Merikivi, J. Bragge, E. Scornavacca, and T. Verhagen, "Binge-watching serialized video content: A transdisciplinary review," *Television & New Media*, vol. 21, no. 7, p. 697–711, 2020. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1527476419848578

[21] L. Pozueco, A. Álvarez, X. García, R. García, D. Melendi, and G. Díaz, "Subjective video quality evaluation of different content types under different impairments," *New Rev. Hypermedia Multimedia*, vol. 23, no. 1, p. 1–28, 2017. [Online]. Available: https://doi.org/10.1080/13614568.2016.1152310

[22] C. Ma, Z. Shi, Z. Lu, S. Xie, F. Chao, and Y. Sui, "A Survey on Image Quality Assessment: Insights, Analysis, and Future Outlook," 2025. [Online]. Available: https://arxiv.org/abs/2502.08540

[23] S. Jamil, "Review of Image Quality Assessment Methods for Compressed Images," *Journal of Imaging*, vol. 10, no. 5, 2024. [Online]. Available: https://www.mdpi.com/2313-433X/10/5/113

[24] I. van der Linde and R. M. Doe, "Influence of affective image content on subjective quality assessment," *J. Opt. Soc. Am. A*, vol. 29, no. 9, pp. 1948–1955, 2012. [Online]. Available: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-29-9-1948

[25] Y. Zhu, I. Heynderickx, and J. A. Redi, "Understanding the role of social context and user factors in video quality of experience," *Computers in Human Behavior*, vol. 49, pp. 412–426, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563215001752

[26] R. Evans, "Why Does Content Desirability Impact Subjective Video Quality Ratings and What Can Be Done About It?" Theses, Rice University, 2011. [Online]. Available: https://hdl.handle.net/1911/64431

[27] P. Kortum and M. Sullivan, "The effect of content desirability on subjective video quality ratings," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 52, no. 1, p. 105–118, 2010. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0018720810366020

[28] J. J. Gross and R. W. L. and, "Emotion elicitation using films," *Cognition and Emotion*, vol. 9, no. 1, pp. 87–108, 1995. [Online]. Available: https://doi.org/10.1080/02699939508408966

[29] K. de Moor, F. Mazza, I. Hupont, M. Ríos Quintero, T. Mäki, and M. Varela, "Chamber QoE: a multi-instrumental approach to explore affective aspects in relation to quality of experience," in *Human Vision and Electronic Imaging XIX*, vol. 9014, San Francisco, CA, United States, 2014, p. 90140U. [Online]. Available: https://hal.science/hal-00952701

[30] M. H. Pinson, *How To Choose Video Sequences For Video Quality Assessment*, 2013. [Online]. Available: https://its.ntia.gov/publications/details.aspx?pub=2694

[31] M. H. Pinson, L. Janowski, and Z. Papir, "Video Quality Assessment: Subjective testing of entertainment scenes," *IEEE Signal Processing Magazine*, vol. 32, no. 1, pp. 101–114, 2015.

[32] M. H. Pinson, M. Barkowsky, and P. Le Callet, "Selecting scenes for 2D and 3D subjective video

quality tests," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, p. 50, 2013. [Online]. Available: https://jivp-eurasipjournals.springeropen.com/articles/10.1186/1687-5281-2013-50

[33] A. Raake, S. Borer, S. M. Satti, J. Gustafsson, R. R. R. Rao, S. Medagli, P. List, S. Göring, D. Lindero, W. Robitza, G. Heikkilä, S. Broom, C. Schmidmer, B. Feiten, U. Wüstenhagen, T. Wittmann, M. Obermann, and R. Bitto, "Multi-Model Standard for Bitstream-, Pixel-Based and Hybrid Video Quality Assessment of UHD/4K: ITU-T P.1204," *IEEE Access*, vol. 8, pp. 193 020–193 049, 2020.

[34] G. Kougioumtzidis, V. Poulkov, Z. D. Zaharis, and P. I. Lazaridis, "A Survey on Multimedia Services QoE Assessment and Machine Learning-Based Prediction," *IEEE Access*, vol. 10, pp. 19 507–19 538, 2022.

[35] Z. Hu, H. Yan, T. Yan, H. Geng, and G. Liu, "Evaluating QoE in VoIP networks with QoS mapping and machine learning algorithms," *Neurocomputing*, vol. 386, pp. 63–83, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:213362421

[36] X. HoangVan and H.-H. Nguyen, "Enhancing Quality for VVC Compressed Videos with Multi-Frame Quality Enhancement Model," in *2020 International Conference on Advanced Technologies for Communications (ATC)*, 2020, pp. 172–176.

[37] X. Meng, X. Deng, S. Zhu, and B. Zeng, "Enhancing Quality for VVC Compressed Videos by Jointly Exploiting Spatial Details and Temporal Structure," in *2019 IEEE International Conference on Image Processing (ICIP)*. Taipei, Taiwan: IEEE, 2019, p. 1193–1197. [Online]. Available: https://ieeexplore.ieee.org/document/8804469/

[38] A. Sackl, R. Schatz, S. Suette, and M. Tscheligi, "From Low Vision to High Quality: Video QoE Enhancement for Visually Impaired Users," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, 2019, pp. 1–6.

[39] J. M. Jose Valanarasu, R. Garg, A. Toor, X. Tong, W. Xi, A. Lugmayr, V. M. Patel, and A. Menini, "Rebotnet: Fast real-time video enhancement," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025, pp. 1424–1435.

[40] N. J. Avanaki, S. Zadtootaghaj, N. Barman, S. Schmidt, M. G. Martini, and S. Moller, "Quality Enhancement of Gaming Content using Generative Adversarial Networks," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. Athlone, Ireland: IEEE, 2020, p. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9123074/

[41] A. Hulus, "The evolution and future of video games in higher education: Critical perspectives and research directions," *Research in Education*, p. 00345237251342128, 2024.

[42] D. Li, T. Jiang, and M. Jiang, "Quality Assessment of In-the-Wild Videos," in *Proceedings of the 27th ACM International Conference on Multimedia*, ser. MM '19. ACM, 2019. [Online]. Available: http://dx.doi.org/10.1145/3343031.3351028

[43] M. A. Gracheva, V. P. Bozhkova, A. A. Kazakova, I. P. Nikolaev, and G. I. Rozhkova, "Subjective assessment of the quality of static and video images from mobile phones," in *International Conference on Machine Vision*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:211554205

[44] F. Wang, F. Wang, J. Liu, R. Shea, and L. Sun, "Intelligent Video Caching at Network Edge: A Multi-Agent Deep Reinforcement Learning Approach," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, 2020, pp. 2499–2508.

[45] X. Min, G. Zhai, J. Zhou, M. C. Q. Farias, and A. C. Bovik, "Study of Subjective and Objective Quality Assessment of Audio-Visual Signals," *IEEE Transactions on Image Processing*, vol. 29, pp. 6054–6068, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:216110143

[46] A. Ahar, T. Birnbaum, M. Chlipala, W. Zaperty, S. Mahmoudpour, T. Kozacki, M. Kujawinska, and P. Schelkens, "Comprehensive performance analysis of objective quality metrics for digital holography," *Signal Processing: Image Communication*, vol. 97, p. 116361, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0923596521001661

[47] X. Jiang, L. Shen, Q. Ding, L. Zheng, and P. An, "Screen content image quality assessment based on convolutional neural networks," *Journal of Visual Communication and Image Representation*, vol. 67, p. 102745, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1047320319303669

[48] B. Mahesh, "Machine Learning Algorithms -A Review," *International Journal of Science and Research*, vol. 9, no. 1, p. 381–386, 2020.

[49] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, 2021.

[50] M. S. Anwar, J. Wang, W. Khan, A. Ullah, S. Ahmad, and Z. Fei, "Subjective qoe of 360-degree virtual reality videos and machine learning predictions," *IEEE Access*, vol. 8, pp. 148 084–148 099, 2020.