Quality Estimation Models for Gaming Video Streaming Services Using Perceptual Video Quality Dimensions

Saman Zadtootaghaj

Quality and Usability lab

TU Berlin

Berlin, Germany

saman.zadtootaghaj@qu.tu-berlin.de

Steven Schmidt

Quality and Usability lab

TU Berlin

Berlin, Germany

steven.schmidt@tu-berlin.de

Saeed Shafiee Sabet

SimulaMet

Simula

Oslo, Norway

saeed@simula.no

Sebastian Möller

Quality and Usability lab

TU Berlin / DFKI Projektbüro

Berlin, Germany

sebastian.moeller@tu-berlin.de

Carsten Griwodz

Department of Informatics

University of Oslo

Oslo, Norway

griff@ifi.uio.no

ABSTRACT

The gaming industry is one of the largest digital markets for decades and is steady developing as evident by new emerging gaming services such as gaming video streaming, online gaming, and cloud gaming. While the market is rapidly growing, the quality of these services depends strongly on network characteristics as well as resource management. With the advancement of encoding technologies such as hardware accelerated engines, fast encoding is possible for delay sensitive applications such as cloud gaming. Therefore, already existing video quality models do not offer a good performance for cloud gaming applications. Thus, in this paper, we provide a gaming video quality dataset that considers hardware accelerated engines for video compression using the H.264 standard. In addition, we investigate the performance of signal-based and parametric video quality models on the new gaming video dataset. Finally, we build two novel parametric-based models, a planning and a monitoring model, for gaming quality estimation. Both models are based on perceptual video quality dimensions and can be used to optimize the resource allocation of gaming video streaming services.

CCS CONCEPTS

• Information systems → Multimedia streaming;

KEYWORDS

Gaming Video Streaming, Quality Assessment, Dataset

Saman Zadtootaghaj, Steven Schmidt, Saeed Shafiee Sabet, Sebastian Möller, Carsten Griwodz. 2020. Quality Estimation Models for Gaming Video Streaming Services Using Perceptual Video Quality Dimensions. *In 11th ACM Multimedia Systems Conference (MMSys '20)*, June 8–11, 2020, Istanbul, Turkey. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/1234567890>

1 Introduction

Gaming as a popular system has recently expanded the associated services by stepping into live streaming services. Many players broadcast their gameplay to thousands of viewers, who are passively watching the user generated content on popular platforms such as Twitch.tv.

While the game industry is growing, more complex games in terms of processing power are getting developed. In order to play these high-end games, players are required to update their end devices every few years. One solution for this is to move the heavy processes such as rendering to the cloud and cut the need for high-end hardware devices for customers. Additionally, cloud gaming offers more flexibility to users by allowing them to play any game on any type of device which is capable of displaying a video and is equipped with a suitable input device. Apart from processing power, additional advantages of cloud gaming are its platform independency, i.e. a game can be played on every client operating system independently to the operating system of the server, and security against piracy as the game content is in the hand of service providers. Although cloud gaming has been seen as a promising application on the top of IP network to grab a huge share of gaming industry revenue, so far only a few companies provide such a service. Cloud gaming as a real-time application suffers from the additional delay due to video encoding and decoding, end-to-end transmission delay as well as other network constraints such as limited bandwidth and packet loss. The Quality of Experience (QoE) of costumers can be negatively affected by these limitations.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

MMSys'20, June 8–11, 2020, Istanbul, Turkey

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-6845-2/20/06 \$15.00

DOI: <https://doi.org/10.1145/3339825.3391872>

In order to optimize the user satisfaction despite these challenges, strategies for resource and network management are highly necessary. As a ground truth for these strategies, subjective user ratings for various system configurations are required. Due to the high costs of subjective tests, the number of available datasets of gaming video quality and with that also proposed models is very limited. Currently, typical quality models for video contents are only candidates for quality prediction of gaming video streaming services. However, video games are artificial and synthetic, which potentially can be used to improve quality predictions. Furthermore, cloud gaming is a highly delay sensitive service which requires the use of hardware accelerated video encoding and decoding such as NVIDIA's encoder (NVENC). This encoding performs differently compared to software encoders [1]. In addition, due to latency constraints of cloud gaming services, a special set of encoding parameters (e.g. llhq, llhp modes of NVENC or fixed macroblock size) might be used to reduce the compression delay. The differences in the process of video encoding for cloud gaming service compared to other video streaming applications raises the question if currently available video quality models can be used for gaming content. This paper focuses on the development of video quality models for gaming content in passive viewing-and-listening paradigm. It must be noted that for an overall gaming QoE model also other quality features related to the interactive nature of such a service have to be considered.

In the following section, we first review the existing quality models for gaming and non-gaming content. In section 3, we provide a large-scale gaming video dataset, which includes popular gaming contents processed with state-of-the-art cloud gaming settings. The performance of popular signal-based and parametric video quality models are presented in section 4 and 5 respectively. In section 6, we propose two parametric-based models for quality prediction of streamed gaming videos. Finally, the discussion and conclusion are presented in section 7 and 8 respectively.

2 Related Work

In this section, we first review the latest works that have been done considering gaming quality assessment and prediction. Next, we review a classification of video quality models.

2.1 Gaming Quality Models

We have seen several attempts to build quality models that predict gaming QoE. While several quality dimensions construct the gaming QoE [2], two important dimensions for gaming QoE are considered in the recently published ITU-T recommendation G.1072 [3]: interaction quality and video quality. Interaction quality, referred to as input quality in [3], takes into account the technical factors that may influence the interaction of users with the game. In addition to the interaction quality, video quality is important which refers to visual aspect of the video stream. It has to be noted that gaming QoE is multidimensional construct

consisting of several dimension such as immersion, flow and presence. Some of these dimensions are mainly influenced by the game content itself, e.g. due to the game design, challenges, and rules, or by preferences of the player. However, from the perspective of a cloud gaming providers, who is not *per se* a content creator, these dimensions are not feasible to be measured or tracked. Therefore, most of the state-of-the-art work, model the gaming QoE based on interaction and video quality.

With respect to interaction quality, several works have been done to model the impact of packet loss and delay on gaming QoE. Schmidt et al. [4] model the effect of delay on gaming QoE for two different games and showed that the delay sensitivity level of games can be different even between two levels of the same game.

Furthermore, a considerable amount of work has focused on the development of models that predict video quality based on encoding parameters or the video signal. Zadtootaghaj et al [5] proposed a NR machine learning-based video quality metric named NR-GVQM for gaming content which is focused on frame-level feature extraction. The idea was to collect low-level image features from the frame and train it over the VMAF score. Göring et al. [6] proposed a NR metrics called Nofu, which is a pixel-based video quality model for gaming content. Nofu uses 12 different frame-based values and a center crop approach for the fast computation of frame-level features. It further uses frame-level features pooling at video-level and feeds the features to a machine learning-based model for the model development.

Finally, a few works have been done which consider both video and interaction quality. Among them, Zadtootaghaj et al. [7] investigate the effect of temporal and spatial video compression on gaming QoE and build a model based on the bitrate and frame rate. Wang et al. [8] has conducted a study, investigating the effect of a wide range of influencing factors on perceived quality and modeled the quality based on the assessed factors. They proposed a model, named GMOS, which is inspired by ITU-T E-Model [9], considering seven influence factors: game type, codec, resolution, frame rate, PSNR, delay, and packet loss. Undoubtedly, the most comprehensive work has been done at ITU Telecommunication Standardization Sector (ITU-T) study group 12, which models the gaming QoE based on several influencing factors ranging from encoding parameters, game types to network degradations. The result is published as recommendation named G.1072 [3]. This recommendation takes into account three quality dimensions for gaming QoE, spatial video quality, temporal video quality and input quality (interaction quality). Since we focus on the video quality, in this paper we only consider the video quality dimension of G.1072.

2.2 Classification of Video Quality Models

We can categorize objective¹ quality assessment techniques according to different aspects. The first classification is based on the amount of source signal information that is needed to run the

¹ We use the term objective quality here, as it is very common in the video quality community. However, it has to be noted that the methods still rely on subjective quality ratings, and thus are not purely objective.

instrumental assessment. Typically, three types of objective quality assessment metrics are considered. No-reference models (NR), Reduced-reference models (RR) and Full-reference models (FR). NR methods require no knowledge from the original signal before transmission. RR methods use some extracted features of the source signal and predict the quality by combining this information with measurements of the received signal. FR methods have full access to the original signal and can compare it to the received video.

Alternatively, the objective quality assessment models can be classified according to the level of information used in regard to the network and transmitted packets. Based on this approach, objective quality assessment metrics are classified into four categories, planning models, bitstream models, signal-based models and hybrid models.

Planning models take the assumed network and client parameters as input and calculate an expected quality score. Bitstream models, which are divided into two categories, use the packet information to predict the quality. The first category of bitstream models predicts the quality by packet-header information (e.g., HTTP or RTP) without accessing the payload information. Therefore, these models do not explicitly consider source and encoding information to evaluate the quality. However, these models try to predict the quality by considering coding and IP network impairments. Examples of such models include the ITU-T P.1201 recommendation series [10].

The second category of the bitstream models uses the payload information. The quality prediction using bitstream payload-based

models is not only achieved by application and transport layer information (e.g., HTTP/TCP or RTP/UDP), but also by extracting and analyzing content features from the coded bitstream. An example of this case is the ITU-T P.1202 and P.1203 family of recommendations [11].

Finally, in a signal-based model, quality is determined by analyzing the decoded media signal, and in hybrid models, decoded signals are combined with information from the bitstream or the packet layer. The above classification is illustrated in Figure 1 according to [12].

These classifications are important for QoS/QoE monitoring since the possibilities for the implementation of monitoring systems heavily depends on the amount of information that is available. For example, the use of a full-reference model for large-scale monitoring of an IPTV system would be practically unfeasible, as the amount of the information that is required to be transmitted from source to the monitoring points over the network is too large. For cloud gaming service, as the cloud gaming providers have access to the bitstream information, bitstream models are one of the practical models as they are less complex compared to signal-based model, while providing a decent prediction.

3 Dataset

There exist many video quality datasets, while to the best knowledge of the authors there are only three video quality datasets using gaming content [13], [14]. For delay sensitive cloud gaming service, currently, most of the providers (e.g. Geforce Now and Parsec) use the hardware accelerated implementation of H.264/MPEG-AVC standard. Due to the lack of available datasets using such a fast encoding setting, we created a new gaming video dataset, which is called CGVDS. Compared to the three other gaming video datasets, CGVDS uses hardware accelerated implementation of H.264/MPEG-AVC (NVENC), a wider spread of video games, more encoding parameters and videos captured at 60 fps. Table 1 gives a short overview of the three available datasets compared to the CGVDS dataset in this paper.

In series of subjective tests with over 100 participants, the influence of three important encoding parameters using 15 different games on the perceived video quality was assessed. The subjective tests were conducted as a passive viewing-and-listening test as described

Figure 1: Overview of objective quality assessment models according to different levels of access to information.

Table 1: Comparison of existing gaming video quality dataset.

	Claypool dataset	GVSET [5]	KUGVD [6]	CGVDS
Influencing Factors	Bitrate	Resolution, Bitrate	Resolution, Bitrate	Resolution, Bitrate, Frame rate
Frame Rate	30 fps	30 fps	30 fps	60, 30, 20 fps
Preset	-	Veryfast	Veryfast	llhq
Encoding mode	CBR	CBR	CBR	CBR
Number of stimuli	30	90	90	360 + anchor stimuli
Encoder	MPEG	FFmpeg, x264	FFmpeg, x264	FFmpeg, NVENC (H.264)
Resolution	800x600	480p, 720p, 1080p	480p, 720p, 1080p	480p, 720p, 1080p

in ITU-T Rec. P.809 [15]. In order to avoid fatigue due to lengthy subjective tests, five studies were designed, each with 3 games and four different bitrate levels and two or three resolution and frame rate levels resulting in total of 72 stimuli per study which requires one-hour test duration per study. Two studies are designed using 3 resolutions and 2 frame rate levels (60 and 30 fps), which is called part-1, while three other studies are created using 2 different resolutions (1080p and 720p) and three frame rate levels, which is called part-2. Such a split was done in order to fit 15 games into 5 series of subjective tests using a within subject design for at least a block of 3 games in each study. Each study was conducted with a minimum of 20 subjects. In order to mix all these studies, three video sequences as anchor conditions were added to the subjective test. Prior to the test, three video sequences are shown to the participants for training purposes and to ensure that the participants understand the used measurement scales (cf. Section 3.3). Table 2 gives a summary of encoding parameters used in the subjective test. In addition, Table 3 shows the selected bitrate per resolution-frame rate pairs.

GPU hardware accelerator encoding (NVENC) was used, as current industry strategies apply this type of encoding to reduce the time for the encoding process and with that, the overall round-trip delay for players. The reference gaming video sequences were captured losslessly in the RGB format at the frame rate of 60 fps. The sequences were recorded using Fraps² and encoded by FFmpeg³.

Table 2: Summary of video encoding parameters.

Parameter	Value
Duration	30 s
Resolution	1080p, 720p, 480p
Frame Rate	60, 30, 20 fps
Encoder	FFmpeg- NVENC (H.264)
Encoding Mode	CBR
Video Compression	Standard H.264, Main 4.0
Preset	llhq (low latency, high quality)

Table 3: Resolution-Frame rate pairs for different bitrates used to obtain distorted (compressed) video sequences.

Resolution	Frame rate	Bitrate (kbps)
1080p	30, 60	6000, 2000, 1000, 300
720p	30, 60	50000, 4000, 2000, 1000
480p	30, 60	6000, 2000, 1000, 300
1080p, 720p, 480p	20	50000, 6000, 2000, 1000, 300

Details about the subjective tests are described in the following sections. The subjective methodology used for the experiments are mainly based on the work summarized in ITU-T Rec. P.809 [15]. However, a few differences applied in the procedure of experiments such as adding three more questions in order to assess sub-dimensions of video quality that are presented in [16] and were also used for the development of ITU-T Rec. G.1072 [3].

3.1 Test Setup

For the subjective tests, the light conditions in the test rooms, its acoustical properties, as well as the viewing distance and position of participants were consistent for all experiments. In general, ITU-T Rec. P.910 [17] and P.911[18] were considered. Participants were offered an adjustable chair and table to sit in a proper position (the chair allows participant to have their feet flat on the floor and their knees equal to, or slightly lower than, their hips). The viewing distance (D) was equal to three times the picture height (H).

For all five series of subjective tests, the following elements remained constant:

- Participants were recruited as described in [15].
- Test duration: maximum of 1 hour for each experiment.
- Video presentation: LCD monitor, 24" display for standard gaming, HD1080 resolution, no G-Sync, no Free-Sync, (others see [18]).
- Audio presentation: With the aim at modeling the video quality, no audio was presented to participants.
- Participants were asked to sit in the center of the video display at the specified viewing distance.

Additionally, the following monitor settings were used:

- Default values (auto) for luminance, contrast and color shade of white were used.
- Brightness was adjusted according to Rec. ITU-T P.910, but not the contrast (it might change the balance of the color temperature).
- The gamma was set to 2.2. and color temperature to 6500 K (default value on most LCDs).
- The refresh rate of the monitor was 60 Hz.

3.2 Participants

Gaming QoE is highly influenced by many user factors. While it is not the aim of this study to accurately predict the user types and their preferences, there are certain characteristics of participants that might be controlled and instructions for the participants should aim to maximize reliability, validity, and objectivity of the results.

For each subjective test, a minimum of 20 participants were participated. All participants in our experiments fulfilled the following criteria:

- Age range of 18 to 35 years (M=25.0 years, SD=4.1 years)
- Gender balance of a 7:3 ratio (out of 101 participants, 71 were males, 28 females, and 2 others)
- Participants have normal (20/30) visual acuity with or without corrective glasses (per Snellen test).
- Participants have normal color vision (per Ishihara test or equivalent).
- Participants have a fair level of language skills (at or above B2 level by Common European Framework of Reference for Languages) to answer the questionnaires.
- Participants do not have relevant neurological disease (e.g., epilepsy) or sensory-motor dysfunctions (e.g., movement disorder).

² <https://www.fraps.com>

³ <https://ffmpeg.org>

Table 4: Perceptual video quality dimensions introduced in [8].

VQD	Name	Description	Example Impairment
I	Fragmentation (FRA)	Fallen apart, torn and blockiness	Low Coding Bitrate
II	Unclearness (UCL)	Unclear and blurry image	Upscaling effect using bicubic function
III	Discontinuity (DIC)	Interruptions in the flow of the video	Low frame rate

Figure 2: Scale for assessment of Video Discontinuity (temporal video quality).

3.3 Test Methodology

In addition to video quality, three perceptual sub-dimensions of video quality (cf. Table 4) as well as an acceptance rating of video quality have been assessed. Each dimension is explained to the participant in an introduction in written form using describing adjectives, and in the form of example videos. Here, typical gaming content is used. The rating scales are continuous, using the dimension name as an item label and antonym pairs to describe the range of the scales. An example for the Discontinuity is given in Figure 2. For more information about the method, the reader is referred to [16], where insight about the usage of the different rating scales as well as the test procedure are given. The dimension discontinuity is considered as temporal video quality, whereas the remaining dimensions, fragmentation and unclearness, form the spatial video quality. The dimension luminosity, as proposed in ITU-T Rec. P.918, was not used, as the video content was not changed in this respect and as the use would have increased the test duration unnecessarily. At the end of the test, a few questions about the importance of different quality dimensions were asked in order to investigate the judgement process of participants better.

3.4 Video Complexity

The selection of game content is a difficult challenge as the research community is still lacking consent on a suitable content classification, which also takes the influence of network impairments or encoding settings into account.

The importance of game characteristics on QoE has been studied and established in work presented in [4] [19]. These characteristics were considered in the selection of material for the dataset. Figure 3 shows how the subjective ratings are distributed for 15 selected video sequences of CGVDS at 1080p and 60 fps which illustrates the proper selection of video games based on their complexity. We used the classification proposed in [19] to determine the complexity

Figure 3: Mean Opinion Score of 15 video sequence under of video games. Therefore, three game characteristics are considered for game classification, Static Area (SA), Degree of Freedom (DoF) and Amount of Camera Movement (ACM). ACM is defined as how often the camera is moving plays an important role in the temporal complexity of a video. DoF is referring the movement of the camera, which may be attached to a character, can have up to six degrees of freedom. SA is defined as the ratio of static pixels between two consequent frames to the overall number of pixels of a frame. We refer the reader to [19] for more details on how to classify the games based on these three characteristics.

3.5 Public Dataset

All video sequences that are allowed to be shared due to copyright, are provided in the shared folder that be accessed through the following link:

<https://github.com/stootaghaj/Cloud-Gaming-Video-Dataset>

Four folders are created in the shared link as follows:

Subject Ratings: Subjective ratings together with additional pre/post-game ratings.

MOS information: The Mean Opinion Score (MOS) information is provided together with raw subjective ratings.

Plots and Metrics: The results of objective metrics are provided in frame level as well as video level together with some scatter plots of each metric and subjective results.

Materials: The raw reference sequence in YUV format together with encoded videos in mp4 header at spatial resolutions are provided in this folder together with command lines used for encoding. The CGVDS contains the following games: Overwatch, Rayman Legends, Black Desert Online, FinalFantasy XV, MapleStory 2, Apex Legends, Fortnite, Fifa 2019, Minecraft, Tekken 7, Dauntless, Grand Theft Auto 5, Bejeweled 3, Teamfight Tactics (LoL), and Worms WMD for various bitrate conditions for

different resolution-frame rate pairs, as well as League of Legends as anchor and Nier: Automata as training conditions.

3.6 Copyright notice

All games that distributed the source are allowed by the publisher company are publicly available in the dataset, while a few might not be publicly available due to unclear copyright state. The details on copyright is provided in the shared link. In addition, the audio of all video sequences is removed due to Intellectual property rights.

4 Performance of Signal-Based Models

We evaluated a total of nine video quality assessment (VQA) metrics on the dataset as follows:

Peak Signal to Noise Ratio (PSNR), is the most widely used VQA metric and relies on the computation of the logarithmic difference between corresponding pixels in the original and impaired frame.

Structural Similarity Index Metric (SSIM) measures the structural similarity between two images and usually provides better video quality predictions compared to PSNR.

Multi-Sale Structural Similarity Index Metric (MS-SSIM) is the refined version of the SSIM which considers the variation of the viewing conditions.

Video Multi-Method Assessment Fusion (VMAF) developed by Netflix, fuses three different metrics together to obtain a single score as the estimation of the video quality [20].

Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) using locally normalized luminance coefficient tries to quantify the possible loss of “naturalness” [21].

Natural Image Quality Evaluator (NIQE), which is based on a space domain NSS model, is a learning-based quality estimation metric [22].

Perception based Image Quality Evaluator (PIQE) is an NR metric that uses cues from human visual system to build a blind image quality model [23].

No Reference Gaming Video Quality Metric (NR-GVQM): A NR machine learning-based video quality metric for gaming content which is trained based on low level image features with the aim of predicting VMAF without having access to a reference video [5].

NDNetGaming (NDG), a deep learning-based model which is trained based on a prominent Convolutional Neural Network (CNN) architecture called Densenet [24]. This metric is trained based on gaming content and has shown very high correlation with subjective test for gaming content [25].

For the computation of PSNR and SSIM, we used the VQMT tool available in [12]. For VMAF calculation, we used the Linux based implementation made available by the developers in [20] (version:

VMAF_VF0.2.4b-0.6.1). PIQE, NIQE and BRISQUE calculations were done using the inbuilt MATLAB function (version: R2019a). NDNetGaming is available under git repository [26]. NR-GVQM is not yet publicly available and the code was obtained directly from the authors. For all metrics, the per-frame values are also available in the dataset. The performance of the VQA metrics for over 360 test sequences with respect to subjective scores is evaluated in terms of Pearson Linear Correlation Coefficient (PLCC) and Spearman's Rank Correlation Coefficient (SRCC) values, as shown in Table 5. Since, most of the metrics are image-based quality metrics and we have different frame rate levels, we evaluated the performance of metrics once only at 60 fps and once on all frame rate levels. Surprisingly, for some metrics, performance goes up if all frame rates are considered. We believe that this is due to more available data for calculation of correlation. We also observed that the three NR metrics, PIQE, NIQE and BRISQUE, perform poorly in presence of blurriness which is in line with the findings of [26]. It can be observed that NDG outperforms all other metrics. However, it should be noted that NDG is a deep learning-based model, which is trained based on gaming datasets, GVSET and KUGVD.

5 Performance of Parametric Models

In the remainder of this section, the performance of the following three standardized video quality models when applied to CGVDS will be described:

- G.1071: Opinion model for network planning of video and audio streaming applications [27]
- G.1072: Opinion Model Predicting Gaming QoE for Cloud Gaming Services [3]
- P.1203: Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport [11]

5.1 Classification of Video Games

Gaming videos are diverse in terms of video complexity in both temporal and spatial domain. Depending on the design of the game, a game could be abstract, which leads to lower spatial complexity, or designed with highly complex texture, which leads to higher spatial complexity. Also, depending on the camera movement and camera position (e.g. third perspective vs omnipresent), the temporal complexity might get affected [19][28]. Without knowledge or assumption on the video complexity of games, it would be difficult to build a planning model that can accurately predict the video quality. Thus, planning models (G.1071, G.1072, and P.1203 mode 0) suffer strongly from varying complexity of video games due to lack of information about the video complexity. Therefore, it is clear that for most of the games, the complexity can

Table 5: Performance of FR and NR signal-based video quality models on CGVD in terms of PLCC and SRCC.

fps	Metric	Full Reference Metrics				No Reference Metrics				
		PSNR	SSIM	MS-SSIM	VMAF	BRISQUE	PIQE	NIQE	NR-GVQM	NDNetGaming
20, 30, 60	PLCC	0.66	0.64	0.71	0.87	-0.48	-0.41	-0.53	0.80	0.92
	SRCC	0.67	0.76	0.79	0.87	-0.46	-0.41	-0.55	0.79	0.93
60	PLCC	0.71	0.61	0.73	0.91	-0.45	-0.35	-0.51	0.89	0.94
	SRCC	0.71	0.72	0.80	0.90	-0.44	-0.36	-0.54	0.88	0.92

be determined by investigating the game design. Thus, we have seen efforts to discover the characteristics that cause a game to become complex in terms of encoding complexity in the research community [19], [28] and standardization bodies [3]. In this paper, we follow the classification that is proposed in [19] for development of the planning model, while the proposed monitoring model does not require any classification.

5.2 ITU-T Planning Models

The ITU-T Rec. G.1071 Annex A, provides models for estimating the quality experienced by a user of audio and video streaming services. The models are designed to be network planning tools hence it only uses properties of the network as parameters. Since the model is trained based on the none-gaming dataset without using hardware accelerator engines and with no knowledge on the game complexity, it was retrained to build a game specific model presented in G.1072, which focuses on predicting the QoE for cloud gaming services, based on a huge gaming dataset. Therefore, both ITU-T planning models, ITU-T G.1071 and G.1072, are using the same structure to predict the video coding impairment.

The video coding impairment takes six input arguments of which two are set constant ($packetLoss = 0$ $slicesPerFrame = 0$, $TSburstinessV = 0$) because they are not present and unknown in the given data set. The remaining parameters bit rate, number of pixels per frame, and frame rate are used to estimate the experienced quality of a video that is being transmitted with these characteristics. The module further depends on 21 coefficients of which only 7 are needed for training the dataset. ITU-T Rec. G.1072 uses the same equation for prediction of coding impairment but the model is retrained based on a huge gaming video dataset build using hardware accelerator engine and based on the three video complexity classes. In Table 6 the change of the coefficients is shown. Significant changes of values can be seen especially for a_{3v} , which is a balancing factor for calculation the content complexity (equation 2). This may be taken as an argument for changing the calculation of the content complexity for gaming content. In G.1072, the high complex class is considered as default mode in case the game complexity class cannot be determined.

The model predicts the video impairment based on an exponential function of a parameter named *BitPerPixel*. *BitPerPixel* is defined as an average amount of bits assigned to a video sequence as it is calculated in equation 2, in which *NumPixelPerFrame* is the product of the width and height of the video resolution. It must be noted that both models considered in this section, ITU-T G.1071 Annex A and G.1072, follow an impairment factor approach similar to the well-known E-model. Therefore, the video quality ratings are transformed from a 5-point ACR (absolute category rating) scale to the R-scale, which ranges from 0 to 100, using the transformation presented in [29]. Since no transmission errors present in the dataset, for the evaluation of the performance of the models, only the coding impairment factors, which estimates the video quality impairment for video compression artefacts on the R-

scale, is considered, and not the video quality impairment for video transmission errors (I_{trav}). This results in the following implementation of the models, which only differ in the coefficients trained on the different train datasets.

$$QV = 100 - I_{VQcod} - I_{trav} \quad (1)$$

where $I_{trav} = 0$ due to no transmission loss.

$$I_{VQcod} = a_{1V} \cdot \exp(a_{2V} \cdot BitPerPixel) + \quad (2)$$

$$a_{3V} \cdot ContentComplexity + a_{4V}$$

$$ContentComplexity = a_{31} \cdot \exp(a_{32} \cdot BitPerPixel) + a_{33} \quad (3)$$

$$BitPerPixel = \frac{Bitrate \cdot 10^6}{NumPixelPerFrame \cdot Framerate} \quad (4)$$

Table 6: Coefficients of G.1071, and G.1072 impairments for each encoding complexity in which class 1, 2 and 3 represent low, medium and high complexity classes respectively.

	G.1071	G.1072		
Coefficient	-	Class 1	Class 2	Class 3
a_{1V}	51.28	52.5052	37.9882	47.7463
a_{2V}	-22.00	-28.017	-13.7208	-12.07
a_{3V}	6.00	-2.68405	8.57837	9.05168
a_{4V}	6.21	5.46648	3.26581	3.41919
a_{31}	3.92	12.4214	6.83276	7.62306
a_{32}	-27.54	-28.0192	-127.997	-167.838
a_{33}	0.26	0.215799	0.479595	0.0760333

5.2.1 Performance of ITU-T G.1071 - G.1072

The performance evaluation result of G.1071 and G.1072 on the presented datasets in terms of PLCC as well as Root Mean Square Errors (RMSE) is shown in Table 7. As it can be observed from Table 7, the classification used to retrain the G.1072 increases the performance of G.1071. We observe that for both models, the performance of low complex class is very low, while G.1072 has the highest performance for high complex video games. It must be noted that both ITU-T models are planning models and they are not supposed to have access to the video complexity information in general. However, the network planner and service provider can make assumption on the type of games they might stream. Therefore, such a classification can be assumed in prediction process.

Table 7: Pearson Correlation and RMSE of G.1071/G.1072 prediction with actual MOS for the presented datasets.

G.1071			G.1072		
PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
0.60	0.56	0.68	0.73	0.69	0.48

5.5 ITU-T Rec. P.1203

The ITU-T P.1203 is a series of models that are designed to predict the audio-visual quality experienced by a user for adaptive and progressive-download-type media streaming that could be placed in end-point locations or at mid-network monitoring points. ITU-T Rec. P.1203 has four modes depending on the level of access to the information due to encryption as shown in Table 8. Mode 0 of ITU-T Rec. P.1203 could be considered as another potential planning model candidate to predict the gaming video quality. We investigate the video quality module of Mode 0 which only uses bitrate, frame rate, and coding resolution. Equation 5 illustrates the function that predicts video quality of Mode 0.

$$\begin{aligned} quant = & a1 + a2 * \log(a3 + \log(bitrate)) \\ & + \log(bitrate) \\ & * bitrate / (coding_res \\ & * framerate) + a4) \end{aligned} \quad (5)$$

$$mos_cod_v = q1 + q2 * \exp(q3 * quant) \quad (6)$$

Due to change of coefficients for different modes and limited space, we did not provide them in details in this paper.

Table 8: ITU-T P.1203 modes of operation.

Mode	Encryption	Input
0	Encrypted payload headers	Meta-data
1	Encrypted media payload	Meta-data and frame size/type
2	No encryption	Meta-data and up-to 2% of the stream
3	No encryption	Meta-data and any information from the video stream

For Mode 1, the equation 5 is used for measuring the quant parameter. However, the coefficients are updated, and the frame size was used instead of video bitrate. It must be noted that the content complexity is taken into account based on the ratio of the I-frames and non-I-frames average size. In addition, the MOS was predicted using a sigmoid function.

Mode 2 can have access to 2% of the media stream. Therefore, the quant equation was measured with higher accuracy using quantization parameters (QP) as follows:

$$quant = \frac{\overline{QP_{PB}}}{51} \quad (7)$$

, where $\overline{QP_{PB}}$ is based on the quantization parameters in the bitstream for non-I-frames. The MOS can be predicted based on the equation 5.

The Mode 3 has access to the full bitstream in which the QP values are parsed for all frames in the measurement window and uses the same equations as Mode 2. The main difference between Mode 2 and Mode 3 is the amount of information from the bitstream that can be accessed for training and predicting the video quality. In Mode 2, due to limited access to the bitstream some QP values are estimated based on the neighboring frames, while in Mode 3 all QP values are used in training and prediction Mode. The result based on the all three modes as well as new fits for Mode 0 is presented in Table 9. It has to be noted that the result for Mode 2 is not provided in Table 9 as the implementation of Mode 2 was not available to the authors. Also, it can be seen that Mode 0 has higher correlation compared to Mode 1, while the RMSE is significantly lower. Therefore, this only implies that Mode 0 prediction follows the VQ distribution slightly better than Mode 1, but overestimates the video quality significantly higher than Mode 1. In general, both modes overestimate the video quality due to higher requirement of bitrate to in NVENC, present of llhq, to reach a certain quality compared to software implementation of H.264.

Table 9: Performance of ITU-T Rec. P.1203 modes on CGVD in terms of RMSE, PLCC and SRCC.

Mode	RMSE	PLCC	SRCC
0	0.98	0.66	0.69
1	0.69	0.58	0.62
3	0.48	0.88	0.88

6. Perceptual Dimension based Gaming Video Quality Model

In this section, we proposed two models, a planning model and monitoring bitstream model, that can predict the overall gaming video quality based on the perceptual video quality dimensions, Fragmentation, Unclearness, and Discontinuity presented in section 3.3. The planning model solely uses the encoding parameters bitrate, frame rate, coding and display resolution, while for the monitoring model, in addition to the encoding parameters, the packet header information such as size of I and P frames is used to predict the perceptual dimensions of video quality. The latter follows a similar approach as ITU-T Rec. P.1203 mode 1 in terms of level of access to the data. We believe that such a structural based model benefits a service provider to update the model easier in case of having new parameters or interest in higher range of parameters. Moreover, such a model can act as a diagnostic model that can explain which degradation causes a potentially low video quality.

Both proposed models share the same impairment factor structure which is inspired by the E-model [30] as illustrated in equation 8.

$$R_{QoE} = R_{max,QoE} - a \cdot I_{VD} - b \cdot I_{VF} - c \cdot I_{VU} \quad (8)$$

6.1. Data Preparation

Before starting the modeling phase, we applied a few steps in order to prepare the data for the next steps. As mentioned in section 3, we followed the ITU-T Rec P.809 for the quality assessment of video games. Thus, the ACR with hidden reference method was used for the subjective tests to collect ratings on a 7-point continuous scale according to [15].

In order to derive suitable impairment factors for the modeling, the extended continuous 7-points ratings had to be transformed to the R-scale (cf. ITU-T Rec. G.107 [30]), which ranges from 0 to 100.

This was done in the following manner:

- Transformation of extended continuous 7-point ratings (EC) to 5-point ACR ratings using the transformation presented in [3]:

$$\widehat{MOS}_{ACR} = -0.0262 \cdot \widehat{MOS}_{EC}^3 + 0.2368 \cdot \widehat{MOS}_{EC}^2 + 0.1907 \cdot \widehat{MOS}_{EC} + 1 \quad (9)$$

- Normalize MOS values as follows, as some condition means were larger than 4.5:

$$MOS_{norm,i} = \frac{MOS_i - 1}{MOS_{max} - 1} \cdot 3.5 + 1 \quad (10)$$

- Calculation of R-value according to ITU-T Rec. 107.
- Calculation of Impairment (differential R-values from reference):

$$I_{factor}(condition) = R_{max,factor} - R_{factor}(condition) \quad (11)$$

The highest MOS of a video quality rating (MOS_{max}) in the dataset is 4.58, which represents a $R_{max,QoE}$ of 100 in Eq. 8. To train the models presented in the following sections, the full CGVDS dataset containing 15 games for various encoding settings was split into a training and test dataset. For the test dataset, we randomly selected three source video sequence (one per complexity class). The remaining games were used as a training dataset to derive the model equations.

6.2 Planning Model

The planning model predicts the video quality based on the impact of compression impairments and targets cloud gaming services or passive gaming video streaming encoded using NVENC. The model is a network planning tool which can be used by network and cloud gaming providers for purposes such as resource allocation and configuration of IP-network transmission settings such as the selection of resolution and bitrates. Each dimension is predicted solely based on parameters relevant for each dimension, which also offers a parameter reduction.

Impairment of Video Discontinuity (I_{VD}): The video discontinuity dimension measures the perceived jerkiness of a video due to low encoding frame rate. Since, VD is only triggered by frame rate, we modeled it based on an exponential function of frame rate as the best fit based on our data. Two classes are driven based on the ITU-T contribution [31] which suggests to used information about the characteristics movements of a virtual camera, frequency of game object movements, and pace of the interaction with the game to classify game content in respect to frame losses. If no information about the class is available, the high complex class should be used as a default mode. VD is predicted using equation 12 using the coefficients per class presented in Table 10.

$$I_{VD} = \exp\left(\frac{d1}{framerate}\right) + d2 \quad (12)$$

Table 10: Coefficients for Equation 12.

Coefficient	Low complex	High complex (default mode)
$d1$	66.29	67.28
$d2$	10.23	3.23

Impairment of Video Fragmentation (I_{VF}): The video fragmentation dimension gets affected by bitrate, resolution and frame rate. Therefore, we define variable "*bitperpixel*" which explains the number of bits that is spent in average for a pixel.

$$bitperpixel = \frac{bitrate}{framerate * height * width} \quad (13)$$

Three classes are driven based on the ITU-T contribution [31] which suggests to used information about the characteristics movements of a virtual camera, texture details, and frequency of movements of game objects to classify game content in respect to its encoding complexity. If no information about the class is available, the high complex class should be used as a default mode. VF is predicted using equation 14 using the coefficients per class presented in Table 11.

$$I_{VF} = e1 + e2 * \log(bitperpixel * bitrate) + e3 * bitrate \quad (14)$$

Table 11: Coefficients for Equation 14.

	Low complex	Medium complex	High complex (default mode)
$e1$	2.1	13.79	11.21
$e2$	-5.426	-8.017	-10.59
$e3$	0.0005258	0.0005442	0.0006314

Impairment of Video Unclearness (I_{VU}): The video unclearness dimension gets affected mainly due to upscaling impairments. However, also strong blockiness can be perceived as unclearness. Thus, we used frame rate and bitrate in addition to the resolution of the video for the prediction of VU. Therefore, we define a variable called “*scaleratio*” which explains the ratio encoding resolution to the display resolution.

$$scaleratio = \frac{height_{encoding} * width_{encoding}}{height_{display} * width_{display}} \quad (15)$$

Three classes are driven in the same way as for VF to consider the importance of the video content. The VU is predicted using equation 16 and the coefficients shown in Table 12.

$$I_{VU} = f1 + f2 * \log(bitperpixel * bitrate) + f3 * \log(scaleratio) \quad (16)$$

Table 12: Coefficients for Equation 16.

	Low complex	Medium complex	High complex (default mode)
$f1$	4.299	18.58	17.13
$f2$	-2.016	-3.422	-15.38
$f3$	-17.99	-4.494	-7.844

Core model: The core model as mentioned in equation 8 follows the structure we introduced in the beginning of section 6. The weighting factors are derived based on the full training set as shown in equation 17.

$$R_{QoE} = R_{max,QoE} - 0.259 \cdot I_{VD} - 0.554 \cdot I_{VF} - 0.341 \cdot I_{VU} \quad (17)$$

6.3 Performance of Proposed Planning Model

The goodness of fit for the prediction of each dimension is reported in Table 13 in terms of RMSE and adjusted R-squared. It has to be noted that the RMSE is reported based on R-scale (ranges from 1 to 100). The performance for the core model is reported based on the MOS scale and based on the test dataset. Table 14 reports the performance of the proposed planning model on the training and test dataset.

Table 13: Goodness of fit for each perceptual dimension in terms of RMSE and Adjusted R-squared.

	Class1		Class2		Class3	
	RMSE	Adj R ²	RMSE	Adj R ²	RMSE	Adj R ²
Fragmentation model	10.43	0.41	6.96	0.60	-	-
Unclearness model	0.59	0.63	9.18	0.75	7.88	0.86
Discontinuity model	6.90	0.83	10.23	0.73	6.90	0.83

Table 14: Performance of proposed planning model on the training and test dataset.

Dataset	Training Set			Test Set		
	RSME	SRCC	PLCC	RMSE	SRCC	PLCC
Score	0.28	0.88	0.89	0.39	0.75	0.78

6.4 Bitstream Model

Bitstream models can be used by both service and network providers to monitor the quality of streamed videos of a certain service. Cloud gaming providers can benefit from bitstream models as these models are in real-time, they require no previous information about the games and all information is accessible on both client and server side with very low processing overhead. The proposed bitstream model requires only simple bitstream information such as frame type, frame size, and encoding parameters. It is comparable to the P.1203 mode 1. With the aim to build the monitoring model, we decided to take a time interval of 5 seconds as the measurement and predication window size. Since the MOS ratings are post-memory judgements over the full duration – in the CGVDS 30 seconds - of a video stimulus, we assign the quality of each 5-second clip of a video sequence based on the distribution of frame-level quality estimation of NDG and final MOS values of the model as shown in equation 18. The selection of NDG was due to high accuracy of this model compared to others which was shown and discussed in section 5. Thus, for each frame the weighted quality is calculated as follows:

$$w_i = \frac{NDG_i}{NDG_{final}} \quad (18)$$

where NDG_i is the measured quality of frame i based on NDNetGaming. NDG_{final} is the final predicted video quality of a certain video based on NDNetGaming.

Lastly, the MOS of an interval duration from frame i to frame j of a video sequence is calculated as follows:

$$\widehat{MOS}_{i,j} = MOS_{video} * \sum_{k=i}^j \frac{w_k}{j-i} \quad (19)$$

, where MOS_{video} is the Mean Opinion Scores of a video sequence. It must be noted that MOS is not only referring to the MOS of video quality, but also the same interval-based approach was used for prediction of MOS for VF and VU.

In order to build the model, we extract multiple features from each 5 second sequence of each video sequence. The following features are used for training the model:

- fr: Frame rate
- sf_{video} : based on Equation 15
- br_{avg} : Average bitrate of five second sequence
- $num_{I-frame}$: Number of I-frames
- br_{avg-I} : Average of I-frames bitrate
- $stat_{P-frame}$: Statistics of P-frames (average, standard deviation, etc)
- $std_{P-frame}$: Standard deviation of P-frames bitrate
- CP_{video} : Video complexity based on as follows $\frac{I_{bitrate}}{P_{bitrate}}$

Video complexity is described using $\frac{I_{bitrate}}{P_{bitrate}}$ as it was used in P.1203 as well. CP_{video} estimates the spatio-temporal complexity of different video gaming contents based on the ratio of I-frame and P-frame size in the measured window. It must be noted that in case of no I-frame in the 5 second window, the first past I-Frame from previous windows of the same video was used. We build the model based on the perceptual video dimensions.

Similar to the proposed planning model, the perceptual dimension based bitstream model, first predicts the impairment of VF, VU, and VD and then predicts the VQ impairment based on the core model, equation 8. In general, each perceptual dimension is trained based on the relevant extracted information using Random Forest regression. The list of features used for training and prediction of each dimension is listed as follows:

Impairment of Video Unclearness (I_{VD}): VD is the perception of jerkiness in the video due to low frame rate, which is affected based on the frame rate, temporal complexity of a video. Among extracted features, two features that might be representative features of temporal video complexity are $std_{p-frame}$, CP_{video} . Therefore, we used fr , $std_{p-frame}$, CP_{video} for prediction of I_{VD} .

Impairment of Video Fragmentation (I_{VF}): VF can be affected based on all extracted features. Therefore, we used all extracted feature to train our model that predicts I_{VF} . We used Recursive Feature Elimination (RFE) method to eliminate those that do not significantly contribute to the model. After applying RFE, the following relevant factors remained in the I_{VF} model: fr , $std_{p-frame}$, CP_{video} , sr , sf_{video} , $stat_{p-frame}$, br_{avg} .

Impairment of Video Unclearness (I_{VU}): VU is affected mainly by scaling ratio. However, as mention before VF can cause the video unclearness. Therefore, all features are added to the model and then similar to I_{VF} model RFE is applied to keep only relevant features. After applying RFE, the following relevant factors remained in our model: fr , $std_{p-frame}$, CP_{video} , sr , sf_{video} , $stat_{p-frame}$.

Core model: we used the equation 8 which the coefficients are updated as it can be seen in equation 20. In order to fairly train and test the model, we split the dataset into five bins which in each bin three games are randomly selected. The model was trained five times, each time one bin was hold-out and the model was trained based on the remaining videos. The performance of the model is calculated based on the average of five-time training phases. Figure 4 shows the performance of the model based on the all videos in a scatter plot. The model achieves RMSE of 0.34 and PLCC of 0.90.

$$R_{QoE} = R_{max,QoE} - 0.243 \cdot I_{VD} - 0.412 \cdot I_{VF} - 0.434 \cdot I_{VU} \quad (20)$$

Figure 4: Scatter plot of predicted MOS and actual MOS for all sequences in the test set.

7. Discussion

The two proposed models are light weight models that can be used for video quality assessment of cloud gaming services for planning and monitoring purposes. Since the could gaming providers have access to the encoding setting as well as bitstream information, we believe that these models are more suitable than signal-based model that require lots of computation and storage. In addition, network planners and providers can benefit from these two models for planning purpose as well as monitoring of the gaming quality. The monitoring model can predict in the interval window of 5-second which allows cloud gaming providers live monitoring of the QoE.

In this work, we only assessed the video quality in passive viewing-and-listening tests. However, the gaming QoE in cloud gaming services is not only video quality but also interaction quality plays a role. While, the gaming QoE is beyond the scope of this paper, in [33] it is shown that passive view-and-listening tests can be used to predict the gaming QoE when only the spatial video quality is affected.

The proposed planning model relies strongly on the knowledge about the game complexity class. It has been shown in ITU-T Rec P.1072 as well as this paper that without such a classification the model strongly suffers from the variation of spatio-temporal video complexity. Therefore, we followed the same proposed game classification which is proposed in [34] and also used for ITU-T Rec G.1072. However, similar to ITU-T Rec 1072, the default mode is considered as high complex class.

8. Conclusion

In this paper, we present the first large scale video quality dataset based on hardware accelerator encoding setting for gaming content. This dataset can be used for building gaming video quality models which can be employed in quality assessment of cloud gaming services. In the presented dataset, we evaluate three perceptual quality dimensions in addition to the video quality which can help the research community to build models for a specific type of distortion such as blurriness or blockiness estimation.

In addition, we investigate the performance of well-known signal-based and parametric models for gaming content which helps the research community in selection of proper quality models for different purposes. Finally, we build two parametric models for planning and monitoring purpose based on perceptual video dimensions. The models outperform the existing the quality models, however, we did not retrain them based on the new dataset which is due to unavailability of implementation for the training the models. Using dimension-based approach is a new approach and offers promising results and analytic insights, while adding minor errors due to two phases of modeling.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871793.

REFERENCES

- [1] R. Arzumanyan, "Turing H.264 Video Encoding Speed and Quality," *NVIDIA Developer Blog*, <https://devblogs.nvidia.com/turing-h264-video-encoding-speed-and-quality/> (accessed Jan. 13, 2020).
- [2] S. Möller, S. Schmidt, and J. Beyer, "Gaming taxonomy: An overview of concepts and evaluation methods for computer gaming qoe," in *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2013, pp. 236–241.
- [3] ITU-T Recommendation G.1072, "Opinion Model Predicting Gaming QoE for Cloud Gaming Services," *Geneva Switz Int Telecommun Union*, 2020.
- [4] S. Schmidt, S. Zadtootaghaj, and S. Möller, "Towards the delay sensitivity of games: There is more than genres," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2017, pp. 1–6, doi: 10.1109/QoMEX.2017.7965676.
- [5] S. Zadtootaghaj, N. Barman, S. Schmidt, M. G. Martini, and S. Möller, "NR-GVQM: A no reference gaming video quality metric," in *2018 IEEE International Symposium on Multimedia (ISM)*, 2018, pp. 131–134.
- [6] S. Göring, R. R. Rao, and A. Raake, "nofu—A lightweight no-reference pixel based video quality model for gaming content," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, 2019, pp. 1–6.
- [7] S. Zadtootaghaj, S. Schmidt, and S. Möller, "Modeling gaming QoE: Towards the impact of frame rate and bit rate on cloud gaming," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 2018, pp. 1–6.
- [8] S. Wang and S. Dey, "Modeling and characterizing user experience in a cloud server based mobile gaming approach," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, 2009, pp. 1–7.
- [9] J. A. Bergstra and C. A. Middelburg, "ITU-T Recommendation G. 107: The E-Model, a computational model for use in transmission planning," 2003.
- [10] ITU-T. Recommendation P. 1201, "Parametric non-intrusive assessment of audiovisual media streaming quality," 2012.
- [11] ITU-T Recommendation P.1203, "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport," 2017.
- [12] A. Raake, M. Garcia, S. Moller, J. Berger, J. Kling, P. List, J. Johann, C. Heidemann, "T-V-model: Parameter-based prediction of IPTV quality," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 1149–1152.
- [13] N. Barman, S. Zadtootaghaj, S. Schmidt, M. G. Martini, and S. Möller, "GamingVideoSET: a dataset for gaming video streaming applications," in *2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)*, 2018, pp. 1–6.
- [14] N. Barman, E. Jammeh, S. A. Ghorashi, and M. G. Martini, "No-reference video quality estimation based on machine learning for passive gaming video streaming applications," *IEEE Access*, vol. 7, pp. 74511–74527, 2019.
- [15] ITU-T Recommendation P.809, "Subjective evaluation methods for gaming quality," *Geneva Switz. Int. Telecommun. Union*, 2018.
- [16] F. Schiffner and S. Moller, "Direct Scaling & Quality Prediction for perceptual Video Quality Dimensions," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 2018, pp. 1–3.
- [17] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," *Geneva Switz. Int. Telecommun. Union*, Apr. 2008.
- [18] ITU-T Recommendation P.911, "Subjective audiovisual quality assessment methods for multimedia applications," *Geneva Switz. Int. Telecommun. Union*, Dec. 1998.
- [19] S. Zadtootaghaj, S. Schmidt, N. Barman, S. Möller, and M. G. Martini, "A classification of video games based on game characteristics linked to video coding complexity," in *2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)*, 2018, pp. 1–6.
- [20] "Perceptual video quality assessment based on multi-method fusion," Jan. 13, 2020. <https://github.com/Netflix/vmaf> (accessed Jan. 13, 2020).
- [21] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.
- [22] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2012.
- [23] "Blind image quality evaluation using perception based features - IEEE Conference Publication." <https://ieeexplore.ieee.org/document/7084843> (accessed Jan. 13, 2020).
- [24] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [25] M. Utke, S. Zadtootaghaj, S. Schmidt, S. Bosse, and S. Möller, "NDNetGaming - Development of a No-Reference Deep CNN for Gaming Video Quality Prediction," *Submitt. Multimed. Tools Appl.*
- [26] N. Barman, S. Schmidt, S. Zadtootaghaj, M. G. Martini, and S. Möller, "An evaluation of video quality assessment metrics for passive gaming video streaming," in *Proceedings of the 23rd Packet Video Workshop*, 2018, pp. 7–12.
- [27] ITU-T Recommendation G.1071, "Opinion model for network planning of video and audio streaming applications," *Geneva Switz. Int. Telecommun. Union*, 2016.
- [28] M. Claypool, "Motion and scene complexity for streaming video games," in *Proceedings of the 4th International Conference on Foundations of Digital Games*, 2009, pp. 34–41.
- [29] F. Köster, D. Guse, M. Wälder, and S. Möller, *Comparison between the discrete ACR scale and an extended continuous scale for the quality assessment of transmitted speech*. 2015.
- [30] ITU-T. Recommendation G. 107, "The E Model, a computational model for use in transmission planning," *Int. Telecommun. Union*, vol. 8, no. 20, 2003.
- [31] ITU-T Contribution C.444, TU Berlin, "Proposal of a Content Classification for Cloud Gaming Services," *Int. Telecommun. Union*, 2019.
- [32] C. Chen, S. Inguva, A. Rankin, and A. Kokaram, "A subjective study for the design of multi-resolution abr video streams with the vp9 codec," *Electron. Imaging*, vol. 2016, no. 2, pp. 1–5, 2016.
- [33] S. Schmidt, S. Möller, and S. Zadtootaghaj, "A comparison of interactive and passive quality assessment for gaming research," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 2018, pp. 1–6.
- [34] S. Zadtootaghaj, S. Schmidt, N. Barman, S. Möller, and M. G. Martini, "A classification of video games based on game characteristics linked to video coding complexity," in *2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)*, 2018, pp. 1–6.