electronics

MDPI

Article

A Hybrid QoS-QoE Estimation System for IPTV Service

Jaroslav Frnda [1,*], Jan Nedoma [2,*], Jan Vanus [3] and Radek Martinek [3]

[1] Department of Quantitative Methods and Economic Informatics, Faculty of Operation and Economics of Transport and Communications, University of Zilina, 01026 Zilina, Slovakia

[2] Department of Telecommunications, Faculty of Electrical Engineering and Computer Science, VSB - Technical University of Ostrava, 70833 Ostrava-Poruba, Czech Republic

[3] Department of Cybernetics and Biomedical Engineering, Faculty of Electrical Engineering and Computer Science, VSB - Technical University of Ostrava, 70833 Ostrava-Poruba, Czech Republic; jan.vanus@vsb.cz (J.V.); radek.martinek@vsb.cz (R.M.)

* Correspondence: jaroslav.frnda@fpedas.uniza.sk (J.F.); jan.nedoma@vsb.cz (J.N.); Tel.: +421-415-133-278 (J.F.); +420-597-326-056 (J.N.)

Received: 24 April 2019; Accepted: 22 May 2019; Published: 27 May 2019

Abstract: The internet protocol television service (IPTV) has become a key product for internet service providers (ISP), offering several benefits to both ISP and end-users. Because packet networks based on internet protocol have not been prepared for time-sensitive services, such as voice or video, packet networks have had to adopt several mechanisms to secure minimal transmission standards in the form of data stream prioritization. There are two commonly used approaches for video quality assessment. The first approach needs an original source for comparison (full-reference objective metrics), and the second one requires observers for subjective evaluation of video quality. Both approaches are impractical in real-time transmission because it is difficult to transform an objective score into a subjective quality perception, and on the other hand, subjective tests are not able to be performed immediately. Since many countries worldwide put IPTV on the same level as other broadcasting systems (e.g., terrestrial, cable, or satellite), IPTV services are subject to regulation by the national regulation authority. This results in the need to prepare service qualitative criteria and monitoring tools capable of measuring end-user satisfaction levels. Our proposed model combines the principles of both assessment approaches, which results in an effective monitoring solution. Therefore, the main contribution of the created system is to offer a monitoring tool able to analyze the features extracted from the video sequence and transmission system and promptly translate their impact into a subjective point of view.

Keywords: IPTV; neural network; QoE; QoS

1. Introduction

The television broadcasting transmitted over packet networks has rapidly become an important service for internet service providers (ISP). Network convergence realized during the 1990s allowed the transportation of different types of services (voice, video, and data) via integrated network infrastructure. However, several issues appeared. Packet networks had not been prepared to transmit time-sensitive services. This problem was solved by adopting a prioritization strategy which focuses on the processing of real-time services by routers with guaranteed preference. Widely-used transport protocol for real-time services (often presented by the marketing term triple play services)—UDP—offers unreliable connectionless communication without any guarantee of delivery. This could potentially lead to packet loss, especially within networks with a high utilization level caused by the lack of bandwidth.

For the purposes of our research we've defined IPTV as the broadcast of television (TV) channels (content is broadcasted at the same time as via other broadcasting systems). The IP/UDP/RTP protocol stack has been used and streamed to a set-top box directly connected to a TV device (or PC). This requires the use of adequate network infrastructure and support from the streaming technology.

Convergence and digitalization have had a significant impact on the regulatory environment. Worldwide triple play services subscriptions reached over 400 million in 2017, and brought considerable benefit to customers in terms of a reasonable service price. Many countries distinguish between regulation covering the distribution and multimedia content, and therefore there is a tendency to prepare regulatory framework dedicated to the quality of IPTV services [1]. It is essential to select parameters and their minimal values to secure an accepted level of Quality of Service (QoS), as well as to be able to monitor these parameters, while being aware of the influence on the final end-user quality perception.

For many years, QoS has represented the main evaluation framework, nevertheless, QoS cannot precisely reflect the user´s perception. Thus, the concept of Quality of Experience (QoE) has been created and standardized step by step by the International Telecommunication Union (ITU) [2,3].

This concept usually requires real observers for subjective video quality analysis, which is a time-consuming operation. Monitoring of network parameters (e.g., packet loss or overall delay) is relatively easy for network providers, however, these parameters are related to the network behavior rather than to the customer. Disadvantages of subjective testing are reflected in the inability to perform and evaluate the tests continuously and immediately.

Hence, our proposed method implements a form of application (Python programming language—works for Windows and Linux) that is able to predict how the human visual system comprehends video quality and distortions introduced within the transmission chain. It allows correction of the interpretation and transformation of the objective score to a subjective point of view. The voice over IP (VoIP) service, as a part of triple play services, includes a standardized mapping function for QoS to QoE conversion (E-model-R factor to user satisfaction expressed in the MOS—Mean Opinion Score scale. However, a QoS to QoE unified transformation function in video quality analysis is still missing. This may cause difficulties for national regulator authorities in terms of customer complaints handling, because the results obtained from objective metrics related to the network behavior cannot be translated correctly to the end-user perception. In relation to this, the main purpose of our system is to offer an application capable of user satisfaction level prediction (by using a mapping function in the form of neural network classificatory), based on the specific video sequence features and network utilization.

2. State-of-the-Art

Video content has been growing consistently and is becoming the dominant portion of all data traffic sent via packet networks.

Typically, an IPTV broadcasting has a one-way direction (communication with the content provider is non-essential in this case), thus, the network delay does not play a significant role in comparison to the voice service. Packet loss caused by the inadequate dimension of the transmission system may, in particular, be a serious network impairment that impacts the overall video quality [4].

Two main topics are usually presented in research papers. Firstly, the robustness of video codecs to packet loss within the network. Secondly, the fact that even when a new digital representant of the video signal is announced, researchers test and verify compression efficiency and complexity, which reflects required computed power [4,5].

The release of video codecs like VP9 (Google) and H.265 (collaboration between MPEG and ITU-T) started the process of their evaluation and comparison with their predecessors.

As mentioned above, QoS metrics are limited when it comes to characterization of the end-user perceptual experience. Using the results obtained by video quality assessment from QoE instead of QoS appears to be a better solution because (i) increasing QoS does not directly improve QoE, and (ii) improving only QoS can increase operating costs remarkably, consequently decreasing the profit of service providers.

Subjective testing needs real observers, hence, the test performing is a time-consuming and costly activity. This leads to a tendency to substitute the testing by objective metrics, which attempt to simulate human perception. These metrics are based on mathematical models and require the original reference video sequence to compute a score of quality for a degraded video sample.

The main motivation behind this work is to come up with a new method of interconnection between the results obtained from both subjective and objective methods. Subjective tests typically use a MOS scale of 1–5 where 5 is the best. On the other hand, every objective metric uses its own scale, and to this day there is no unified mapping function to interpret objective results into a point of view of subjective score that could be expressed by the MOS scale. The PSQA (pseudo subjective quality assessment) is the first attempt to use a neural network as a tool of artificial intelligence [6].

In that work, the authors used qualitative video parameters, like packet loss or bitrate, as the input, and a trained neural network for subjective score calculation. Although only an old codec MPEG-2, low resolution (352 × 288), and a small-size test set were used, this paper served as the basic concept for our present work. In the report of D. Valderrama and colleagues [7], different video attributes were selected. Authors performed the training process of a neural network with inputs such as different length of GOP (group of pictures), prioritization policies (BestEffort and DiffServ), or they created bottlenecks on the testing network topology. Pearson´s coefficient reached more than 0.9, but again, only one low resolution (720 × 480) and several packet loss scenarios (1%, 5%, 10%) were used. Works of research teams led by D. Botia and J. Søgaard [8,9] proposed regression functions for video quality calculation. Based on the video content (dynamic or static scenes), the correlation coefficient oscillated between 0.7 and 0.9. The paper of D. Mocanu [10] provided a summary of applicability for several machine learning tools, and the performed tests showed that the neural network obtained the highest correlation coefficient.

Authors of the papers [11,12] used regression functions for subjective score prediction only for the newer video codec H.265 but with UHD resolution absented, and Pearson´s coefficient was over 0.92. The research team of M. Alreshoodi [13] created their own fuzzy rules to QoE prediction. They extracted specific qualitative video sequence attributes, namely spatial and time information. Their fuzzy interface system allowed them to estimate QoE results in the MOS scale with a correlation of more than 0.95. Software-defined networks (SDN) as a cloud computing technology also serve as a way of multimedia content transport. T Abar et al. [14] tested several machine learning tools and tried to find the best correlation. They used resolutions ranging from 320 × 240 to 1280 × 720, with several bitrates and packet losses. The best results were obtained by using a decision tree with an M5P algorithm, where there was a correlation coefficient slightly over 0.81.

All the above-mentioned works tried to propose a computational model capable of deriving the subjective perception of video quality from objective parameters. The next section brings a comprehensive view of all video parameters, such as codec and scene type, bitrate, resolution, together with their influence on prediction accuracy. Section 4 describes the whole process of application making, as well as verification of the predicted outputs. The Discussion and Conclusion section provides a summary of work and future plans related to this topic.

3. Methodology

3.1. Video Processing

TV broadcasting consists of several resolutions, bitrates, and codecs. Nowadays, the most used compression standard presented in a wide spectrum of applications, ranging from mobile video to broadcasting in high definition (HDTV), is MPEG-4 H.264/AVC (MPEG Part 10). This codec improves the key features of its predecessor, such as changeable size of motion compensation blocks, or multiple reference frame motion estimation [5]. High-Efficiency Video Coding (HEVC), referred to as H.265, offers a similar level of picture quality, with a bitrate of about the half of the size, or it increases the visual quality at the same bitrate as H.264. It supports resolutions up to 8 K UHD (8192 × 4320). The main improvements for HEVC include motion compensated prediction (various coding areas ranging from 16 × 16 to 64 × 64 pixels), better accuracy for motion vectors calculation (up to 35 directions, compared to nine in H.264), or sample adaptive offset, which reduces artifacts at block edges.

The disadvantage of better efficiency is reflected in a higher need for computation power. The complexity of HEVC requires more powerful hardware to encode the same video quality as H.264. [15].

TV broadcast bitrate typically ranges between 10 and 15 Mbps (better for premium channels), and, according to the tests performed by public service broadcasters, 15 Mbps looks like a standard bitrate for HEVC for the near future [5,15].

High-definition (HD) broadcasting provides an image resolution that is at least twice as high as the standard-definition offered by television. HD is the current standard video format for terrestrial and cable TV broadcasting, Blu-ray discs, and streaming videos over the internet. The following resolutions were used: HD (1280 × 720), FullHD (1920 × 1080) and UltraHD (3840 × 2160).

3.2. Video Transmission

To transfer the video files via transmission infrastructure based on IP protocol, basically unreliable protocols are used. Communication between two nodes leaves the sender without feedback regarding the missing parts of data. There are also some exceptions. Some special video streaming services (video on demand—VoD), e.g., Netflix or YouTube, use the reliable transmission protocol TCP instead of the unreliable UDP.

First, the main purpose of these streaming services is not to provide real-time broadcasting. TCP uses an error-check function, and in the case of packet drop, TCP forwards missing packets again. These services also use prefetching and buffering to realize perfect play-out. On the other hand, TV broadcasting is a live streaming during which the process of buffering and waiting for potential retransmission of missing parts adds an undesirable play-out delay in comparison to, for example, terrestrial TV broadcasting. UDP provides only the most basic transport layer functionality, it is used with higher layer protocols such as RTP (real-time transport protocol) or RTCP (RTP control protocol), which secure correct incoming packet ordering or statistical and control information. These VoD platforms also support the streaming technology HTTP-DASH (dynamic adaptive streaming over HTTP), which allows for better adaptation to network congestion and changes the video quality dynamically (e.g., reduction of resolution). HTTP needs reliable communication, which again disqualifies UDP, but it is not appropriate for real-time TV broadcasting [16].

Packet loss typically takes place when the network is congested. Internal buffers of network elements are full and incoming packets drop. Our previous works proved that packet loss higher than 1% caused significant deterioration of video quality [4,17], thus we focused primarily on packet loss as the major network impairment on overall video service quality.

Our research was principally focused on live TV streaming, which needs to be monitored by the national regulation authority. Many countries (e.g., EU countries, New Zealand) differentiate between transmission at the schedule time (subject to broadcasting and content regulations) and VoD platforms enabling a user to select what and when they watch. Quality of service evaluation is likely to be an issue that the regulator will consider. IPTV providers can control QoS because they use their own privately-managed network. It will be important for regulators to define minimal quality criterions and make sure that consumer interests are protected [18].

Our hybrid system can help with the creation of an IPTV regulatory framework, and it also serves as the baseline for minimal quality standard criteria selection.

3.3. Methods for Video Quality Assessment

At first, we had to select an objective and subjective assessment method. The objective method SSIM offers high correlation with human perception and is widely used. As depicted in Figure 1, the calculation process is derived from an analysis of contrast, brightness, and structure similarity with the reference video sequence. Results are shown in the form of a quality score for the investigated video, where 0 means no similarity with the reference and 1 means two identical sequences or images.

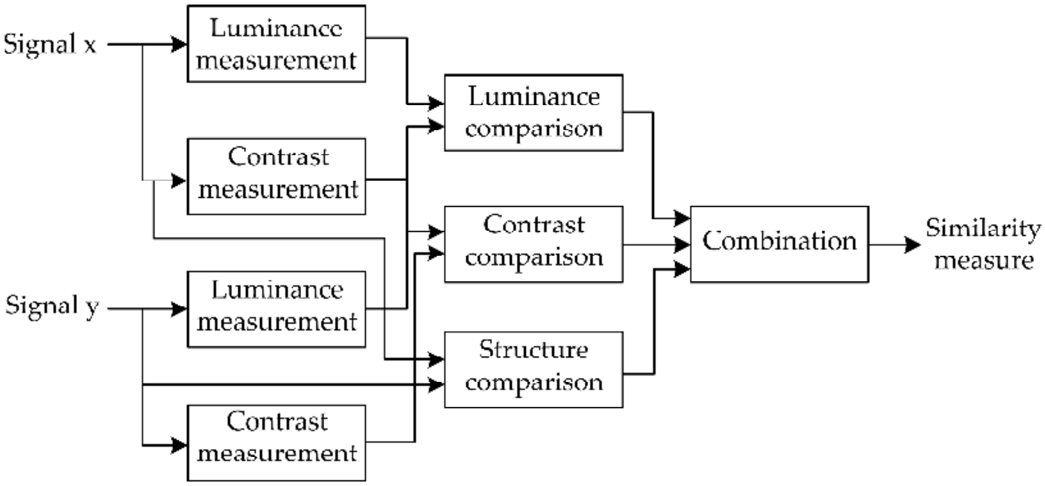


Figure 1. The block diagram of the SSIM index metric.

SSIM is a full-reference metric that improves older techniques based on the ratio between the maximum possible value of a signal and the power of distorting noise, the so-called peak signal-to-noise ratio (PSNR). SSIM calculation differs by evaluating the structural distortion instead of the error rate.

The final SSIM score is a combination of three parameters, with reference signal x and encoded test signal y being defined as follows [4]:

$$\mathrm{SSIM}(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma}, \tag{1}$$

- Element l(x,y) compares the brightness differences between x and y;
- Element c(x,y) compares the contrast of the signal;
- Element s(x,y) measures the structural similarity;
- $\alpha > 0$, $\beta > 0$, $\gamma > 0$ measures the weight of individual elements.

ITU-T standardized two subjective methods for video quality assessment. The first one is called the DCR (degradation category rating) method. Sequences are presented in pairs—the first sequence presented in each pair is always the reference video sequence, while the second stimulus is the test video sequence [2].

The absolute category rating method (ACR) is a category judgment where the test sequences are presented one at a time. This approach requires that after each presentation the observers are asked to evaluate the quality of the sequence shown (MOS scale). The time schedule for the stimulus presentation is illustrated by Figure 2. The ACR method was selected by us, in order to reflect the real situation where the end-user cannot compare the delivered quality of video with the original video sequence [2].

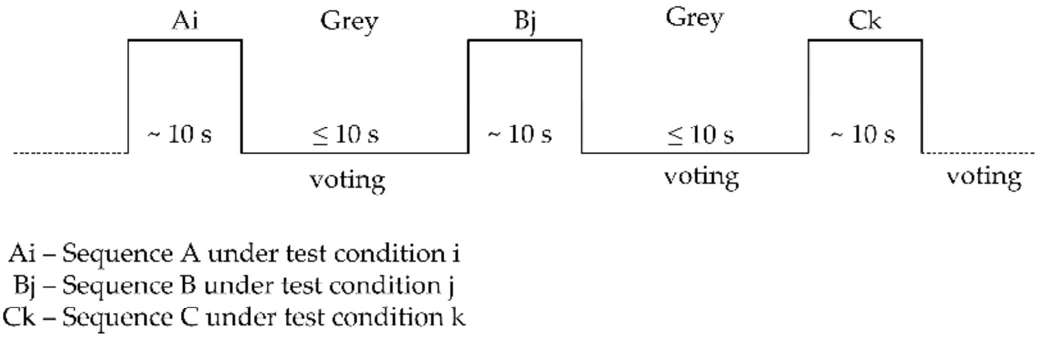

Ai
Grey
Bj
Grey
Ck
~ 10 s
≤ 10 s
~ 10 s
≤ 10 s
~ 10 s
voting
voting
voting
Ai – Sequence A under test condition i
Bj – Sequence B under test condition j
Ck – Sequence C under test condition k

Figure 2. Stimulus presentation in the absolute category rating (ACR) method.

4. Making of Testing Video Sequences

In order to create a neural network able to estimate subjective quality based on the objective score and selected video attributes, we needed to prepare testing video sequences in the required quality. None of the mentioned models in State-of-the-art can predict H.264 and H.265 concurrently by one model.

The recommendation by ITU-T P.910 [2] defines several categories of video content according to the time and spatial (SI and TI) information. These two parameters describe the character of the scene (e.g., sport, action movie, static TV news).

The temporal perceptual information (TI) is based on the differences in motion expressed by the feature, M_n (i, j), that represent the change between the pixel at the same location in space for following times or frames. M_n (i, j) as a function of time (n) is expressed as:

$$M_n(i,j) = F_n(i,j) - F_{n-1}(i,j), \tag{2}$$

where F_n (i, j) is the pixel at the i-th row and j-th column of n-th frame in time. The calculation of the temporal information (TI) is computed as the maximum over time ($\max_{time}$) of the standard deviation over space (std_{space}) of M_n (i, j) over all (i) and (j):

$$TI = \max_{time} std_{space}[M_n(i,j)]. \tag{3}$$

The spatial perceptual information (SI) is based on the Sobel filter. For each video frame within the time n (F_n), the Sobel filter [Sobel(F_n)] is applied first. The standard deviation over the pixels (std_{space}) in each Sobel-filtered frame is then calculated. This action is repeated for each frame in the video sequence and results in a time series of spatial information regarding the scene. The maximum value in the time series ($\max_{time}$) has been selected to characterize the spatial information content of the scene. This method can be represented as follows:

$$SI = \max_{time}\left\{std_{space}[Sobel(F_n)]\right\}. \tag{4}$$

A research team from Shanghai Jiao Tong University released its own uncompressed UHD video sequences in YUV format for testing purposes [19]. This public database is available to the research community free of charge.

Video sequences are 10 seconds long with frame rate 30, therefore all videos contain 300 frames and offer various scenes in a range of spatial and temporal parameters, as depicted in Figure 3.

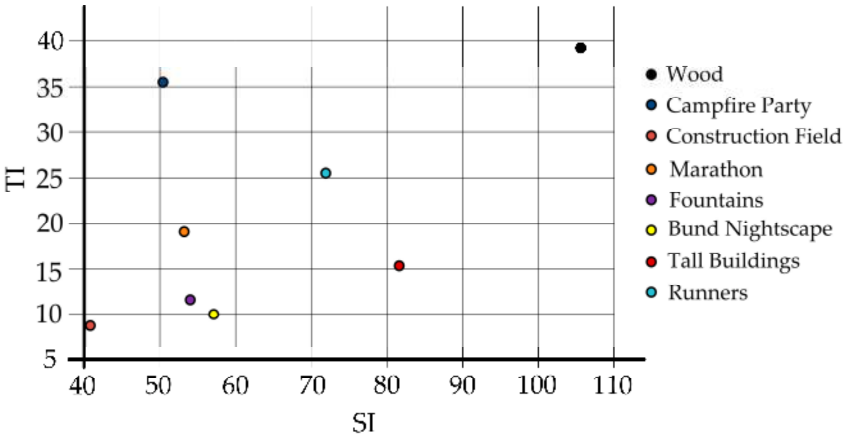


Figure 3. Spatial information (SI) and temporal information (TI) values for UHD video sequences [19].

As per the ITU-T recommendation, we decided to use four sequences—one from each of the four quadrants of the spatial–temporal information graph [2]. Our database was made from these types of video sequences:

- Construction Field—A construction vehicle is surrounded by buildings under construction. Dynamic objects include an excavator and walking workers. The scene is captured as a static shot;
- Runners—A marathon race, static shot again, dynamic movement of racers;
- Wood—Forest scenery, the dynamic camera is moving from left to right and the motion accelerates in the sequence. The highest values for spatial and temporal information;
- Campfire Party—Night scene, relatively static people near the bonfire. Fast change of flame moving.

All sequences have been encoded in the UHD resolution (3840 × 2160); 4:2:0 colour sampling and 8 bits per sample (16.7 million colours) correspond with the typical TV broadcasting profile. The measurement procedure consisted of the following steps.

First, selected sequences were downloaded from the webserver [19] in YUV format (the uncompressed format) and used as the reference sequences. Afterwards, all video sequences were encoded to both MPEG compression standards, H.264/AVC and H.265/HEVC, via the FFmpeg tool (includes ×264 and ×265 encoders). The target bitrates were set to 5, 10, and 15 Mbps, and the group of picture format (GOP) was set to the half of the framerate, i.e., M = 3, N = 15. Finally, the quality between these sequences (encoded back to YUV format) and the reference (uncompressed) sequence was compared and evaluated. This was done by using the MSU Video Quality Measuring Tool. A simulation of quality disruption by packet dropping was performed by FFmpeg and VLC Player software, the first serving as a streaming server and the second one as playout. We captured and saved the broadcast stream via the local computer interface using VLC Player. During the streaming process, we set the packet loss to 0.1% on the local interface. Then, we repeated this step for packet loss in increments of 0.2%, 0.3%, 0.5%, 0.75%, and 1%. The streaming process was based on the RTP/UDP/IP method with MPEG-TS (Transport Stream), thus we completely adopted the mechanism of IPTV transmission over the IP network [20]. A total of 432 testing video sequences were created. In order to obtain a reference output for our designed model, we needed to have a subjective evaluation of the testing video sequences. We prepared a testing room with a TV screen, meeting the conditions stated in ITU-T BT.500-13 [3].

There were 60 observers in the age range 18–35. Men dominated the ratio 38:22. ACR was selected as the subjective assessment method. The whole measuring process is described in Figure 4.

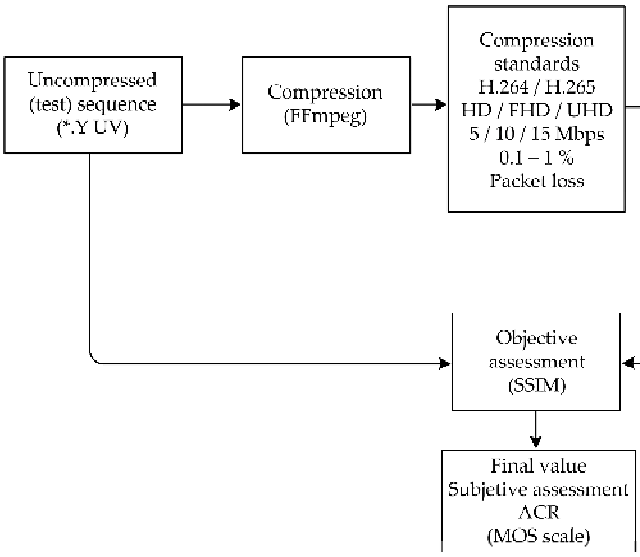


Figure 4. The whole procedure of creation and evaluation of testing video sequences.

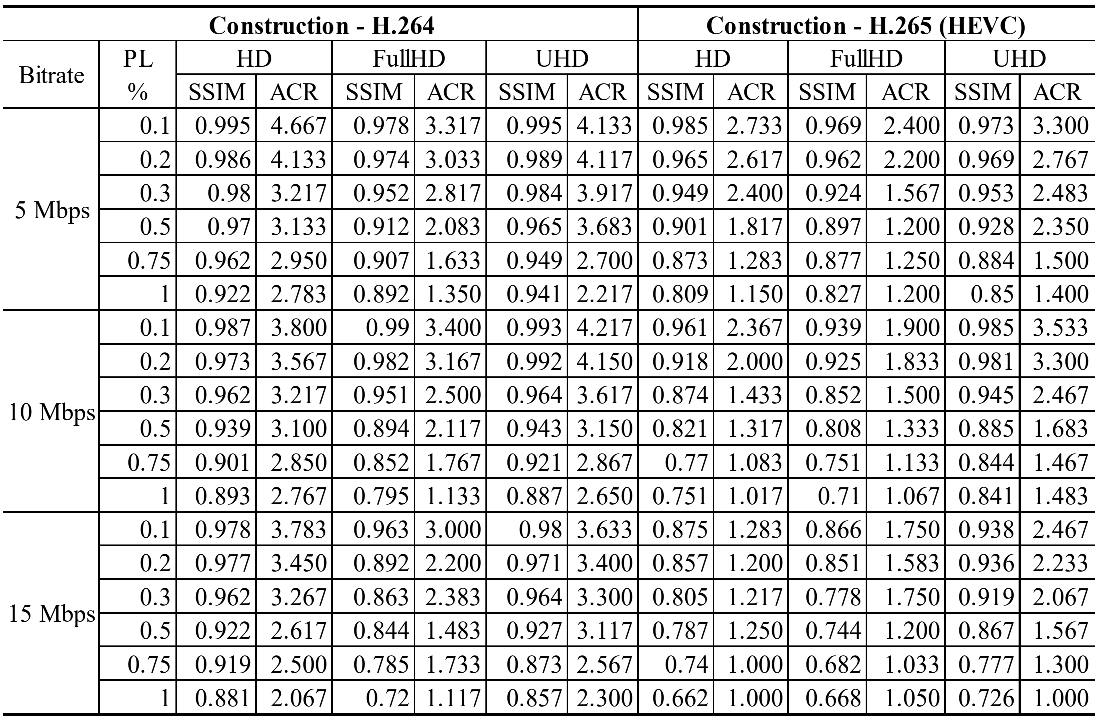
4.1. Database of Subjective and Objective Tests Results

Subjective evaluation can be affected by high variability. Due to this fact, we needed to verify how precisely the calculated mean represents the subjective perception. For every test sequence, the variation coefficient was calculated. The variation coefficient is defined as the ratio of the standard deviation (σ) to the mean. If the value of this statistical parameter is higher than 50%, the arithmetic average cannot be used for data representative purposes due to its significant dispersion of collected data. We computed the variation coefficient for all obtained subjective evaluations, and discovered that only 50 from a total of 432 testing video sequences had a variation coefficient higher than 35%, and none of them exceeded 40%. All results are shown below in Figures 5 and 6.

Campfire - H.264								Campfire - H.265 (HEVC)					
Bitrate	PL	HD		FullHD		UHD		HD		FullHD		UHD	
	%	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR
5 Mbps	0.1	0.997	4.017	0.985	3.117	0.993	2.617	0.987	3.050	0.99	2.817	0.991	3.633
	0.2	0.995	3.700	0.981	2.900	0.991	2.417	0.978	3.100	0.975	2.333	0.966	2.700
	0.3	0.991	3.500	0.978	2.850	0.982	1.667	0.893	2.433	0.923	1.500	0.963	2.517
	0.5	0.985	3.150	0.964	2.083	0.964	1.450	0.869	1.800	0.914	1.267	0.942	2.017
	0.75	0.974	3.133	0.959	2.067	0.963	1.317	0.837	1.467	0.905	1.150	0.932	1.917
	1	0.959	2.883	0.922	1.683	0.946	1.167	0.809	1.333	0.881	1.100	0.853	1.467
10 Mbps	0.1	0.987	4.083	0.985	2.867	0.99	3.250	0.965	3.233	0.969	2.150	0.984	3.200
	0.2	0.973	3.533	0.975	2.417	0.981	3.133	0.938	3.117	0.946	2.033	0.981	3.000
	0.3	0.927	3.033	0.961	2.200	0.929	2.950	0.918	2.700	0.917	1.450	0.919	2.583
	0.5	0.907	2.967	0.945	2.000	0.904	2.700	0.874	1.933	0.907	1.167	0.862	1.833
	0.75	0.887	2.217	0.92	1.700	0.9	2.633	0.853	1.733	0.887	1.167	0.855	1.700
	1	0.886	1.517	0.904	1.367	0.884	1.933	0.772	1.417	0.86	1.200	0.836	1.400
15 Mbps	0.1	0.944	4.233	0.971	3.117	0.972	2.583	0.921	2.533	0.968	2.117	0.979	2.967
	0.2	0.913	3.600	0.954	2.517	0.92	2.183	0.919	2.417	0.967	2.067	0.952	2.533
	0.3	0.904	2.900	0.937	2.400	0.894	2.100	0.892	2.067	0.914	1.467	0.947	2.417
	0.5	0.869	2.650	0.89	1.700	0.863	1.317	0.76	1.450	0.898	1.200	0.898	2.100
	0.75	0.813	2.467	0.884	1.600	0.853	1.250	0.754	1.267	0.873	1.200	0.829	1.400
	1	0.768	2.067	0.861	1.333	0.842	1.100	0.734	1.233	0.809	1.033	0.764	1.067

(a)

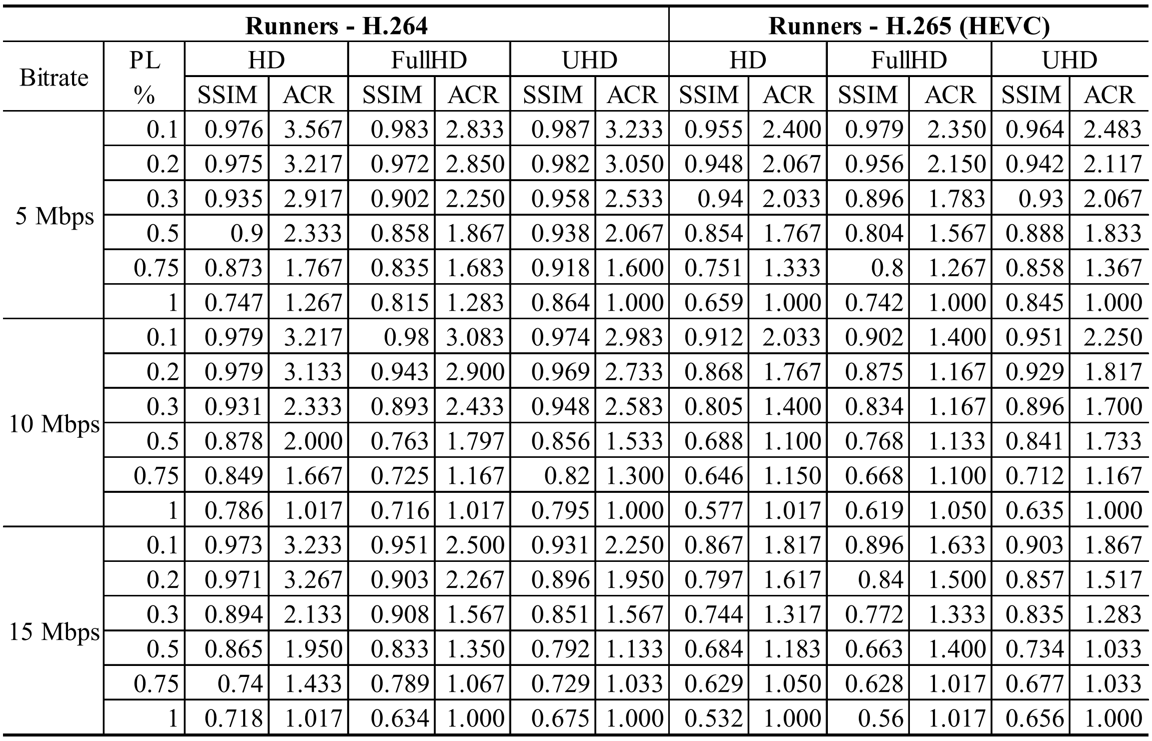
Figure 5. *Cont.*



Construction - H.264								Construction - H.265 (HEVC)					
Bitrate	PL	HD		FullHD		UHD		HD		FullHD		UHD	
	%	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR
5 Mbps	0.1	0.995	4.667	0.978	3.317	0.995	4.133	0.985	2.733	0.969	2.400	0.973	3.300
	0.2	0.986	4.133	0.974	3.033	0.989	4.117	0.965	2.617	0.962	2.200	0.969	2.767
	0.3	0.98	3.217	0.952	2.817	0.984	3.917	0.949	2.400	0.924	1.567	0.953	2.483
	0.5	0.97	3.133	0.912	2.083	0.965	3.683	0.901	1.817	0.897	1.200	0.928	2.350
	0.75	0.962	2.950	0.907	1.633	0.949	2.700	0.873	1.283	0.877	1.250	0.884	1.500
	1	0.922	2.783	0.892	1.350	0.941	2.217	0.809	1.150	0.827	1.200	0.85	1.400
10 Mbps	0.1	0.987	3.800	0.99	3.400	0.993	4.217	0.961	2.367	0.939	1.900	0.985	3.533
	0.2	0.973	3.567	0.982	3.167	0.992	4.150	0.918	2.000	0.925	1.833	0.981	3.300
	0.3	0.962	3.217	0.951	2.500	0.964	3.617	0.874	1.433	0.852	1.500	0.945	2.467
	0.5	0.939	3.100	0.894	2.117	0.943	3.150	0.821	1.317	0.808	1.333	0.885	1.683
	0.75	0.901	2.850	0.852	1.767	0.921	2.867	0.77	1.083	0.751	1.133	0.844	1.467
	1	0.893	2.767	0.795	1.133	0.887	2.650	0.751	1.017	0.71	1.067	0.841	1.483
15 Mbps	0.1	0.978	3.783	0.963	3.000	0.98	3.633	0.875	1.283	0.866	1.750	0.938	2.467
	0.2	0.977	3.450	0.892	2.200	0.971	3.400	0.857	1.200	0.851	1.583	0.936	2.233
	0.3	0.962	3.267	0.863	2.383	0.964	3.300	0.805	1.217	0.778	1.750	0.919	2.067
	0.5	0.922	2.617	0.844	1.483	0.927	3.117	0.787	1.250	0.744	1.200	0.867	1.567
	0.75	0.919	2.500	0.785	1.733	0.873	2.567	0.74	1.000	0.682	1.033	0.777	1.300
	1	0.881	2.067	0.72	1.117	0.857	2.300	0.662	1.000	0.668	1.050	0.726	1.000

(**b**)

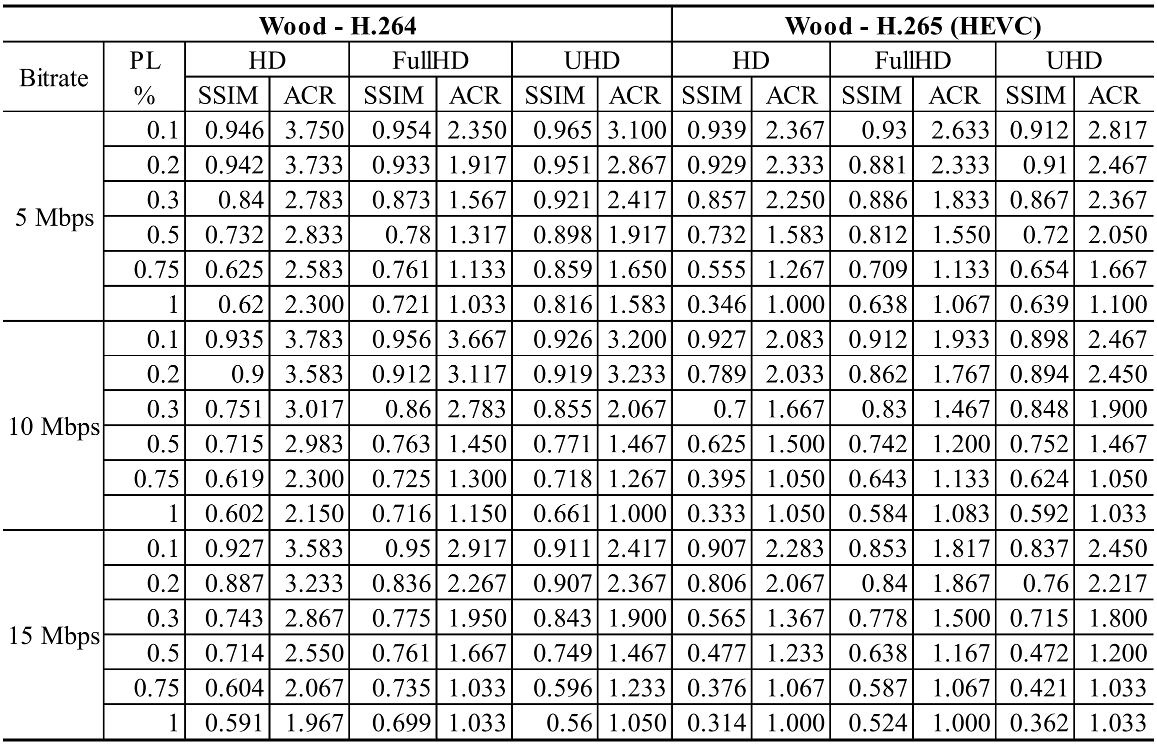
Figure 5. Subjective and objective metric results for the Campfire (**a**) and Construction (**b**) scenes.



Runners - H.264								Runners - H.265 (HEVC)					
Bitrate	PL	HD		FullHD		UHD		HD		FullHD		UHD	
	%	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR
5 Mbps	0.1	0.976	3.567	0.983	2.833	0.987	3.233	0.955	2.400	0.979	2.350	0.964	2.483
	0.2	0.975	3.217	0.972	2.850	0.982	3.050	0.948	2.067	0.956	2.150	0.942	2.117
	0.3	0.935	2.917	0.902	2.250	0.958	2.533	0.94	2.033	0.896	1.783	0.93	2.067
	0.5	0.9	2.333	0.858	1.867	0.938	2.067	0.854	1.767	0.804	1.567	0.888	1.833
	0.75	0.873	1.767	0.835	1.683	0.918	1.600	0.751	1.333	0.8	1.267	0.858	1.367
	1	0.747	1.267	0.815	1.283	0.864	1.000	0.659	1.000	0.742	1.000	0.845	1.000
10 Mbps	0.1	0.979	3.217	0.98	3.083	0.974	2.983	0.912	2.033	0.902	1.400	0.951	2.250
	0.2	0.979	3.133	0.943	2.900	0.969	2.733	0.868	1.767	0.875	1.167	0.929	1.817
	0.3	0.931	2.333	0.893	2.433	0.948	2.583	0.805	1.400	0.834	1.167	0.896	1.700
	0.5	0.878	2.000	0.763	1.797	0.856	1.533	0.688	1.100	0.768	1.133	0.841	1.733
	0.75	0.849	1.667	0.725	1.167	0.82	1.300	0.646	1.150	0.668	1.100	0.712	1.167
	1	0.786	1.017	0.716	1.017	0.795	1.000	0.577	1.017	0.619	1.050	0.635	1.000
15 Mbps	0.1	0.973	3.233	0.951	2.500	0.931	2.250	0.867	1.817	0.896	1.633	0.903	1.867
	0.2	0.971	3.267	0.903	2.267	0.896	1.950	0.797	1.617	0.84	1.500	0.857	1.517
	0.3	0.894	2.133	0.908	1.567	0.851	1.567	0.744	1.317	0.772	1.333	0.835	1.283
	0.5	0.865	1.950	0.833	1.350	0.792	1.133	0.684	1.183	0.663	1.400	0.734	1.033
	0.75	0.74	1.433	0.789	1.067	0.729	1.033	0.629	1.050	0.628	1.017	0.677	1.033
	1	0.718	1.017	0.634	1.000	0.675	1.000	0.532	1.000	0.56	1.017	0.656	1.000

(**a**)

Figure 6. *Cont.*



Wood - H.264								Wood - H.265 (HEVC)					
Bitrate	PL	HD		FullHD		UHD		HD		FullHD		UHD	
	%	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR	SSIM	ACR
5 Mbps	0.1	0.946	3.750	0.954	2.350	0.965	3.100	0.939	2.367	0.93	2.633	0.912	2.817
	0.2	0.942	3.733	0.933	1.917	0.951	2.867	0.929	2.333	0.881	2.333	0.91	2.467
	0.3	0.84	2.783	0.873	1.567	0.921	2.417	0.857	2.250	0.886	1.833	0.867	2.367
	0.5	0.732	2.833	0.78	1.317	0.898	1.917	0.732	1.583	0.812	1.550	0.72	2.050
	0.75	0.625	2.583	0.761	1.133	0.859	1.650	0.555	1.267	0.709	1.133	0.654	1.667
	1	0.62	2.300	0.721	1.033	0.816	1.583	0.346	1.000	0.638	1.067	0.639	1.100
10 Mbps	0.1	0.935	3.783	0.956	3.667	0.926	3.200	0.927	2.083	0.912	1.933	0.898	2.467
	0.2	0.9	3.583	0.912	3.117	0.919	3.233	0.789	2.033	0.862	1.767	0.894	2.450
	0.3	0.751	3.017	0.86	2.783	0.855	2.067	0.7	1.667	0.83	1.467	0.848	1.900
	0.5	0.715	2.983	0.763	1.450	0.771	1.467	0.625	1.500	0.742	1.200	0.752	1.467
	0.75	0.619	2.300	0.725	1.300	0.718	1.267	0.395	1.050	0.643	1.133	0.624	1.050
	1	0.602	2.150	0.716	1.150	0.661	1.000	0.333	1.050	0.584	1.083	0.592	1.033
15 Mbps	0.1	0.927	3.583	0.95	2.917	0.911	2.417	0.907	2.283	0.853	1.817	0.837	2.450
	0.2	0.887	3.233	0.836	2.267	0.907	2.367	0.806	2.067	0.84	1.867	0.76	2.217
	0.3	0.743	2.867	0.775	1.950	0.843	1.900	0.565	1.367	0.778	1.500	0.715	1.800
	0.5	0.714	2.550	0.761	1.667	0.749	1.467	0.477	1.233	0.638	1.167	0.472	1.200
	0.75	0.604	2.067	0.735	1.033	0.596	1.233	0.376	1.067	0.587	1.067	0.421	1.033
	1	0.591	1.967	0.699	1.033	0.56	1.050	0.314	1.000	0.524	1.000	0.362	1.033

(**b**)

Figure 6. Subjective and objective metric results for the Runner (**a**) and Wood (**b**) scenes.

According to these results, we can make the following deductions:

- Video codec H.265 (HEVC) gets worse results than its predecessor H.264 when packet loss appears;
- Higher bitrate means lower resistance to packet loss due to the bigger data transmission;
- An MOS value of at least 4 (good quality) was reached for the lowest SSIM value 0.945. An MOS value of 3 (average quality) was reached for the first time for SSIM value 0.887, which corresponds with the paper´s conclusion in [21];
- The scenes Construction (slow-motion) and Campfire (dark night background) gained better evaluations than high dynamic scenes Runners and Wood. Algorithms for missing data mask the work more efficiently when there is no dynamic change of motion and colors among the following frames [17,22].

4.2. Using the Database for Neural Network Modeling

The database of subjective and objective evaluations is used for neural network (NN) inputs, as well as a reference for finding, testing, and validating NN topology and activation function. As can be seen, many scenarios and network situations were performed. Obtained results serve as NN inputs, namely packet loss, resolution, bitrate, codec type, SSIM, and characteristic of the scene (static, dynamic, sport, and night). As an output, the MOS value would be calculated and compared with a reference value obtained from the dataset.

5. The Neural Network as a Classification for Subjective Video Quality Estimation

The artificial neural network is a computing system inspired by the biological neural network. NN currently provides the best solutions to many problems, such as recognition issues (images, emotions), or for the creation of prediction models (e.g., voice/video quality approximation).

5.1. Neural Network Characteristics Modeling

The multi-layer perceptron (MLP) is a feedforward neural network consisting of one input layer, at least one hidden layer, and one output layer. Feedforward means that data move from the input to the output layer. This type of network is trained by the back-propagation learning algorithm. MLPs are widely used for prediction, pattern classification, recognition, and estimation. MLP can resolve problems which are not separable linearly. The principal benefit of neural networks lies in an iterative learning process, in which the dataset is loaded to the network one at a time, and the weights associated with the input values are changed each time. After presenting all cases, the process often starts over again. During this learning stage, the network learns by calibrating the weights, which allows us to predict the proper outcome of input samples. Advantages of neural networks involve their adaptability to noisy data, as well as their capability to classify patterns on which they have not been trained.

Decisions regarding the number of layers and the number of neurons per layer are crucial. These settings to a feedforward, back-propagation topology show the "art" of the network designer. In our case, we tested more than 140 topologies (up to eight layers and a maximum of 230 neurons per layer).

Activation functions are an extremely important feature of the artificial network modeling process. They decide whether a neuron should be activated or not, and whether the information that the neuron is receiving is relevant for the given information or should be ignored. The activation function is the generally non-linear mapping function between inputs and the response variable. Because of this, finding the best fitting activation function is the second step of NN modeling. During the training of the NN, we had set several well-known activation functions (sigmoid, hyperbolic tangent, ReLu), and chosen the best one based on the measured values (correlation, RSME, and computing time).

The proposed model was implemented in Python language that supported several tools for modeling and computation, e.g., library Keras, packages NumPy, Scipy or Theano.

The dataset creation resulted in 25,920 (432 × 60) observations for training, validation, and testing of the neural network. After that, approximately 30% of the randomly chosen observations from the training set formed a validation set, which was then used to confirm the ability of the neural network to estimate the video quality. The training and validation were repeated 10 times. The proposed system achieved a high correlation with Pearson's Coefficient around 0.989 and RMSE of 0.2155 (MOS), which corresponds to an error of approximately 7% (related to the middle of the MOS scale). More than 73% of all observations lay within 10% relative error, as depicted in Figure 7. Since packet loss has a great impact on video quality, many of the observations in all sets are below the threshold of 2.5. Figure 8 shows four testing video sequences [19].

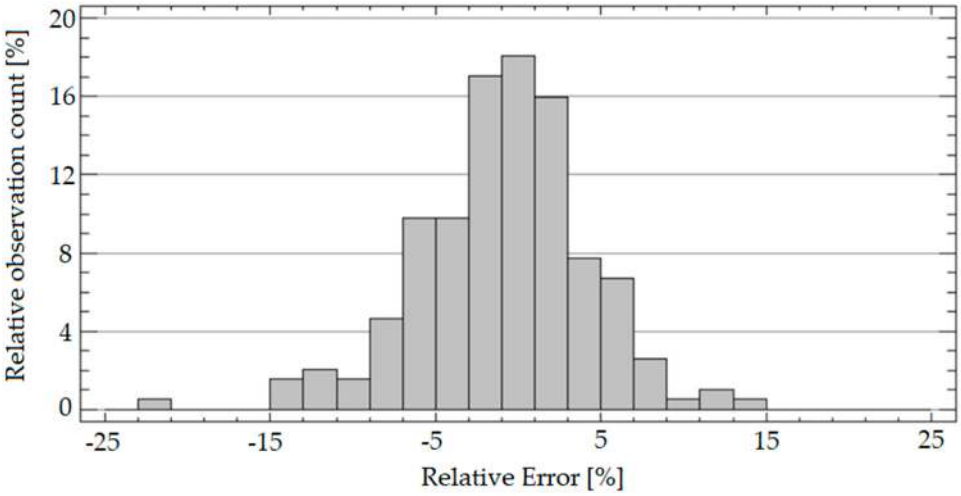


Figure 7. The relative error distribution.

Figure 8. Four testing video sequences [19].

5.2. Verification of the Proposed Model

The proposed model works with two video codecs, namely H.264 and H.265. The testing dataset which had not participated in the training and validation phase was divided into two equal smaller groups, unique for codecs H.264 and H.265. This step helped us recognize if the prediction accuracy was similar for both codecs. All statistics methods were performed with a significance level of 0.05.

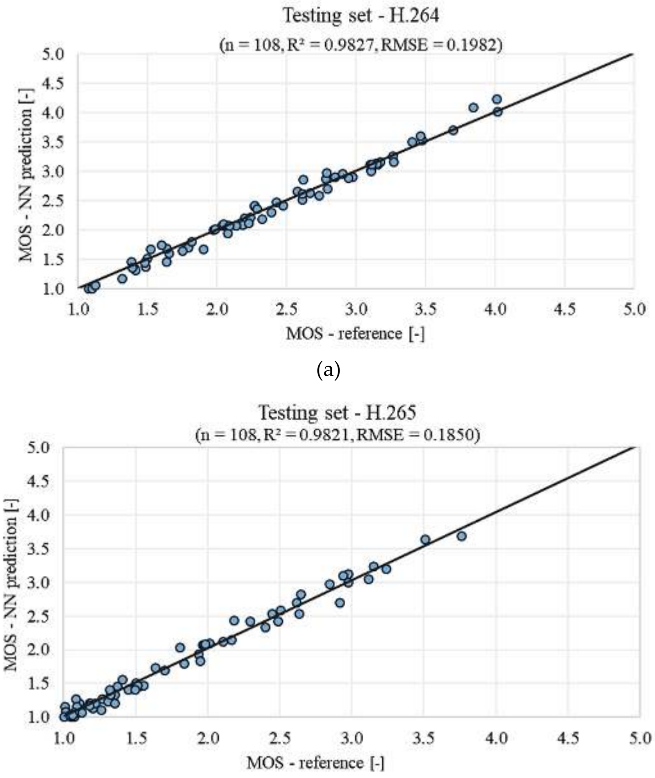
Table 1 shows the comparison (prediction of subjective perception in MOS) of median, modus, confidential intervals, and standard derivation for the reference and testing set for both video codecs. As described, the reference and model datasets were very comparable. Because the Shapiro–Wilk test of normality rejected the hypothesis of the datasets' normal distribution, a nonparametric test, called the Mann–Whitney–Wilcoxon U test, was selected as the next statistical method. This test could be used to determine whether two independent samples came from the same population. It figured out whether two sample means were equal or not. Calculated values coming from this test are shown in Table 2. Because the p-value of the Mann–Whitney–Wilcoxon test is higher than the critical value 0.05, it will be assumed that there is no significant difference between the reference and model datasets. Correlation diagrams are depicted in Figure 9. Error function RMSE is even better for all testing datasets, and the two model outputs greatly correlated with their reference outputs.

Table 1. Results of exploratory statistical parameters.

Codec	Dataset	σ_{min} (-)	σ_{max} (-)	Median (-)	Mean (-)
H.264	Reference	0.748	0.980	2.509	2.467
	Model (NN)	0.722	0.945	2.567	2.497
H.265	Reference	0.556	0.728	1.742	1.822
	Model (NN)	0.541	0.708	1.722	1.812

Table 2. Results of the Mann–Whitney–Wilcoxon test.

Codec	Dataset	Rank Sum	W	*p*-Value
H.264	Reference	107.139	5979	0.7497
	Model (NN)	109.861		
H.265	Reference	108.736	5806.5	0.9566
	Model (NN)	108.264		



(a)

(b)

Figure 9. Correlation diagrams for H.264 (**a**) and H.265 (**b**).

The last phase of verification was a comparison with the models published so far. As shown in Table 3, our prediction model based on a back propagation neural network (BPNN) reached better results in all cases except for the first one (for the nowadays outdated codec MPEG-2) that has created the basis for our research and motivation.

Table 3. Comparison of the proposed model with published models for subjective prediction of video quality.

Research Paper	No. of Scenes	Resolution	Codec	BR	FR	PLR	MLT	r (-)	RMSE (-)
[6]	ND	CIF (352 × 288)	MPEG-2	X	X	X	RNN	0.981	0.07
[7]	4	HD	H.264	X	-	X	BPNN	0.945	0.26
[8]	3	HD	H.264	X	-	X	Regression	0.9265	ND
[9]	3 free available databases	HD	H.264	-	-	-	Elastic Net	0.9	0.58
[10]	5	HD, FullHD	H.264	X	-	X	Regression, RNN, DBN	0.77, 0.85, 0.83	0.63, 0.51, 0.54
[11]	16	HD, FullHD	H.265	-	X	X	Regression	0.934	ND
[12]	3	FullHD	H.265	-	-	X	Regression	0.92	0.23
[13]	4	UHD	H.265	X	-	X	FIS	0.935	0.17
[14]	2	CIF, HD	H.264	X	X	X	Decision Tree	0.81	0.57
[23]	4 and 3 for testing set	HD, FullHD	H.264	-	-	-	ABFR	0.924	0.53
[24]	4	HD	H.264	-	-	-	RNN	ND	0.37
Our Model	**4**	**HD, FullHD, UHD**	**H.264, H.265**	**X**	**-**	**X**	**BPNN**	**0.983** **0.982**	**0.20** **0.19**

ND—not defined, X—presented, BR—bitrate, FR—framerate, PLR—packet loss rate, RNN—recurrent neural network, DBN—deep belief network, ABFR—adaptive basis function regression, FIS—fuzzy interface system, MLT—Machine learning tool.

Our model can estimate subjective perceptions of video quality not only for mentioned codecs, but also for the three resolutions, which have not been offered by any of the below-mentioned models.

6. Discussion

In Table 3 we can see that several machine learning methods were used. Some models do not use any extracted video features (e.g., codec type or bitrate) as an input for mapping function. The high regression value of Pearson´s correlation coefficient indicates how good our proposed model is. The obtained value of more than 0.98 suggests that the neural network has gained a high level of accuracy. Due to the usage of video features and packet loss (major network impairment for video service) we can simulate a high scale of scenarios and adapt service behavior more precisely than any other mentioned model. We plan to continue with research in this field and prepare an extended version of the application. We want to incorporate comments and feedback obtained from real deploying, and add more scenes for accuracy improvement. We are also discussing the implementation of other video codecs, such as VP9, AVI, or MP4, that are not used primarily for IPTV, but which are met by end-users in VoD services. Finally, because smartphones have become a powerful tool, we want to make a version for the Android environment.

Research in this field is still interesting and actual. Authors in [25] created a video service recovery mechanism that performs network load balancing according to the MOS value. Paper [26] dealt with picture quality enhancements. Poor quality images evaluated by objective metrics PSNR and SSIM are improved by computer vision techniques and then analyzed again. The proposed framework used a temporal information feature and should be an interesting way of improving video service quality in an unreliable IP environment. In [27], the research team prepared a subjectively evaluated dataset. As a main video sequence feature, they used SI and TI information. Their video quality assessment dataset should eliminate the need for frequently subjective test performances.

7. Conclusions

This paper described an application that offers a prediction of subjective video quality related to the specific video attributes and objective quality score. Network (content) providers have to monitor and solve problems appearing in transmission infrastructure. They have to know not only the objective score, but also realize the impact of network behavior on subjective customer perceptions of quality. We tested many scenarios, including two codecs, three bitrates and resolutions, four scenes, and packet loss up to 1%. The proposed prediction model, based on machine learning principles, is unique because it offers subjective quality estimation for H.264 and H.265 by one common mapping function. Our monitoring tool aims to be a helpful tool for video quality estimation, and it can serve as a baseline for creating a regulatory framework devoted to QoS for IPTV providers.

Author Contributions: J.F. proposed the system idea and edited the manuscript. J.F., J.N., J.V. and R.M. developed, tested and validated data. J.F. and J.N. wrote the manuscript. J.F. critically evaluated the quality of the research data and experimental methods used to generate/acquire them as well as the soundness/validity of the scientific and engineering techniques, wrote the manuscript, and performed its final edits.

Funding: This article was supported by the Ministry of Education of the Czech Republic (Projects No. SP2019/80, SP2019/85). This article was also supported by the European Regional Development Fund in the Research Centre of Advanced Mechatronic Systems project, project number CZ.02.1.01/0.0/0.0/16_019/0000867 within the Operational Programme Research, Development and Education. This work was supported by the European Regional Development Fund in Research Platform focused on Industry 4.0 and Robotics in Ostrava project, CZ.02.1.01/0.0/0.0/17_049/0008425 within the Operational Programme Research, Development and Education.

Acknowledgments: We appreciate the editors and peer reviewers for their valuable comments and suggestions. The authors would like to thank MDPI Institutional Open Access Program (IOAP) and University of Zilina for membership in this program.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Latal, J.; Wilcek, Z.; Kolar, J.; Vojtech, J. Measurement of IPTV Services on a Hybrid Access Network. In Proceedings of the 20th International Conference on Transparent Optical Networks (ICTON), Bucharest, Romania, 1–5 July 2018.
2. International Telecommunications Union. *ITU-T P.910. Subjective Video Quality Assessment Methods for Multimedia Applications*; International Telecommunications Union: Geneva, Switzerland, 2008.
3. International Telecommunications Union. *ITU-T BT.500-13. Methodology for the Subjective Assessment of the Quality of Television Pictures*; International Telecommunications Union: Geneva, Switzerland, 2012.
4. Frnda, J.; Voznak, M.; Rozhon, J.; Mehic, M. Prediction Model of QoS for Triple Play Services. In Proceedings of the 21st Telecommunications Forum (Telfor), Belgrade, Serbia, 26–28 November 2013; pp. 733–736.
5. Bienik, J.; Uhrina, M.; Kuba, M.; Vaculik, M. Performance of H. 264, H. 265, VP8 and VP9 Compression Standards for High Resolutions. In Proceedings of the 19th International Conference on Network-Based Information Systems (NBiS), Ostrava, Czech Republic, 7–9 September 2016; pp. 246–252.
6. Mohamed, S.; Rubino, G. A Study of Real-Time Packet Video Quality Using Random Neural Networks. *IEEE Trans. Circuits Syst. Video Technol.* **2002**, *12*, 1071–1083. [CrossRef]
7. Valderrama, D.; Gómez, N. Nonintrusive Method Based on Neural Networks for Video Quality of Experience Assessment. *Adv. Multimed.* **2016**, *2016*. [CrossRef]
8. Botia, D.; Gaviria, N.; Menedez, J.; Jimenez, D. An approach to correlation of QoE metrics applied to VoD service on IPTV using a Diffserv Network. In Proceedings of the 4th IEEE Latin-American Conference on Communications, Cuenca, Ecuador, 7–9 November 2012.
9. Søgaard, J.; Forchhammer, S.; Korhonen, J. Video quality assessment and machine learning: Performance and interpretability. In Proceedings of the 7th International Workshop on Quality of Multimedia Experience (QoMEX), Pylos-Nestoras, Greece, 26–29 May 2015.
10. Mocanu, D.C.; Pokhrel, J.; Garella, P.; Seppänen, J.; Liotou, E.; Narwaria, M. No-reference video quality measurement: added value of machine learning. *J. Electron. Imaging* **2016**, *24*, 6. [CrossRef]
11. Cheng, Z.; Ding, L.; Huang, W.; Yang, F.; Qian, L. A unified QoE prediction framework for HEVC encoded video streaming over wireless networks. In Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Cagliari, Italy, 7–9 June 2017; pp. 1–6.
12. Anegekuh, L.; Sun, L.; Jammeh, E.; Mkwawa, I.H.; Ifeachor, E. Content-Based Video Quality Prediction for HEVC Encoded Videos Streamed Over Packet Networks. *IEEE Trans. Multimed.* **2015**, *17*, 1323–1334. [CrossRef]
13. Alreshoodi, M.; Adeyemi-Ejeye, A.O.; Woods, J.; Walker, S.D. Fuzzy logic inference system-based hybrid quality prediction model for wireless 4k UHD H.265-coded video streaming. *IET Netw.* **2015**, *4*, 296–303. [CrossRef]
14. Abar, T.; Ben Letaifa, A.; El Asmi, S. Machine learning based QoE prediction in SDN networks. In Proceedings of the 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, 26–30 June 2017; pp. 1395–1400.
15. Tan, T.T.; Weerakkody, R.; Mrak, M.; Ramzan, N.; Baroncini, V.; Ohm, J.R.; Sullivan, G.J. Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 76–90. [CrossRef]
16. Sayit, M.; Cetinkaya, C.; Yildiz, H.U.; Tavli, B. DASH-QoS: A scalable network layer service differentiation architecture for DASH over SDN. *Comput. Netw.* **2019**, *154*, 12–25. [CrossRef]
17. Frnda, J.; Sevcik, L.; Uhrina, M.; Voznak, M. Network Degradation Effects on Different Codec Types and Characteristics of Video Streaming. *Adv. Electr. Electron. Eng.* **2014**, *12*, 377–383. [CrossRef]
18. ITU News. IPTV: New challenges for regulators. *ITU News*, October 2008; pp. 31–32, ISSN 1020-4148.
19. Song, L.; Tang, X.; Zhang, W.; Yang, X.; Xia, O. The SJTU 4K video sequence dataset. In Proceedings of the 5th International Workshop on Quality of Multimedia Experience, Klagenfurt am Worthersee, Austria, 3–5 July 2013.
20. International Telecommunications Union. *ITU-T Y.1910. IPTV Functional Architecture*; International Telecommunications Union: Geneva, Switzerland, 2008.

21. Zinner, T.; Abboud, O.; Hohlfeld, O.; Hossfeld, T.; Tran-Gia, P. Towards QoE Management for Scalable Video Streaming. In Proceedings of the 21st ITC Specialist Seminar on Multimedia Applications—Traffic, Performance and QoE, Miyazaki, Japan, 2–3 March 2010.
22. Adeyemi-Ejeye, A.O.; Alreshoodi, M.; Al-Jobouri, L.; Fleury, M.; Woods, J. Packet loss visibility across SD, HD, 3D, and UHD video streams. *J. Vis. Commun. Image Represent.* **2017**, *45*, 95–106. [CrossRef]
23. Narwaria, M.; Lin, W. Machine Learning Based Modeling of Spatial and Temporal Factors for Video Quality Assessment. In Proceedings of the 18th IEEE International Conference on Image Processing (ICIP), Brussels, Belgium, 11–14 September 2011; pp. 2513–2516.
24. Singh, G.K.D.; Hadjadj-Aoul, Y.; Rubino, G. Quality of experience estimation for adaptive Http/Tcp video streaming using H.264/AVC. In Proceedings of the 9th Anual IEEE Consumer Communications and Networking Conference Multimedia and Entertaiment Networking and Services, Las Vegas, NV, USA, 14–17 January 2012; pp. 127–131.
25. Zhang, H.; Wang, R.Y.; Liu, H. Video Service Recovery Mechanism Based on Quality of Experience-Aware in Hybrid Wireless-Optical Broadband-Access Network. *Mob. Netw. Appl.* **2018**, *23*, 664–672. [CrossRef]
26. Jammal, S.; Tillo, T.; Xiao, J. Multiview video quality enhancement without depth information. *Signal Process. Image Commun.* **2019**, *75*, 22–31. [CrossRef]
27. Aldahdooh, A.; Masala, E.; Van Wallendael, G.; Lambert, P.; Barkowsky, M. Improving relevant subjective testing for validation: Comparing machine learning algorithms for finding similarities in VQA datasets using objective measures. *Signal Process. Image Commun.* **2019**, *74*, 32–41. [CrossRef]