AN OBJECTIVE EVALUATION OF CODECS AND POST-PROCESSING TOOLS FOR 8K VIDEO COMPRESSION

Charles Bonnineau[1,2,3], Jean-Yves Aubié[1], Wassim Hamidouche[1,3],
Olivier Déforges[3], Jean-François Travers[2], Naty Sidaty[2]

[1] IRT b<>com France [2] TDF, France [3] IETR-INSA Rennes, France

ABSTRACT

With the deployment of the latest Ultra High Definition Television (UHDTV) system, Quality of Experience (QoE) of users is expected to be improved through the introduction of new features to the existing High Definition Television (HDTV) system, such as High Dynamic Range (HDR), wider color gamut, High Frame-Rate (HFR) and higher spatial resolutions including 4K (3840x2160) and 8K (7680x4320). The delivery of such video formats on current broadcast infrastructures is a real challenge and requires efficient compression methods to reach the available bandwidth while ensuring high video quality. On the other hand, with the outstanding performance of AI-based spatial up-scalers, TV manufacturers have recently included them as a feature in their products. As 8K contents are still uncommon, it can be useful to reconstruct high-quality 8K from a lower resolution to exploit the screen capabilities. Moreover, these filters can also be used to reduce the high bitrate requirements of 8K by downscaling the original 8K signal before encoding and reconstruct it after decoding.

In this paper, we propose to evaluate different video coding scenarios for 8K compression using VVC and HEVC standards through an objective study. Tested configurations include: 8K source encoded with HEVC, and 8K and 4K sources encoded with VVC and then upscaled with two methods: a Lanczos filter and a deep-learning-based Super-Resolution method called SRFBN. All configurations are tested on a set of 8K sequences using the verification model of the VVC and HEVC standards.

INTRODUCTION

With the advent of new immersive video formats, High Dynamic Range (HDR), HFR (High-Frame Rate), and UHD (Ultra-High Definition) are among the main improvements that are currently and progressively introduced in recent deployments of television networks. The ultimate phase of the migration to the new video formats, a.k.a. UHD-2 [1], considers the introduction of the 8K definition (7680x4320), in addition to existing formats like 4K (3840x2160) or HD (1920x1080). By increasing the number of pixels, a finer signal digitization is possible, bringing more details into the image, and thus increasing the Quality of Experience (QoE) for the users. However, high bandwidth capacities are required to transmit 8K estimated to 80Mbps with HEVC encoder [2]. Contributions to compression standards like HEVC or its successor VVC, scheduled to be released in Q4 2020, aims at reducing these requirements, but such bitrate values are still hard to reach in a broadcast context.

Recently, AI-based up-scalers [6] [7] [8] have shown outstanding performance over classical interpolation filters like lanczos [9] and bicubic [10]. These algorithms allow high-resolution details to be convincingly reconstructed from lower resolutions. In a broadcast context, the use of this type of approach would allow a high-quality 8K reconstruction to be made through the transmission of a single 4K bitstream.

In this paper we focus on evaluating, through an objective study, different coding approaches allowing to deliver 8K. For our tests we used the reference software models that provide the best implementation of the standards with a limited degree of optimization. We divided our tests into two parts:

1. The evaluation of VVC over HEVC using VVC Test Model (VTM) [11] and HEVC Test Model (HM) [12] implementations.

2. The evaluation of State-of-the-Art upscaling techniques over single layer 8K coding using VVC.

For post processing, we selected the lanczos interpolation filter as naive approach, and one of the State-of-the-Art of Super Resolution techniques called Feedback Network for Super-Resolution (SRFBN) [13] .

The remainder of this paper is organized as follows. We first present the different approaches that would allow 8K video delivery. Then we detail the tested configurations and the experimental settings. Results are then presented and analyzed. Finally, the last section concludes this paper.

8K VIDEO DELIVERY

Single-layer coding

Although, HEVC has brought significant bitrate reduction for 4K applications, its efficiency is not enough for the delivery of 8K contents. Indeed, some studies [2] [3] have demonstrated that 8K 60Hz and 120Hz (temporally scalable) would require bitrates up to 80Mbps. In practice, HEVC codecs have been used for satellite broadcasting in Japan, where HEVC codec for 8K 120Hz has been developed [4]. For satellite transmission with DVBS2X, bandwidth in the range 70-80Mb/s can be reached with the use of a complete transponder or multiple bonded transponders. For terrestrial transmission, such bandwidth requirements prevent the deployment of 8K with backward-compatibility, as practical DVB-T2 [5] channels offer bandwidth in the range 30-40Mb/s over an 8MHz channel.

With the recent progress of the Joint Video Exploration Team (JVET) of ITU-T and MPEG, in developing the next video coding standard, a brighter solution may be on the horizon for

8K services. The next generation standard, called Versatile Video Coding (VVC) should make delivery of this type of content more affordable as it is predicted to be able to encode 8K video below 50Mbps. The VVC standardization is scheduled to be finalized in July 2020.

Super-resolution

To cover a wide range of compatible UHD-1 or UHD-2 receivers, one solution would be to apply down-scaling and up-scaling operations to the signal outside the coding pipeline, as described in Figure 1. Thus, the bandwidth is limited to broadcast 4K only while both resolutions can be displayed by the receiver. In the image processing field, the process of estimating a high-resolution version of a low-resolution content is referred to as super-resolution.

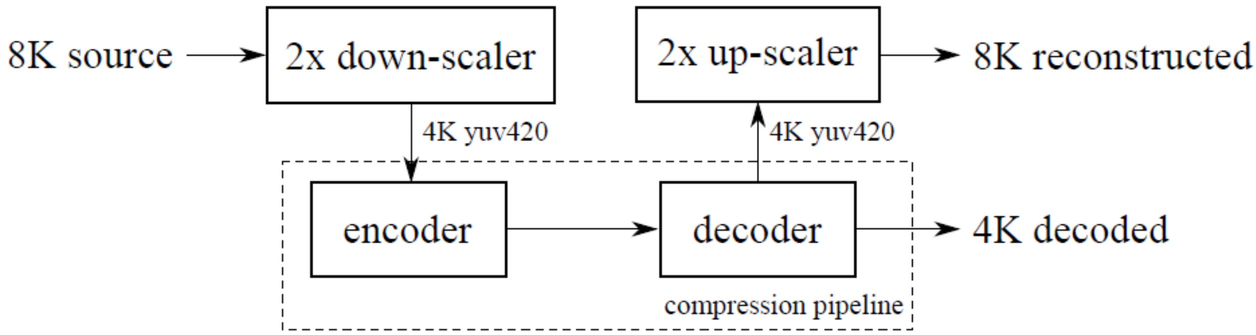

Figure 1 : Pre & Post-processing coding pipeline

Recently, traditional up-scaling methods, like bicubic or Lanczos, have been outperformed by learning-based Super-Resolution approaches [3] [4] [5]. These methods aim at learning the non-linear mapping between low-resolution (LR) and high-resolution (HR) images or videos by analysing deep features. For this study, we selected SRFBN [13]. This neural network corresponds to an end-to-end Convolutional Neural Network (CNN) coupled with a feedback mechanism that allows providing high performance in both visual quality enhancement and runtime. This method is initially optimized to recover details from uncompressed LR single images generated with a known downscale (e.g. bicubic) by minimizing the L1-norm between the reconstruction and the ground truth over training iterations. Initially, the data-sets used to train the network are the publicly-available super-resolution image data-sets Flickr2K and DIV2K [14]. However, the low-resolution sequences used for our study may be degraded by compression, making the baseline SRFBN network not adapted to the target task. Thus, to perform a fair evaluation, we have trained the model by using the initial data-sets Flickr2K and DIV2K encoded with VTM in All-Intra coding configuration.

EXPERIMENTS ON 8K COMPRESSION: EXPERIMENTAL SETTINGS

Configurations

These experiments aim to evaluate different possible scenarios for 8K broadcast. First, the performance of the compression standards HEVC and VVC are evaluated on 8K contents using their respective reference software HM and VTM. The use case of transmitting coded 4K-downscaled stream before applying super-resolution after decoding to reach 8K on the display is then tested against 8K single layer coding with VVC.

For this study, the visual quality is evaluated by using the Peak Signal to Noise Ratio (PSNR), the Structural SIMilarity (SSIM) and Video Multimethod Assessment Fusion (VMAF) [15] which provides good correlation with subjective scores. This metric relies on a Machine Learning (ML) model trained to estimate a score from different other metrics (motion, spatial, texture) that maximizes the correlation with MOS scores resulted from subjective tests conducted on 4K. To quantify the average gain in bitrate or visual quality offered by a tested method over an anchor, we use the Bjøntegaard delta (BD) method [16].

Initially, SRFBN is trained to recover the high-resolution version of uncompressed low-resolution data. This low complex Single Image method is applied frame by frame on the video sequences. In our case, we focus on assessing this method on compressed videos. Thus, to propose a fair evaluation of the model, we have trained the baseline network using a compressed versions of the initial image datasets DIV2K and Flickr2K. First, we have generated pairs of LR/HR images by applying a bicubic down-scaling filter with a scale-factor of 2 from HR images. Then, each LR image has been encoded using the All Intra configuration of the VTM with five QP values, including 22, 27, 32 and 37 to cover a large panel of distortions. These coded samples, representing different levels of compression distortions, are used to fine-tune the network. To evaluate Super-resolution, we also used Lanczos filter as a naive approach.

Test Sequences

The sequence used for our experiments were provided by The Institute of Image Information and Television Engineers. Screenshots of the sequences are illustrated in Figure 3. These sequences are in 8K (7680x4320) 60fps, YUV420 10 bits. The spatial and temporal information (SI-TI) [17] of these sequences is plotted in Figure 2. This 2D-plan shows that the contents selected for the study have various spatial and temporal features. The objective is to have a diversity of contents and analyse the coding efficiency regarding their features. The 4K versions of the sequences are generated by a bicubic down-scale.

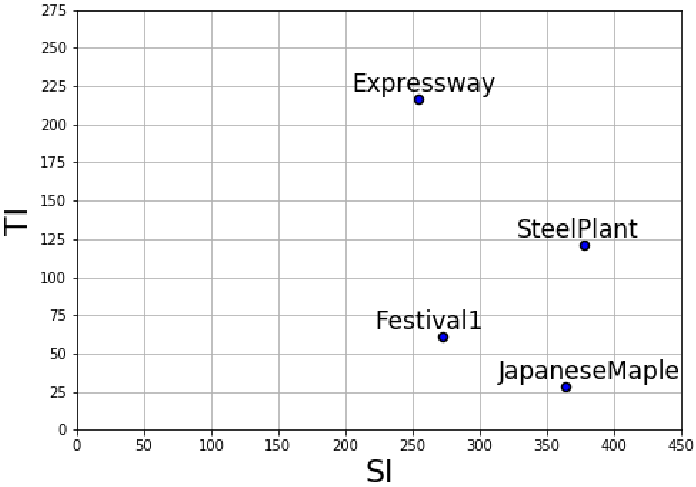

Figure 2 : SI-TI graph of the tested sequences

JapaneseMaple

Festival1

SteelPlant

Expressway

Figure 3 : 8K test sequences

Verification models

HEVC verification models (HM) and VVC verification model are developed by the Joint Collaborative Team on Video Coding (JCT-VC) and the Joint Video Exploration Team (JVET), respectively. These models provide a reference implementation the standards to simulate their performance in optimal conditions.

Table 1 : Codec specifications

	VVC	**HEVC**
Reference Soft.	VTM-5.0	HM16.20
Profile	next	main10
GOP size	16	16
Intra Period	64	64

For the experiment, the Common Test Conditions for VTM and HM are used to provide a fair rate/distortion estimation. The coding configurations are summarized in Table 1 for both codecs. Each test sequence is encoded in Random Access coding configuration with Quantization Parameter (QP) values of 22, 27, 32, and 37.

RESULTS

In this section, we analyse the results of our experiments. First, single layer codecs performance is compared on the 8K sequences using their respective reference implementations. Then, we show the benefits of pre and post-processing in a coding pipeline.

8K coding : VVC vs HEVC

To compare the performance of both reference software, we first compute rate distortion curves for all sequences using the different objective quality metrics, represented on Figures 4, 5 and 6. These graphs demonstrate that the performance differs from a sequence to another. For example, VVC allows higher bitrate saving for the sequence Festival1 than for the sequence JapaneseMaple. Moreover, for SteelPlant and Expressway, both curves are converging at high bitrate regarding VMAF. We can notice that although being classified as a non-trivial sequence by the SI-TI metrics, Expressway reaches high PSNR values at relatively low bitrate compared to the other videos. This sequence represents a camera panning of a road with high global motion, thus causing a loss of spatial definition. On the contrary, JapaneseMaple, SteelPlant and Festival1 are static contents with many details. These observations were validated by experts during visualisation sessions. These content particularities highlight the inability of current cameras to capture the 8K definition of high-moving scenes. Moreover, as all elements have linear trajectories in the scene, Expressway is easy to predict by the encoder, generating few residuals.

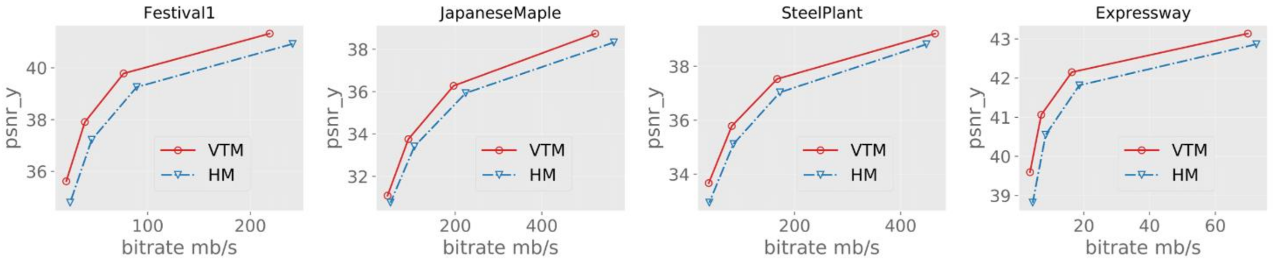

Figure 4 : PSNR

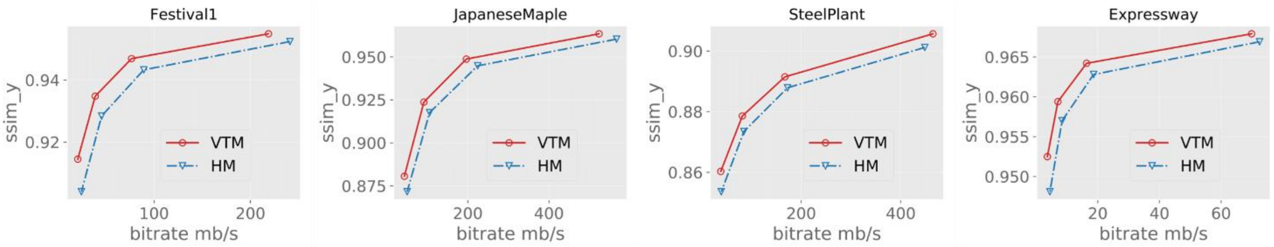

Figure 5 : SSIM

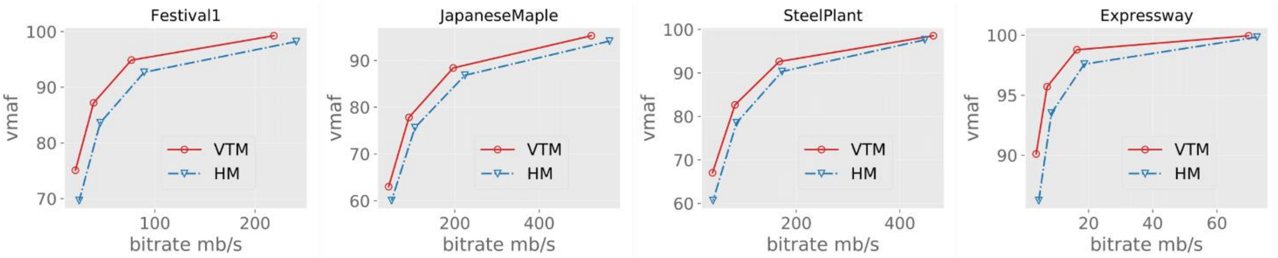

Figure 6 : VMAF

To provide more accurate performance indicators, we have also computed the BD-rate values corresponding to each curve. These scores are represented in Table 2. A negative value corresponds to a gain (in %) of bitrate for the same quality (regarding the corresponding objective metric) for VVC over HEVC. This table show that the gains vary based on the used metric. Indeed, 42.09% of bitrate is saved in average for VVC regarding SSIM against only 32.90% regarding VMAF.

Table 2 : BD-rate (%) resultats for VVC against HEVC (anchor)

	PSNR	**SSIM**	**VMAF**
Festival1	-45.40	-48.45	-43.68
JapaneseMaple	-27.81	-31.97	-30.99
SteelPlant	-30.36	-31.03	-25.84
Expressway	-57.46	-56.91	-31.07
Average	-40.26	-42.09	-32.90

8K coding : pre and post-processing

In this section, we evaluate the post processing methods applied on 4K VVC signal over 8K single layer coding with VVC. Rate distortion curves are represented in Figures 7, 8 and 9. We can notice that for some sequences, SRFBN provide significant improvement over the Lanczos filter. However, even re-trained on compressed data, the presented SRFBN gains are far from the performance announced in the related contribution [13]. Indeed, most of the State-of-the-Art Super-Resolution models are designed for upscaling low-resolutions images or videos generated without any additional degradations. This study demonstrates that there is a need for achitectures adapted to “real-world” use cases. Regarding 4K up-scaled against 8K, upscaling shows generally better performance than single-layer 8K coding, especially at low bitate. For the most spacially complex sequences, SRFBN proposes gain over single-layer 8K coding on a larger range of bitrate, especially regarding SSIM and VMAF scores.

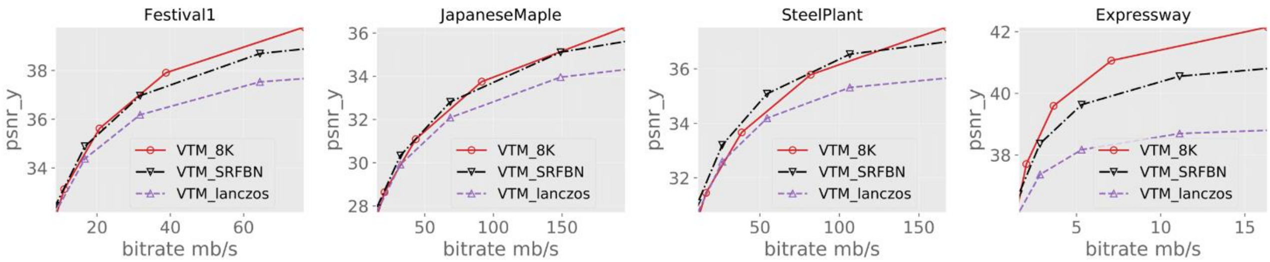

Figure 7 : PSNR

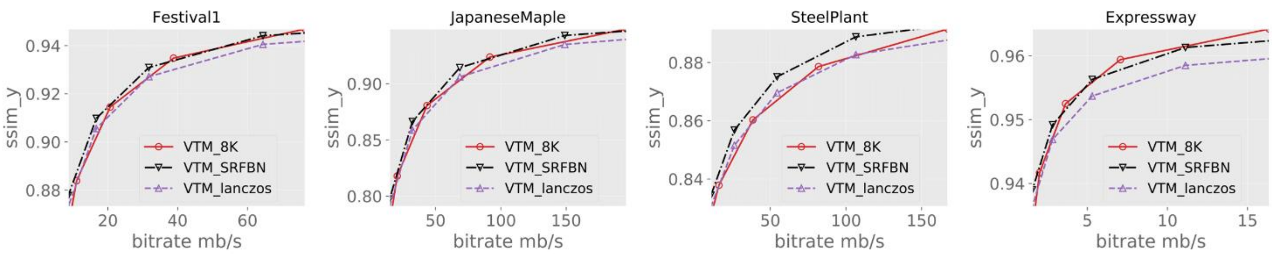

Figure 8 : SSIM

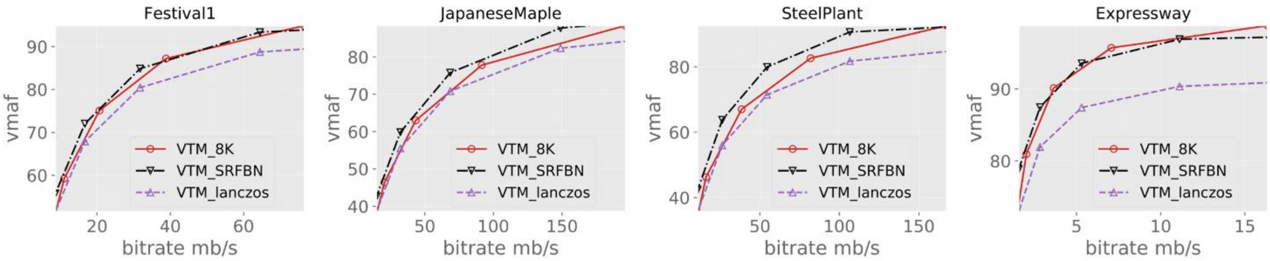

Figure 9 : VMAF

CONCLUSION

In this paper, we have presented and evaluated different coding tools that would enable efficient delivery of 8K services. First, single layer codecs VVC and HEVC were evaluated on the set of 8K sequences using rate-distortion curves and BD-rate scores. Then, a pre and post processing coding scheme was considered against single layer coding using VVC. We have shown that VVC is an important step for 8K transmission as it significantly reduces the bitrate required to encode such kind of signal for the same visual quality. Moreover, deep-learning tools show promising performance for this use-case as they give the benefit of delivering content at bitrate of lower resolution. However, the gain presented is far from the theoretical performance of the model on uncompressed data. This shows that there is a need for “real word” super-resolution models. However, the complexity of deep-learning-based tools is to be considered as it is added at the decoder-side. Future works will consider subjective evaluations to consolidate the results of objective evaluation.

REFERENCES

[1] ITU-R, *Recommendation BT.2020-1: Parameters Values of Ultra-High Definition Television Systems for Production and International Programme Exchange.*

[2] A. Ichigaya and Y. Nishida, "Required bit rates analysis for a new broadcasting service using HEVC/H. 265," *IEEE Transactions on Broadcasting,* vol. 62, p. 417–425, 2016.

[3] Y. Sugito, S. Iwasaki, K. Chida, K. Iguchi, K. Kanda, X. Lei, H. Miyoshi and K. Kazui, "A Study on the Required Video Bit-rate for 8K 120-Hz HEVC/H. 265 Temporal Scalable Coding," in *2018 Picture Coding Symposium (PCS)*, 2018.

[4] Y. Sugito, K. Iguchi, A. Ichigaya, K. Chida, S. Sakaida, H. Sakate, Y. Matsuda, Y. Kawahata and N. Motoyama, "HEVC/H. 265 codec system and transmission experiments aimed at 8K broadcasting," 2015.

[5] ESTI, *Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications; Part 2: DVB-S2 Extensions (DVB-S2X).*

[6] C. Dong, C. C. Loy, K. He and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence,* vol. 38, p. 295–307, 2015.

[7] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[8] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang and others, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.

[9] C. E. Duchon, "Lanczos filtering in one and two dimensions," *Journal of applied meteorology,* vol. 18, p. 1016–1022, 1979.

[10] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE transactions on acoustics, speech, and signal processing,* vol. 29, p. 1153–1160, 1981.

[11] F. Bossen, J. Boyce, K. Suehring, X. Li and V. Seregin, *JVET common test conditions and software reference configurations for SDR video,* 2019.

[12] F. Bossen, *Common test conditions and software reference configurations,* 2012.

[13] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon and W. Wu, "Feedback Network for Image Super-Resolution," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[14] E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017.

[15] A. A. Zhi Li, "Toward A Practical Perceptual Video Quality Metric," in *Netflix TechBlog*, 2016.

[16] G. Bjontegaard, *Document VCEG-M33: Calculation of Average PSNR Differences Between RD- Curves,* 2001.

[17] ITU-R, *Recommendation BT.910: Subjective video quality assessment methods for multimedia applications.*