Enhancing Quality for VVC Compressed Videos with Multi-Frame Quality Enhancement Model

1st Xiem HoangVan
Faculty of Electronics and Telecommunications
University Of Engineering and Technology
Vietnam National University
xiemhoang@vnu.edu.vn

2nd Huu-Hung Nguyen
Institute of System Integration

Le Quy Don Technical University
hungnh.isi@lqdtu.edu.vn

***Abstract*—Versatile Video Coding (VVC) is the most recent video coding standard, released in July 2020 with two major purposes: (1) providing a similar perceptual quality as the current state-of-the-art High Efficiency Video Coding (HEVC) solution at around half the bitrate and (2) offering native flexible, high-level syntax mechanisms for resolution adaptivity, scalability, and multi-view. However, despite of the compression efficiency, the decoded video obtained with VVC compression still contains distortions and quality degradation due to the nature of the hybrid block and transform based coding approach. To overcome this problem, this paper proposes a novel quality enhancement method for VVC compressed videos where the most advanced deep learning-based multi-frame quality enhancement model (MFQE) is employed. In the proposed QE method, the VVC decoded video is firstly segmented into the peak quality and non-peak quality pictures. After that, a Long-short term memory and two sub-networks are created to achieve better quality video pictures. Experimental results show that, the proposed MFQE based VVC quality enhancement method is able to achieve important quality improvement when compared to the original VVC decoded video.**

***Index Terms*—Versatile Video Coding, Multi-Frame Quality Enhancement, High Efficiency Video Coding**

I. Introduction

The most recent High Efficiency Video Coding (HEVC) was developed by jointly cooperating between the ITU-T Video Coding and the ISO/IEC Moving Picture Experts Group. In 2003, the first version of HEVC was released [1], providing 50% bit-rate reduction as compared to its predecessor, the H.264 / MPEG-4 Advanced Video Coding (AVC) standard [2]. Recently, the main increases in the reach and speed of broadband internet services leads to the share of video data traffic in global continuing growth. It is already around 80% and is increasing monotonously [3]. Additionally, the proportion of high-resolution 4K (3840×2160) TV is steadily growing, and these higher-resolution TVs request the higher quality video content for reaching their complete potential. Although a HEVC decoder is equipped with almost every 4K TV to playback high quality 4K video, the data rates required to deliver that video stream are still rather high. It is necessary for a new compression that is even more efficient than the current HEVC standard. To meet that needs, VCEG and MPEG are working together to develop a new video coding standard (Versatile Video Coding - VVC or H266) from 2018 with two main purposes: aiming at another 50% reduction in bit-rate together supporting and/or improving a wide range of additional functionalities. Even though VVC is designed to maintain the compressed video in high quality with additional coding tools. It still inevitably suffers from compression artifacts, which may lead to the decline in the quality of experience (QoE). Therefore, it is necessary to enhance the QoE of VVC compressed video/images at the decoder side.

Several works has been done for enhancing the visual quality of compressed images. Specifically, Liew et al. [4] introduced an over complete wavelet representation to enhance compressed images by reducing the blocking artifacts. The method proposed by Foi et al. [5] applied point-wise Shape-Adaptive Discrete Cosine Transform (SA-DCT) to reduce the blocking and ringing effects of JPEG compressed videos. Several sparse coding approaches were presented to remove JPEG compression artifacts [7], [8]. Moreover, JPEG image de-blocking was obtained by exploiting of Regression Tree Fields (RTF) [6].

For videos, several works have also been done to handle the compression artifacts for the HEVC videos by improving the coding efficiency at both the encoder and decoder sides. To deal with this issue, Rate-Distortion Optimization (RDO) at the encoder was proposed in [16]. The visual quality of HEVC videos was improved by adding an in-loop filter after the original one [9]. Applying Structure-driven Adaptive Nonlocal Filter (SANF) at both the encoder and decoder sides of HEVC was proposed by Zhang [10] which applied together with the Deblocking Filter (DF) as well as sample adaptive offset (SAO) filter.

Recently, applying deep learning has also been succeeded in improving the visual quality at both encoder and decoder sides. The re-trained SR-CNN [12] was replaced by the HEVC SAO filter to enhance the video quality. Later, the extension of AR-CNN [14], VRCNN [11], was proposed by Dai as an advanced in-loop filter in HEVC intra-coding without bit-rate increase. Nevertheless, it is necessary to modify the HEVC encoder, hence unsuitable for existing HEVC bitstreams. Most recently, Wang introduced a deep network at the decoder side to improve the visual quality of HEVC decoded videos, which possibly is applied to the existing video streams [13].

Yang et al. proposed quality enhancement convolutional neural network method named QE-CNN [17] to reduce the distortion of HEVC I and P/B frames. This author also developed Multi-Frame Quality Enhancement (MFQE) [18] including a Support Vector Machine (SVM) classifier to detect Peak Quality Frames (PQFs) and a novel Motion-Compensation subnet (MC-subnet) to compensate the temporal motion between neighboring frames. Zhenyu Guan [19] improved it by developing a novel Multi-Frame Convolutional Neural Network (MF-CNN) with a Bidirectional Long Short-Term Memory (BiLSTM) detector, called MFQE 2.0.

To the best of our knowledge, there is no method for the visual quality enhancement for versatile video coding streams. This paper proposes a deep learning approach for improving the quality of compressed videos using the recent VVC standard by applying MFQE 2.0 framework at the decoder side. The experimental results prove that the proposed approach is possible to enhance the visual quality of the compressed videos by VVC standard.

The rest of this paper is organized as follows. Section II presents a brief introduction VVC as well as MFQE. Section III describes the . The experimental results are given in Section IV. Finally, Section V concludes this paper.

II. Related work

A. Background work on Versatile Video Coding

This subsection briefly describes the versatile video coding standard (VVC). Several system functionalities for a wide range of applications are also added together with the standard:

- **Support for high resolutions from 4K to 16K video**: Larger and more flexible block structures are extended to support for higher resolutions together with a luma adaptive de-blocking filter designed for HDR video characteristics.
- **Support for 360-Degree Video**: VVC includes an efficient coding tool of immersive video including 360-Degree Video.
- **Computer Generated or Screen Content**: VVC is expected to extend the HEVC screen content coding by intra block copy (IBC) block-level differential pulse code modulation (BDPCM), adaptive color transform (ACT) and palette mode coding as well as full-sample adaptive MV precision.
- **Ultra-low Delay Streaming**: Gradual Decoder Refresh (GDR) integrated into VVC to avoid bit-rate peaks allows smooth the bit-rate of a bit-stream by distributing intra-coded slices or blocks in multiple pictures as opposed intra-coding entire pictures.
- **Resolution Allowance**: VVC supports an adaptive stream by taking advantage of reference picture re-sampling.
- **Scalability Support**: VVC also assists the multi-layer coding by using a single-layer-friendly approach.

The VVC encoding framework is depicted in 1. Although the same coding framework as HEVC standard is applied, a lot of novel coding tools are added in each module of VVC to further enhance the compression ratio summarized as follows:

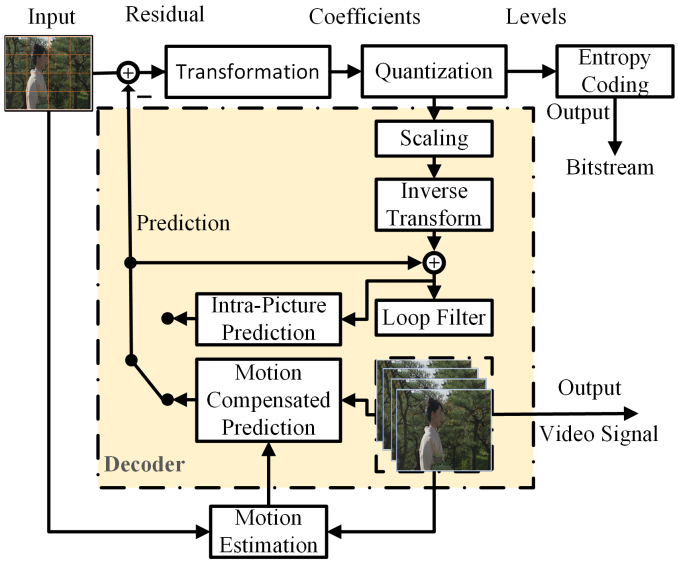

Fig. 1. Block diagram of VVC encoding framework

- **Block Partitioning**: The quad-tree with multiple partition unit types in HEVC is replaced by a quad-tree with nested multi-type tree using binary and ternary splits partitioning structure. Additionally, the maximum Coding Tree Unit size and transform length are increased to 128×128 and 64, respectively.
- **Intra-Picture Prediction**: Intra prediction from adjacent reference samples are obtained from neighboring blocks in the same image. VVC expands intra prediction by increasing to 67 types with 65 directional angular modes from 35 and 33 from HEVC, respectively together with the more planar, DC, cross-component linear model (CCLM) and matrix weighted intra prediction (MIP).
- **Inter-Picture Prediction**: Inter coding features from HEVC is modified by adding the two essential motion information coding methods: the merge mode and the advanced MV prediction (AMVP). Moreover, Sub-block based motion inheritance is introduced in VVC to divide the current CU into sub-blocks with equal size. Additionally, VVC introduces MV refinement and bi-directional optical flow methods to improve the motion compensation result in the further enhancement of prediction quality.
- **Transforms and Quantization**:The better energy compaction for the residual signals of large sized smooth areas thanks to the extension the maximum transform size to 64×64. To obtain better energy compaction of the residual signals, multiple transform cores using a predefined subset of sinusoidal transforms is applied in VVC with the transform selection signaled at CU level.
- **Entropy Coding**: Compared to the CABAC design in HEVC, two major changes are included in VVC: 1) A binary arithmetic encoder applied with the high accuracy multi-hypothesis probability estimate 2) Transform coefficients coding with improved context modeling, and

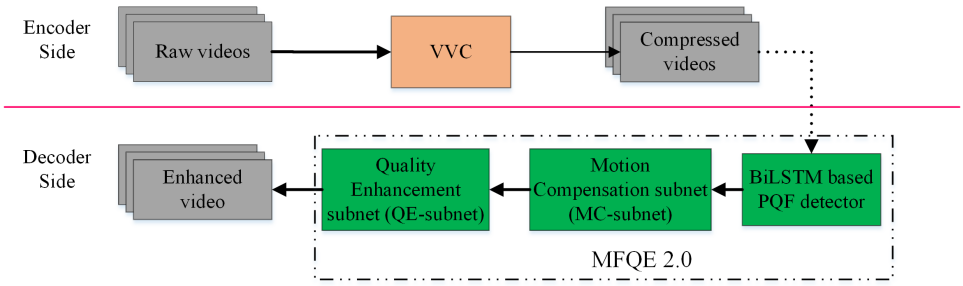

Fig. 2. The proposed quality enhancement for VVC videos

coding combination of chroma residuals of both Cb and Cr components.

- **In-Loop Filter**: The recent VVC standard designs an adaptive loop filter (ALF) after deblocking and SAO filters to reduce the potential distortion introduced by quantization and transform process. It also introduces a new luma mapping with chroma scaling (LMCS) that modifies the input signal before deblocking.

B. Multi-frame Quality Enhancement

As mentioned above, several quality enhancement approaches possibly are applied to the video streams at decoder side. Yang proposed Quality Enhancement Convolutional Neural Network (QE-CNN) method [17] for reducing the distortion of both I and P/B frames. This is first CNN model developed to determine distortion features of intra/inter coding result in enhancing effectively the quality of both I and P/B frames. A multiple model including QE-CNN-I and QE-CNN-P is designed to concatenate the intra- and inter-coding distortion features to enhance the compression quality. To meet the real-time requirement, a time-constrained quality enhancement optimization scheme was proposed to trade-off the computational complexity and the QoE. Yang also utilized the existence of large quality fluctuation across frames to develop MFQE 1.0 [18] for compressed video with three main components: SVM based PQF detector, MC-subnet and QE-subnet. A Support Vector Machine classifier was trained to detect Peak Quality Frames (PQFs) by extracting 36 spatial one-dimensional features from the five consecutive frames. MC-subnet was designed to compensate for the temporal motion existing between non-PQFs and PQFs across frames. The architecture of our MC-subnet was developed based on Spatial Transformer Motion Compensation (STMC) method for multi-frame super-resolution [23]. The compensated PQFs of MC-subnet can be enhanced by third component QE-subnet. Its architecture is designed to handle the spatio-temporal information. Zhenyu Guan [19] improved MFQE by replacing SVM based PQF detector by another detector based Bidirectional Long Short-Term Memory (BiLSTM) and adding the multi-scale strategy, batch normalization as well as dense connection to QE-subnet, called MFQE 2.0.

III. Proposed VVC Quality Enhancement

In this section, we present a deep learning approach to improving the VVC compressed videos, namely MFQE_VVC. As shown in [19], MFQE 2.0 successfully applied to the HEVC video to increase the PSNR around 0.5dB on average. It achieves the state-of-the-art quality enhancement performance for compressed images. In the most recent VVC standard, several coding tools are added with an expectation of enhancing the visual quality and reducing the bit-rate. Even though, the visual quality of VVC compressed video increases compared to that of HEVC, it still un-avoid compression artifacts. To obtain the higher visual quality at the decoder side, a quality enhancement method for VVC is necessary to develop, obviously. To the best of our knowledge there is no solution for this problem of recent VVC standard. To deal with this issue, we take advantage capability of MFQE 2.0 in quality enhancement that will be applied to the HEVC videos successfully.

A. Proposed MFQE_VVC Framework

Our proposed approach is described specifically in Fig.2. We focus on improving the compressed video coded by recent VVC standard. At encoder side, the raw video go through VVC to be the compressed video with size reduction. At decoder side, the compressed videos as input of the quality enhancement framework including BiLSTM based PQF detector, MC-subnet and QE-subnet. The BiLSTM network detect PQFs without reference that extract the long- and short-term correlation between PQF and non-PQF. The role of MC-subnet is to compensate for the temporal motion of between the current non-PQF and its nearest PQFs in advance. As mentioned above, QE-subnet architecture is designed to handle the spatio-temporal information, therefore the quality of the compressed video is enhanced. The performance of our approach is validated in the next section.

B. Multi-frame Quality Enhancement Model

Note that, the quality of compressed videos fluctuate across frames dramatically. In general, a compressed video is a sequence of images including key-frame in which its quality is higher than that of neighboring frame. Therefore, the high-quality frames (PQFs) are used to enhance the quality of their neighboring frames in low-quality (non-PQFs) possibly.

In video quality enhancement, raw images are unavailable so PQFs and non-PQFs are unknown. The effectiveness and generalization ability of MFQE approach in advancing the state-of-the-art compressed video quality enhancement method thanks to three essential techniques: a BiLSTM-based PQF/non-PQF detector for extracting the dependencies from both backward and forward directions, a MC-subnet for the temporal motion across frames and QE-subnet for quality enhancement. BiLSTM network extracts the "long- and short-term correlation" between PQF and non-PQF based the quality fluctuation frequent appearance in compressed video. 38 features are extracted for each image are the input to BiLSTM in form of a 38-dimension vector including number of assigned bits, quantization parameters and 36 features at pixel domain extracted by the non-reference quality assessment method [20]. Since PQFs are detected, the quality of non-PQFs was enhanced by taking advantage of their neighboring PQFs. Several coding tools have been added into VVC standard to enhance the compressed frames, therefore the quality of both PQFs and non-PQFs increases. MC-subnet based on the CNN method of Spatial Transformer Motion Compensation was developed to compensate for the temporal motion between neighboring frames. To handle large scale motion, STMC estimated the two levels down-scale $x4$ and $x2$ motion vector maps. However, the down-scale leads to the accuracy reduction of motion vector estimation. For that reason, in addition to STMC, MC-subnet developed a concatenation layer for pixel-wise motion estimation that is a convolutional layer concatenating non-PQF and PQF. The input of QE-subnet includes the compensated previous and subsequent PQFs as well as the non-PQFs. The spatial and temporal features of these three frames are extracted and fused to gain advantageous information in the adjacent PQFs for enhancing the quality of the non-PQFs.

IV. Performance Evaluation

This section presents the experimental results to validate the effectiveness of our proposed approach, MFQE_VVC. Different from previous method proposing MFQE 2.0 on HEVC compressed video, in this paper we apply MFQE 2.0 on VVC compressed video. Specifically, we evaluate the quality enhancement performance of $MFQE\_VVC$ method in terms of △PSNR (Peak signal-to-noise ratio), which calculates the PSNR difference between the enhanced and original compressed sequences via the mean squared error (MSE). In this case, the signal is the original data, and the noise is the error introduced by compression. We evaluate the proposed approach on 8 standard test sequences created by Joint Collaborative Team on Video Coding (JCT-VC) [21] with different setting the Quantization Parameters (QPs) to 22, 27, 32 and 37, respectively. These sequences are video high definition including HD and FHD resolution. The experimental results show that MFQE_VVC improves the visual quality of the compressed video. Currently, we use the optimal setting of MFQE for HECV videos for testing the VVC videos. That mean, we keep all the training and testing parameters such as epoch number, learning rate from MFQE 2.0 method. The LSTM length is set to 8. The input of MF-CNN network is the raw and compressed videos segmented into 64×64 patches.The batch size is set to be 128. The initial learning rate for MC-subnet is set oversize 0.001. Two parameters a, b of QE-subnet are set $a = 0.01$ and $b = 1$, respectively in order that the QE-subnet can converge fast.

To show the quality enhancement performance, the △PSNR is defined by the following equation

$$\triangle PSNR = PSNR_{MFQE\_VVC} - PSNR_{VVC} \quad (1)$$

Firstly, we measure △PSNR of our quality improvement method for 8 test sequences. The results for different QPs are shown detail in Table I. The results in this table indicate that the PSNR increases 0.20 dB on average. It also reveals the increase trend along different QPs for all tested sequences. In general, the higher PSNR gain is at QP 37 for each sequence.

TABLE I
PSNR GAIN WITH THE PROPOSED APPROACH

Sequence	QP			
	37	32	27	22
PeopleOnStreet_2560x1600_150	**0.246**	0.158	0.123	0.077
Kimono_1920x1080_240	**0.203**	0.143	0.135	0.094
ParkScene_1920x1080_240	**0.25**	0.214	0.237	0.144
PartyScene_832X480_500	**0.264**	0.196	0.264	0.233
RaceHorses_832x480_300	0.046	**0.068**	0.046	0.025
BasketballPass_416X240_500	**0.299**	0.297	0.299	0.269
BlowingBubbles_416x240_500	**0.41**	0.317	0.41	0.395
RaceHorses_416x240_300	0.169	0.169	**0.174**	0.168
Average	**0.236**	0.196	0.211	0.175

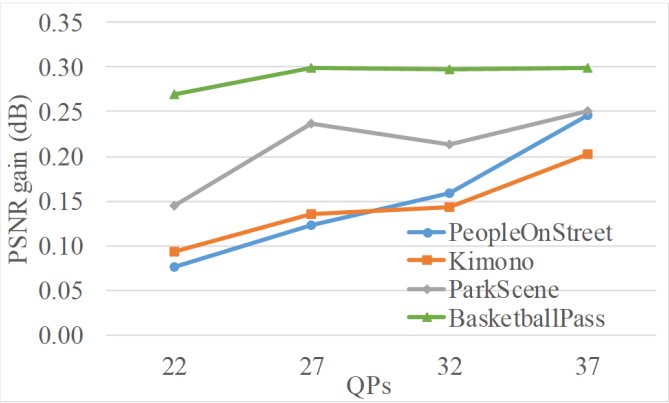

Fig. 3. PSNR curves of for video sequence enhanced by MFQE_VVC

To imagine easily the increasing trend of the different PSNR, we visualize the results in the table for four sequences PeopleOnStreet, Kimono, ParkScene and BasketballPass along

with the QPs in Fig. 3. The PSNR gains for four sequences monotonously increase in general. Four chosen sequences including all motions may lead to quality degradation such as slow motion, fast motion, complex motion...etc. The PSNR gain with the proposed method is around 0.2 dB that is similar to that of the method exploiting temporal structure and spatial details for VVC compressed videos [24],

The results in Table I and Fig. 3 verify that our proposed approach can enhance the visual quality of the VVC compressed video.

V. Conclusion

This paper introduced a quality improvement approach for VVC compressed video. The novelty of our approach lies in utilizing the multiple frame quality enhancement based deep learning framework for the recent VVC standard. The performance comparison to the origin VVC in experimental results reinforce the applicability of the proposed method. In the future, we plan to redesign PQF detector and MC-subnet more suitable for VVC compressed video to further increase the visual quality. And then, more video sequences are extended to evaluate.

Acknowledgment

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2020.15.

References

[1] ITU-T and ISO/IEC JTC 1, High Efficiency Video Coding, Rec. ITU-T H.265 and ISO/IEC 23008-2 (HEVC), April 2013 (and subsequent editions).

[2] ITU-T and ISO/IEC JTC 1, Advanced Video Coding for generic audio-visual services, Rec. ITU-T H.264 and ISO/IEC 14496-10 (AVC), May 2003 (and subsequent editions).

[3] Cisco Systems, "Cisco Visual Networking Index: Forecast and Trends, 2017–2022", Cisco Systems White Paper, 2019.

[4] A.-C. Liew and H. Yan, "Blocking artifacts suppression in block-coded images using overcomplete wavelet representation," IEEE TCSVT, pp. 450–461, 2004.

[5] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," IEEE TIP, 2007.

[6] J. Jancsary, S. Nowozin, and C. Rother, "Loss-specific training of non-parametric image restoration models: A new state of the art," in ECCV,2012.

[7] H. Chang, M. K. Ng, and T. Zeng, "Reducing artifacts in JPEG decompression via a learned dictionary," IEEE TSP, pp. 718–728, 2014.

[8] C. Jung, L. Jiao, H. Qi, and T. Sun, "Image deblocking via sparse representation," Signal Processing: Image Communication, pp. 663–677,2012.

[9] Q. Han and W.-K. Cham, "High performance loop filter for HEVC," in ICIP, 2015.

[10] J. Zhang, C. Jia, N. Zhang, S. Ma, and W. Gao, "Structure-driven adaptive non-local filter for high efficiency video coding (HEVC)," in DCC, 2016.

[11] W.-S. Park and M. Kim, "CNN-based in-loop filtering for coding efficiency improvement," in IVMSP, 2016.

[12] Y. Dai, D. Liu, and F. Wu, "A convolutional neural network approach for post-processing in HEVC intra coding," in MMM, 2017.

[13] T. Wang, M. Chen, and H. Chao, "A novel deep learning-based method of improving coding efficiency from the decoder-end for hevc," in Data Compression Conference, 2017.

[14] C. Dong, Y. Deng, C. Change Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in ICCV, 2015.

[15] C. Dong, C. L. Chen, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in ECCV, 2014.

[16] S. Li, C. Zhu, Y. Gao, Y. Zhou, F. Dufaux, and M.-T. Sun, "Lagrangian multiplier adaptation for rate-distortion optimization with inter-frame dependency," IEEE TCSVT, pp. 117–129, 2016.

[17] R. Yang, M. Xu, Z. Wang, and Z. Guan. Enhancing quality for HEVC compressed videos. arXiv preprint arXiv: 1709.06734, 2017.

[18] R. Yang, M. Xu, Z. Wang, and T. Li, "Multi-frame quality enhancement for compressed video," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6664–6673.

[19] Z. Guan, Q. Xing, M. Xu, R. Yang, T. Liu and Z. Wang, "MFQE 2.0: A New Approach for Multi-frame Quality Enhancement on Compressed Video," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2019.2944806.

[20] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," IEEE transactions on image processing, vol. 19, no. 6, pp. 1427–1441, 2010.

[21] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standardsincluding high efficiency video coding (hevc)," IEEE Transactions on circuits and systems for video technology, vol. 22, no. 12, pp. 1669– 1684, 2012.

[22] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[23] J. Caballero, C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, and W. Shi. Real-time video super-resolution with spatio-temporal networks and motion compensation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[24] X. Meng, X. Deng, S. Zhu and B. Zeng, "Enhancing Quality for VVC Compressed Videos by Jointly Exploiting Spatial Details and Temporal Structure," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 1193-1197, doi: 10.1109/ICIP.2019.8804469.