



Video Compression using Neural Weight Step and Huffman Coding Techniques

Faten Ahmed Jebur Al-Sudani

Department of Computer Systems Techniques, Institute of Administration Rusafa, Middle Technical University, fatenahmed152@gmail.com, Baghdad, Iraq.

*Corresponding author email: fatenahmed152@gmail.com
; mobile: 07702505122

ضغط الفيديو باستخدام خطوة الوزن العصبي وتقنيات ترميز هوفمان

فاتن احمد جبر السوداني*

قسم تقنيات أنظمة الحاسوب، معهد الإدارة الرصافة، الجامعة التقنية الوسطى، fatenahmed152@gmail.com، بغداد، العراق

Received: 14/12/2022 Accepted: 7/5/2023 Published: 30/6/2023

ABSTRACT

Background:

This paper proposes a Hierarchical Video Compression Scheme (HVCS) method with three hierarchical layers of quality with Recurrent Quality Enhancement (RQEN) network. Image compression techniques are used to compress frames in the first layer, where frames have the highest quality. Using high-quality frame as a reference, the Bi-Directional Deep Compression (BDC) network is proposed for frame compression in the second layer with considerable quality. In the third layer, low quality is used for frame compression using adopted Single Motion Compression (SMC) network, which proposes the single motion map for motion estimation within multiple frames. As a result, SMC provide motion information using fewer bits. In decoding stage, a weighted Recurrent Quality Enhancement (RQEN) network is developed to take both bit stream and the compressed frames as inputs. In RQEN cell, the update signal and memory are weighted using quality features to positively influence information of multi-frame for enhancement. In this paper, HVCS adopts hierarchical quality to benefit the efficiency of frame coding, whereas high-quality information improves frame compression and enhances the low-quality frames at encoding and decoding stages, respectively. Experimental results validate that proposed HVCS approach overcomes the state-of-the-art of compression methods.

Materials and Methods:

Tables 1& 2 illustrate representing yielded values for rate-distortion on both video datasets. As aforementioned, PSNR and MS-SSIM are used for quality evaluation, where bit-rates are calculated using bits per pixel (bpp). Table 1 illustrates PSNR performance, where they show better PSNR performance for the proposed compression model than other methods such as Chao et al [7] or optimized methods [1]. In addition, they outperform applying H.265 on standard JCT-VC dataset. On the other side, proposed compression scheme yielded better bit-rate performance than applying H.265 on UVG. As in Table 2, the MS-SSIM evaluation provided better performance of proposed scheme than all other learned approaches, where it reached better performance than H.264 and H.265. Due to bit-rate performance on UVG, Lee et al. [11] has comparable performance, and Guo et al [10] yielded lower performance than H.265. Applying on JCT-VC, DVC [10] is only comparable with H.265. On the opposite, the performance of HVCS-rate-distortion have obvious better performance than H.265.

Furthermore, BjØntegaard Delta Bit-Rate (BDBR) [47] is also computed depending on H.265. A BDBR measure computes the average difference of bit-rate considering H.265 anchor, where better performance is indicated on lower values of BDBR [48]. In Table 3, BDBR performance is illustrated depending on PSNR and MS-SSIM, in which, bit-rate reduction considering the anchor is indicated by showed negative numbers. Such results outperform H.265 performance, where bold numbers represent best



yielded results by learned methods. Table 3 provided a fair comparison on (MS-SSIM & PSNR) optimized techniques DVC [10] considering the anchor H.265.

Results:

Tables 1 & 2 illustrate representing yielded values for rate-distortion on both video datasets. As aforementioned, PSNR and MS-SSIM are used for quality evaluation, where bit-rates are calculated using bits per pixel (bpp). Table 1 illustrates PSNR performance, where they show better PSNR performance for the proposed compression model than other methods such as Chao et al [7] or optimized methods [1]. In addition, they outperform applying H.265 on standard JCT-VC dataset. On the other side, proposed compression scheme yielded better bit-rate performance than applying H.265 on UVG. As in Table 2, the MS-SSIM evaluation provided better performance of proposed scheme than all other learned approaches, where it reached better performance than H.264 and H.265. Due to bit-rate performance on UVG, Lee et al. [11] has comparable performance, and Guo et al [10] yielded lower performance than H.265. Applying on JCT-VC, DVC [10] is only comparable with H.265. On the opposite, the performance of HVCS-rate-distortion have obvious better performance than H.265.

Furthermore, BjØntegaard Delta Bit-Rate (BDBR) [47] is also computed depending on H.265. A BDBR measure computes the average difference of bit-rate considering H.265 anchor, where better performance is indicated on lower values of BDBR [48]. In Table 3, BDBR performance is illustrated depending on PSNR and MS-SSIM, in which, bit-rate reduction considering the anchor is indicated by showed negative numbers. Such results outperform H.265 performance, where bold numbers represent best yielded results by learned methods. Table 3 provided a fair comparison on (MS-SSIM & PSNR) optimized techniques DVC [10] considering the anchor H.265.

Conclusion:

This work proposes a learned video compression scheme utilizing the hierarchical frame quality with recurrent enhancement. Specifically, this work proposes dividing frames into hierarchical levels 1, 2 and 3 in decreasing quality. For the first layer, image compression methods are proposed, while proposing BDC and SMC for layers 2 and 3 respectively. RQEN network is developed with frame quality compressed frames and bit-rate information as inputs for multi-frame enhancement. Experimental results validated the efficiency of proposed HVCS compression scheme.

Similarly with other compression techniques, frame structure is manually set the in this scheme. A promising recommendation for future work can be accomplished by developing DNN networks which are automatically learned for the prediction and hierarchy.

Key words:

Hierarchical Video Compression, Recurrent Quality Enhancement, compress frames, Bi-Directional Deep Compression, adopted Single Motion Compression, Huffman coding.



الخلاصة

مقدمة:

تقترح هذه الورقة طريقة مخطط ضغط الفيديو الهرمي (HVCS) مع ثلاث طبقات هرمية من الجودة مع شبكة تحسين الجودة المتكررة (RQEN). تستخدم تقنيات ضغط الصور لضغط الإطارات في الطبقة الأولى، حيث تتمتع الإطارات بأعلى جودة. باستخدام إطار عالي الجودة كمرجع، تم اقتراح شبكة الضغط العميق ثنائي الاتجاه (BDC) لضغط الإطار في الطبقة الثانية بجودة كبيرة. في الطبقة الثالثة، يتم استخدام جودة منخفضة لضغط الإطار باستخدام شبكة ضغط الحركة الواحدة (SMC) المعتمدة، والتي تقترح خريطة الحركة الواحدة لتقدير الحركة داخل إطارات متعددة. نتيجة لذلك، يوفر SMC معلومات الحركة باستخدام عدد أقل من البتات. في مرحلة فك التشفير، يتم تطوير شبكة تحسين الجودة المتكررة (RQEN) المرجحة لأخذ كل من تدفق البتات والإطارات المضغوطة كمدخلات. في خلية RQEN، يتم ترجيح إشارة التحديث والذاكرة باستخدام ميزات الجودة للتأثير بشكل إيجابي على معلومات الإطارات المتعددة ...

طرق العمل:

يوضح الجدولان 1 و 2 تمثيل القيم الناتجة لتشويه المعدل في مجموعتي بيانات الفيديو. كما ذكرنا سابقاً، يتم استخدام PSNR و MS-SSIM لتقييم الجودة، حيث يتم حساب معدلات البتات باستخدام بت لكل بكسل (bpp). يوضح الجدول 1 أداء PSNR، حيث يظهرون أداء PSNR أفضل لنموذج الضغط المقترح من الطرق الأخرى مثل [7] Chao et al أو الطرق المحسنة [1]. بالإضافة إلى ذلك، يتفوقون في تطبيق H.265 على مجموعة بيانات JCT-VC القياسية. على الجانب الآخر، أسفر مخطط الضغط المقترح عن أداء معدل بت أفضل من تطبيق H.265 على UVG. كما هو الحال في الجدول 2، قدم تقييم MS-SSIM أداء أفضل للمخطط المقترح من جميع النهج المستفاد الأخرى، حيث وصل إلى أداء أفضل من H.264 و H.265. نظراً لأداء معدل البت على UVG، يتمتع Lee et al. [11] بأداء مماثل، وحقق Guo et al [10] أداء أقل من H.265. التقديم على JCT-VC، [10] DVC يمكن مقارنته فقط ب H.265. على العكس من ذلك، فإن أداء تشويه معدل HVCS له أداء أفضل واضح من H.265. علاوة على ذلك، يتم حساب معدل بت دلتا (BDBR) [47] Bjøntegaard أيضاً اعتماداً على H.265. بحسب مقياس BDBR متوسط الفرق في معدل البت مع الأخذ في الاعتبار مرسة H.265، حيث يشار إلى أداء أفضل على القيم المنخفضة ل [48] BDBR. بحسب مقياس BDBR متوسط الفرق في معدل البت مع الأخذ في الاعتبار مرسة H.265، حيث يشار إلى أداء أفضل على القيم المنخفضة ل [48] BDBR. في الجدول 3، يتم توضيح أداء BDBR اعتماداً على PSNR و MS-SSIM، حيث يشار إلى تخفيض معدل البتات بالنظر إلى المرسة بأرقام سالبة معروضة. تتفوق هذه النتائج على أداء H.265، حيث تمثل الأرقام الجريبة أفضل النتائج التي تم تحقيقها من خلال الأساليب المستفاد. قدم الجدول 3 مقارنة عادلة حول تقنيات DVC المحسنة [10] (MS-SSIM & PSNR) مع الأخذ في الاعتبار المرسة H.265.

الاستنتاجات:

يقترح هذا العمل مخطط ضغط فيديو مستفاد باستخدام جودة الإطار الهرمي مع التحسين المتكرر. على وجه التحديد، يقترح هذا العمل تقسيم الإطارات إلى مستويات هرمية 1 و 2 و 3 في انخفاض الجودة. بالنسبة للطبقة الأولى، يتم اقتراح طرق ضغط الصور، مع اقتراح BDC و SMC للطبقات 2 و 3 على التوالي. تم تطوير شبكة RQEN بإطارات مضغوطة بجودة الإطار ومعلومات معدل البت كمدخلات لتحسين الإطارات المتعددة. أثبتت النتائج التجريبية كفاءة مخطط ضغط HVCS المقترح. وبالمثل مع تقنيات الضغط الأخرى، يتم تعيين هيكل الإطار يدوياً في هذا المخطط. يمكن تحقيق توصية واعدة للعمل المستقبلي من خلال تطوير شبكات DNN التي يتم تعلمها تلقائياً للتنبؤ والتسلسل الهرمي.

الكلمات المفتاحية: ضغط الفيديو الهرمي، تحسين الجودة المتكررة، إطارات الضغط، الضغط العميق ثنائي الاتجاه، ضغط الحركة الواحدة، ترميز هوفمان.



INTRODUCTION

In last decade, exchanging videos across the World Wide Web (WWW) witnessed wider propagation. About (70- 80) % of data traffic in mobile applications is currently consumed, due to Cisco Forecast, by video exchange [1]. Such consumption is increasingly caused by high quality (video resolution). To utilize bandwidth-limited Internet for higher-efficiency videos transmission with high quality, videos must be, necessarily, compressed with higher-performance compression techniques. Over the recent decades, researchers proposed plenty standards for video compression, like H.265 [2] and H.264[3], and many others[4]. Yet, such compression codecs were traditionally handcrafted, which decreases the opportunity of optimizing them depending on end-to-end manners.

Recent image and video compression studies adopting Deep Learning (DL) expressed considerable potential of DL to improve compression performance regarding rate-distortion[5][6]. Therefore, it was expectable that interest, in DL based video compressing techniques, is increasing[7][8][9][10][11]. For example, an optical flow was adopted in motion compensation, where auto encoders were adopted for the compression of flow and residual[10]. After that, a new video compression technique was proposed using 3D auto-encoder combined by autoregressive prior [11]. Such methods trained their proposed models using one loss function, which was applied to all of the frames. As a result, layers were not considered due to the hierarchical quality, which utilizes frames with high quality in the compression algorithm and then in post-processing other frames.

In this paper, three layers regarding a hierarchical quality are adopted with recurrent enhancement network to build the proposed Hierarchical Video Compression Scheme (HVCS). As Figure 1 illustrates, in the three layers 1, 2 and 3, frames are handled using hierarchical levels of quality high, medium and the low, respectively. Hierarchical- quality compression provides two types of benefits: Firstly, frames with high quality enhance the performance of compression process for other frames in coding stage, and at the same time they provide references with high quality. Secondly, frames with high quality provide advantageous information to enhance the decoding process of other frames with low quality, which can be utilized due to high correlations between adjacent frames.

Such improvements enhance the quality without exaggerating the bit-rate, which improves the performance of rate- distortion. As example in Figure 1, compression of frames F3 and F8 (from layer 3) is accomplished using low quality and bit-rate. After that, the proposed recurrent enhancement network applies significant improvements on their quality, utilizing other high-quality frames like frames (f0 and f5). As a result, f3 and f8 provide considerable quality comparing with the frame (f5) from layer 2, but with less bit-rate consumption. This indicates considerable efficiency provided by our video compression scheme.

In this work, a compression method is used for layer 1 compression. Then for layer 2, a predictive video compression technique is proposed (Bi-Directional Compression (BDC)), where compressed frames in layer 1 are considered as references for bi-directional compression. Then, utilizing how adjacent frames are related, this work compresses layer 3 frames using our



compression network (Single Motion Compression (SMC)). Such network adopts single motion maps for detecting a motion within set of frames, which reduces bit-rates to encode motion maps. Finally, this work enhances the proposed algorithm by developing the weighted Recurrent Quality Enhancement (RQEN) network depending on [12], where the authors weighted recurrent cells by quality features for applying information of multiple frames for recurrent enhancement. Experimental results showed that proposed HVCS scheme yields a considerable performance as due to the published state-of-the-art in video compression researches, which provided higher performance than other techniques.

I. Related works

In recent decades, plenty researchers proposed different standards for image compression like Better Portable Graphics (BPG) [13], Joint Photographic Experts Group (JPEG) [2], and JPEG 2000 [14]. Then, Deep Neural Networks (DNNs) were also been significantly adopted to enhance the performance of compression techniques [5][6][15][16][17][18][19][20][21][22][23]. Several end-to-end DNN schemes were proposed for deep image compression in [5][6], where they applied the factorized-prior and hyper-prior models respectively for entropy estimation. Besides, other researchers proposed recurrent structures as image compression techniques [16][17][22]. Then entropy models were proposed for more enhancements in the performance of rate-distortion such as hierarchical prior model and context-adaptive model as in [21][23] respectively. Due to their-time state-of-the-art, their compression schemes outperformed published compression schemes. Depending on published standards in image compression, different standards for video compression were proposed like H.265 [2], H.264 [3] MPEG [24] many others [4]. Recently, DNN networks attracted increasing attention to build efficient video compression algorithms. Different researchers tended to modify their approaches traditional video algorithms to DNN ones [25][26][27][28][29]. As example, Jiaying et al. [26] modified the motion compensation by utilizing a DNN to enhance the fractional interpolation, while DNNs were utilized in improving in-loop filters [29][8][25]. Yet, such methods could have performance improvements in each particular module individually, where the ability of optimizing such compression frameworks jointly was unavailable.

Using end-to-end concepts, DNN was also adopted to propose several video compression techniques [1][7][9][11][8]. Specifically, Wu et al. [7] predicted the frames by depending on reference frames for interpolation, and they adopted the proposed compression scheme by George, et al [16] for residual compression. Guo et al. [10] proposed a DNN Compression method, which used optical flow for temporal-motion estimate, where they adopted 2 auto-encoders for motion and residual compression. Meanwhile, Zhengxue, et al [9] added energy compaction (spatial temporal) to the loss function to enhance compression-scheme performance. An auto-encoder was proposed for rate-distortion, which handle the video entropy using an autoregressive prior for coding [11]. Due to the best of our knowledge, there was only one published research which adopts hierarchical prediction [7]. Yet we didn't find a research that used learns for video compression regarding a hierarchical quality. As a result, they didn't provide references with high quality to



compress the other frames, which may not utilize the significant information provided by the post processing of multiple frames.

It's inevitable that any algorithm of lossy compression degrades the video quality and artifacts, therefore different researchers focused on improving compressed-video quality[12][30][31][32][33][34][35]. Among these works, the authors of[33][34][35]. Built their schemes on single-frame issue with one-frame input each time. Then, a multi-frame compression schemes with quality enhancement were proposed by Yang et al.[36][12], which utilized correlation between inter-frames. In addition, Guo, et al [35] proposed using Deep Kalman (DK) filter for compression-artifacts reduction.

Nevertheless, all aforementioned techniques were designed by adopting the post-processing concept as video coding technique. As a result, multi-frame compression techniques[36][12], it was not easy to obtain high accuracy for frame quality, but it was estimated under acceptable prediction error. In the proposed HVCS compression scheme and for each frame, the compression quality is encoded into the result stream of bits, which will be an input to the enhancement network combined by compressed frames. Such that, improvements are supported by the quality of high-accuracy frames and will be utilized by the deep decoder of the compression scheme.

Materials and Methods

• Subjects

a. System Framework

The proposed HVCS framework is shown in Figure 2 applied on the first Picture Group (PG), where it's similarly applied on each other PG. HVCS scheme compresses the frames considering (3) hierarchical layers of quality (layer1, layer2 and layer3) in descending order of quality as in Figure 2.

Layer1. Frames in first layer (single-line gray-blocks) are encoded by image compression method (using BPG [13] and Jooyoung et al. [22]for PSNR and MS-SSIM respectively), and y_i^c represents a compressed frame of order (i). As for "I-frames" in traditional methods of coding [3][2], the highest bit-rates are consumed on compressed frames in layer 1 providing the highest quality for compressed frames. Such that, they provide the ability of stopping error propagation while encoding/ decoding the video. Furthermore, high-quality information provided by such frames supports the enhancement and compression of adjacent frames.

Layer2. In the second layer (2), the frames (double-line light brown-blocks) are found between 2 frames from layer1. A BDC network is proposed to compress frames of layer 2, which considers previous and next frames (compressed from layer1) as bi-directional references. Medium-quality technique is used to compress layers from layer2, and to provide supporting information for enhancing and compressing a low-quality frame from layer 3. Section 3.2 illustrates more detailed information about BDC.



Layer3. Frames, not compressed in Layer1 & Layer2, in layer 3 (triple-line green-blocks), will be compressed using low-level quality contributing the lowest level of bit-rate. In recent techniques of DNN video compression, one-motion map (at least) was required to compensate motion for each frame [7][22]. Due to the motion-correlations between successive frames, therefore redundancy is led to where one motion map is encoded for each related frame. Therefore, an SMC network is proposed in this work, in which a single motion map is applied for describing motions between successive frames, which leads to bit-rate reduction. It's noticeable that the frames (y_6 till y_n) have the same manner of compression as for (y_1 till y_4) frames, so they are excluded (hidden) in Figure 2. Section 3.3 introduces the SMC network. manner of compression as for (y_1 till y_4) frames, so they are excluded (hidden) in Figure 2. Section 3.3 introduces the SMC network.

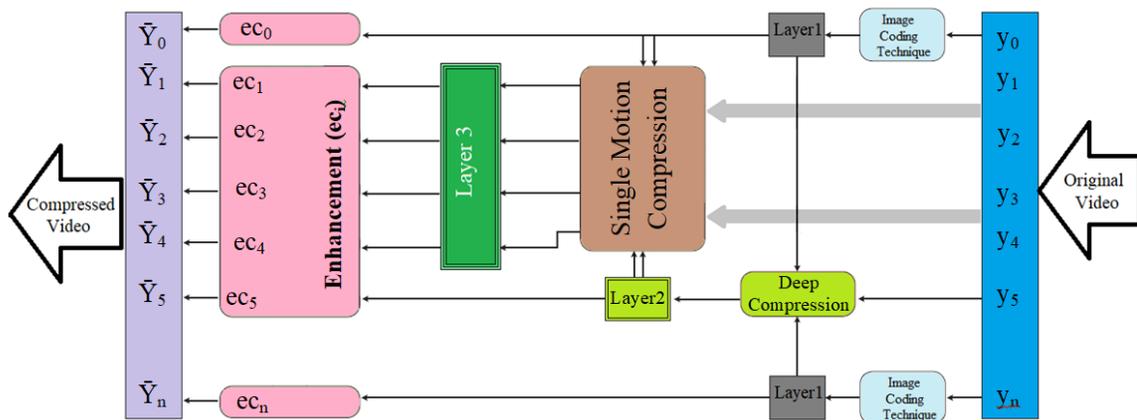


Figure 2: General diagram of proposed HVCS approach explains video compression due to hierarchical layers of quality adopting BDC and SMC networks (employing RQEN in video decoding).

Because of high correlations between successive video frames [33], a weighted Recurrent Quality Enhancement (RQEN) network were developed and adopted in this work. It weights recurrent cells with quality features for reasonable leverage of multi-frame information. In particular, frame-quality, in layer3, can have significant enhancements by utilizing information from high-quality frames in layer1 and layer2. In this case, additional information is not required to be kept to enhance the quality, which means less bit-rate, specifically when frames have low quality. RQEN is considered as a part of the proposed decoder, where quality information combined with compressed frames (encoded in the bit stream) represent the inputs of decoder side. Section 3.4 discusses more details about the RQEN.

مجله جوباس للعلوم التطبيقية والعلوم الحاسوبية، جامعة بابل، العراق، 2023، المجلد 31، العدد 2، الصفحة 65-70

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com | ISSN: 2312-8135 | Print ISSN: 1992-0652

b. Bi-directional Deep Compression (BDC)

The BDC network used to compress layer2 frames is illustrated in Figure 3, where the 1st PG was also used as an example. Firstly, the subnet Motion Estimation (ME) was firstly used in BDC to detect temporal motions between target and reference frames. Since there is long interval between layer1 and layer2 frames (5 frames as in Fig 3), Anurag and Michael [35] was followed for applying a hierarchical network. Such pyramid handles large motions, which also utilizes the wide receptive field. It's noted in this approach that backward warping is used, and therefore backward motions were estimated. In Figure 3, for example, adopted- subnet (ME) outputs are the motions from y_5 to y_0^c (referred to as $f_{5 \rightarrow 0}$) and from y_5 to y_{10}^c (referred to as $f_{5 \rightarrow 10}$), respectively.

Considering estimated motions, we utilize an auto-encoder for Motion Compression (MC), where

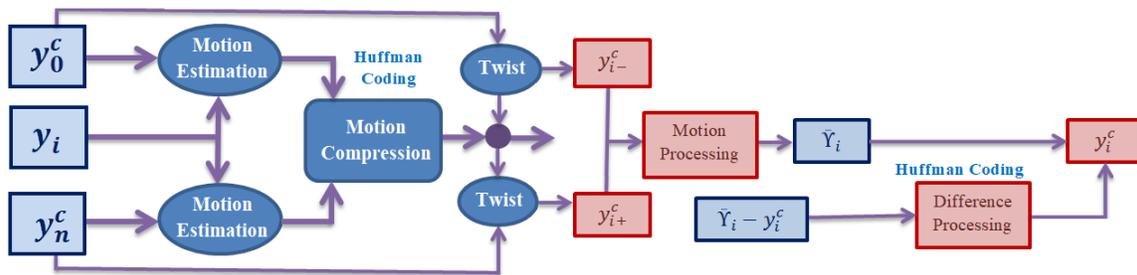


Figure 3: proposed diagram for our HVCS approach illustrates adopting hierarchical quality layers for video compression and depending on BDC and SMC networks

similarities between video frames are considered. Because of the existing correlations between motions in different frames, this work proposes concatenating bi-directional motions (referred to as $[\cdot, \cdot, \dots]$) as inputs to Enj encoder. This leads to input transforming to the latent q_j representation. After that, we quantize q_j to ϕ_j , and then feed ϕ_j into the decoder D_j to reach the compressed motion, where Huffman Coding [40] is used to encode ϕ_j into bits. Similarly, compressed motions will be defined as $\mathcal{E}_{5 \rightarrow 0}$ and $\mathcal{E}_{5 \rightarrow 10}$, and we can formulate MC subnet as:

$$\theta_j = \text{round} \left(\text{En}_j([\mathcal{E}_{5 \rightarrow 0}, \mathcal{E}_{5 \rightarrow 10}]) \right) \dots \dots 1$$

$$[\mathcal{E}_{5 \rightarrow 0}, \mathcal{E}_{5 \rightarrow 10}] = D_j(\theta_j) \dots \dots 2$$

In the next, compressed motions are used to wrap reference frames y_0^c, y_{10}^c to the target, and then for motion compensation, the subnet Motion Post- processing (PM) is used to merge warped frames. Where the operation of backward warping is defined as BKW, we can formulate motion compensation as:

$$y_{0 \rightarrow 5}^c = \text{BKW}(y_0^c, \mathcal{E}_{5 \rightarrow 0}) \dots \dots 3$$

$$y_{10 \rightarrow 5}^c = \text{BKW}(y_{10}^c, \mathcal{E}_{10 \rightarrow 5}) \dots \dots 4$$

$$y_5 = P_m(y_{0 \rightarrow 5}^c, y_{10 \rightarrow 5}^c, \mathcal{E}_{5 \rightarrow 0}, \mathcal{E}_{5 \rightarrow 10}) \dots \dots 5$$

مجله تخصصي علوم الحاسب والذكاء الاصطناعي
 Journal of Pure and Applied Sciences (JUBPAS)
 ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com | ISSN: 2312-8135 | Print ISSN: 1992-0652



Where γ_5 represents frame compensation. Finally, the subnet of Residual Compression (RC) is used to compress the residual of raw frame (y_5) and compensated frame (γ_5). As in MC subnet, the encoder (Enr)/ decoder (Dcr) networks are in RC too. Where φ_r denotes quantized latent representation, we can write the RC subnet as:

$$\begin{aligned}\Theta_r &= \text{round}(En_r(y_5 - \gamma_5)) \dots \dots 6 \\ y_5^c &= Dc_r(\Theta_r) + \gamma_5 \dots \dots 7\end{aligned}$$

Where y_5^c refers to the compressed (y_5) frame. In RC, standard Huffman coding is adopted to encode φ_r into stream of bits, which, combined by encoded φ_m , contributes the stream of bits in layer2. Compression quality, referred to as Q5, is computed and involved in the stream of bits, which is more explained in Section 3.4 with proposed deep decoder. This paper evaluates results quality using Signal-to-Noise Ratio (PSNR) and Multi-Scale Structural Similarity (MS-SSIM) adopted by Zhou et al[37].

c. Single Motion Compression (SMC)

SMC network is proposed to compress remaining frames as layer3, where references are used from nearest compressed frames in layers1 and layers2. Proposed SMC network uses single motion map to compress two frames, i. e. as in Figure 2, a single motion map is used to compress two frames y_1 and y_2 with y_0^c as reference, and y_5^c is used as reference to compress y_3 and y_4 .

Proposed architecture for SMC network is explained in Figure 4 using y_1 and y_2 as examples. In this figure, a DNN is used to firstly compress y_2 frame with architecture similar to BDC, which contains the four subnets (RC, MC, ME, and MP); then the compressed y_2^c frame is obtained. As aforementioned and regarding correlations between multiple adjacent frames, the motion among y_1 and y_0^c or y_2^c is predicted using the motion among y_0^c and y_2 . Accordingly, y_0^c or y_2^c are used as references frames to compress y_1 without consuming bits for motion map, which improves the performance of rate distortion.

In the proposed SMC, using the inverse motion is proposed to predict the motion. Practically, we can define a motion map as:

$$f(n, m) = [\nabla n(n, m), \nabla m(n, m)] \dots \dots 8$$

Where: n & m represent corresponding coordinates, and ∇n & ∇m represent motion maps over n & m coordinates respectively. The inverse motion, for Equation (8), can be computed as:

$$f^{-1}(n + \nabla n(n, m), m + \nabla m(n, m)) = -f(n, m) \dots \dots 9$$

In Equation (9), $f(n, m)$ illustrates that, at each (n, m) pixel, the motion is to the new $(n + \nabla n(n, m), m + \nabla m(n, m))$ position, where the value of $-f(n, m)$ is assigned to the new position of $(f-1)$. To be simplified.

As illustrated this approach adopts backward warping, so that y_2 is compressed using the motion from y_2 , to y_0^c which is $(f_2 \rightarrow 0)$. Consequently, y_1 is compressed using the motion from y_1 to y_0^c from y_1 to y_0^c

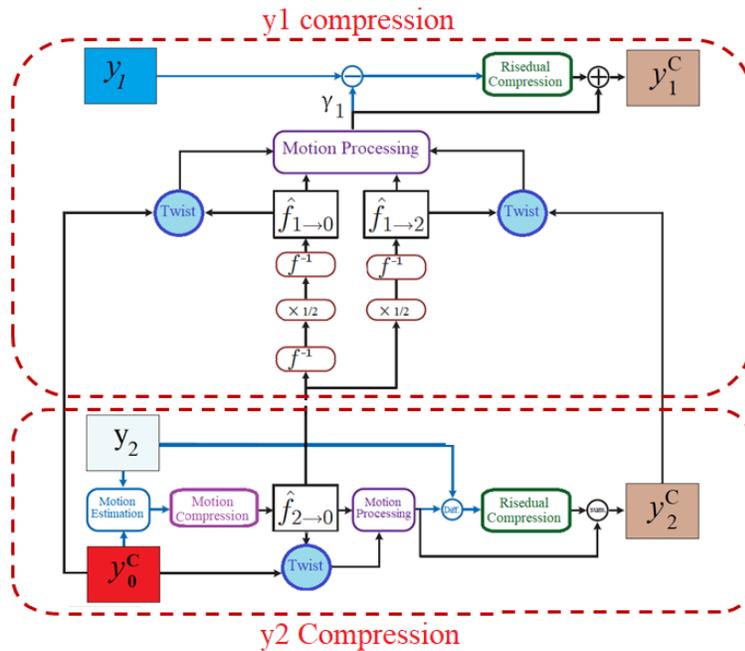


Figure 4: Proposed architecture for SMDC network. RC and MC are used to generate the bit stream, which includes the quality of compression. Bit stream is omitted for simplicity.

$(\mathcal{E}_{1 \rightarrow 0})$ and the reference frames y_0^c & y_2^c . Whereas raw frame y_2 is unavailable at decoding stage, recovering $f_{2 \rightarrow 0}$ is not possible. Therefore, it's possible to use compressed motions $\mathcal{E}_{2 \rightarrow 0}$ for predicting $\mathcal{E}_{1 \rightarrow 0}$ and $\mathcal{E}_{1 \rightarrow 2}$. Using $\mathcal{E}_{2 \rightarrow 0}$ and f^{-1} operation $\mathcal{E}_{1 \rightarrow 0}$ is predicted as:

$$\mathcal{E}_{1 \rightarrow 0} = \underbrace{(0.5 \times (\mathcal{E}_{2 \rightarrow 0})^{-1})^{-1}}_{\mathcal{E}_{0 \rightarrow 2}} \dots \dots 10$$

$$\underbrace{\hspace{10em}}_{\mathcal{E}_{0 \rightarrow 1}}$$

Similarly, $\mathcal{E}_{1 \rightarrow 2}$ is obtained as:

Then, similarly to (3 & 4), we warp reference frames with the predicted motions and feed them into the MP subnet in order to generate the frame γ_1 (motion compensation). Finally, the compressed frame y_1^c is obtained by compressing the differences between $(y_1 - \gamma_1)$ by the RC subnet. Here, result bit streams include the compressed quality (Q_1 & Q_2) to utilize them in the proposed RQEN network.

$$\mathcal{E}_{1 \rightarrow 2} = \underbrace{(0.5 \times \mathcal{E}_{2 \rightarrow 0})^{-1}}_{\mathcal{E}_{2 \rightarrow 1}} \dots \dots 11$$



d. Weighted Recurrent Enhancement (RQEN)

At last, at decoding stage, RQEN is used to enhance the quality, which is designed utilizing standard method (QG-ConvLSTM)[12], which use quality-gated cell and a temporal-spatial structure to exploit correlations between adjacent frames.

Unlike the followed technique[12], this work uses residual blocks in reconstruction networks and spatial feature extraction, and skip connections is employed to enhance the performance. Furthermore, as illustrated in[33][12], frame significance, for quality-enhancement of other frames, is significantly related to its relative quality regarding the others. In addition, the significant quality of each frame is not reachable at the decoding stage[33][12]. On the contrary, since we encode compression-quality in the bit stream, proposed compression scheme accesses Q_i (compression quality) and B_i (bit-rate) from the bit stream, where $F_i = \{Q_i, B_j\}_{j=i-1}^{i+2}$ was utilized as a “quality feature”. Weights generator is fed with F_i to produce $(w_i = [w_i^M, w_i^S])$ weights, which represent the input of quality-gated cell [12] combined by $(S(y_i^C))$ spatial features.

refreshing previous memory with the current information is accomplished by learning the weights (w_i) to reasonably control M_i . In specific, the memory M_i in high-quality frames is supposed to be weighted by a small w_i^M value to refresh previous information with low quality, where update (w_i^S) weights are supposed to be big, to enhance other frames to support the memory by its information with high quality. On the contrary, big w_i^M values and small w_i^S are potential within low quality frames. In addition, where w_i^M values represent sigmoid-function output (w_i^M less than 1 holds), and then previous-frames the information is decreasing in the memory where frame distance is increasing. This supports the, lower correlations for farer frames, fact, which provide less quality enhancement. Therefore, for quality-gated cell, the different-quality frames enhance the memory M_i with different levels of significance. This enables the proposed RQEN network to provide positive influence on multi-frame information in quality enhancement.

e. Training strategy

Within training stage, the density model, as in[5], is used to anticipate the bit-rate to encode φ_j and φ_r in (1) and (6), and use $R(\cdot)$ for defining the estimated bit-rate. Then, this work follows [10][23] in formulating the loss (L):

$$L = \hat{h} D + R \dots \dots 12$$

Where \hat{h} is the hyper-parameter for controlling D (distortion) and (bit-rate) R trade-off.

It's clear in (12) that, for trained models, the compression quality depends on \hat{h} , i.e., less \hat{h} value provide lower quality and bit-rate. Therefore, for a hierarchical-quality compression in the proposed HVCS scheme, we applied different \hat{h} values in SMC and BDC networks for layers2 and layers3 compression, respectively. Specifically, given Equation (12) with the estimated R , the loss function is set in the proposed BDC network as:

$$L_{BD} = \hat{h}_{BD} + D(y_5, y_5^C) + R(\varphi_j) + R(\varphi_r) \dots \dots 13$$

While for SMC the loss is:

$$L_{SM} = \hat{h}_{SM} + D(y_1, y_1^C) + D(y_2, y_2^C) + R(\varphi_j) + R(\varphi_{r1}) + R(\varphi_{r2}) \dots \dots 14$$

As in Equation (14), φ_{r1} and φ_{r2} represent the RC networks of x_1 and x_2 , respectively. In Equations (13) and (14), Mean Square Error (MSE) is used as a distortion measure, whereas $MSE(j, k)$ represents $D(j, k)$, while training the proposed HVCS scheme for Peak signal-to-noise ratio (PSNR). For the optimization of MS-SSIM, the distortion is applied as:



$$D(j, k) = 1 - MS - SSIM(j, k) \dots \dots 15$$

For more importance, \hat{h}_{BD} was set to be larger than \hat{h}_{SM} in (11& 12) Equations to learn proposed compression scheme on compressing layer2-frames with higher quality than layer3 ones, which provide better applying of hierarchical-quality compression. Proposed RQEN network is trained by the minimization of the loss function of the case when N represents RQEN step length. Due to the bi-directional structure of RQEN, bigger N valued results more delay in decoding response leading to more training time. As a result, N value was restricted to 11, to represent frames interval in layer1, in both phases of training and inference.

f. Experimental Settings

The training stage of proposed (BDC & SMC) networks was conducted using standard Vimeo-90k dataset [38], where 142 videos were collected from VQEG [38] and Xiph [39] datasets for the training of proposed RQEN. For testing stage, we adopted UVG [40] dataset combined by Classes B, C and D from standard JCT-VC dataset [41]. Note that training and testing datasets are separated datasets. Different levels of resolution were available such as 1920×1080 (UVG and JCT-VC Class B), which is high resolution. Fair resolutions were also adopted like 832×480 and 416×240 (JCT-VC Classes C and D) respectively. Following [1] for a comparison, all frames were used to test UVG videos against 100 frames to test JCT-VC videos. PSNR and MS-SSIM were used for quality evaluation. For model training, we set \hat{h}_{SM} values as (256, 512, 1024 & 2048 for PSNR and as (8, 16, 32 & 64) for MS-SSIM. For hierarchical quality, \hat{h}_{BD} was set as ($4 \times \hat{h}_{SM}$). For benchmarking HVCS was compared with recent ANN-based compression methods. Furthermore, standards H.264 [3] and H.265 [2] coding techniques were also included in the benchmarking.

Results and Discussion

a. Results

Tables 1& 2 illustrate representing yielded values for rate-distortion on both video datasets. As aforementioned, PSNR and MS-SSIM are used for quality evaluation, where bit-rates are calculated using bits per pixel (bpp). Table 1 illustrates PSNR performance, where they show better PSNR performance for the proposed compression model than other methods such as Chao et al [7] or optimized methods [1]. In addition, they outperform applying H.265 on standard JCT-VC dataset. On the other side, proposed compression scheme yielded better bit-rate performance than applying H.265 on UVG. As in Table 2, the MS-SSIM evaluation provided better performance of proposed scheme than all other learned approaches, where it reached better performance than H.264 and H.265. Due to bit-rate performance on UVG, Lee et al. [11] has comparable performance, and Guo et al [10] yielded lower performance than H.265. Applying on JCT-VC, DVC [10] is only comparable with H.265. On the opposite, the performance of HVCS-rate-distortion have obvious better performance than H.265.



Table 1: Comparison for PSNR performance by state different methods on both datasets

Compression Methods	Bits per Pixels							
	0.01-		0.1-		0.2-		0.3- 0.4	
	JCTVC	UVG	JCTVC	UVG	JCTVC	UVG	JCTVC	UVG
H.256	28.5	31.1	30.2	36.5	34	37.7	34.8	38.4
H2.64	28.1	30.1	31.2	35.5	33.1	37.1	34	37.3
DVC	28.2	28.4	31	35.4	32.3	37.5	33.6	23
Proposed Method	29.2	32.5	31.5	37.7	35.3	37.8	35	38.6

Furthermore, BjØntegaard Delta Bit-Rate (BDBR) [42] is also computed depending on H.265. A BDBR measure computes the average difference of bit-rate considering H.265 anchor, where better performance is indicated on lower values of BDBR [43]. In Table 3, BDBR performance is illustrated depending on PSNR and MS-SSIM, in which, bit-rate reduction considering the anchor is indicated by showed negative numbers. Such results outperform H.265 performance, where bold numbers represent best yielded results by learned methods. Table 3 provided a fair comparison on (MS-SSIM & PSNR) optimized techniques DVC [10] considering the anchor H.265.

Table 2: Comparison for SSIM performance by state different methods on both datasets

Compression Methods	Bits per Pixels							
	0.01-		0.1-		0.2-		0.3- 0.4	
	JCTVC	UVG	JCTVC	UVG	JCTVC	UVG	JCTVC	UVG
H.256	28.5	29.9	32.2	35.7	36.4	37.5	35.6	38.7
H2.64	26.3	31.1	29.8	36.8	35.1	36.9	37.1	38.1
DVC	26.4	28.4	32.8	34.5	36.7	37.7	36.6	35.4
Proposed Method	29.5	31.5	35.8	36.9	37.1	37.7	38.9	38.9

As shown in Table 3, PSNR results of the proposed scheme outperforms other models performance due to MS-SSIM values (average = -6.04). On Class C, proposed-model PSNR shows obvious better performance than MS-SSIM values for optimized method by Cheng et al. [9]. Furthermore, adopted MS-SSIM in this work provided better results than all other learned techniques on MS-SSIM, which provides less bit-rate than H.265 by about (36%) on average. Considerably, our no-RQEN-MS-SSIM yielded significant performance (BDBR = -28%), which still yields significantly better results than all other methods.



Table 3: BDBR results considering H.265 anchor, where C-B, C & D represent class b, c & d respectively. Bold numbers represent best yielded results

Dataset	BDBR Percentage (PSNR)			BDBR Percentage (MS-SSIM)					
	PSNR Optimization			MS-SSIM Optimization			PSNR Optimization		
	Best of Others		HVCS-RQEN	Best of Others		HVCS	Best of Others		HVCS
	Method	Results		Method	Results		Method	Results	
UVG	DVC[10]	8.89	-1.37	Lee [11]	-21.94	-30.12	DVC[10]	8.05	11.24
C-B	HVCS without RQEN	-2.61	-11.75	HLVC without RQEN	-35.32	-37.44		-2.74	-8.11
C-C		20.44	7.83		-20.87	-23.63		-6.88	-9.10
C-D		-1.52	-12.57	Chen[9]	-32.72	-52.56		-18.51	-18.44
Average			6.43	-4.46		-27.77		-35.94	-5.02

Obviously in Table 3, the BDBR is computed by PSNR. On average, yielded bit-rate results are less than H.265 by 4.46%, in addition to the superior results (7.83% bit-rate) over JCT-VC C-C. Over the studied video-dataset (20 videos) HVCS yielded better PSNR results than H.265 for 15 videos. Despite that RQEN provided considerable enhancements, Table 3 shows that no-RQEN--PSNR results are still better than latest PSNR techniques (DVC [10]). As a summary, the general performance of proposed HVCS scheme outperforms all techniques regarding PSNR, and its performance yielded better results than H.265.

b. Ablation studies

Ablation study is generally conducted to discuss the efficiency of inner components of any proposed system [44], and we conducted such study to evaluate the proposed HVCS scheme. In this study, the baseline (bl) model is defined to be the proposed scheme without the principal enhancements such as:

- Hierarchy in frames quality (training models using single \hat{h} value with each frame) referred to as (HQ)
- applying Single Motion approach (using same motion map for all frames) referred to as (SM)
- Proposed enhancement network referred to as (RQEN) .

Then, the performance is analyzed for baseline model with adding the enhancements successively like:

- bl+HQ
- bl+HQ+SM
- the total scheme (bl+HQ+SM+RQEN(")

In addition, the enhanced non-hierarchical (bl+RQEN) is also discussed. Table 4 illustrates the results of ablation study.

- bl+HQ:** in Table 4 combining the base-line with HQ yields obviously better results than baseline alone, which indicates the gained enhancement by applying HQ on the compression performance. Furthermore, Table 4 illustrates the differences in bit-rate and PSNR values and their effects on low quality frames (layer3) and high quality frames (layer1 and layer2) between (bl) and (bl+HQ). It's also clear that, on layer1 and layer2, employing HQ increases PSNR and



bit-rate both. Applying (b1 +HQ) on layer3 frames yields better PSNR results than (b1) even with small bit-rate. This is a logical result when layer1 and layer2 provide high quality references for the compression of layer 3 frames in (b1+HQ). Since layer3 contains most of video frames, adding HQ to b1 provides significant improvement on scheme performance.

- **bl+HQ+SM:** using SMC strategy to enhance compression performance adds more improvements, as shown in Table 4, to the proposed scheme. As a comparison to (bl+HQ), using (bl+HQ+SM) strategy reduces used bits in the motion maps. As an example, in (bl+HQ), the bit-rate yielded (0.018 bpp as average) for motion information is at $\hat{h} = 256$, whereas the total bit-rate yielded 0.097 bpp. Using SMC reduces bits consumption, for motion, to 0.014 bpp, which is (about 23%) lower than (b1+HQ), where total bit-rate was decreased to 0.097 bpp. On the other hand, using (bl+HQ+SM) improves PSNR (29.26 dB to 29.47 dB), because, on residual coding, more bits are allocated. This provides a validation about proposed SMC to reduce the video-motion redundancy, and enhance compression performance.
- **bl+HQ+SM+RQEN:** It's obvious in Table 4 that adding the proposed RQEN network provides significant enhancements to the quality performance of (bl+ HQ+SM) as in Table 4. The enhancements with highest significance were provided to low quality frames, i. e., PSNR enhancements were about (1 dB) for frame3 and frame9. Table 4 shows w_i^S & w_i^M , where proposed RQEN learns for bigger w_i^S and smaller w_i^M on high quality frames, which decreases previous memory to update the significant information, and vice versa for frames with low quality. Moreover, Table 4 visually shows that frame3 suffers from considerable distortions since bit-rate is low, while higher quality frame (6) ha highly correlations with frame 3. Thus, in RQEN, high w_i^S in frame 6 updates large proportion of the information to the memory. Therefore, there is an ability of recovering lost information of frame 3 providing significant enhancement to the quality.

Table 4: Illustrates exclusion comparison for the inner components of the proposed scheme

Compression Methods	Bits per Pixels							
	PSNR	BPP	PSNR	BPP	PSNR	BPP	PSNR	BPP
B1	28.05	0.1	31.11	0.2	32.81	0.3	33.5	0.4
B1+RQEN	28.11		31.24		32.89		33.49	
B1+HQ	28.12		31.02		32.91		33.86	
B1+HQ+SM	28.32		31.31		33.21		34.01	
All of them	28.25		31.82		35.01		34.31	

Advantages of Hierarchical Quality: Due to Table 4, enhancements in frame quality from bl to (bl+ RQEN) are considerable less than (bl+HQ+SM) to the proposed HVCS scheme. A major cause of such variation is the similarity in frame quality, which is handled in the hierarchical scheme. Hierarchy utilizes the high quality reference to enhance low quality of other frames. As a summary, all components of the proposed compression scheme yielded significant enhancements the compression framework achieves considerable results benchmarking with other state-of-the-art video compression methods.



Acknowledgments:

We thank the Computer System Technic Department, Institute of Administration Rusafa, Middle Technical University, Baghdad-Iraq, and Thanks a lot with our study. I'd like to express my gratitude to the anonymous reviewers and editor for their insightful remarks and evaluations.

Conflict of interests.

There are non-conflicts of interest.

References

- [1] Cisco, "Cisco visual networking index (VNI) global mobile data traffic forecast update, 2017-2022 white paper," *Computer Fraud & Security*, 2019. .
- [2] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal Process. Mag.*, vol. 18, no. 5, 2001, doi: 10.1109/79.952804.
- [3] C. Poppe, S. De Bruyne, T. Paridaens, P. Lambert, and R. Van de Walle, "Moving object detection in the H.264/AVC compressed domain for video surveillance applications," *J. Vis. Commun. Image Represent.*, vol. 20, no. 6, 2009, doi: 10.1016/j.jvcir.2009.05.001.
- [4] S. Pandit, P. K. Shukla, A. Tiwari, P. K. Shukla, M. Maheshwari, and R. Dubey, "Review of video compression techniques based on fractal transform function and swarm intelligence," *International Journal of Modern Physics B*, vol. 34, no. 8. 2020, doi: 10.1142/S0217979220500617.
- [5] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," 2017.
- [6] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," 2018.
- [7] C. Y. Wu, N. Singhal, and P. Krähenbühl, "Video compression through image interpolation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11212 LNCS, doi: 10.1007/978-3-030-01237-3_26.
- [8] Z. Chen, T. He, X. Jin, and F. Wu, "Learning for Video Compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 2, 2020, doi: 10.1109/TCSVT.2019.2892608.
- [9] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Learning image and video compression through spatial-temporal energy compaction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, vol. 2019-June, doi: 10.1109/CVPR.2019.01031.
- [10] G. Lu, W. Ouyang, D. Xu, X. Zhang, C. Cai, and Z. Gao, "Dvc: An end-to-end deep video compression framework," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, vol. 2019-June, doi: 10.1109/CVPR.2019.01126.
- [11] S. Lee, J. B. Jeong, and E. S. Ryu, "Atlas level rate distortion optimization for 6DoF immersive video compression," 2022, doi: 10.1145/3534088.3534354.
- [12] "https://bellard.org/bpg/ ." .
- [13] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consum. Electron.*, vol. 38, no. 1, 1992, doi: 10.1109/30.125072.
- [14] G. Toderici *et al.*, "Variable rate image compression with recurrent neural networks," 2016.
- [15] G. Toderici *et al.*, "Full resolution image compression with recurrent neural networks," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-January, doi: 10.1109/CVPR.2017.577.
- [16] M. Li, W. Zuo, S. Gu, D. Zhao, and D. Zhang, "Learning Convolutional Networks for Content-Weighted Image Compression," 2018, doi: 10.1109/CVPR.2018.00339.
- [17] E. Agustsson *et al.*, "Soft-to-hard vector quantization for end-to-end learning compressible representations," in *Advances in Neural Information Processing Systems*, 2017, vol. 2017-



- December.
- [18] L. Theis, W. Shi, A. Cunningham, and F. Huszár, “Lossy image compression with compressive autoencoders,” 2017.
- [19] D. Minnen, J. Ballé, and G. Toderici, “Joint autoregressive and hierarchical priors for learned image compression,” in *Advances in Neural Information Processing Systems*, 2018, vol. 2018-December.
- [20] N. Johnston *et al.*, “Improved Lossy Image Compression with Priming and Spatially Adaptive Bit Rates for Recurrent Networks,” 2018, doi: 10.1109/CVPR.2018.00461.
- [21] J. Lee, S. Cho, and S. K. Beack, “Context-adaptive entropy model for end-to-end optimized image compression,” 2019.
- [22] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool, “Conditional Probability Models for Deep Image Compression,” 2018, doi: 10.1109/CVPR.2018.00462.
- [23] D. J. Le Gall, “The MPEG video compression algorithm,” *Signal Process. Image Commun.*, vol. 4, no. 2, 1992, doi: 10.1016/0923-5965(92)90019-C.
- [24] T. Li, M. Xu, C. Zhu, R. Yang, Z. Wang, and Z. Guan, “A Deep Learning Approach for Multi-Frame In-Loop Filter of HEVC,” *IEEE Trans. Image Process.*, vol. 28, no. 11, 2019, doi: 10.1109/TIP.2019.2921877.
- [25] J. Liu, S. Xia, W. Yang, M. Li, and D. Liu, “One-for-All: Grouped Variation Network-Based Fractional Interpolation in Video Coding,” *IEEE Trans. Image Process.*, vol. 28, no. 5, 2019, doi: 10.1109/TIP.2018.2882923.
- [26] M. Xu, T. Li, Z. Wang, X. Deng, R. Yang, and Z. Guan, “Reducing complexity of HEVC: A deep learning approach,” *IEEE Trans. Image Process.*, vol. 27, no. 10, 2018, doi: 10.1109/TIP.2018.2847035.
- [27] Y. Dai, D. Liu, and F. Wu, “A convolutional neural network approach for post-processing in HEVC intra coding,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10132 LNCS, doi: 10.1007/978-3-319-51811-4_3.
- [28] T. Li, M. Xu, R. Yang, and X. Tao, “A DenseNet Based Approach for Multi-frame In-loop Filter in HEVC,” in *Data Compression Conference Proceedings*, 2019, vol. 2019-March, doi: 10.1109/DCC.2019.00035.
- [29] T. Chen, H. Liu, Q. Shen, T. Yue, X. Cao, and Z. Ma, “DeepCoder: A deep neural network based video compression,” in *2017 IEEE Visual Communications and Image Processing, VCIP 2017*, 2018, vol. 2018-January, doi: 10.1109/VCIP.2017.8305033.
- [30] R. Yang, M. Xu, and Z. Wang, “Decoder-side HEVC quality enhancement with scalable convolutional neural network,” 2017, doi: 10.1109/ICME.2017.8019299.
- [31] R. Yang, M. Xu, T. Liu, Z. Wang, and Z. Guan, “Enhancing Quality for HEVC Compressed Videos,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 7, 2019, doi: 10.1109/TCSVT.2018.2867568.
- [32] R. Yang, M. Xu, Z. Wang, and T. Li, “Multi-frame Quality Enhancement for Compressed Video,” 2018, doi: 10.1109/CVPR.2018.00697.
- [33] T. Wang, W. Xiao, M. Chen, and H. Chao, “The multi-scale deep decoder for the standard HEVC bitstreams,” in *Data Compression Conference Proceedings*, 2018, vol. 2018-March, doi: 10.1109/DCC.2018.00028.
- [34] G. Lu, W. Ouyang, D. Xu, X. Zhang, Z. Gao, and M. T. Sun, “Deep kalman filtering network for video compression artifact reduction,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11218 LNCS, doi: 10.1007/978-3-030-01264-9_35.
- [35] A. Ranjan and M. J. Black, “Optical flow estimation using a spatial pyramid network,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-January, doi: 10.1109/CVPR.2017.291.
- [36] A. S. Arif, S. Mansor, H. Abdul Karim, and R. Logeswaran, “Lossless compression of Fluoroscopy



- medical images using correlation and the combination of Run-length and Huffman coding,” 2012, doi: 10.1109/IECBES.2012.6498149.
- [37] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multi-scale structural similarity for image quality assessment,” in *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, 2003, vol. 2, doi: 10.1109/acssc.2003.1292216.
- [38]. “<https://media.xiph.org/video/derf/>. .” .
- [39] F. Bossen, “Common Conditions and Software Reference Configurations, document JCTVC-H1100,” 2013.
- [40] T. Xue, B. Chen, J. Wu, D. Wei, and W. T. Freeman, “Video Enhancement with Task-Oriented Flow,” *Int. J. Comput. Vis.*, vol. 127, no. 8, 2019, doi: 10.1007/s11263-018-01144-2.
- [41] B. GISLE, “Calculation of Average PSNR Differences between RD-curves,” *ITU-T VCEG-M33*, April. 2001, 2001, Accessed: May 03, 2023. [Online]. Available: <https://cir.nii.ac.jp/crid/1572824499807909504.bib?lang=en>.
- [42] M. Quach, G. Valenzise, and F. Dufaux, “Learning Convolutional Transforms for Lossy Point Cloud Geometry Compression,” in *Proceedings - International Conference on Image Processing, ICIP*, 2019, vol. 2019-September, doi: 10.1109/ICIP.2019.8803413.
- [43] L. Du, “How Much Deep Learning does Neural Style Transfer Really Need? An Ablation Study,” in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020, pp. 3139–3148, doi: 10.1109/WACV45572.2020.9093537.
- [44] “<http://ultravideo.cs.tut.fi/#testsequences/>. .” .