POLITECNICO DI TORINO
Repository ISTITUZIONALE

Improving Learned Video Codec Performance in Low-Latency Lossy Communication Scenarios

Original
Improving Learned Video Codec Performance in Low-Latency Lossy Communication Scenarios / Masala, E.. - STAMPA. - (2026). (20th IEEE International Conference on Application of Information and Communication Technologies, AICT 2026 Baku (Azerbaijan) 14-16 Oct 2026).

Availability:
This version is available at: 11583/3016915 since: 2026-10-23T14:17:48Z

Publisher:
IEEE

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

04 November 2026

Improving Learned Video Codec Performance in Low-Latency Lossy Communication Scenarios

Enrico Masala
Control and Computer Engineering Department
Politecnico di Torino
Turin, Italy
enrico.masala@polito.it

***Abstract*—Learned video codecs based on conditional coding architectures, e.g., DCVC, have demonstrated rate-distortion performance competitive with state-of-the-art classical standards, yet their behavior in presence of data loss is still largely unknown. The tight inter-frame dependency introduced by deep temporal context models causes far more severe problems than in classical hybrid codecs in presence of data loss, questioning the possibility of their actual usage in real-world communication systems. This paper presents a feedback-based error control mechanism for a learned video codec capable of encoding both intra-coded (I) and inter-coded (P) frames. When the receiver detects a lost frame and notifies the encoder through a feedback channel, the encoder inserts an I-frame at the earliest opportunity, restoring decoder synchronization and halting decoding issues. The approach requires no modification to the codec's training procedure. Experimental results on standard video sequences over simulated lossy channels show that the proposed mechanism substantially reduces quality degradation, while the learned codec retains its rate-distortion advantage over AVC and HEVC coded video even at substantially high loss rate. Moreover, it can achieve the same or higher quality at a significantly less total bitrate cost.**

***Index Terms*—Learned Video Compression, Robust Video Communication, Deep Neural Networks, Video Quality Evaluation**

I. Introduction

Resilience to data loss in real-time video communication over packet networks has been studied for decades. Classical hybrid codecs such as AVC and HEVC were designed with this problem in mind, and such standards include well-established mechanisms and tools such as, for instance, periodic intra-refresh, reference picture selection, and feedback-driven intra insertion, that can effectively limit the temporal spread of errors caused by lost packets [1], [2].

However, the emergence of learned video codecs introduces a new set of challenges in this context. Learned video codecs such as the ones from the DCVC family [3] are currently able to achieve compression efficiency superior to classical standards in pure compression tests. This can be attributed to deep temporal context models that exploit inter-frame redundancy more tightly than block-based motion compensation. However, this tighter coupling is precisely what makes them far more sensitive to missing data: when a reference frame is lost, the error propagates through the temporal context chain and creates much more serious errors than in classical residual-based codecs. In some cases, even the basic entropy decoding process can be severely impaired since it relies on precise context modeling. In addition, learned codecs are typically optimized exclusively for pure compression performance thus, unlike their classical counterparts, they lack built-in tools to contain the negative effect of data loss.

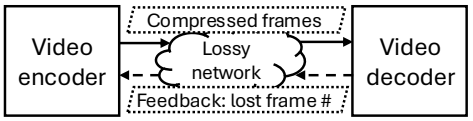


Fig. 1. Block diagram of the proposed scheme.

This work focuses on real-time video communication systems since this scenario is often very limited in bandwidth, hence the usefulness of high compression. However, this scenario is challenging because data loss may frequently happen and, due to latency constraints, losses cannot be easily recovered through, e.g., retransmission as it happens in on-demand streaming systems. However, such systems have access to a feedback path from receiver to sender. This return channel allows the receiver to notify the encoder of lost frames, and in traditional hybrid video codec deployments it can trigger intra-coded frame insertion to stop further error propagation. Whether an analogous feedback-based control mechanism can be effective (and how much) for learned video codecs has, to the best of our knowledge, not been previously investigated.

This paper presents a first investigation of feedback-based error control for a learned video codec (DCVC-RT [4]) capable of producing both intra-coded (I) and inter-coded (P) frames. Upon knowing of a lost frame, we instructed the encoder to insert an I-frame at the earliest opportunity, restoring decoder synchronization and halting error propagation. Fig. 1 shows the block diagram of the system. This raises questions specific to the learned codec setting, in particular how much more expensive is an I-frame relative to a P-frame compared to the classical codecs case, and what is the net effect of this strategy on quality over a range of loss rates.

The main contributions of this work can be summarized as: (i) a feedback-based error control architecture integrating loss notification with I-frame insertion in a learned codec; (ii) a characterization of the rate overhead of forced I-frame insertion in a learned codec compared to AVC and HEVC; and (iii) experimental results on standard sequences over simulated lossy channels showing the quality, measured using the Peak

Signal-to-Noise Ratio (PSNR), as a function of loss rate using a periodic intra refresh coding schemes with and without feedback-driven error control.

The remainder of this paper is organized as follows. Section II reviews related work. Section III overviews the considered system and its practical implementation. Section IV presents the experimental methodology and results, followed by Section V that draws conclusions.

II. Related Work

Error resilience in classical hybrid video codecs has been extensively studied. For instance, the AVC and HEVC standards even include codec-level mechanisms such as intra-refresh, flexible macroblock ordering and reference picture selection which are specifically designed for this purpose, i.e., to limit error propagation in lossy networks [1], [2], [5]. When a return channel is available, the encoder can respond to receiver-reported loss events by changing its encoding behavior, for instance by inserting intra-coded content on demand, that is a well-established strategy in real-time communication systems [6], [7].

The natural tendency to improve transmission performance also goes in the direction of improving the rate-distortion performance of the codec itself, thus requiring less and less bandwidth to achieve the same quality, allowing transmission even when bandwidth is severely limited. Learned video codecs are a promising direction since they already achieved performance that is competitive with the most recent video coding standards. In practice, they adopt end-to-end optimized approaches [8], [9] that replace hand-crafted transforms such as the DCT with neural networks. In addition, conditional coding architectures such as the ones in the DCVC family [10]–[12] further improved efficiency by conditioning each frame's latent representation on a deep temporal context extracted from the past frames.

However, the behavior of learned video codecs under transmission impairments is still largely unexplored. Some work focused on using deep learning techniques together with joint source-channel coding for transmission over wireless channels, respectively for image [13] and video [14] coding. However, they do not employ learned image/video codecs. Instead, their focus is on either improving the pixel-to-symbols mapping for a specific wireless channel transmission [13] or extracting semantic information by means of deep neural networks to enhance the quality of the reconstructed video in presence of transmission errors [14].

To the best of our knowledge, this is the first work to focus on exploring the performance of learned video codecs when transmitted over standard packet lossy network systems. In this case, the fundamental issue is that blocks of data are either received correctly or not received at all. Traditional video coding systems relies on differential encoding, hence missing data may cause temporal and spatial error propagation. One of the simplest method to recover from such a condition is intra refresh, either periodic or demand-driven, proposed many years ago [15] but still used especially in low-latency video communication systems [16]. This is the mechanism adopted in this work: upon knowing about a lost frame arrived at the decoder, the encoder inserts a full I-frame, resynchronizing the decoding process without requiring any modification to the codec's architecture or training procedures, which is a significant simplification for learned coding systems. Also note that, depending on the way the learned video codec works, this could be the only possibility to resume correct entropy decoding after a loss, as explained in more details in Section III-A. However, it is possible that other strategies, e.g., reference picture selection, may also work if the codec is modified.

III. System Description

A. Overview

The system considered in this work aims at implementing a videoconferencing system where video data must be acquired, transmitted with low latency through a lossy network to a receiver, and then decoded and displayed to the user. The system can be divided, at the logical level, in four main components: a video encoder, a channel with feedback, a video decoder, and an error control module.

The video encoder has the possibility to encode frames in either self-contained mode (so called *intra* mode) or differentially with respect to the previous frame. This naturally maps on I frames and P frames in traditional video codecs based on the so-called hybrid model, such as AVC and HEVC. The learned video codec considered in this work (DCVC-RT [3]) also provides this possibility, since it can encode a video frame in both ways. The encoding mode can be modified at every new frame to be encoded. However, most of the times, for simplicity, a fixed coding scheme is used, such as an initial I frame followed by a fixed number N of P frames, then the scheme repeats again with another I frame and other N P frames etc. This is what we will refer to as a Periodic Intra Refresh (PIR) scheme. The main advantage of periodically inserting P frames is to have a resynchronization point to start from on the decoder side in case some errors happen during the decoding phase, typically due to missing data. This is useful for all differential encoding schemes, including the ones of traditional video codecs for which losses cause both temporal and spatial error propagation that heavily affect the quality of the decoded content.

For the case of learned video codecs such as DCVC-RT, the following summarizes the points of interest for the purpose of the present work. The reader is referred to more specialized works [4] to understand their inner workings, since that is out of the scope of the present work. In particular, here we discuss the effect of data loss on DCVC-RT. In fact, this is a problem also for the DCVC-RT learned codec architecture considered in this work [10], but with some important differences compared to the traditional codecs case. To achieve high compression efficiency, the DCVC-RT architecture requires that a conditioning context, derived from previous frames, is provided to the decoder in order to perform even just entropy decoding. This is different from traditional

codecs such as AVC and HEVC that can correctly perform entropy decoding for a frame if all the bits of that frame have been correctly received: they can correctly determine, for instance, macroblock types, quantization parameters, value of the motion vectors and value of the residuals. Unfortunately, this is not the case for DCVC-RT. Therefore, in case of loss of a frame, attempting to decode subsequent frames which are not I frames lead to catastrophic results in terms of reconstruction or it is not possible at all. In such a case, the only way to resume correct decoding is to restart from the next I frame.

The component modeling the channel with feedback is the same for all experiments and, on the basis of a random variable, can decide to leave the data corresponding to the frame unaltered, or to remove it completely, i.e., the frame is not passed to the decoder itself. It is also assumed that the channel is bidirectional, i.e., it can transport some metadata from the receiver to the transmitter. In particular, it is assumed that the information about which frame did not arrive to the receiver can be conveyed back to the transmitter, so that it can reach the error control module that decides which frame type should be used by the video encoder.

The video decoder component operates normally. However, since in this case it is possible that some frames are missing, the decoder handles such missing frames in the following way. It is assumed that the decoder is informed of the fact that a frame did not arrive to the destination. This is possible, in practice, by using suitable multimedia transport protocols such as RTP [17] that number packets so that it is easy to identify missing data. Since most of the time decoders cannot easily decode a compressed stream when part of the data is missing, as highlighted before, in this preliminary investigation we assumed that whenever data from a frame is missing, the decoder waits until the reception of the next I frame to resume decoding. However, the decoding process cannot stop, therefore the decoder is assumed to produce, as output, a copy of the last decoded frame every time a new frame is needed for playback. This can produce a so called "freeze", which is accounted for in the quality computation in the result section.

Finally, the error control module component is in charge to decide what to do when a frame has been lost. In practice, it is supposed to detect the missing data on the receiver side, to communicate such information back to the transmitter using the feedback channel, and to instruct the encoder to insert an I frame during the encoding process, if not already needed for other reasons, such as the periodic I frame type insertion. In this preliminary investigation, for simplicity's sake, we assume that the feedback information is never lost, but it may take some time to reach the encoder. During such time, the encoder operates as if there were no losses in the communication, since it is unaware of the communication issue. Note that no retransmission is attempted, since the system operates under the assumption that the end-to-end latency budget of typical interactive video conferencing systems does not permit the additional delay that retransmission would introduce in the system, i.e., the necessity for a larger playout buffer on the receiver side to wait for the retransmitted data.

B. Implementation

In this work we specifically refer to and use the DCVC-RT variant of DCVC, whose implementation, including the weights of the trained neural networks (one set for I frames and one for P frames), is publicly available [3]. This one is the fastest variant of the DCVC family of codecs which have been published until now. We slightly modified the high-level python code that is provided in the public implementation [3] to separate the encoding and decoding parts into two different programs. Also, the encoder can take the type of frames to use during encoding from an external source. The information about the used frame type and its representation in terms of bits are already handled by the software itself in a binary format that can be directly read by the decoder as-is.

Note also that all the considered encoders, both the DCVC-RT and the traditional AVC and HEVC, can produce higher or lower bitrates for the same source content. In this work, for simplicity and also to avoid additional confounding factors, the encoding is performed using a constant quality parameter, equal for all the frames of the sequence, regardless of their type. This is the so-called "constant QP" mode for traditional standards such as AVC and HEVC. In such standards, higher QP values correspond to higher compression (and lower quality), and vice versa. The DCVC-RT software also provides a similar parameter that, however, works in the opposite way (higher values yield higher qualities and vice versa). Suitable QP values have been selected for the encoded content so that the quality and bitrates are in a reasonable range for the chosen application, i.e., videoconferencing. Rate-distortion curves for the chosen encoding points are described later in Section IV-B and they are graphically shown in Fig. 2.

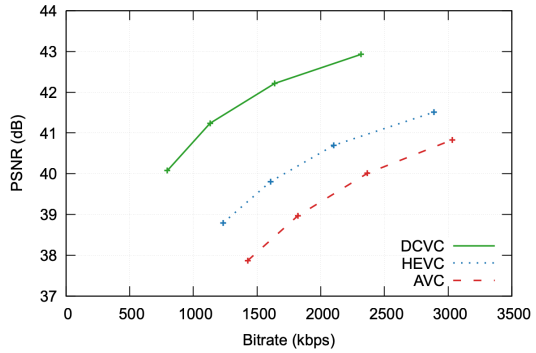
The rest of the system has been implemented in the form of a simulator written in the python language. Loss events are simulated using loss traces. For each loss condition, i.e., loss rate, many traces are generated and the transmission experiment is repeated many times, then the quality results are averaged in order to obtain statistically significant results.

Since this is a simulation setting, it is possible to directly measure the quality of the video arriving at the receiver by comparing it with the original uncompressed video fed into the video encoder. The quality evaluation has been carried out by means of the PSNR measure, as it has been done since the beginning in most of the research works investigating video quality. We are aware of the fact that other video quality estimation measures exists, notably VMAF [18]. However, they were typically not designed for lossy scenarios. Therefore, in this preliminary investigation we decided to use only the PSNR as a first quality indicator, since it is also well understood by the research community and it facilitates comparisons with other works.

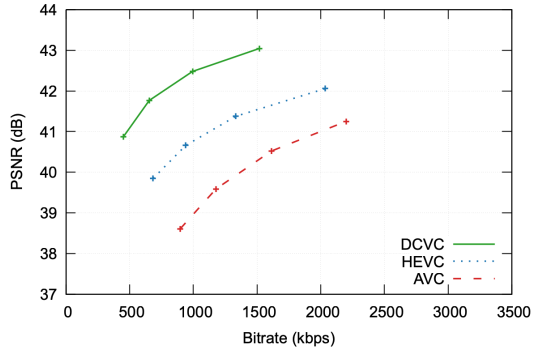
IV. Experimental Methodology and Results

A. Experimental Setup

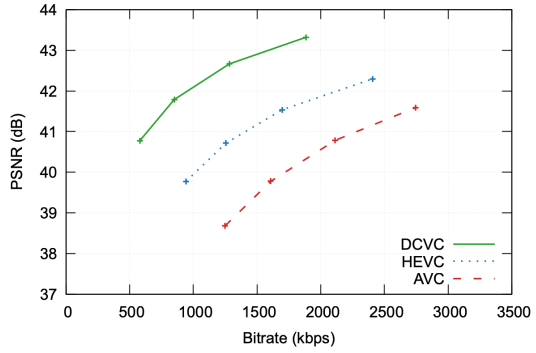
All experiments were conducted using the simulator described in Section III. Video encoding was performed using



(a) *FourPeople*



(b) *Johnny*



(c) *KristenAndSara*

Fig. 2. Rate-distortion curves (PSNR vs. bitrate) for the three codecs on the 1280×720 test sequences under no-loss channel conditions (PIR encoding).

the DCVC-RT learned video codec [3], slightly modified as described in Section III, as well as using the AVC and HEVC video coding standards. For the latter two cases, we used the codec implementation from the libx264 and libx265 libraries (as linked in the ffmpeg v. 6.1.1 software), respectively.

Due to the demanding amount of computation of the learned video codec, as recommended by the software authors, encoding and decoding for the DCVC-RT codec were performed by means of a GPU. In particular, we relied on a workstation equipped with an Nvidia GeForce RTX 5090 GPU with 32 GB of VRAM. Note that in this setup the time required for frame encoding and decoding was in the order of tens of milliseconds, therefore compatible with the real-time requirements of a videoconferencing application. These results are in line with the ones reported by the authors [4].

TABLE I
PSNR, BITRATE AND QUALITY CONTROL PARAMETER (QP) FOR THE KRISTENANDSARA SEQUENCE, PERIODIC INTRA REFRESH (PIR) ENCODING, LOSS-FREE CASE.

AVC			**HEVC**			**DCVC-RT**		
QP	Bitrate (kbps)	PSNR (dB)	QP	Bitrate (kbps)	PSNR (dB)	QP	Bitrate (kbps)	PSNR (dB)
26	2744	41.59	26	2407	42.30	63	1883	43.32
28	2113	40.78	28	1696	41.54	56	1282	42.66
30	1607	39.78	30	1254	40.71	48	849	41.79
32	1249	38.68	32	942	39.78	40	579	40.78

Frame loss was modeled as an independent Bernoulli process with loss probability 0.01, 0.03, 0.05, 0.10, 0.15, 0.20 to span the range from lightly congested to severely impaired network conditions.

Experiments were conducted using standard video sequences drawn from the HEVC common test conditions class E sequences [19] (spatial resolution 1280×720, 60 fps), namely *KristenAndSara*, *Johnny*, and *FourPeople*, whose scenes well represents typical videoconferencing content. All sequences were 10 seconds in duration to ensure that loss events and recovery cycles could be observed across multiple occurrences within a single sequence.

Two error control strategies were evaluated, each applied to three codecs: the learned codec DCVC-RT, the AVC codec and the HEVC codec. The encoding configurations were:

1) **Periodic intra refresh (PIR).** Each encoder inserts an I-frame every 32 frames unconditionally, regardless of channel conditions. No feedback from the receiver is used.
2) **Periodic intra refresh with error control (PIR-EC).** Upon detecting a lost frame, it is assumed that the receiver notifies the encoder, which inserts an I-frame. The notification may take time to reach the encoder. Two delay values are evaluated: 2 frames and 3 frames. These model the feedback round-trip time in frame units and it is expected to be in line with a range of values representative of low-latency interactive communications.

Rate-distortion operating points were selected to yield a nominal (loss-free) bitrate ranging from about 500 to 3,000 kbps, for the three different video sequences. For each configuration and each loss condition, results are averaged over 100 independent loss realizations of the communication channel, then averaged across sequences. It is assumed that, if a frame is lost, the concealment strategy consists in using the previous frame as the output (frame "freeze"), until an I-frame arrives at the decoder.

Reconstruction quality was measured using the PSNR, in dB, computed between the original uncompressed frame and the decoded (or concealed) frame at the receiver. The PSNR of the whole sequence is computed as the mean of the PSNR of each single frame.

B. Rate-Distortion Performance Under No Loss Conditions

As a reference point, Fig. 2 shows rate-distortion curves for the three sequences employed in the subsequent experiments. The same frame type encoding scheme is used for all codecs,

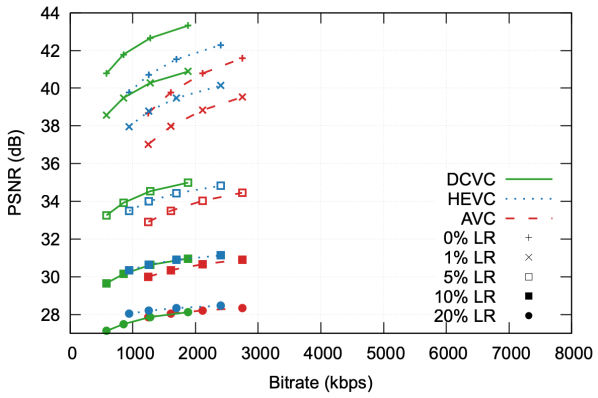


Fig. 3. Average PSNR for different codecs and loss probabilities when encoded with periodic intra refresh (PIR) only, *KristenAndSara* sequence.

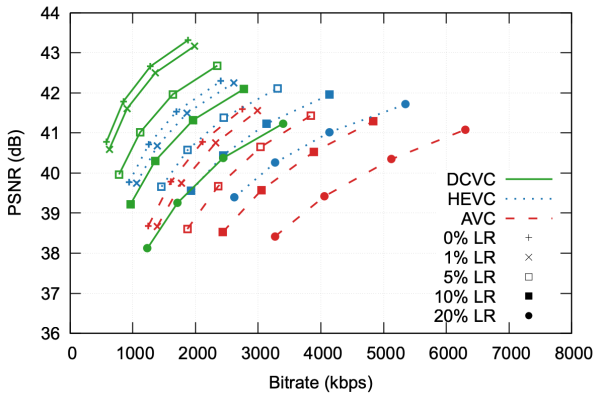


Fig. 4. Average PSNR for different codecs and loss probabilities when the error control strategy (PIR-EC) inserts I frames with a 2-frame delay, *KristenAndSara* sequence.

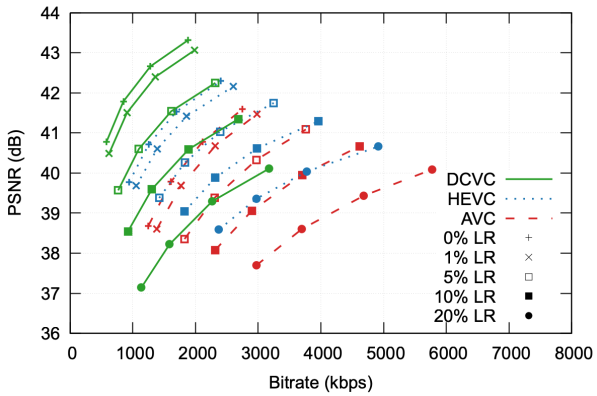


Fig. 5. Average PSNR for different codecs and loss probabilities when the error control strategy (PIR-EC) inserts I frames with a 3-frame delay, *KristenAndSara* sequence.

i.e., periodic intra refresh (PIR) every 32 frames. For better readability, Table I also quantitatively reports the loss-free rate-distortion performance and the quality control parameters (QP) for the three codecs for one sample sequence (*KristenAndSara*, Fig. 2c). Note however that QP cannot be easily related mathematically to quality.

Fig. 2 shows that, on average, the DCVC-RT codec provides from 1.5 to 2 dB PSNR gains over HEVC. The gain is even more pronounced with respect to AVC, reaching up to 3 dB PSNR. For the same PSNR value, the DCVC-RT bitrate saving ranges from 50% up to 66% over HEVC depending on the considered PSNR level. Regarding AVC, bitrate reductions can be up to 75% in the most favorable conditions. These results are consistent with previously reported comparisons between the DCVC family of codecs and classical standards [4].

C. Periodic Intra Refresh Without Error Control

Fig. 3 shows the average PSNR over all channel realizations, as a function of the sequence bitrate, for different codecs and loss probabilities for the *KristenAndSara* sequence.

Without any error control strategy the quality rapidly decreases when the loss rate increases. Of course, in this case, no additional bandwidth is used, because no modifications to the type of frames chosen during encoding is performed. Hence, for each codec, the horizontal position of the points does not change when moving towards the bottom. The better rate-distortion performance at encoding time for the DCVC-RT with respect to the other codecs (i.e., about 2 dB PSNR with respect to HEVC and 3 in the case of AVC) allows the DCVC-encoded video under 1% loss rate to perform almost on par with the zero-loss HEVC. In this sense, it can be said that the higher quality provided by the DCVC-RT encoding can compensate for a small loss rate compared to other codecs. However, this difference shrinks as the loss rate increases, and it is almost not existent for 10% or higher loss rates (compare curves with different colors but same point type).

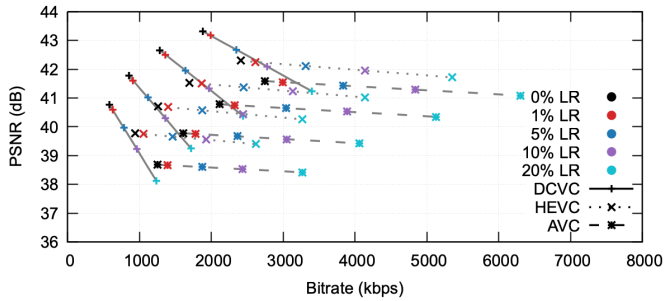
D. Periodic Intra Refresh with Error Control

Fig. 4 shows the average PSNR as a function of the sequence bitrate using the error control strategy based on additional I frame insertion (PIR-EC), described in Section IV-A, for the *KristenAndSara* sequence when the insertion delay is equal to 2 frames. Note that the bitrate increases as the loss rate increases because it also accounts for the bitrate needed to transmit the additional I frames. As in Fig. 3, different codecs and loss probabilities are considered. It can be noticed that traditional codecs such as AVC and HEVC tend to be more resilient, i.e., their PSNR decreases less rapidly when the loss rate increase with respect to the DCVC-RT case. However, this comes at the expense of a much higher increase in bandwidth. The latter is due to the fact that the average ratio of the size of I and P frames is much higher in the case of AVC and HEVC than the case of DCVC-RT. Table II reports such ratio for the considered sequences and coding conditions, ranging from about 8 to 13 for DCVC-RT and from about 5 to 33 for traditional codecs. The same effect happens, even more pronounced, when the I frame insertion delay is equal to 3 frames, as shown in Fig. 5.

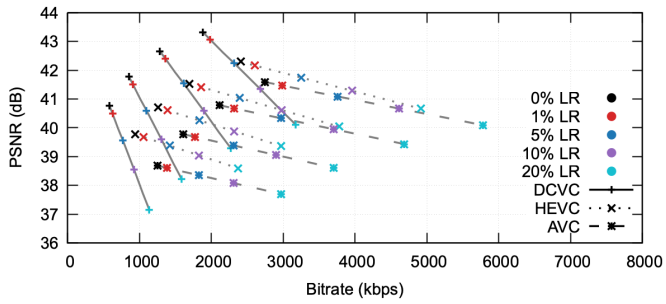
To better visualize such effect, Fig. 6 shows the same data but connecting the points that belong to the same original encoding. The leftmost point correspond to the loss-free condition, while the right one to the maximum loss rate condition (20% in our experiments). The difference in slope is particularly evident.

TABLE II
AVERAGE I/P FRAME-SIZE RATIO AFTER ENCODING.

Codec	QP	Sequence		
		FourPeople	*Johnny*	*KristenAndSara*
DCVC-RT	63	9.15	9.17	8.29
DCVC-RT	56	10.18	10.64	9.38
DCVC-RT	48	11.33	12.09	10.73
DCVC-RT	40	12.50	13.18	12.16
HEVC	20	6.64	5.72	6.10
HEVC	24	12.33	10.97	10.63
HEVC	28	19.63	21.40	17.27
HEVC	32	25.67	33.33	24.05
AVC	20	7.81	6.66	6.79
AVC	24	15.60	14.02	12.28
AVC	28	21.45	23.10	17.09
AVC	32	25.41	29.62	20.56



(a) 2-frame refresh delay



(b) 3-frame refresh delay

Fig. 6. Lines highlight, for the same encoding, the performance change when the loss rate changes. *KristenAndSara* sequence.

V. CONCLUSION

This paper has investigated a feedback-based error control strategy for one of the most recent learned video codec (DCVC-RT) supporting both intra and inter coding modes. Simulations that exploits a return channel to trigger I-frame insertion upon frame loss detection showed that the proposed mechanism can prevent the strong quality decrease that would be experienced with a simple, standard, periodic intra refresh. Simulation results show that the better rate-distortion compression performance of the learned video codec is essentially preserved with respect to competing traditional codecs such as AVC and HEVC. This also happens at a substantially lower bitrate increase since the I/P frame ratio for the learned video codec is typically lower than the one of AVC and HEVC in the considered experimental settings. These results seem a promising direction to better understand the behavior of learned video codecs in lossy conditions, thus making them more suitable for low-latency communications scenarios.

REFERENCES

[1] S. Wenger, "H.264/AVC over IP," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 645–656, Jul. 2003.

[2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[3] Microsoft, "DCVC: Deep contextual video compression," https://github.com/microsoft/DCVC, (commit 75d9d1a8 from July 2025, retrieved Jun 2, 2026).

[4] Z. Jia, B. Li, J. Li, W. Xie, L. Qi, H. Li, and Y. Lu, "Towards practical real-time neural video compression," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025, pp. 12 543–12 552.

[5] P. Lambert, W. De Neve, Y. Dhondt, and R. Van de Walle, "Flexible macroblock ordering in H.264/AVC," *Journal of Visual Communication and Image Representation*, vol. 17, no. 2, pp. 358–375, Apr. 2006.

[6] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," *Proceedings of the IEEE*, vol. 87, no. 10, pp. 1707–1723, Oct. 1999.

[7] J. Ott, S. Wenger, N. Sato, C. Burmeister, and J. Rey, "Extended RTP profile for real-time transport control protocol (RTCP)-based feedback (RTP/AVPF)," IETF, RFC 4585, Jul. 2006.

[8] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," in *Proc. International Conference on Learning Representations (ICLR)*, Toulon, France, 2017.

[9] D. Minnen, J. Ballé, and G. D. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, Montréal, Canada, 2018.

[10] J. Li, B. Li, and Y. Lu, "Deep contextual video compression," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[11] J. Li, B. Li, and Y. Lu, "Hybrid spatial-temporal entropy modelling for neural video compression," in *Proc. ACM International Conference on Multimedia (ACM MM)*, Lisbon, Portugal, 2022, pp. 1503–1511.

[12] J. Li, B. Li, and Y. Lu, "Neural video compression with diverse contexts," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, Canada, 2023, pp. 22 616–22 626.

[13] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, Sep. 2019.

[14] S. Wang, J. Dai, Z. Liang, K. Niu, Z. Si, C. Dong, X. Qin, and P. Zhang, "Wireless deep video semantic transmission," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 214–229, Jan. 2023.

[15] J. Y. Liao and J. D. Villasenor, "Adaptive intra update for video coding over noisy channels," in *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 3, Lausanne, Switzerland, 1996, pp. 763–766.

[16] Z. Ma, L. Song, R. Xie, and W. Zhang, "Motion adaptive intra refresh for low delay HEVC encoding," in *International Forum on Digital TV and Wireless Multimedia Communications*, 2019, pp. 379–390.

[17] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications," IETF, RFC 3550, Jul. 2003.

[18] C. G. Bampis, Z. Li, and A. C. Bovik, "Spatiotemporal feature integration and model fusion for full reference video quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 8, pp. 2256–2270, 2019.

[19] F. Bossen, B. Bross, K. Sühring, and D. Flynn, "HEVC complexity and implementation analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1685–1696, Dec. 2012.