The 38th Annual AAAI Conference on Artificial Intelligence

FEBRUARY 20-27, 2024 | VANCOUVER, CANADA

(Paper ID: 7069)

Offline and Online Optical Flow Enhancement for Deep Video Compression

Chuanbo Tang Xihua Sheng Zhuoyuan Li Haotian Zhang Li Li Dong Liu

University of Science and Technology of China

1. Contributions

- Proposing offline enhancement on the optical flows with the guidance of MV of VTM.
- Enhancing the adaptivity of the optical flows by online optimizing the latent features of the optical flows according to the contents of different coding sequences in the inference stage.
- Superior compression performance on two state-of-the-art schemes DCVC and DCVC-DC without increasing the model or computational complexity of the decoder side.

2. Motivation and Analysis

I. Motivation

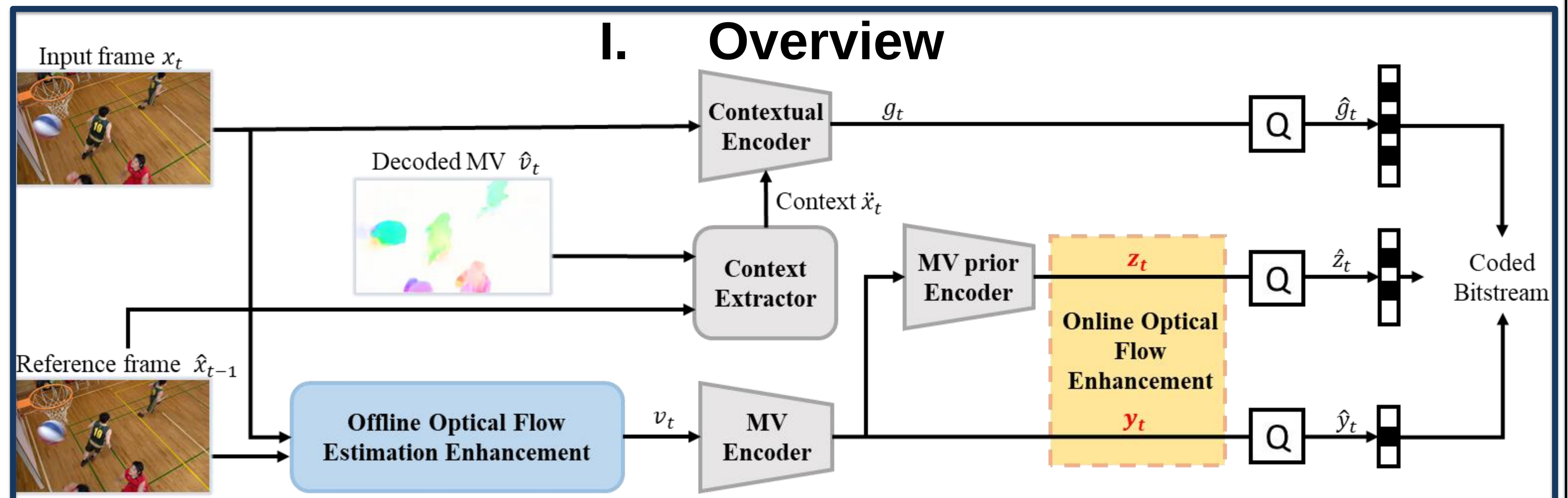
- Mainstream deep video compression networks often adopt pre-trained optical flow estimation networks as motion estimation module, which may be less suitable for video compression.
- The pre-trained optical flow estimation networks are trained to perform inter-frame prediction as accurately as possible, but the optical flows themselves may cost too many bits to encode.
- The optical flow estimation networks are trained on synthetic data, and may not generalize well enough to real-world videos.
- In the inference stage, the motion information is obtained by a simple forward pass through the motion estimation and encoder.

II. Analysis

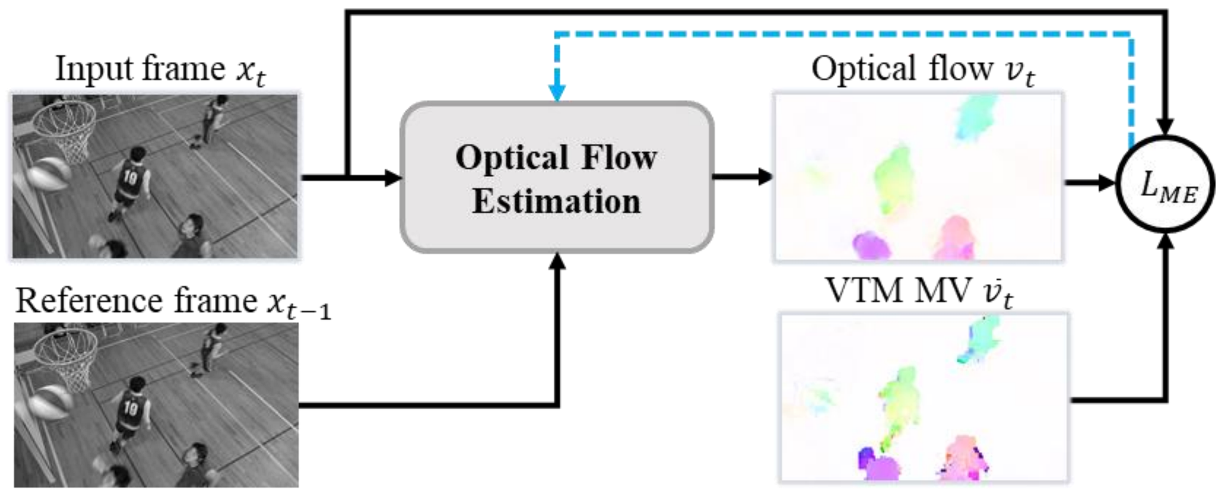
- MV of VTM, searched for the best rate-distortion (RD) performance for each coding sequence, is believed to achieve a better rate-distortion trade-off.
- The online searching strategy in VTM, rate-distortion-optimization (RDO), can achieves content-adaptive video compression.

3. Method

I. Overview



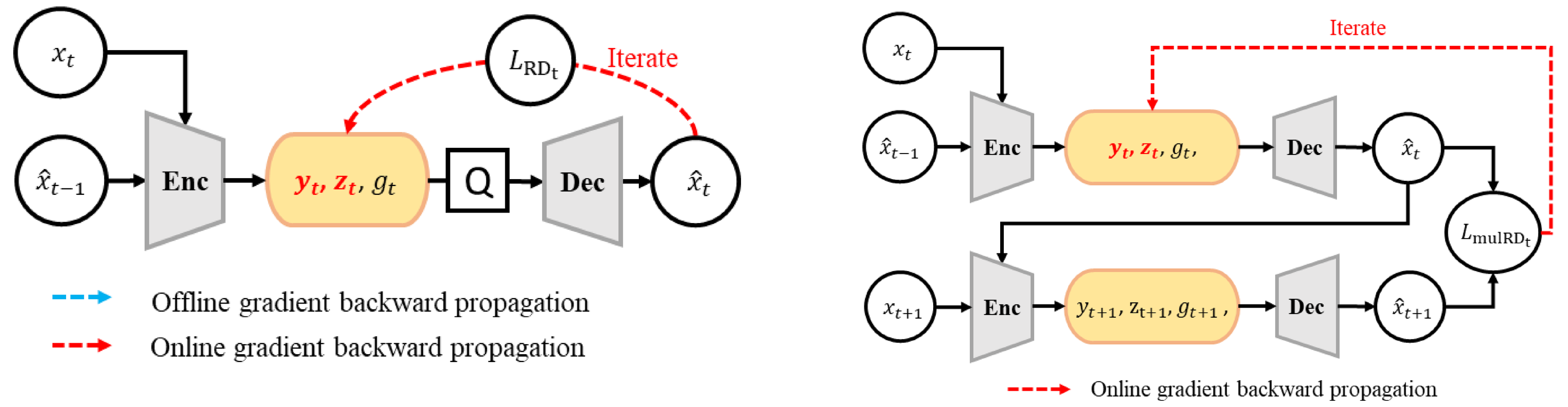
II. Offline Enhancement



- We fine-tune the pre-trained Spynet under the guidance of the extracted MV $\overline{v}_t$. The loss function including the End Point Error (EPE) loss between optical flows and MV and Mean Squared Error (MSE) loss between the input frame and the warp frame $\breve{x}_t$.

$$L_{ME} = \frac{1}{mn}\sum_{i,j}\sqrt{(v_i-\overline{v}_i)^2+(v_j-\overline{v}_j)^2}+\lambda_{ME}*d(x_t,\breve{x}_t)$$

III. Online Enhancement



(a) Single-Frame Online Optimization **(b) Multi-Frame Online Optimization**

- In the inference stage, we online optimize the latent features of the optical flows with a gradient descent-based algorithm minimizing the RD loss in single-frame level and multi-frame level.

$$\tilde{L}^i_{RD_t} = \sum_{j=t}^{W}\alpha_j[\lambda d(x_j,\tilde{x}^i_j)+H(\tilde{y}^i_j)+H(\tilde{z}^i_j)+H(\tilde{g}^i_j)]$$

4. Experiment

I. Comparison with Baseline and SOTA Methods

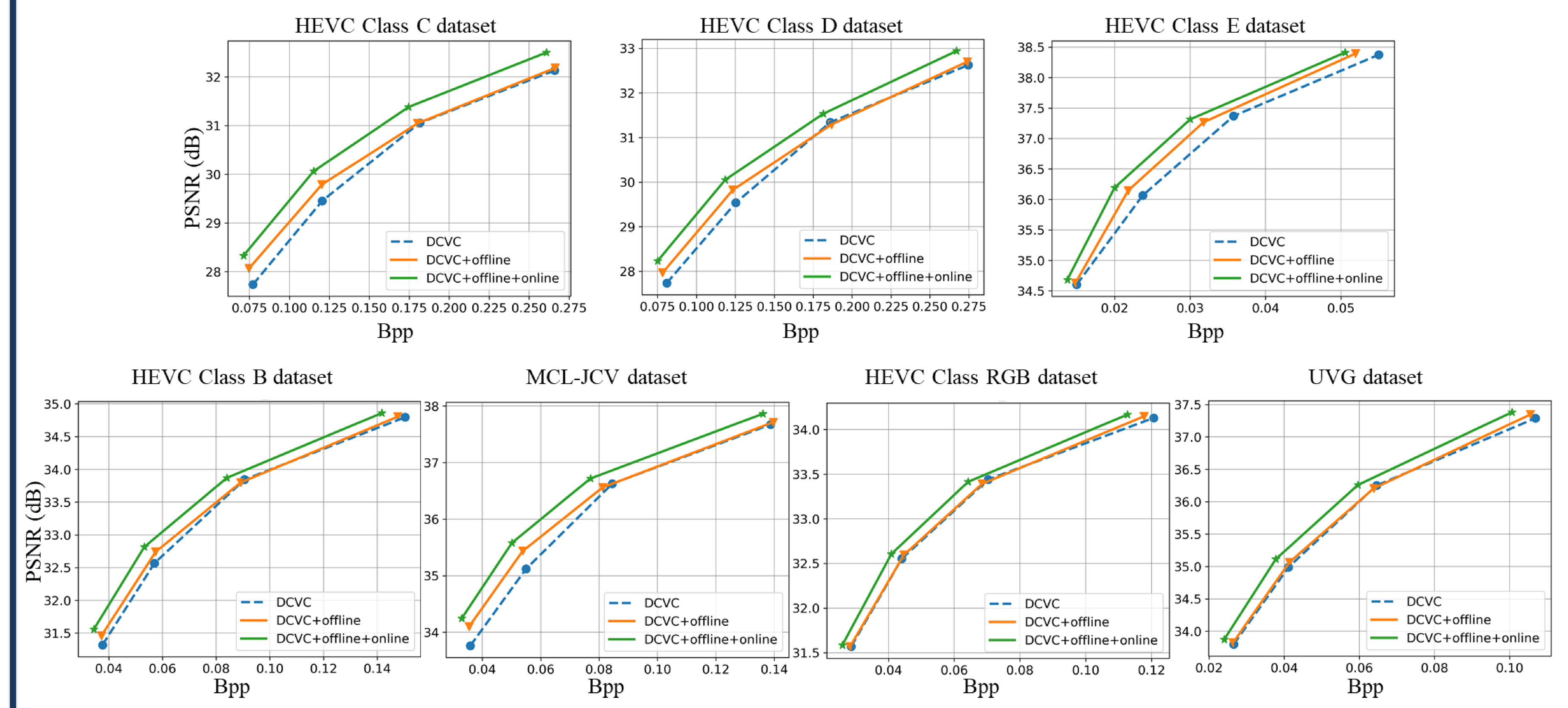
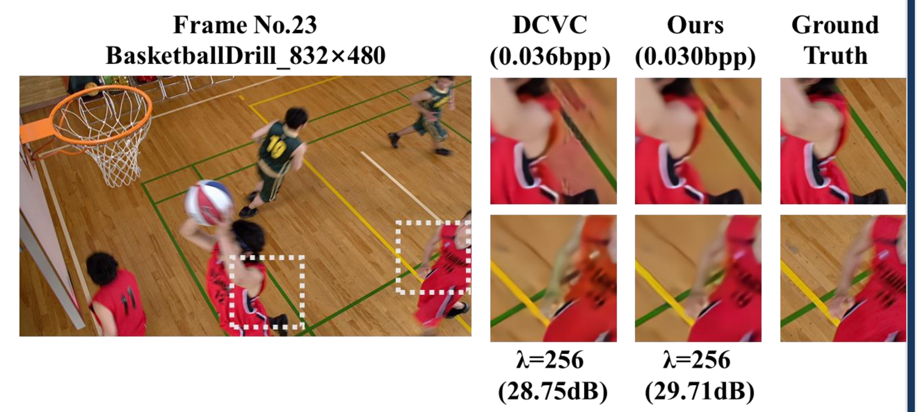


Table 1: Effectiveness of the offline and online enhancement on SOTA method DCVC-DC. BD-Rate(%) comparison for PSNR. Negative values in BDBR represent the bitrate saving.

	B	C	D	UVG	Average
DCVC-DC	0.0	0.0	0.0	0.0	0.0
DCVC	66.6	79.7	76.7	78.7	75.4
DCVC-DC + offline	-0.7	-1.0	-2.1	-0.4	-1.1
DCVC-DC + offline + online	-2.8	-4.9	-4.6	-4.2	-4.1



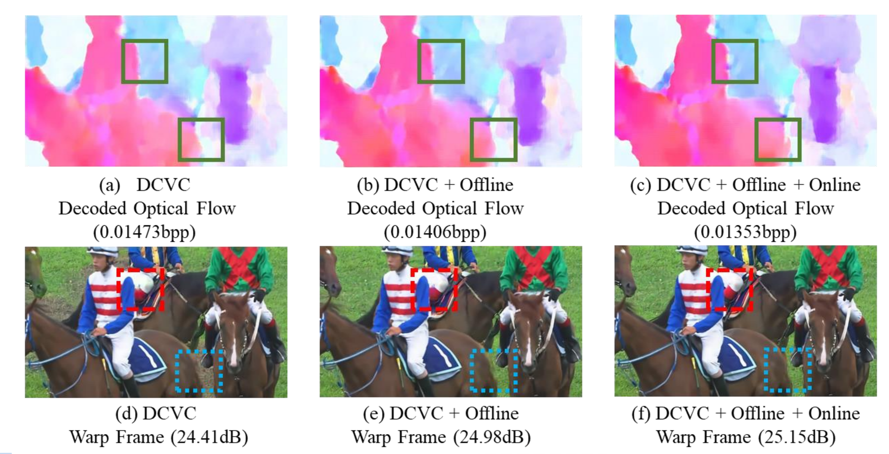
II. Ablation Study

Ablation study of Offline Enhancement and Online Enhancement

Offline	Online	B	C	D	E	RGB	UVG	MCL	Average
✗	✗	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
✓	✗	-3.0	-5.9	-4.4	-7.9	-0.7	-1.3	-6.7	-4.3
✗	✓	-10.7	-14.3	-11.1	-9.0	-8.5	-10.1	-11.3	-10.7
✓	✓	-12.0	-17.1	-13.1	-15.3	-8.8	-10.5	-16.9	-13.4

Ablation Study of Online updating Times

U	C	D	ENC_T C(s)	DEC_T C(s)	ENC_T D(s)	DEC_T D(s)
0	0.0	0.0	2.71	6.94	0.70	1.91
100	-6.1	-5.1	28.15	6.84	10.42	1.90
500	-9.6	-7.9	132.78	6.95	48.99	1.87
1000	-10.8	-8.6	269.20	6.73	92.58	1.89
1500	-11.2	-8.7	388.73	6.86	141.03	1.91
2000	-11.5	-9.1	530.10	6.84	190.64	1.89
2500	-11.6	-9.2	674.54	6.89	239.05	1.88



Ablation Study of Online updating Frames

W	C	D	ENC_T C(s)	DEC_T C(s)	ENC_T D(s)	DEC_T D(s)
2	0.0	0.0	518.25	6.84	187.82	1.89
3	-0.5	-0.4	1631.35	6.84	546.99	1.88
4	-0.7	-0.7	2187.58	6.82	683.04	1.88
5	-0.8	-0.8	2706.39	6.87	874.56	1.86