Neural Video Compression with Context Modulation

Chuanbo Tang   Zhuoyuan Li   Yifan Bian   Li Li   Dong Liu

CVPR Nashville JUNE 11-15, 2025

1. Contributions

1. Proposing **context modulation** in the conditional coding-based framework to generate high-quality temporal context exploiting the reference information in both pixel and feature domain.

2. Proposing **flow orientation** to mine inter-frame correlation between the reference frame and prediction frame.

3. Proposing **context compensation** to modulate the propagated temporal context with oriented context.

4. **SOTA compression performance** compared with both H.266/VVC (**22.7%**) and previous SOTA neural video codec DCVC-FM (**10.1%**).

2. Motivation and Analysis

I.   Motivation

1. Previous conditional coding-based neural video codecs (NVCs) adopted the inherent reference propagation structure mechanism in temporal context generation with input of propagated adjacent reference feature F_{t-1} and reconstructed MV $\hat{v}_t$.

2. DCVC-FM has mitigated the error propagation in NVCs by manually switching the input of context generation from the propagated reference feature F_{t-1} to reference frame $\hat{x}_{t-1}$.
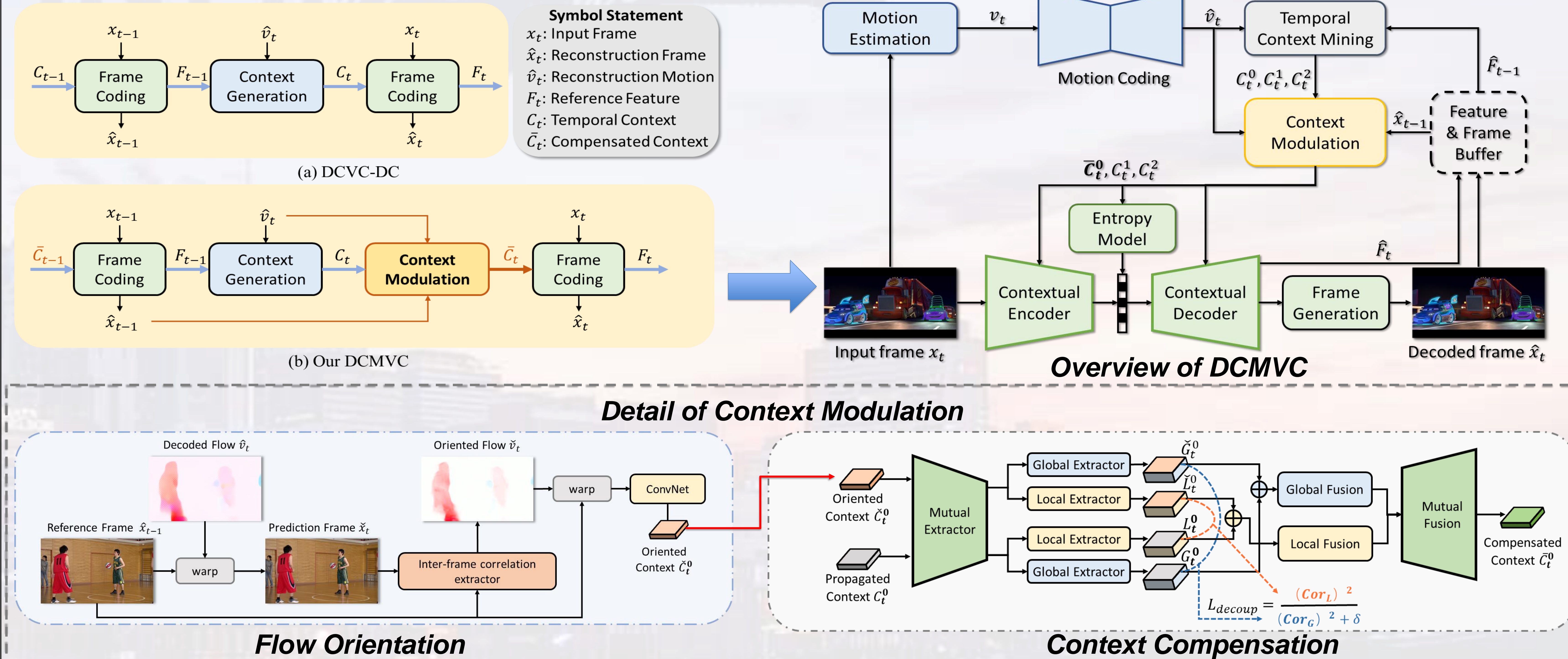
II.   Analysis

1. **Reference frame** $\hat{x}_{t-1}$, constrained by the distortion loss in the rate-distortion (RD) loss function, though contains less information than propagated reference feature F_{t-1}, yet conveys less irrelevant information in the prediction chain.

2. Further exploiting the potential of reference frame $\hat{x}_{t-1}$ in context generation can enhance the reference mechanism, enabling better temporal context modeling.

Old context C_t^0 Our context $\check{C}_t^0$ Modulated context $\bar{C}_t^0$

Visualization of Different temporal contexts

3. Framework (*Proposed DCMVC*)

Symbol Statement
x_t: Input Frame
$\hat{x}_t$: Reconstruction Frame
$\hat{v}_t$: Reconstruction Motion
F_t: Reference Feature
C_t: Temporal Context
$\bar{C}_t$: Compensated Context

(a) DCVC-DC

(b) Our DCMVC

Overview of DCMVC

Detail of Context Modulation

Flow Orientation

$$L_{decoup} = \frac{(Cor_L)^2}{(Cor_G)^2 + \delta}$$

Context Compensation

4. Experiment Results of *Proposed DCMVC*

Comparison with SOTA Methods
(BD-Rate(%) for RGB-PSNR)

		UVG	MCL-JCV	HEVC B	HEVC C	HEVC D	HEVC E	USTC-TD	Average
IP32	VTM	0.0	0.0	0.0	0.0	0.0	0.0	**0.0**	0.0
	DCVC [25]	133.9	106.6	119.6	152.5	110.9	274.8	139.6	148.3
	DCVC-TCM [47]	23.1	38.2	32.8	62.1	29.0	75.8	75.3	48.0
	DCVC-HEM [26]	−17.2	−1.6	−0.7	16.1	−7.1	20.7	20.6	4.4
	SDD [48]	−19.7	−7.1	−13.7	−2.3	−24.9	−8.4	7.7	−9.8
	DCVC-DC [27]	−25.9	−14.4	−13.9	−8.8	−27.7	−19.1	10.8	−14.1
	DCVC-FM [28]	−20.4	−8.1	−10.3	−8.4	−25.8	−21.9	25.4	−9.9
	Our DCMVC	**−30.6**	**−17.3**	**−14.5**	**−14.4**	**−31.6**	**−28.1**	0.4	**−19.4**
IP-1	VTM	0.0	0.0	0.0	0.0	0.0	0.0	**0.0**	0.0
	DCVC [25]	259.5	160.0	212.2	254.8	180.6	858.4	168.2	276.2
	DCVC-TCM [47]	61.6	55.5	61.7	99.6	50.4	213.9	90.2	90.4
	DCVC-HEM [26]	1.2	4.9	10.0	30.0	−1.1	68.6	27.2	20.1
	SDD [48]	−5.5	−0.3	−2.2	16.9	−18.2	46.4	12.3	7.1
	DCVC-DC [27]	−21.2	−13.0	−10.8	−0.1	−24.2	−7.7	11.9	−9.3
	DCVC-FM [28]	−24.3	−12.5	−11.7	−8.2	−28.5	−26.6	23.9	−12.6
	Our DCMVC	**−35.6**	**−22.8**	**−16.8**	**−14.8**	**−34.6**	**−36.5**	1.9	**−22.7**

Ablation Study

	M_a	M_b	M_c	M_d	M_e	M_f	M_g
Flow orientation		✓		✓	✓		✓
Context compensation			✓	✓	✓		✓
Decoupling loss					✓		✓
Long-sequence training						✓	✓
BD-Rate (%)	0.0	−1.9	−3.5	−4.4	−5.4	−4.3	−10.3

Oriented flow $\check{v}_t$ Decoded flow $\hat{v}_t$ Estimated flow v_t

Visualization of Different flows

HEVC Class C BasketballDrill

Quality and bitrate cost comparison