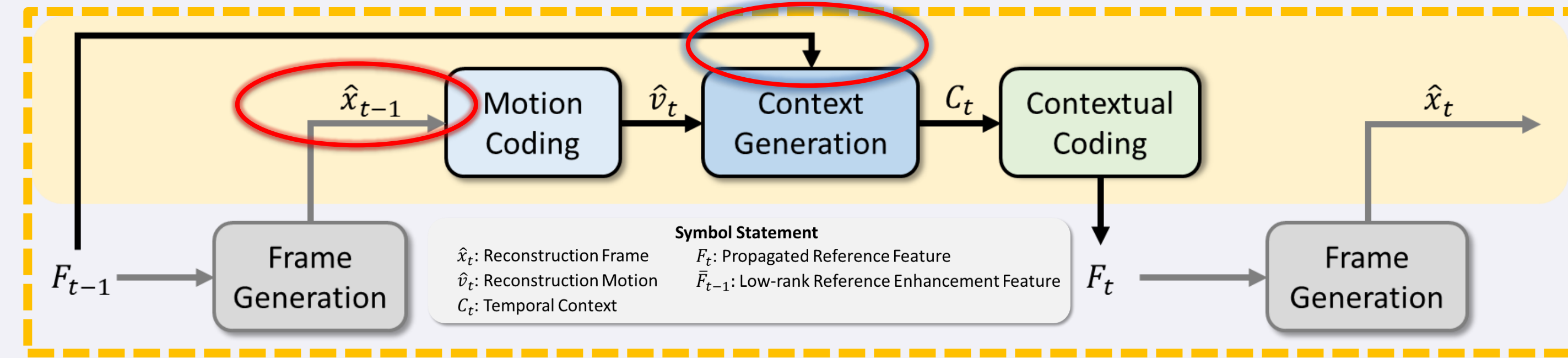


1. Contributions

- Unified framework DRHVC**: form a multi-frame feature-domain reference loop for NVC.
 - **Hierarchical Reference Structure (HRS)**: manage the multi-frame temporal contexts in the propagated reference features.
 - **HRS-guided feature-wise inter-frame prediction**: align the reference for motion estimation and compensation.
- SOTA Performance**: compared with both H.266/VVC (26.0%) and previous SOTA neural video codec DCVC-DC (28.2%).

2. Motivation

Previous methods (e.g. DCVC-DC and DCVC-FM):

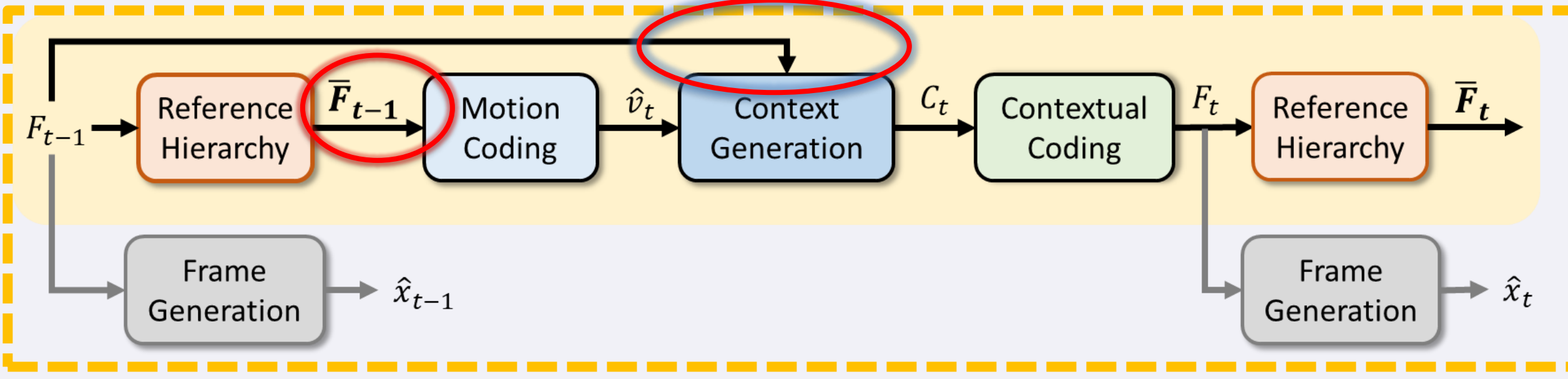


- Lack multi-frame reference for motion estimation.
- Mismatch reference in motion estimation and compensation.

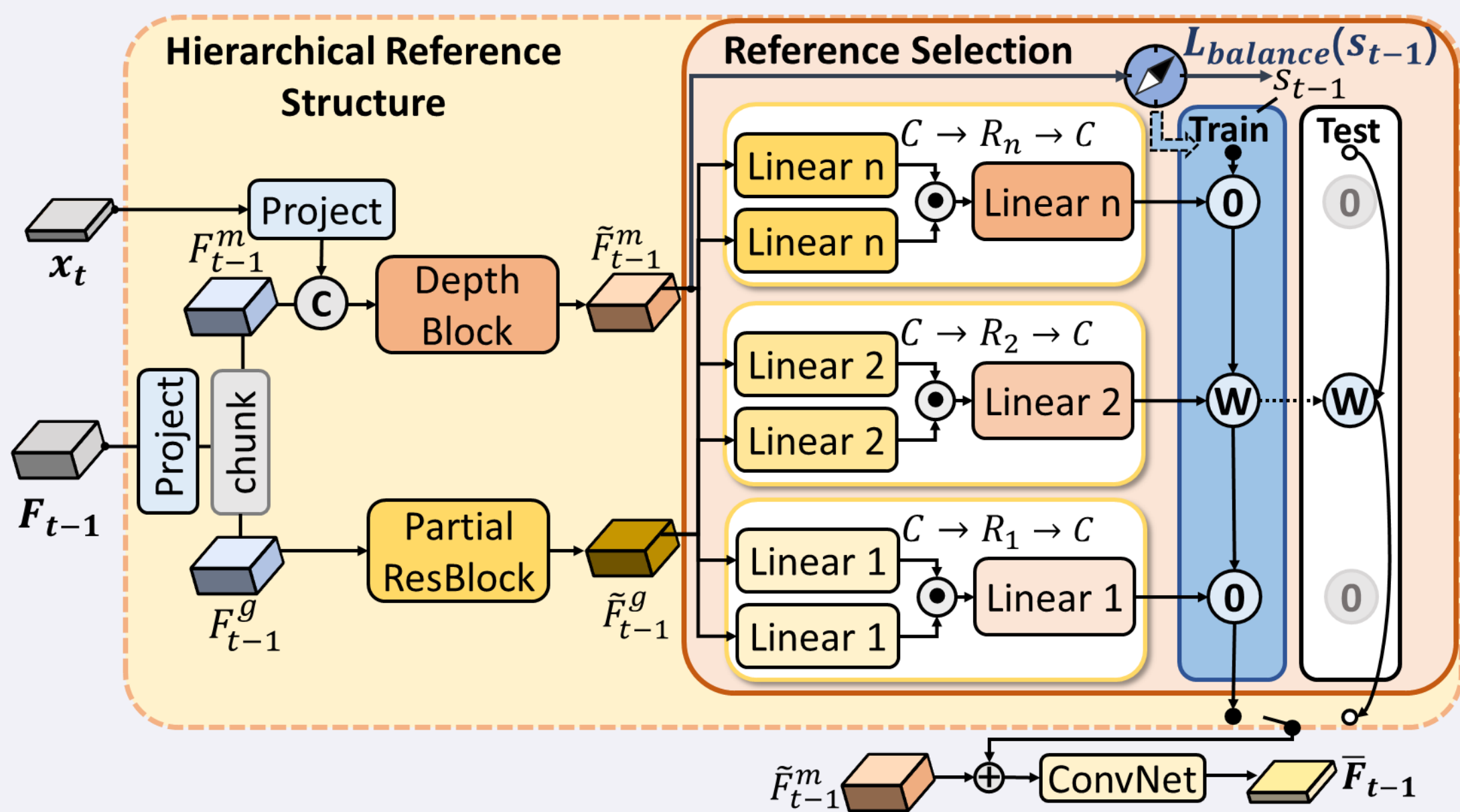
↓
 Weaken the coherence of inter-frame prediction

3. Framework (Proposed DRHVC)

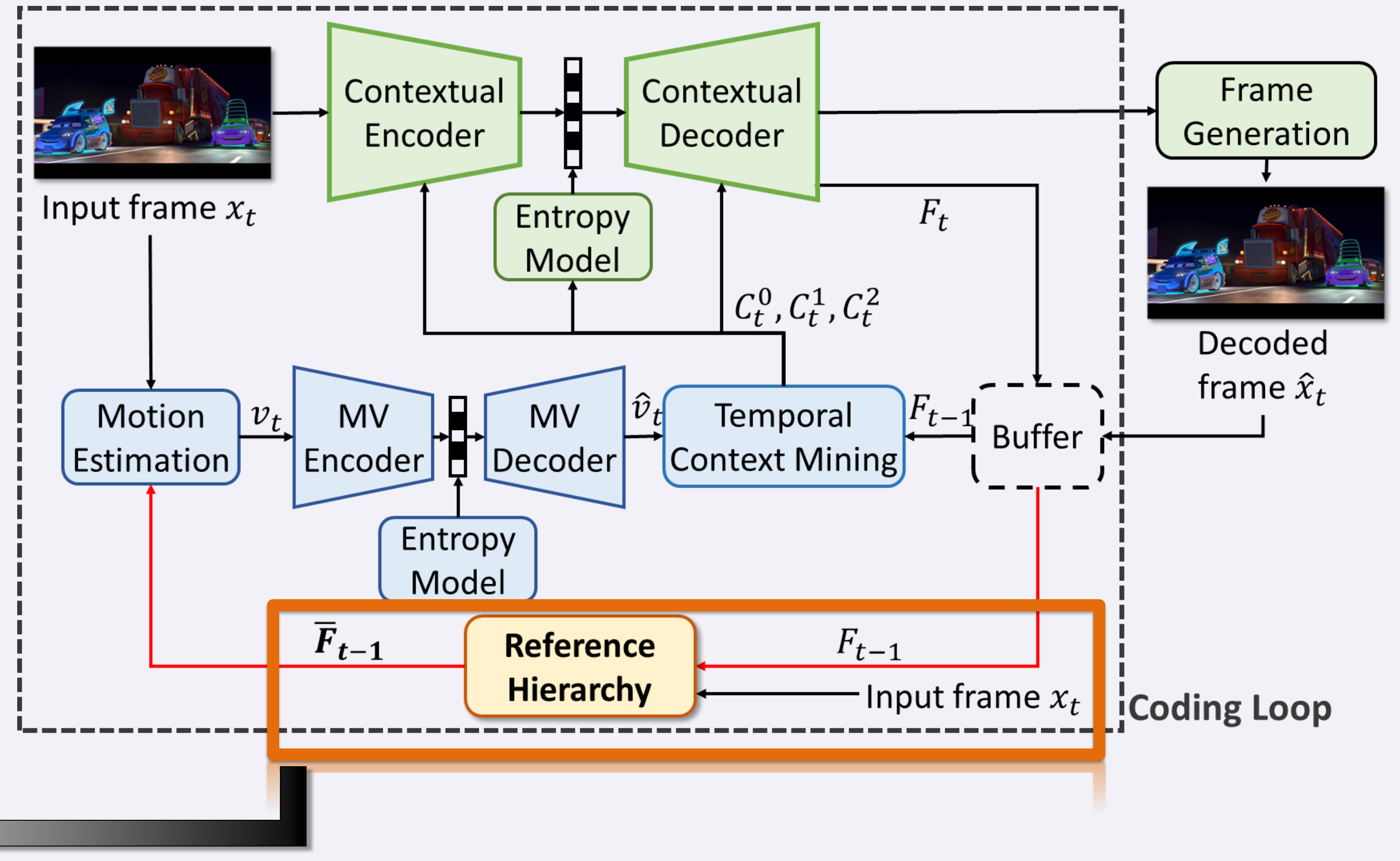
Prediction chain



Detail of Reference Hierarchy



Framework Overview



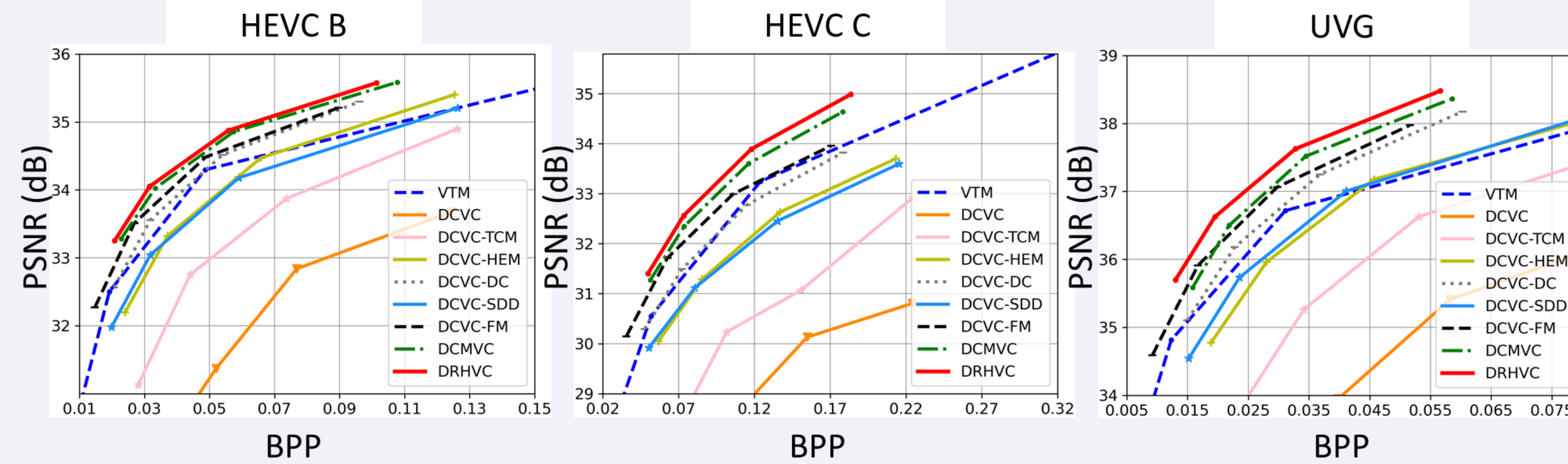
4. Experimental Results of DRHVC

Comparison with SOTA Methods (IP-1 for RGB-PSNR)

	UVG	MCL-JCV	HEVC B	HEVC C	HEVC D	HEVC E	USTC-TD	Average
VTM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DCVC	314.1	185.1	291.7	266.9	224.6	2083.2	221.2	512.4
DCVC-TCM	109.8	69.9	106.5	132.2	81.3	492.7	99.3	156.0
DCVC-HEM	23.8	10.5	27.5	33.4	11.3	139.1	24.2	38.5
SDD	19.1	5.7	30.7	35.3	-6.7	234.6	15.4	47.7
DCVC-DC	-7.8	-10.0	-0.9	6.1	-15.0	33.3	9.4	2.2
DCVC-FM	-19.8	-10.9	-12.2	-12.9	-27.1	-27.7	16.3	-13.5
DCMVC	-25.2	-20.9	-17.8	-18.0	-31.4	-23.5	-6.3	-20.4
Ours	-36.2	-24.0	-22.6	-24.2	-31.3	-38.8	-5.0	-26.0

Ablation Study

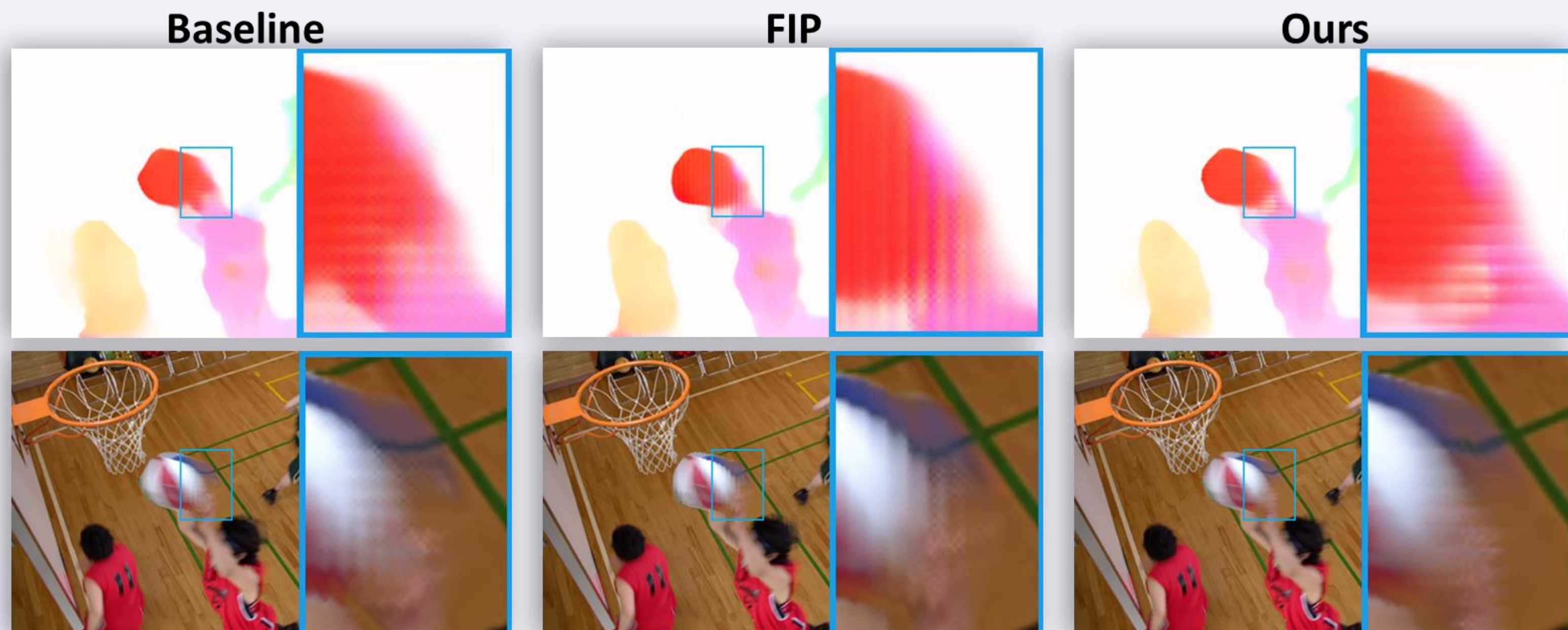
	FIP	HRS	Balance loss	Long train	BD-rate
M_a					0.0
M_b	✓				-3.9
M_c	✓	✓			-7.4
M_d	✓	✓	✓		-8.8
M_e				✓	-4.3
M_f	✓	✓	✓	✓	-13.4



➤ Reference hierarchy can **amplify** the benefits of long-sequence training.

5. Analysis

Decoded flows and corresponding warp frames



1) Decoded flow exhibits **better moving objects boundaries**.

2) Low-rank reference enhancement feature **contains more structural details and temporal information**.

References comparison (grayscale)



	MAC	Params	Enc Time	Dec Time
DCVC-DC	2786G	19.78M	663ms	557ms
DCVC-FM	2354G	18.57M	587ms	495ms
SDD	4775G	21.77M	968ms	775ms
DCMVC	4133G	20.98M	932ms	810ms
DRHVC	2953G	18.86M	697ms	558ms

Only 6% encoding complexity increasing on DCVC-DC