基于深度学习的视频压缩技术
Learning-based Video Compression

杨韧 / 苏黎世联邦理工学院 (ETH Zürich)

1920 x 1080 = 2,073,600 pixels

Uncompressed image: 2,073,600 x 1.5 = 3 MB

Uncompressed video (50 fps): 3 MB x 50 = 150 MBps (1.2 Gbps)

(1.5 GB for 10 seconds)

Lossless compression (x265): 67.8 MBps (542 Mbps)

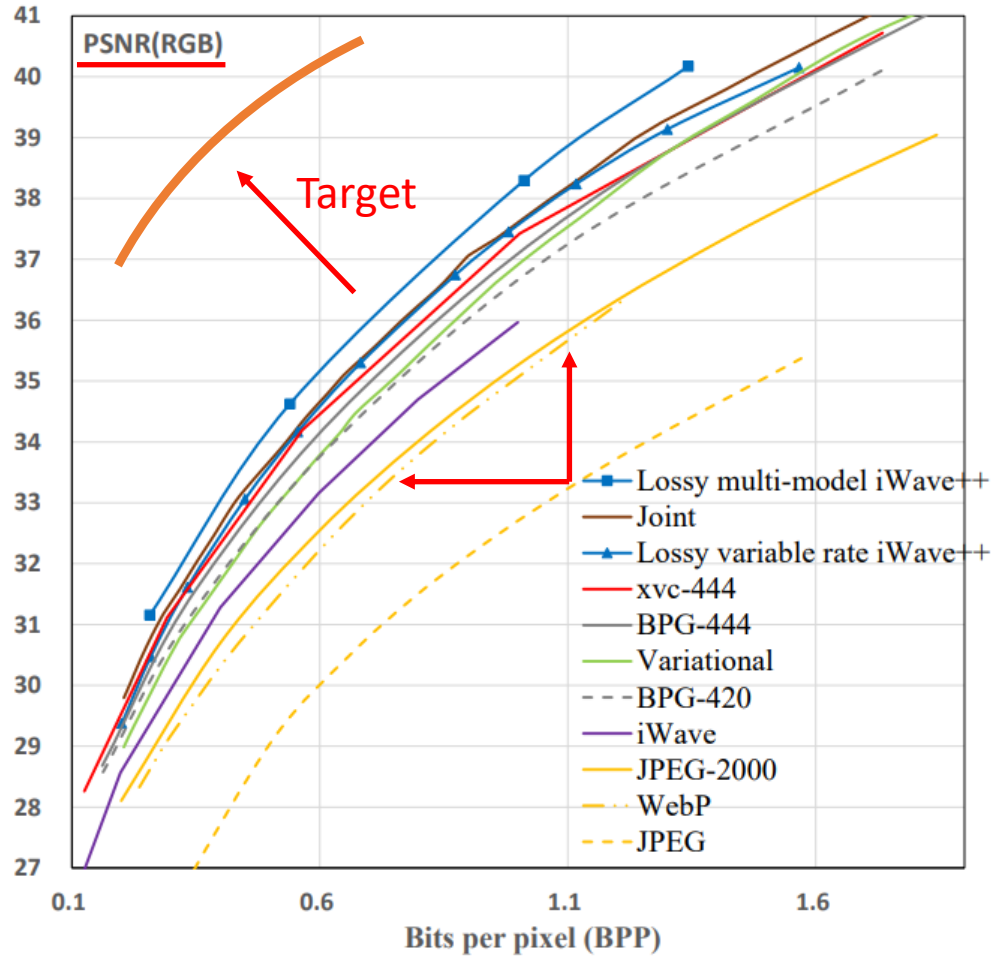
Lossy compression: 500 kbps, 1 Mbps, 3 Mbps, …

Image/video compression plays an important role in multimedia streaming, online conference, data storage, etc.

Rate-distortion trade-off

Metrics: PSNR, (MS-)SSIM, NIMA, LPIPS, user studies, etc.

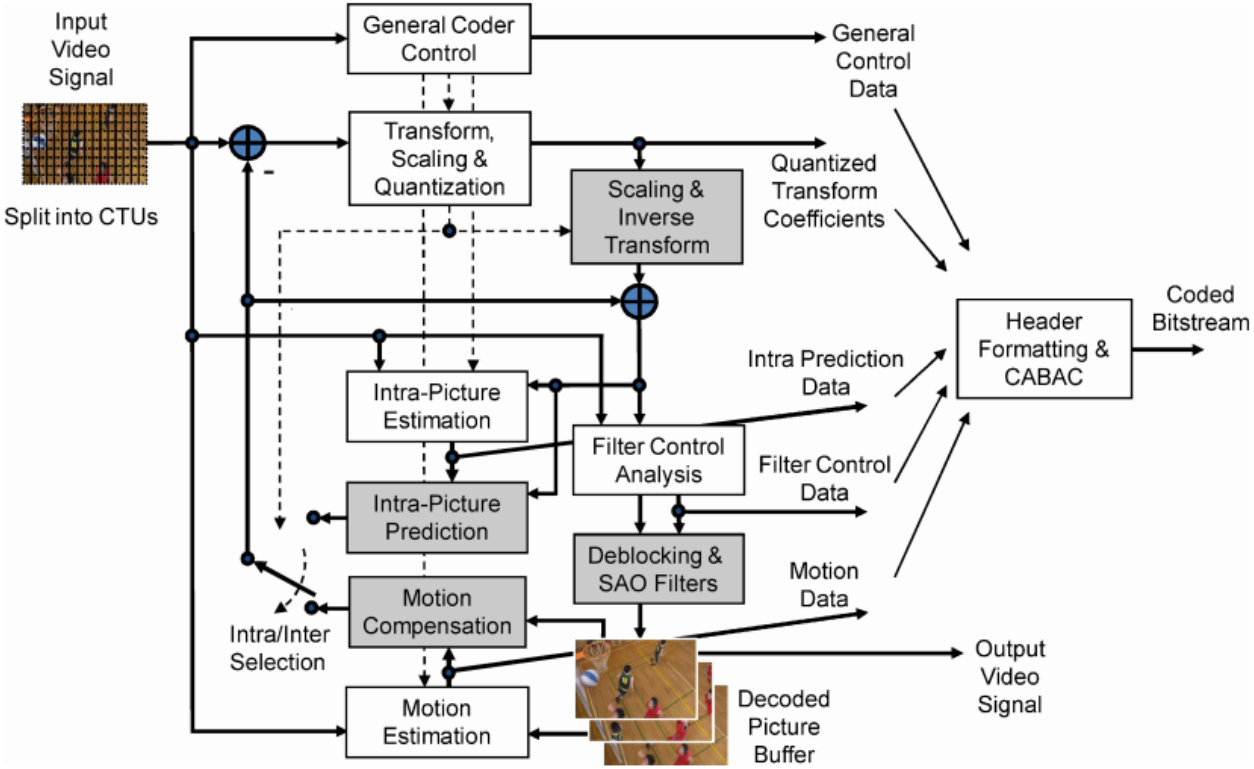
Higher quality
(lower distortion)



Better quality at the same bit-rate

Lower bit-rate at the same quality

Higher bit-rate (file size)



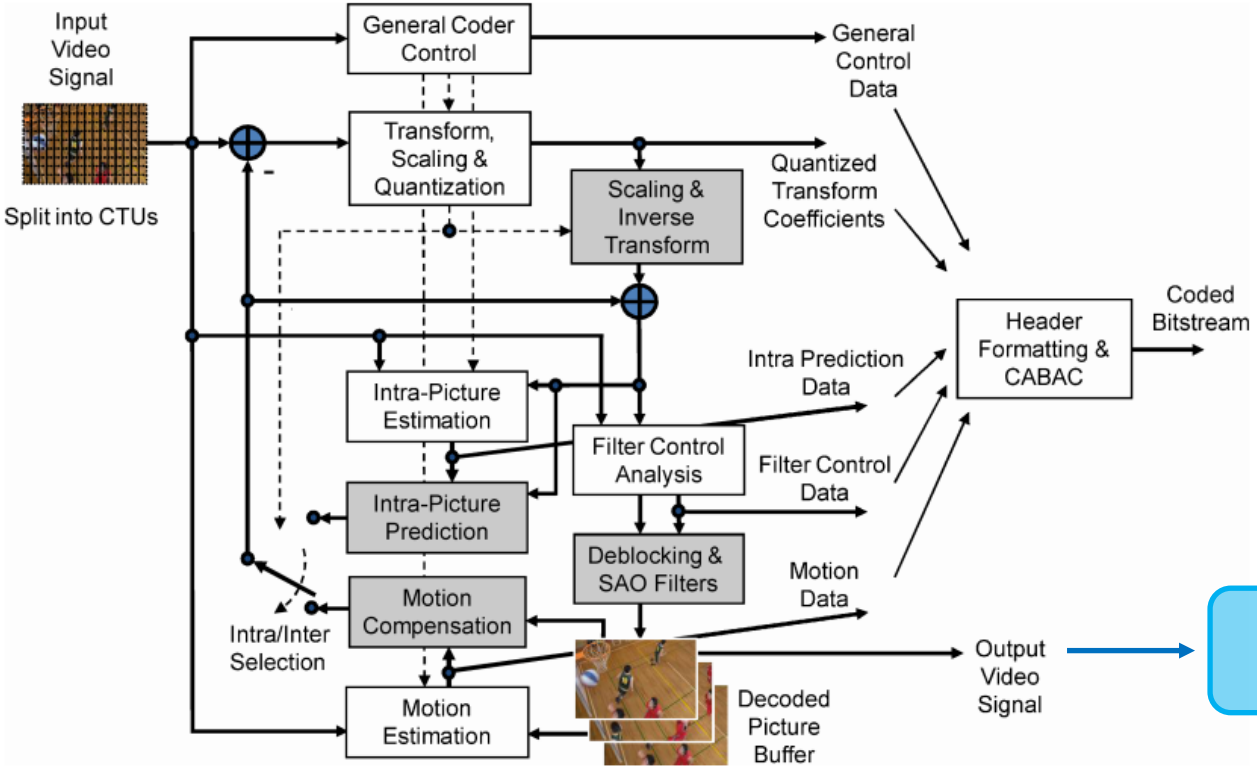
Traditional video compression framework (HEVC) [1]
Gray: modules in both encoder and decoder; White: modules only in encoder

基于深度学习的视频压缩技术
Learning-based Video Compression

- 基于深度学习的压缩视频质量增强研究
  DNN-based enhancement for compressed video
- 结合深度网络的传统视频压缩方法
  DNN-based handcrafted video compression
- 端到端优化的视频压缩深度网络
  End-to-end deep video compression network
- 兼容一般播放器的深度学习压缩方法
  Learning for compression with standard decoder

[1] Sullivan, Gary J., et al. "Overview of the high efficiency video coding (HEVC) standard." IEEE T-CSVT. 2012.

1. DNN-based enhancement for compressed video



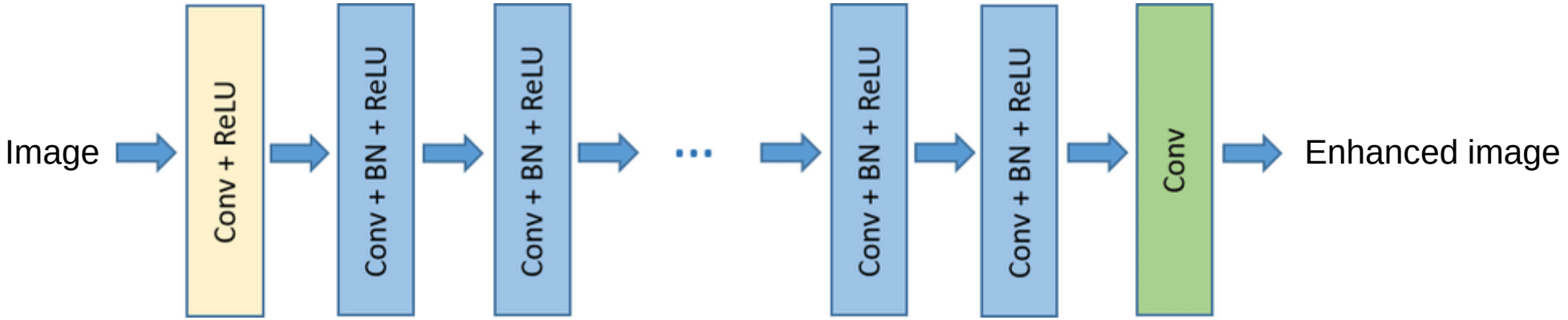
Traditional video compression framework (HEVC) [1]

Gray: modules in both encoder and decoder; White: modules only in encoder

[1] Sullivan, Gary J., et al. "Overview of the high efficiency video coding (HEVC) standard." IEEE T-CSVT. 2012.

1. DNN-based enhancement for compressed video

DNN-based image enhancement (e.g., DnCNN [2])



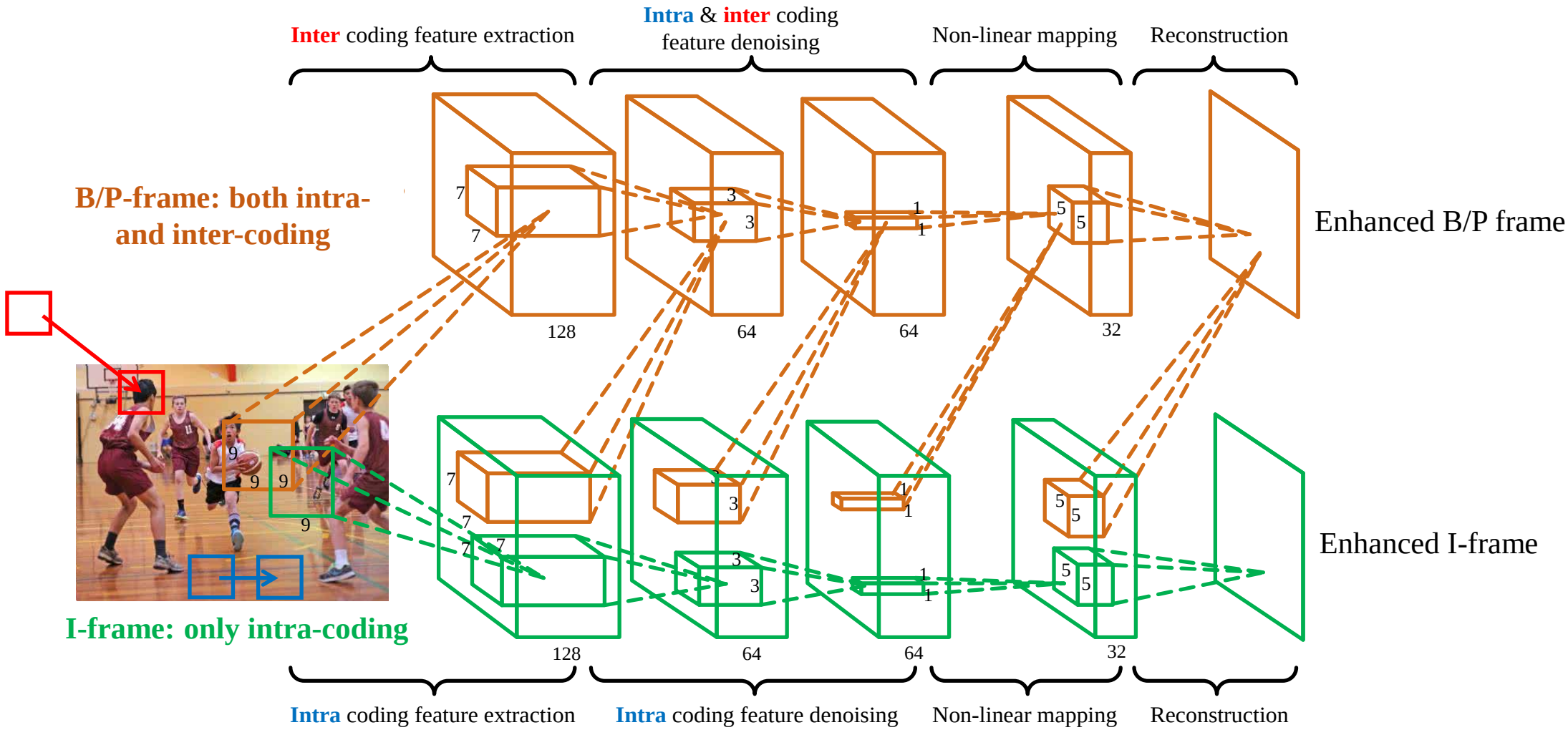
DNN-based video enhancement

- Different prediction mode (intra- and inter-prediction) in different coding units (CUs)
- Temporal correlation among consecutive video frames

[2] Zhang, Kai, et al. "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising." IEEE T-IP. 2017.

1. DNN-based enhancement for compressed video

Single-frame video compression: QE-CNN [3, 4]



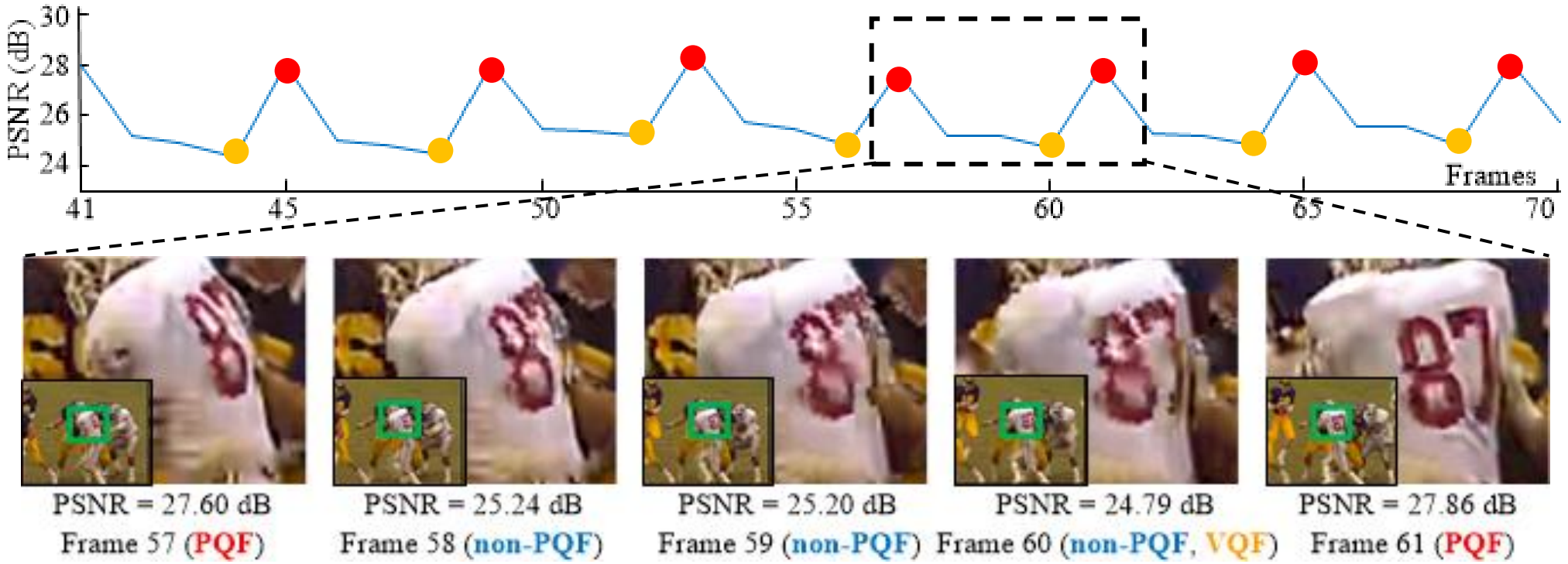
[3] Yang, Ren, et al. "Decoder-side HEVC quality enhancement with scalable convolutional neural network." IEEE ICME. 2017.
[4] Yang, Ren, et al. "Enhancing quality for HEVC compressed videos." IEEE T-CSVT. 2018.

1. DNN-based enhancement for compressed video

Multi-frame video compression: MFQE [5, 6]

Motivation: High quality neighboring frames may help to enhance low quality frames

● Peak Quality Frame (PQF) ● Valley Quality Frame (VQF)



Large quality fluctuation

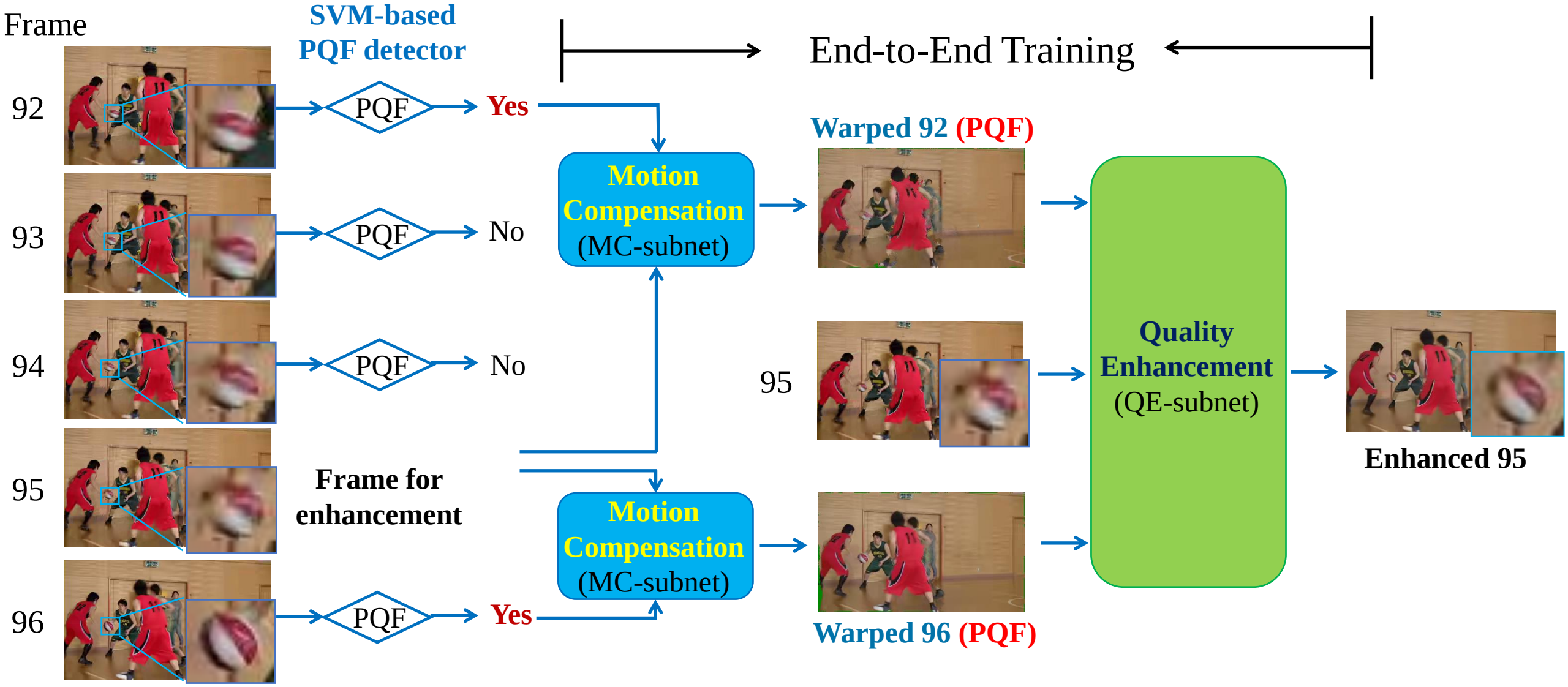
High similarity

[5] Yang, Ren, et al. "Multi-frame quality enhancement for compressed video." in CVPR. 2018.

[4] Guan, Zhenyu, et al. "MFQE 2.0: A new approach for multi-frame quality enhancement on compressed video." IEEE T-PAMI. 2020.

1. DNN-based enhancement for compressed video

Multi-frame video compression: MFQE [5, 6]

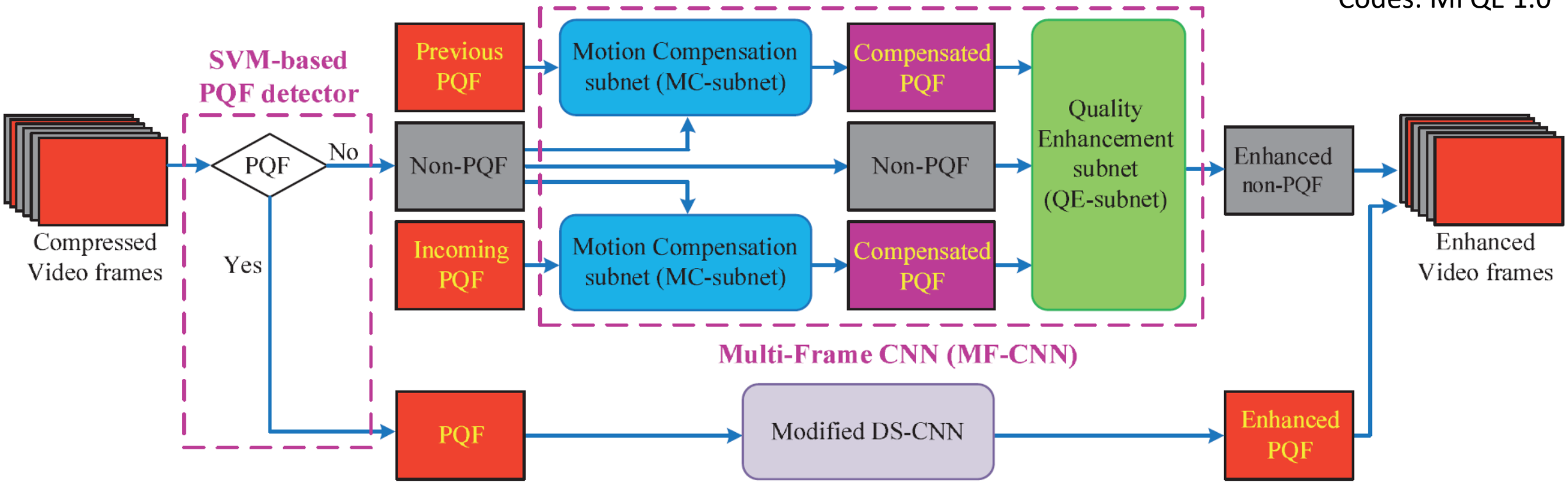


1. DNN-based enhancement for compressed video

Multi-frame video compression: MFQE [5, 6]

MFQE 1.0 (CVPR 2018)

Codes: MFQE 1.0

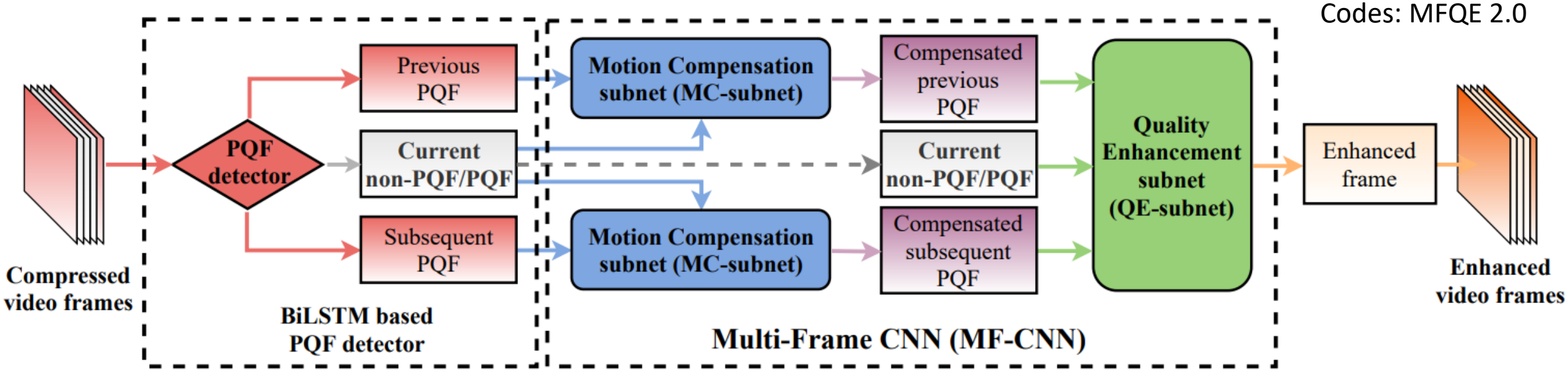


1. DNN-based enhancement for compressed video

Multi-frame video compression: MFQE [5, 6]

MFQE 2.0 (IEEE T-PAMI 2019)

Codes: MFQE 2.0

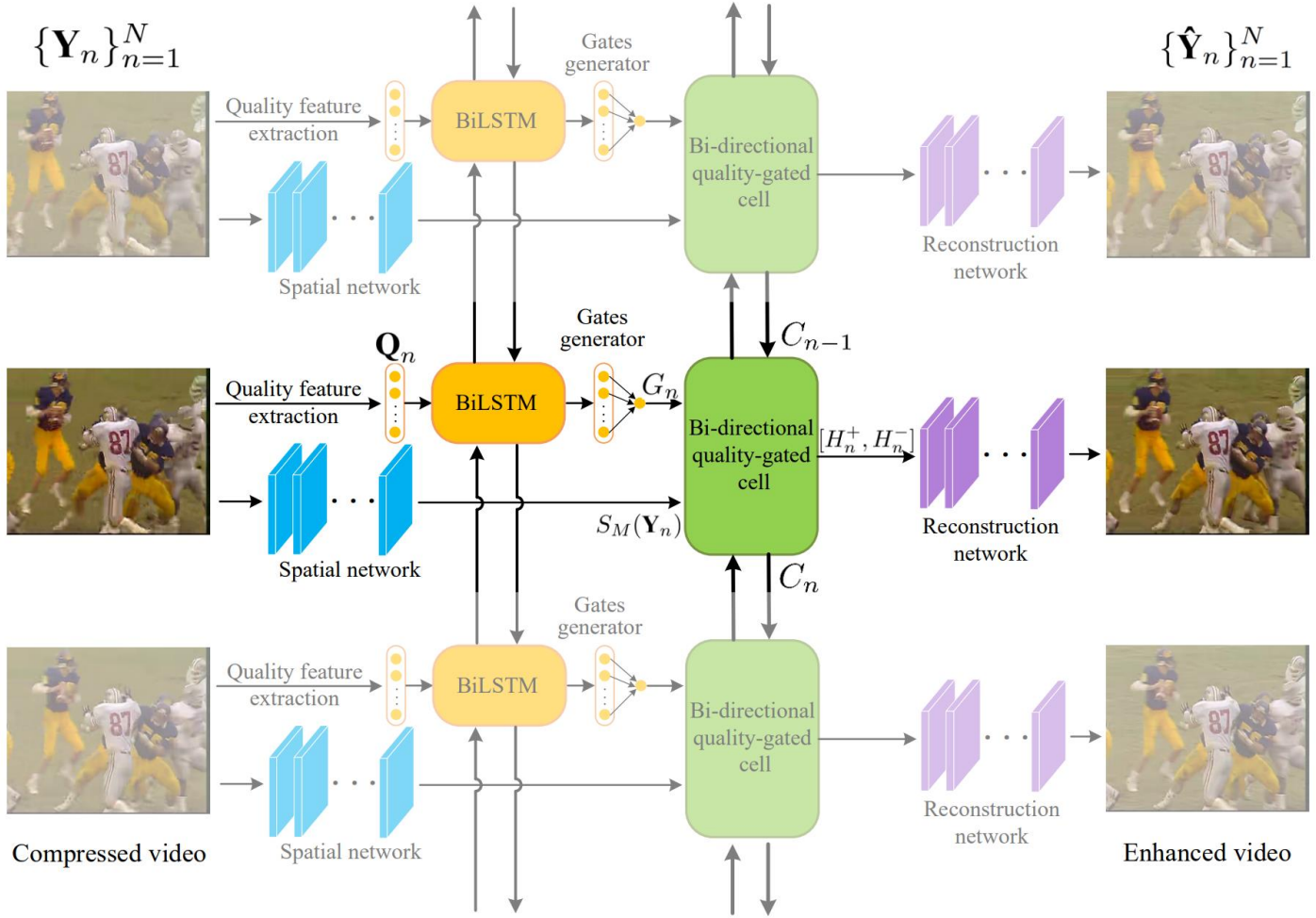


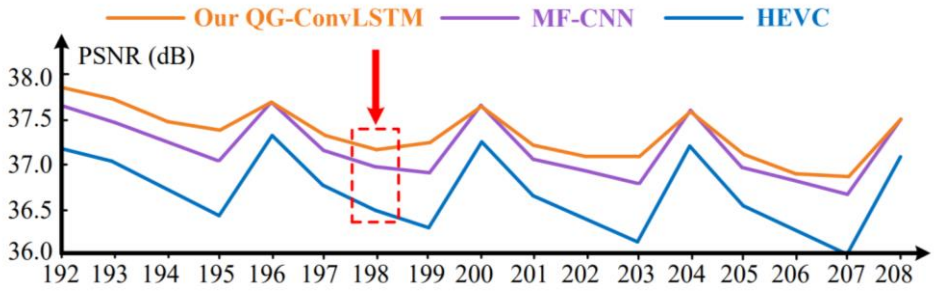
How to explore more temporal information?

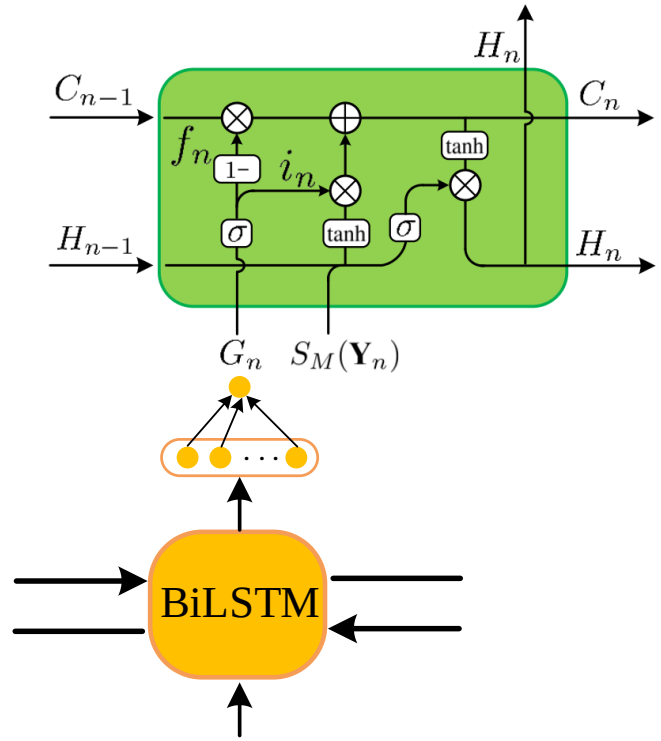
1. DNN-based enhancement for compressed video

Recurrent multi-frame video compression: QG-ConvLSTM [7]

- ✓ **Leverage information in a large range of frames**
- ✓ **Take quality fluctuation into consideration**







[38 quality-related features*, bit-rate, QP]

*http://live.ece.utexas.edu/research/quality/BRISQUE_release.zip (Mittal et al., IEEE T-IP 2012)

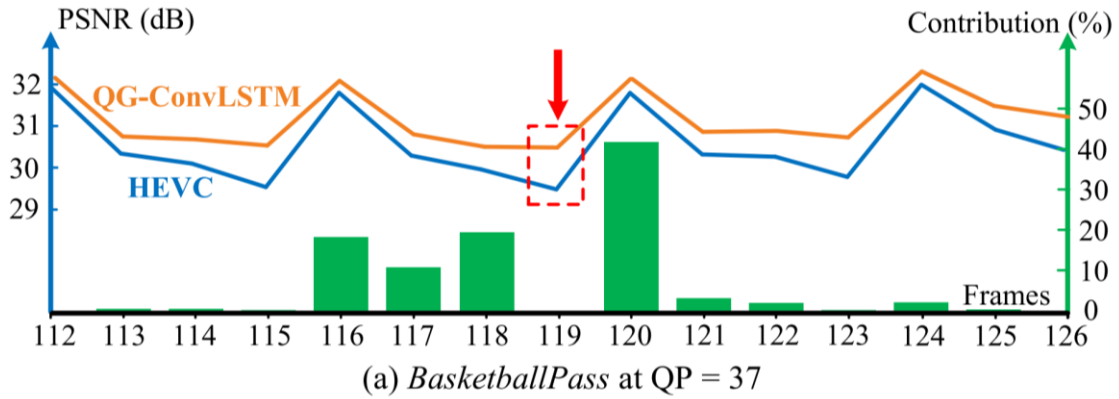
[7] Yang, Ren, et al. "Quality-gated convolutional LSTM for enhancing compressed video." in IEEE ICME. 2019.

1. DNN-based enhancement for compressed video

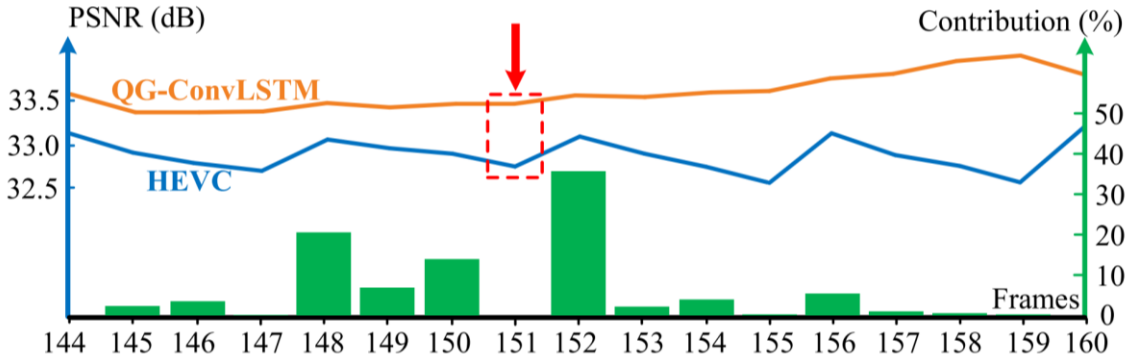
Recurrent multi-frame video compression: QG-ConvLSTM [7]

The contribution from the (*n*-*k*)-th frame: $\frac{1}{2} \cdot i_{n-k} \cdot \prod_{j=n-k+1}^{n} f_j$

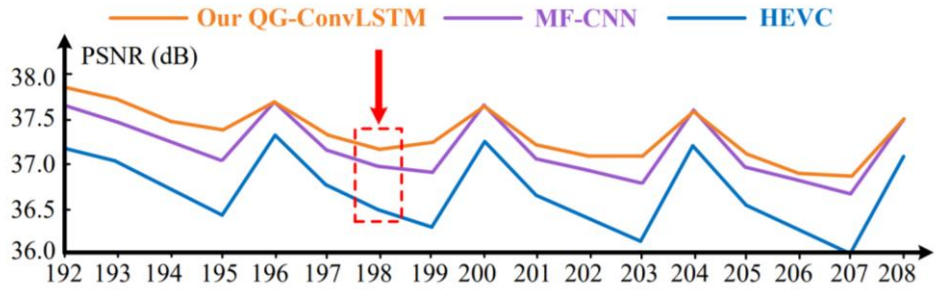
The contribution from the (*n*+*k*)-th frame: $\frac{1}{2} \cdot i_{n+k} \cdot \prod_{j=n}^{n+k-1} f_j$



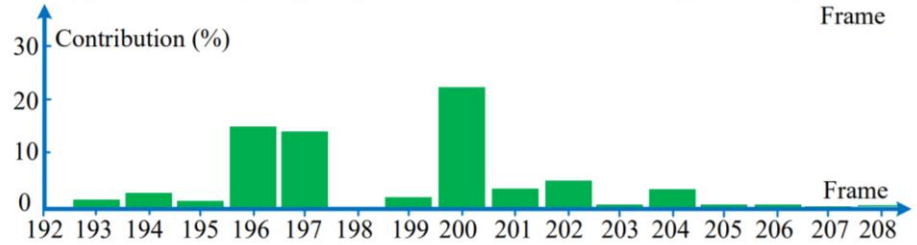
(a) *BasketballPass* at QP = 37



(b) *Vidyo3* at QP = 42



(a) Performance on quality enhancement of HEVC sequence *Vidyo1*



(b) Contribution of each frame on enhancing the 198-th frame

Codes: QG-ConvLSTM

[7] Yang, Ren, et al. "Quality-gated convolutional LSTM for enhancing compressed video." in IEEE ICME. 2019.

1. DNN-based enhancement for compressed video

Table 1. Quality enhancement performance in terms of ΔPSNR (dB)

QP	Resolution	Seq.	AR-CNN [4] (ICCV'15)	Li *et al.* [8] (ICME'17)	DnCNN [7] (TIP'17)	DCAD [11] (DCC'17)	QE-CNN [13] (TCSVT'18)	MF-CNN [15] (CVPR'18)	Proposed QG-ConvLSTM
42	352 × 288	1	0.2275	0.2242	0.2523	0.2140	0.2664	0.3968	**0.5686**
	416 × 240	2	0.1509	0.1876	0.1970	0.1556	0.1986	0.4506	**0.6066**
	416 × 240	3	0.1899	0.2266	0.2301	0.1954	0.2180	0.3908	**0.4062**
	1280 × 720	4	0.2334	0.2789	0.2999	0.2262	0.4161	0.4647	**0.6352**
	1280 × 720	5	0.1457	0.2018	0.2013	0.1358	0.4169	0.4225	**0.5824**
	1280 × 720	6	0.1715	0.2094	0.2233	0.0444	0.3370	0.3359	**0.4750**
	1920 × 1080	7	0.1073	0.1024	0.1391	0.1164	0.3032	0.4609	**0.5468**
	1920 × 1080	8	0.0766	0.1550	0.1211	0.0032	0.1396	0.5378	**0.7412**
	1920 × 1080	9	0.1864	0.1817	0.2001	0.0589	0.2129	0.4223	**0.5371**
	2560 × 1600	10	0.1382	0.1619	0.2092	0.1324	0.5609	0.7280	**0.9077**
	AVERAGE		0.1627	0.1930	0.2073	0.1282	0.3070	0.4610	**0.6007**
37	**AVERAGE**		0.1364	0.2717	0.2188	0.1556	0.3738	0.5102	**0.5871**

1: *MaD* 2: *BasketballPass* 3: *RaceHorses* 4: *Vidyo1* 5: *Vidyo3* 6: *Vidyo4* 7: *Kimono* 8: *TunnelFlag* 9: *BarScene* 10: *PeopleOnStreet*

ARCNN, Li et al., DnCNN, DCAD: Single-frame enhancement with single model;

QE-CNN: Designed for enhancement quality of both intra- and inter-coding;

MF-CNN: Taking advantage of the nearest high quality frames;

QG-ConvLSTM: Exploring the temporal correlation in a large range of frames.

Codes: QG-ConvLSTM

Examples

Compressed

Examples

Enhanced

Examples

Compressed

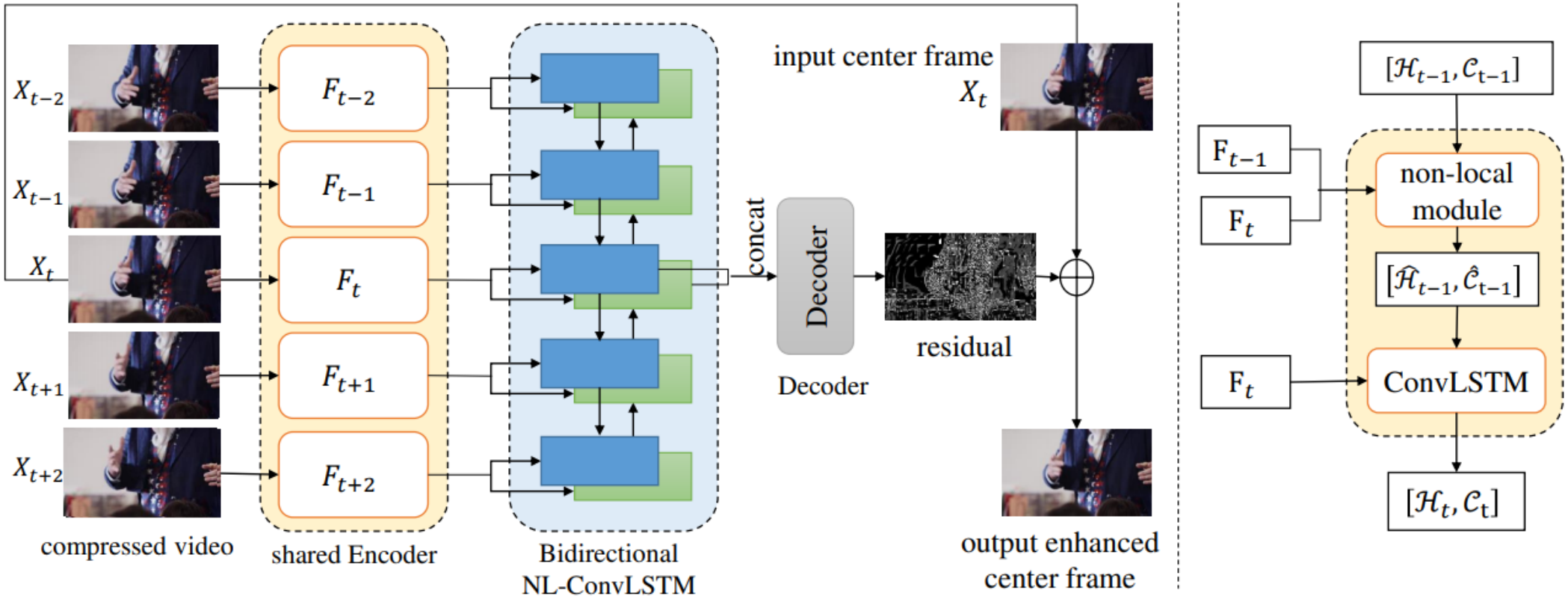
Examples

Enhanced

1. DNN-based enhancement for compressed video

Other works:

[8] Lu, Guo, et al. "Deep Kalman filtering network for video compression artifact reduction." in ECCV. 2018.
[9] Xu, Yi, et al. "Non-local ConvLSTM for video compression artifact reduction." in ICCV. 2019.



1. DNN-based enhancement for compressed video

Other works:

[8] Lu, Guo, et al. "Deep Kalman filtering network for video compression artifact reduction." in ECCV. 2018.
[9] Xu, Yi, et al. "Non-local ConvLSTM for video compression artifact reduction." in ICCV. 2019.

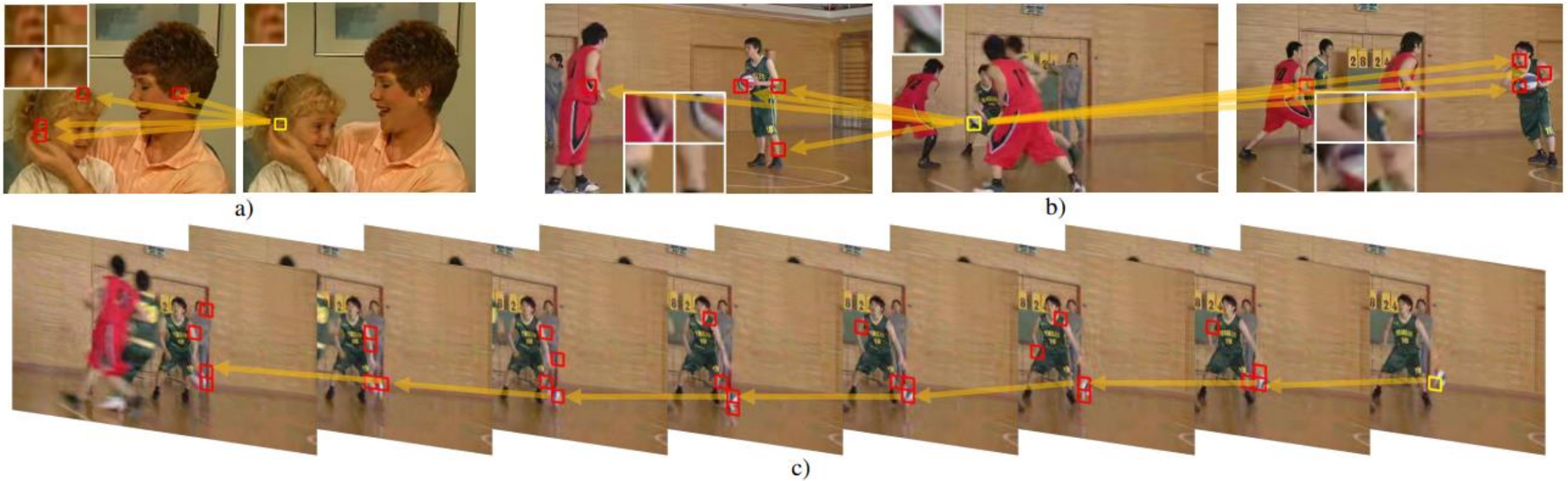
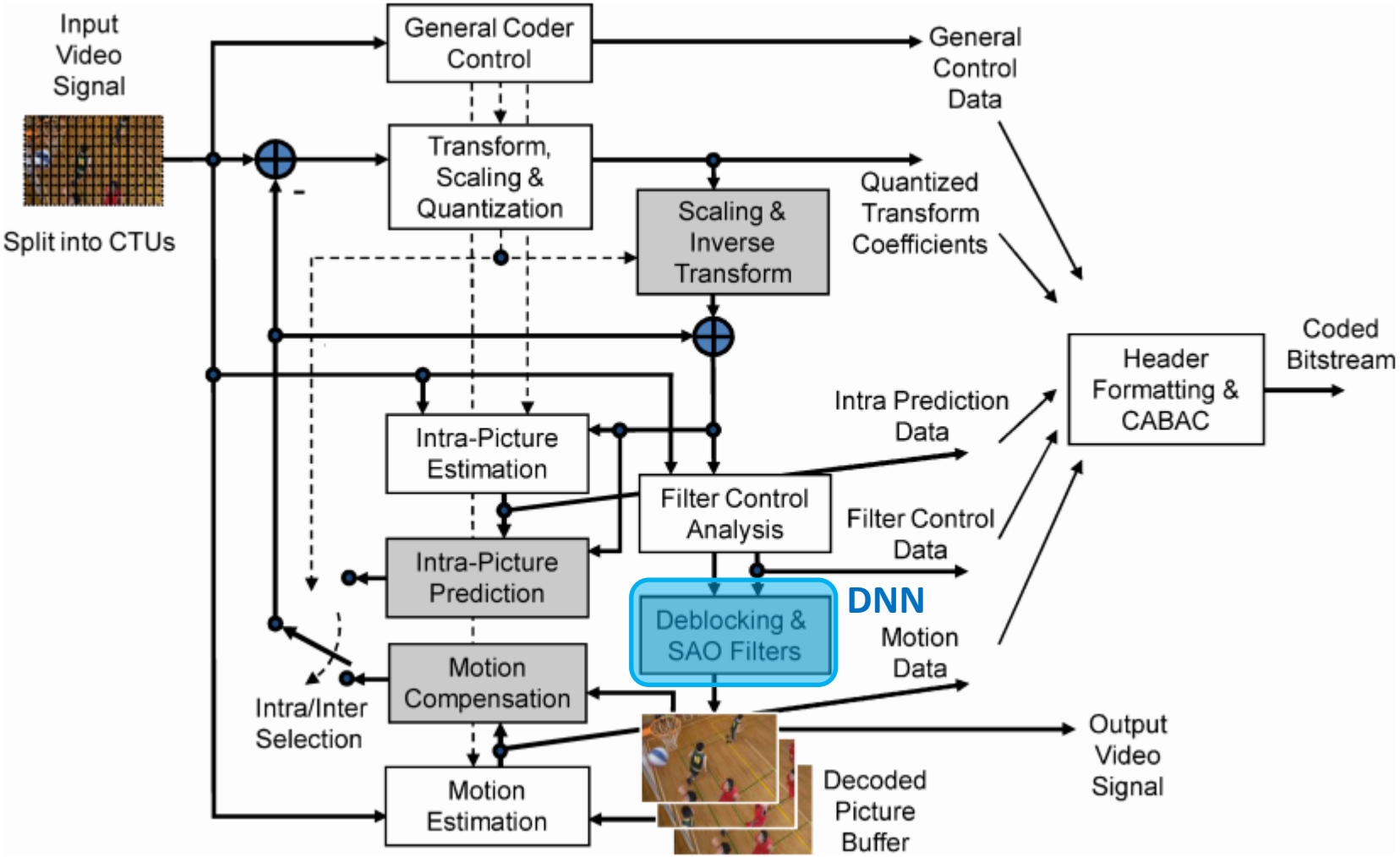


Figure 4. Examples of blocks after pre-filtering in the first stage of our method. Images are from Yang *et al.*'s dataset. The red blocks are the top-4 most similar blocks with respect to the yellow block in another frame.

2. DNN-based handcrafted video compression



Traditional video compression framework (HEVC) [1]

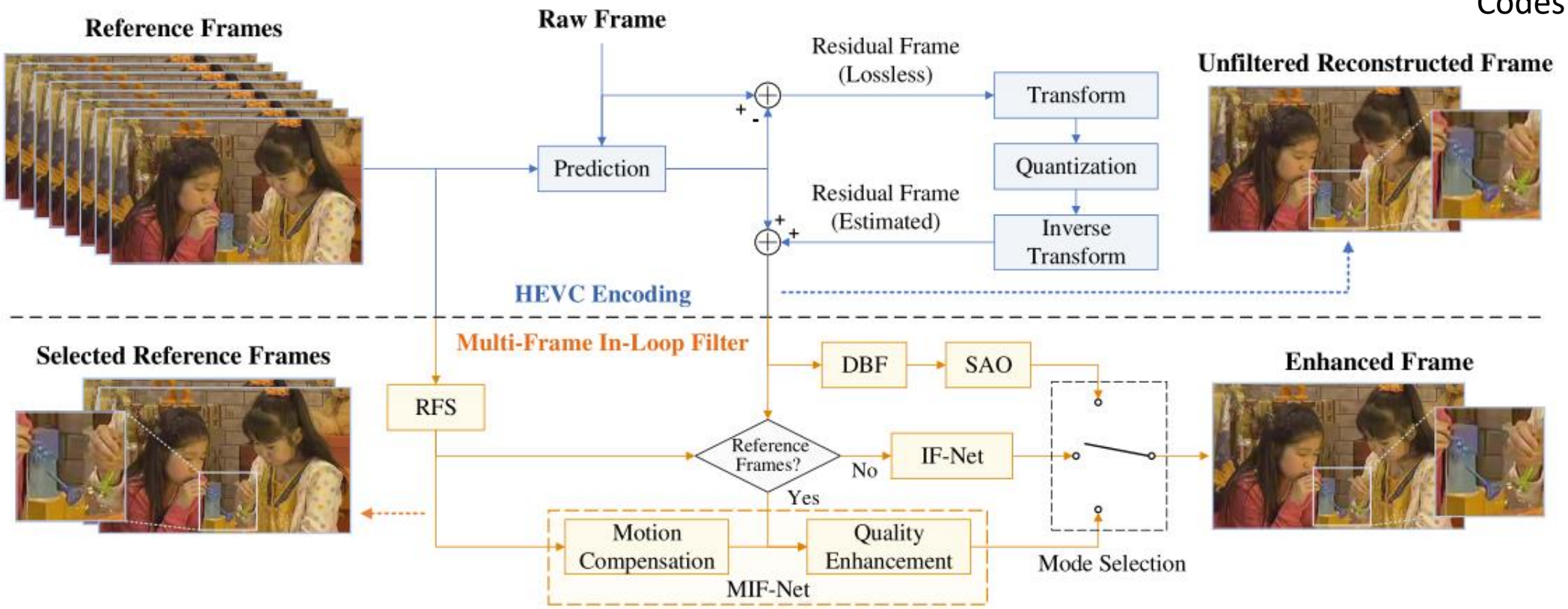
Gray: modules in both encoder and decoder; White: modules only in encoder

[1] Sullivan, Gary J., et al. "Overview of the high efficiency video coding (HEVC) standard." IEEE T-CSVT. 2012.

2. DNN-based handcrafted video compression

Deep in-loop filter [10, 11]

Codes: MIF



[10] Li, Tianyi, Mai Xu, Ce Zhu, Ren Yang, et al. "A deep learning approach for multi-frame in-loop filter of HEVC." IEEE TIP. 2019.
[11] Li, Tianyi, Mai Xu, Ren Yang, Xiaoming Tao. "A DenseNet based approach for multi-frame in-loop filter in HEVC." in DCC. 2019.

2. DNN-based handcrafted video compression

Deep in-loop filter [10, 11]

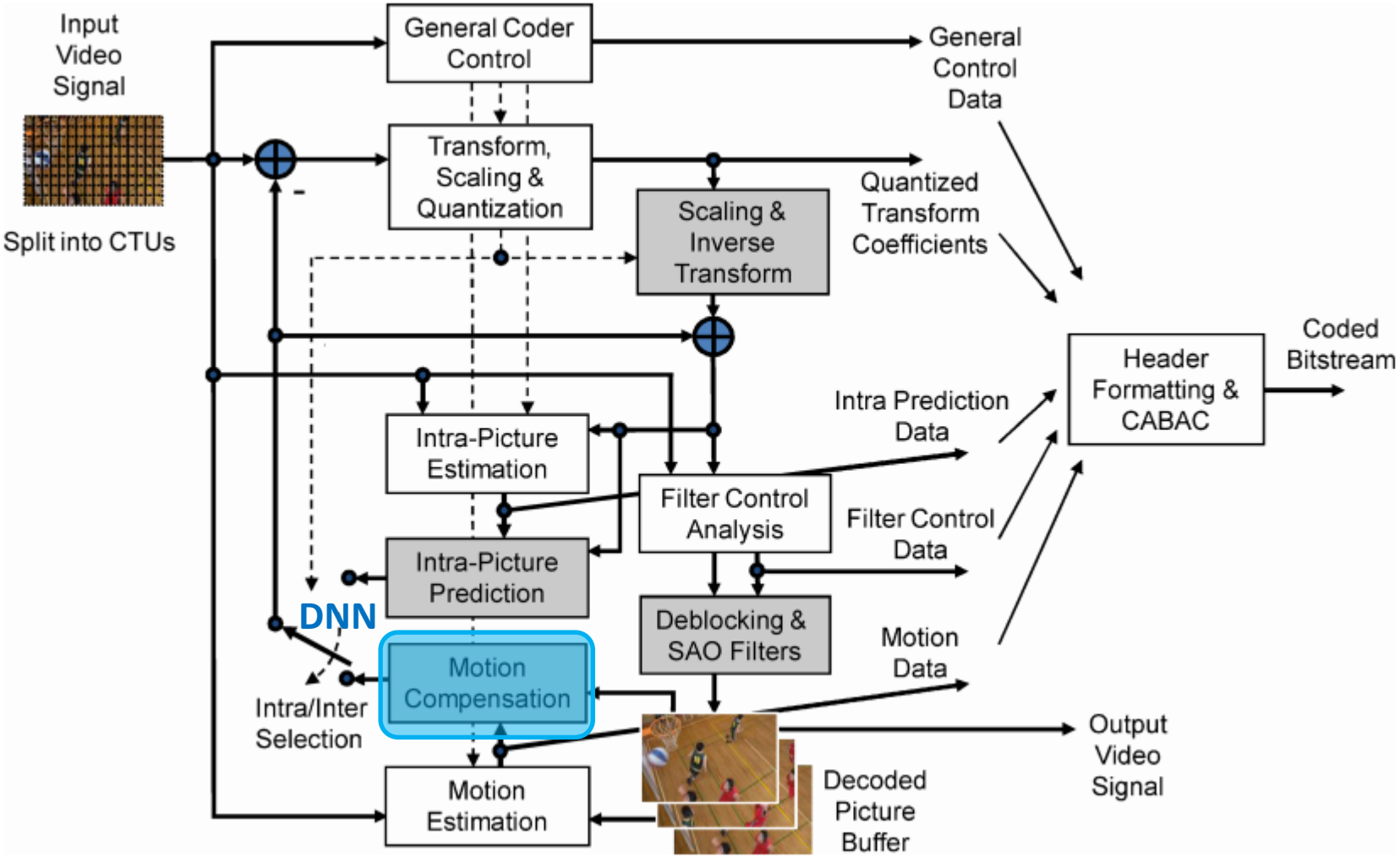
RD PERFORMANCE OF IN-LOOP FILTERS (RA CONFIG.). (a) PERFORMANCE ON THE JCT-VC TEST SET. (b) PERFORMANCE ON THE SUPPLEMENTARY TEST SET

Class	Sequence	DBF and SAO		[10]		[20]		Proposed MIF	
		BD-BR (%)	BD-PSNR (dB)	BD-BR (%)	BD-PSNR (dB)	BD-BR (%)	BD-PSNR (dB)	BD-BR (%)	BD-PSNR (dB)
A	*PeopleOnStreet*	-8.287	0.368	-12.025	0.540	-12.484	0.568	**-16.824**	**0.777**
	Traffic	-5.348	0.162	-6.169	0.188	-9.816	0.304	**-12.152**	**0.383**
B	*BasketballDrive*	-6.655	0.148	-8.838	0.198	-11.053	0.250	**-14.870**	**0.347**
	BQTerrace	-7.149	0.111	-11.397	0.173	-14.365	0.228	**-17.126**	**0.271**
	Cactus	-7.540	0.157	-8.904	0.189	-12.518	0.272	**-15.829**	**0.349**
	Kimono	-7.536	0.220	-9.261	0.273	-10.482	0.311	**-12.235**	**0.368**
	ParkScene	-3.679	0.112	-4.083	0.124	-5.940	0.182	**-7.994**	**0.249**
C	*BasketballDrill*	-5.017	0.206	-5.393	0.222	-7.818	0.326	**-10.324**	**0.434**
	BQMall	-3.933	0.150	-4.451	0.170	-7.663	0.296	**-9.376**	**0.367**
	PartyScene	-1.054	0.044	-1.224	0.051	-2.417	0.100	**-4.159**	**0.173**
	RaceHorses	-6.154	0.222	-7.082	0.257	-10.403	0.386	**-12.736**	**0.476**
D	*BasketballPass*	-3.852	0.182	-4.324	0.204	-7.700	0.370	**-9.984**	**0.484**
	BlowingBubbles	-0.829	0.034	-0.831	0.034	-3.072	0.126	**-3.980**	**0.164**
	BQSquare	-0.053	0.002	0.010	-0.000	-3.263	0.123	**-4.401**	**0.165**
	RaceHorses	-4.441	0.199	-4.797	0.215	-8.860	0.407	**-10.992**	**0.510**
E	*FourPeople*	-7.018	0.262	-8.489	0.319	-13.937	0.538	**-16.480**	**0.644**
	Johnny	-5.599	0.143	-8.034	0.208	-11.649	0.302	**-14.370**	**0.378**
	KristenAndSara	-6.410	0.203	-8.011	0.254	-12.637	0.406	**-15.340**	**0.501**
Average		-5.031	0.162	-6.295	0.201	-9.227	0.305	**-11.621**	**0.391**

Codes: MIF

[10] Li, Tianyi, Mai Xu, Ce Zhu, Ren Yang, et al. "A deep learning approach for multi-frame in-loop filter of HEVC." IEEE TIP. 2019.
[11] Li, Tianyi, Mai Xu, Ren Yang, Xiaoming Tao. "A DenseNet based approach for multi-frame in-loop filter in HEVC." in DCC. 2019.

2. DNN-based handcrafted video compression



Traditional video compression framework (HEVC) [1]

Gray: modules in both encoder and decoder; White: modules only in encoder

[1] Sullivan, Gary J., et al. "Overview of the high efficiency video coding (HEVC) standard." IEEE T-CSVT. 2012.

2. DNN-based handcrafted video compression

Learning-based motion compensation [12]

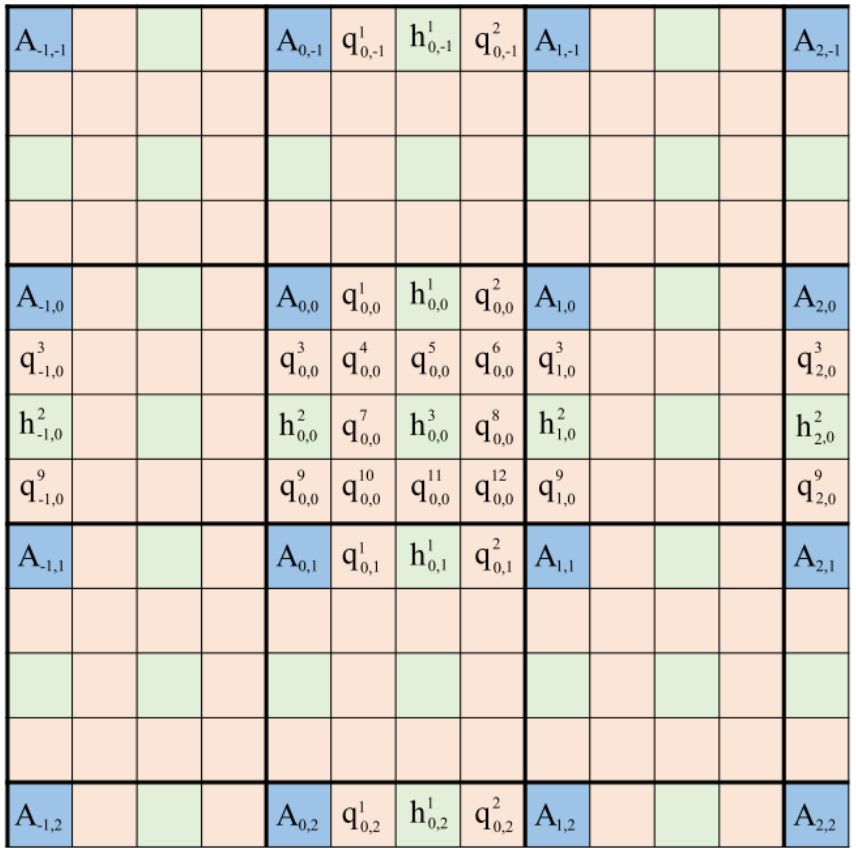
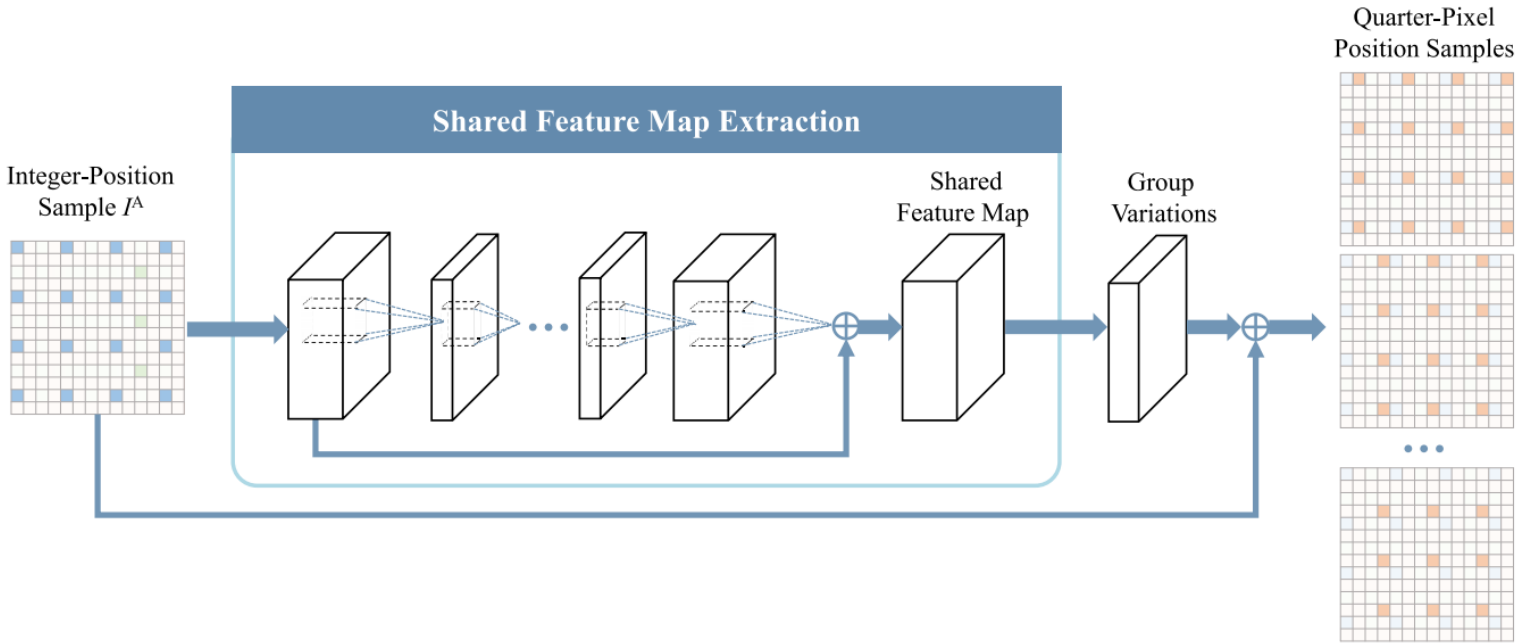


Fig. 1. Positions of different fractional pixels. Blue, green and pink blocks indicate respectively the integer- ($A_{i,j}$), half- ($h^1_{i,j}, h^2_{i,j}, \ldots, h^3_{i,j}$) and quarter- ($q^1_{i,j}, q^2_{i,j}, \ldots, q^{12}_{i,j}$) pixel positions for luma interpolation.



[12] Liu, Jiaying, et al. "One-for-All: grouped variation network-based fractional interpolation in video coding." IEEE T-IP. 2019.

2. DNN-based handcrafted video compression

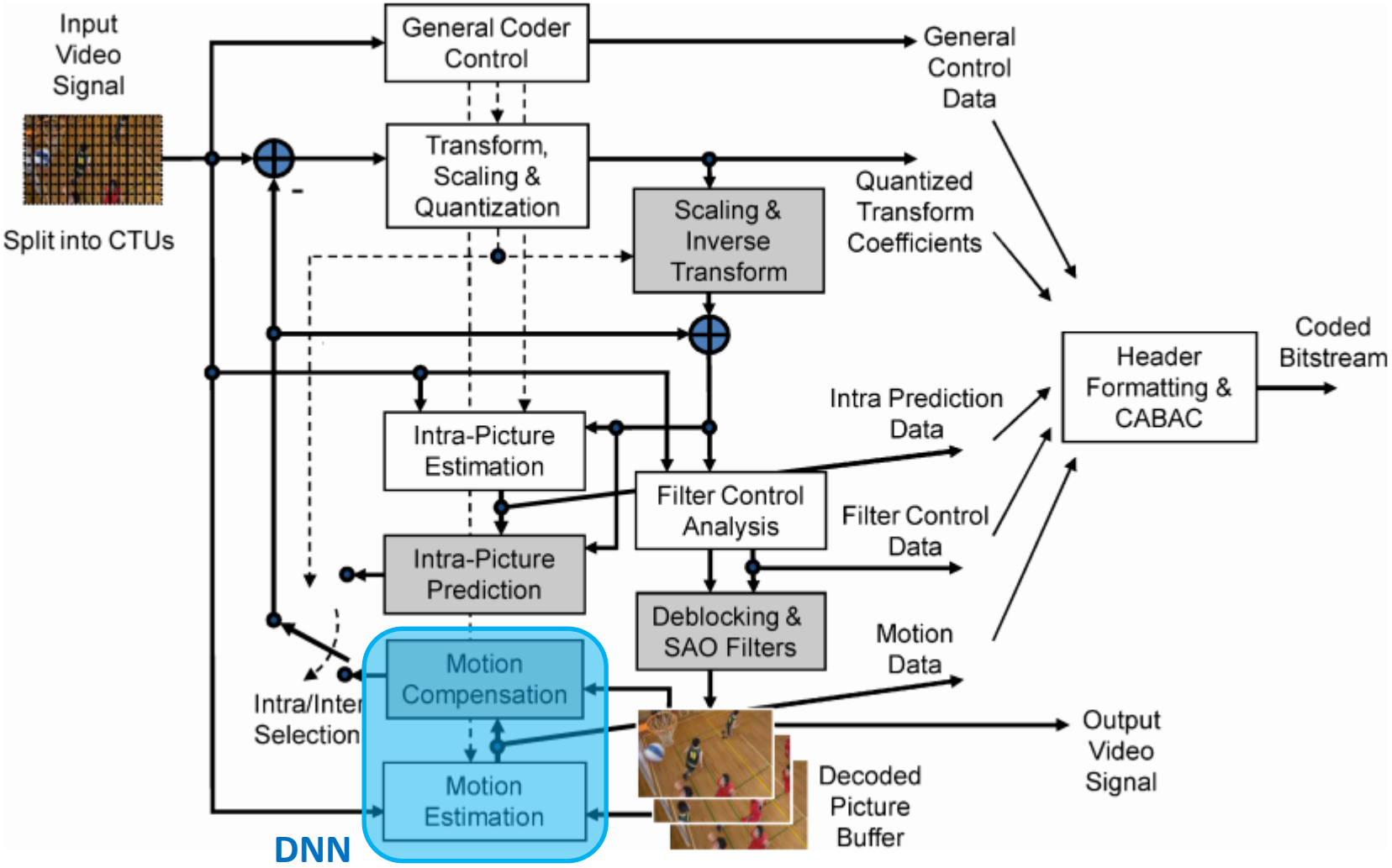
Learning-based motion compensation [12]

BD-RATE REDUCTION OF THE PROPOSED METHOD COMPARED TO HEVC

Class	Sequence	BD-rate of LDP			BD-rate of LDB			BD-rate of RA		
		Y	U	V	Y	U	V	Y	U	V
Class A	Traffic	-	-	-	-	-	-	-1.1%	0.1%	0.3%
	PeopleOnStreet	-	-	-	-	-	-	-0.9%	-0.1%	-0.8%
	Nebuta	-	-	-	-	-	-	-0.1%	-0.3%	-0.4%
	SteamLocomotive	-	-	-	-	-	-	-0.2%	-0.5%	-0.5%
	Average	-	-	-	-	-	-	-0.6%	-0.2%	-0.3%
Class B	Kimono	-4.1%	2.1%	1.6%	-1.7%	0.9%	0.4%	-1.3%	0.2%	-0.2%
	BQTerrace	-5.2%	-3.4%	-3.9%	-1.3%	0.3%	0.2%	-2.5%	0.0%	-1.1%
	BasketballDrive	-3.3%	0.2%	-0.5%	-0.5%	0.5%	0.1%	-1.4%	-0.8%	-0.8%
	ParkScene	-1.3%	0.0%	-0.5%	-0.8%	0.7%	0.2%	-0.7%	0.3%	-0.5%
	Cactus	-2.5%	-0.5%	-0.8%	-1.0%	0.0%	-0.1%	-1.1%	-0.2%	-1.0%
	Average	-3.3%	-0.3%	-0.8%	-1.1%	0.8%	0.4%	-1.4%	-0.1%	-0.7%
Class C	BasketballDrill	-2.2%	-1.3%	-0.6%	-1.0%	0.0%	-0.1%	-0.7%	0.0%	0.1%
	BQMall	-2.9%	-2.1%	-1.7%	-1.3%	-0.9%	-1.0%	-1.1%	-0.2%	-1.0%
	PartyScene	-1.6%	-0.5%	-1.4%	-0.9%	-0.2%	-0.4%	-0.7%	-0.4%	-1.1%
	RaceHorsesC	-2.0%	-1.4%	-1.6%	-1.5%	-0.4%	-0.8%	-1.6%	-1.5%	-1.3%
	Average	-2.2%	-1.3%	-1.3%	-1.1%	-0.4%	-0.6%	-1.0%	-0.5%	-0.8%
Class D	BasketballPass	-3.3%	-1.7%	-1.4%	-1.8%	-0.9%	-1.5%	-0.9%	-1.0%	-1.5%
	BlowingBubbles	-2.1%	-0.9%	-0.3%	-1.0%	0.2%	-0.2%	-0.8%	-0.7%	0.2%
	BQSquare	-0.6%	1.2%	3.1%	-0.7%	1.6%	0.9%	-0.7%	-0.3%	-0.5%
	RaceHorses	-2.7%	-1.7%	-0.8%	-1.9%	0.0%	-0.9%	-1.4%	-0.9%	-1.1%
	Average	-2.2%	-0.7%	0.2%	-1.4%	0.2%	-0.4%	-1.0%	-0.7%	-0.7%
Class E	FourPeople	-1.6%	-0.5%	-0.2%	-2.8%	5.9%	0.0%	-	-	-
	Johnny	-2.9%	0.2%	0.4%	-1.2%	1.1%	1.4%	-	-	-
	KristenAndSara	-2.2%	1.1%	0.2%	-1.3%	1.6%	1.2%	-	-	-
	Average	-2.2%	0.3%	0.2%	-1.6%	2.5%	0.8%	-	-	-
Class F	BasketballDrillText	-1.8%	-0.9%	0.1%	-1.0%	-0.6%	-0.8%	-0.9%	0.2%	-0.5%
	ChinaSpeed	-1.4%	-1.9%	-1.4%	-1.0%	-1.3%	-1.5%	-1.4%	-2.0%	-1.8%
	SlideEditing	0.0%	-0.1%	-0.2%	0.0%	-0.1%	-0.1%	0.6%	0.9%	0.9%
	SlideShow	-0.5%	0.2%	-0.6%	-0.9%	0.5%	1.0%	-0.8%	0.3%	-0.9%
	Average	-0.9%	-0.7%	-0.5%	-0.7%	-0.9%	-0.6%	-0.6%	-0.2%	-0.6%
All Sequences	Overall	-2.2%	-0.6%	-0.5%	-1.2%	0.4%	-0.1%	-0.9%	-0.3%	-0.6%

[12] Liu, Jiaying, et al. "One-for-All: grouped variation network-based fractional interpolation in video coding." IEEE T-IP. 2019.

2. DNN-based handcrafted video compression



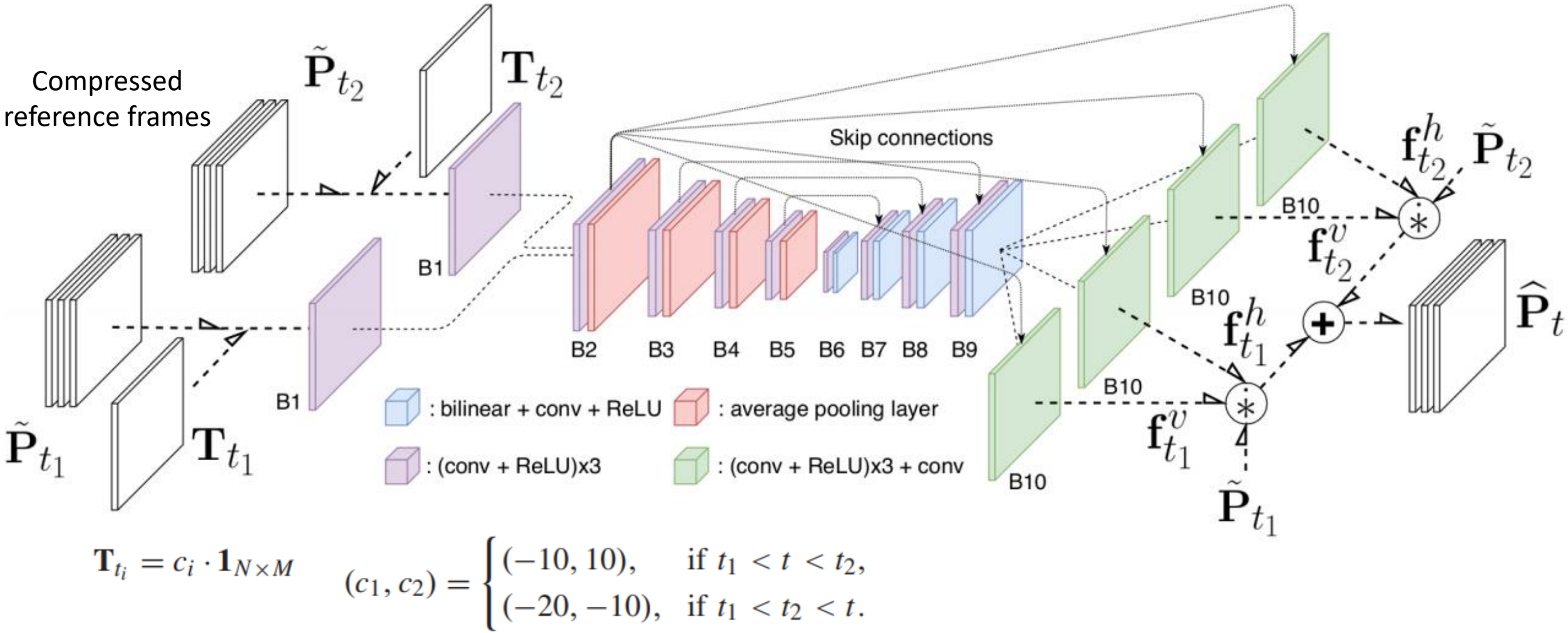
Traditional video compression framework (HEVC) [1]

Gray: modules in both encoder and decoder; White: modules only in encoder

[1] Sullivan, Gary J., et al. "Overview of the high efficiency video coding (HEVC) standard." IEEE T-CSVT. 2012.

2. DNN-based handcrafted video compression

Learning-based frame prediction [13]

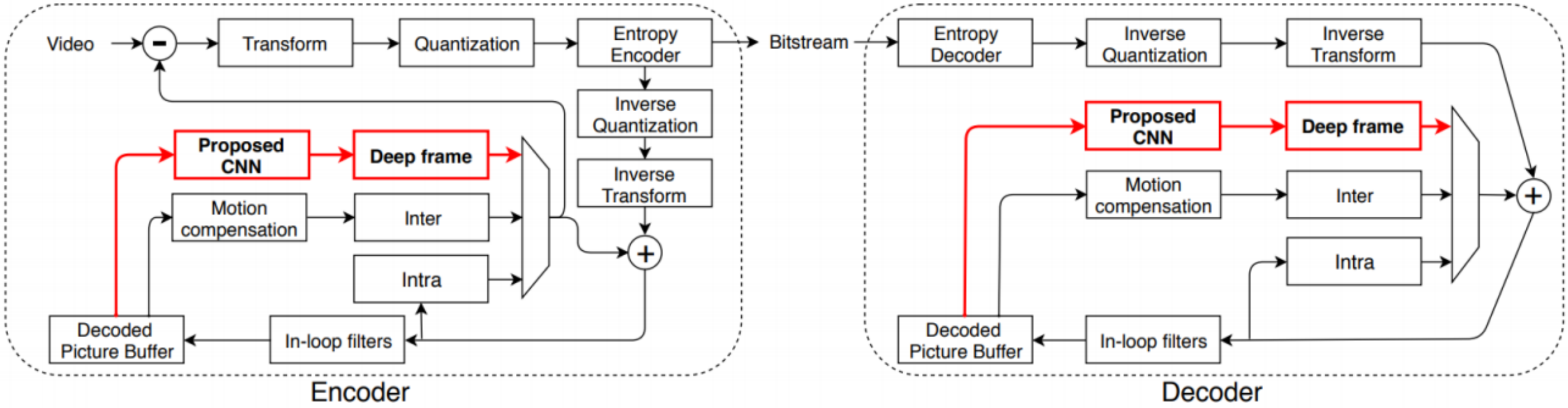


$$\mathbf{T}_{t_i} = c_i \cdot \mathbf{1}_{N \times M} \qquad (c_1, c_2) = \begin{cases} (-10, 10), & \text{if } t_1 < t < t_2, \\ (-20, -10), & \text{if } t_1 < t_2 < t. \end{cases}$$

[13] Choi, Hyomin, et al. "Deep frame prediction for video coding." IEEE T-CSVT. 2020.

2. DNN-based handcrafted video compression

Learning-based frame prediction [13]



$$J^* = \min(J_{\text{Intra}}, J_{\text{Inter}}, J_{\text{DNN}}).$$

[13] Choi, Hyomin, et al. "Deep frame prediction for video coding." IEEE T-CSVT. 2020.

2. DNN-based handcrafted video compression

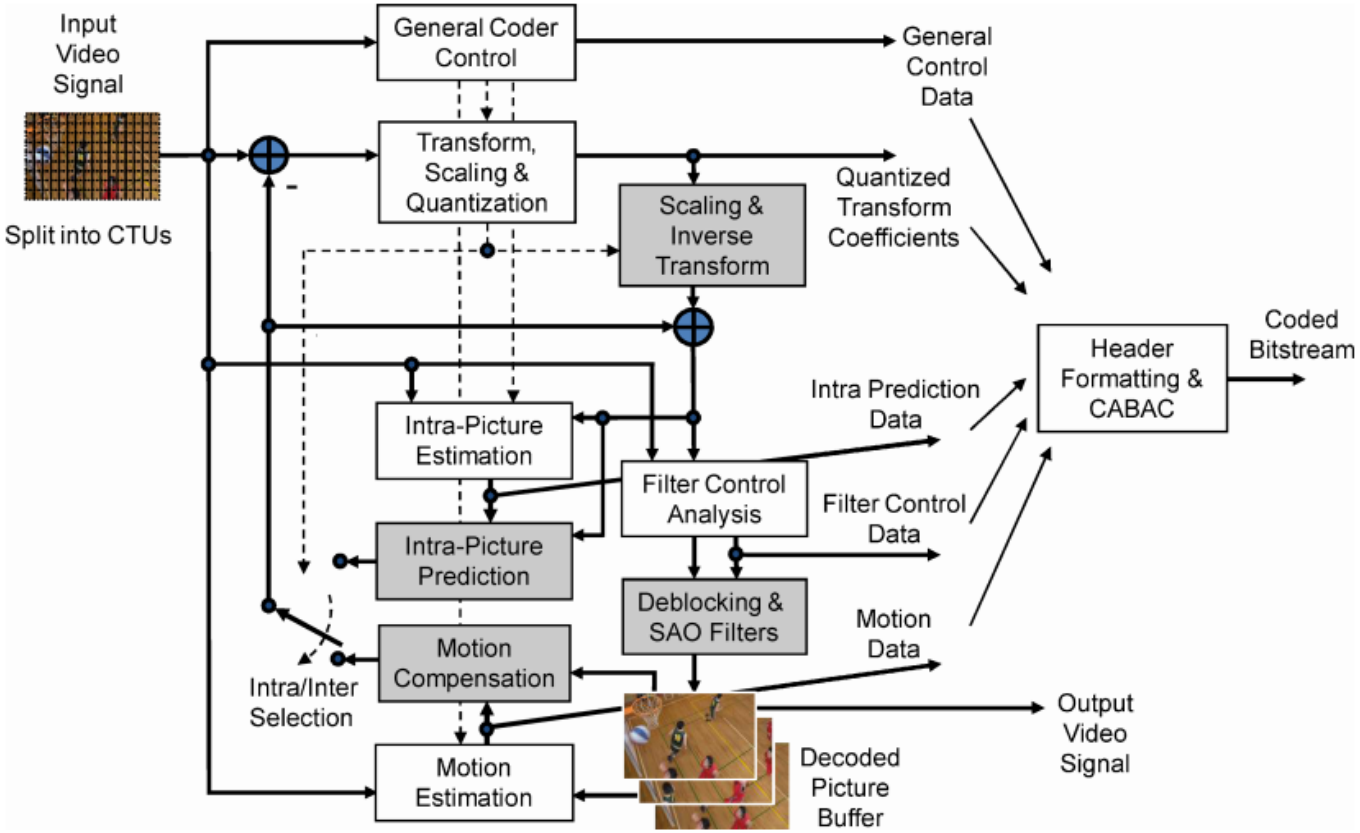
Learning-based frame prediction [13]

BD-BITRATE RELATIVE TO HM-16.20 OVER THREE COMMON TEST CONDITIONS

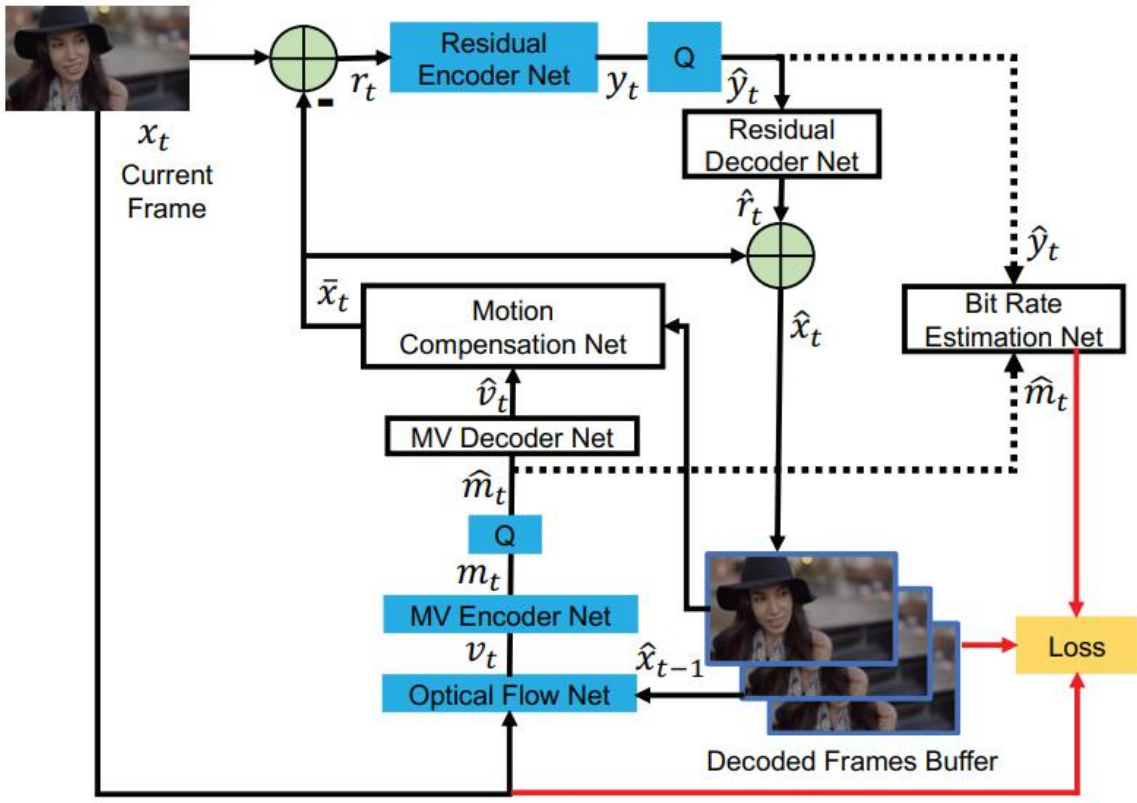
Class	Sequence	fps	Low delay P (LP)						Low delay (LD)						Random access (RA)					
			Sep. (%)			Comb. (%)			Sep. (%)			Comb. (%)			Sep. (%)			Comb. (%)		
			Y	U	V	Y	U	V	Y	U	V	Y	U	V	Y	U	V	Y	U	V
A	PeopleOnStreet	30	–5.8	–4.0	–4.4	–5.2	–3.5	–4.8	–4.4	–2.2	–2.7	–3.7	–2.1	–3.3	–4.0	–5.7	–5.6	–4.2	–5.7	–5.9
	Traffic		–3.6	–3.8	–3.1	–3.5	–4.0	–2.9	–2.5	–2.4	–2.3	–2.1	–2.7	–2.5	–2.5	–2.7	–2.3	–2.2	–2.5	–2.0
B	BQTerrace	60	–4.2	–3.6	–1.4	–4.0	–2.2	0.1	–1.3	–2.0	–1.4	–1.1	–1.2	0.7	–2.0	–0.5	–0.4	–1.6	0.0	0.0
	BasketballDrive	50	–2.8	–5.6	–3.7	–2.8	–5.1	–3.5	–1.1	–2.6	–1.8	–0.9	–2.0	–1.3	–1.2	–2.0	–1.5	–1.0	–1.7	–1.2
	Cactus		–6.9	–10.3	–6.0	–6.0	–9.1	–5.5	–3.7	–7.1	–4.7	–2.8	–5.7	–3.4	–2.8	–5.2	–3.4	–3.0	–4.9	–3.3
	Kimono	24	–5.1	–7.3	–3.7	–5.7	–7.8	–4.3	–2.1	–4.4	–2.5	–1.8	–3.9	–2.5	–0.9	–1.6	–1.0	–0.9	–1.6	–0.9
	ParkScene		–3.3	–4.4	–2.8	–3.2	–5.0	–2.8	–2.1	–3.0	–2.5	–1.9	–3.4	–2.2	–1.9	–2.4	–1.8	–1.6	–2.0	–1.6
C	BQMall	60	–6.0	–7.3	–6.7	–5.3	–7.1	–6.3	–4.5	–6.3	–5.7	–3.6	–5.8	–4.6	–4.2	–4.9	–4.7	–3.5	–4.2	–3.9
	BasketballDrill	50	–2.9	–5.2	–3.0	–2.9	–5.2	–3.4	–1.6	–3.4	–1.6	–1.4	–3.1	–1.8	–1.5	–2.8	–2.6	–1.5	–2.7	–2.4
	PartyScene		–3.3	–3.9	–3.8	–2.9	–3.1	–3.1	–2.0	–3.0	–3.1	–1.6	–2.3	–2.3	–4.5	–4.0	–3.8	–3.7	–3.8	–3.4
	RaceHorsesC	30	–1.1	–1.5	–1.9	–1.0	–1.5	–1.9	–0.8	–1.1	–1.5	–0.6	–0.8	–1.2	–0.7	–1.2	–1.5	–0.6	–1.0	–1.2
D	BQSquare	60	–3.2	–0.6	1.5	–1.4	–2.7	–0.3	–1.6	–0.6	–3.0	–0.8	–2.1	–1.8	–3.3	–0.7	–0.4	–2.3	–0.1	0.0
	BasketballPass	50	–4.4	–5.7	–4.6	–4.2	–5.7	–4.2	–3.2	–4.3	–3.6	–2.9	–3.7	–3.1	–4.1	–5.6	–4.7	–3.5	–4.7	–3.4
	BlowingBubbles		–4.1	–5.2	–5.3	–3.4	–5.0	–4.9	–2.7	–4.5	–5.4	–2.2	–3.8	–4.4	–4.3	–4.0	–3.8	–3.7	–3.8	–3.4
	RaceHorses	30	–1.8	–2.7	–3.2	–1.6	–2.4	–2.5	–1.3	–2.3	–2.6	–1.1	–2.1	–1.9	–1.6	–2.8	–3.0	–1.5	–2.4	–2.5
E	FourPeople	60	–10.1	–11.9	–11.6	–9.8	–11.3	–10.5	–7.0	–9.3	–10.1	–6.5	–8.5	–8.7		–			–	
	Johnny		–8.6	–9.1	–7.5	–7.8	–8.8	–8.3	–4.9	–7.2	–7.4	–4.2	–6.4	–5.5		–			–	
	KristenAndSara		–8.7	–22.2	–9.1	–7.9	–10.6	–9.7	–5.4	–8.5	–7.4	–4.6	–7.4	–5.9		–			–	
	Average		–4.8	–5.7	–4.5	–4.4	–5.6	–4.4	–2.9	–4.1	–3.8	–2.4	–3.7	–3.1	–2.6	–3.1	–2.7	–2.3	–2.7	–2.3
	ΔT_{Enc}		163%			164%			146%			145%			150%			149%		
	ΔT_{Dec}		16,563%			16,562%			15,415%			15,410%			11,389%			11,488%		

[13] Choi, Hyomin, et al. "Deep frame prediction for video coding." IEEE T-CSVT. 2020.

3. End-to-end deep video compression network



Traditional video compression framework (HEVC) [1]



End-to-end deep video compression framework (DVC) [14]

[1] Sullivan, Gary J., et al. "Overview of the high efficiency video coding (HEVC) standard." IEEE T-CSVT. 2012.
[14] Lu, Guo, et al. "DVC: An end-to-end deep video compression framework." in CVPR. 2019.

Entropy coding

Entropy:

$$\mathrm{H}(X) = \mathrm{E}[\mathrm{I}(X)] = \mathrm{E}[-\log(\mathrm{P}(X))]$$

$$\mathrm{H}(X) = -\sum_{i=1}^{n} \mathrm{P}(x_i) \log_b \mathrm{P}(x_i)$$

Cross entropy:

$$H(p, q) = -\sum_{x \in \mathcal{X}} p(x) \log q(x) \quad \text{(Eq.1)}$$

real estimated

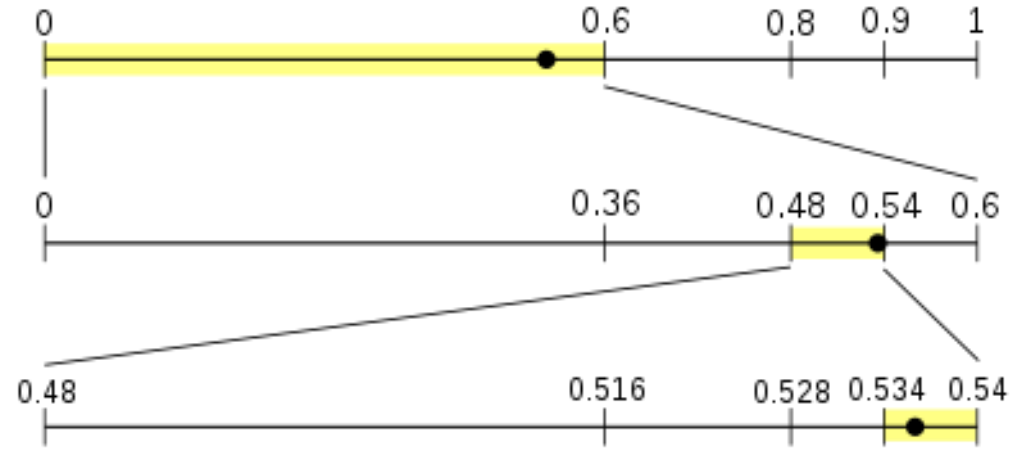
(Adaptive) arithmetic coding is theoretically able to losslessly compress data at

- bit-rate ≅ cross entropy (with little overhead)

Arithmetic coding:

- 60% chance of symbol NEUTRAL
- 20% chance of symbol POSITIVE
- 10% chance of symbol NEGATIVE
- 10% chance of symbol END-OF-DATA.

NEUTRAL NEGATIVE END-OF-DATA message



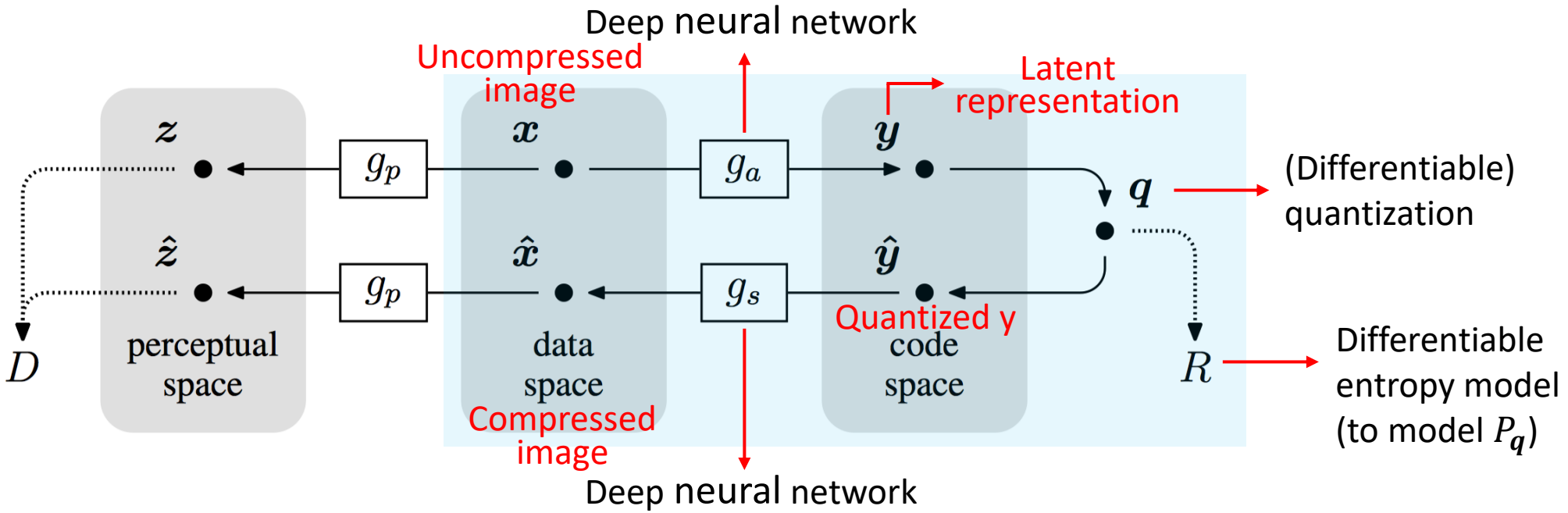
0.10001010 **8 bits**
(0.5390625 decimal)

Adaptive arithmetic coding:

Changing the frequency (or probability) tables while processing the data.

Learned Image Compression

- Basic architecture [15]: **End-to-end trainable**

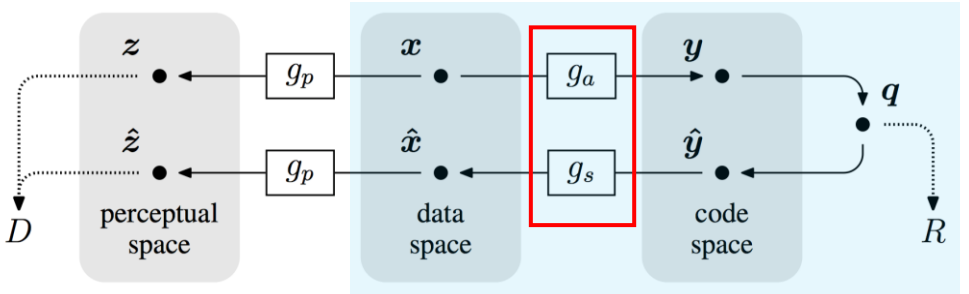


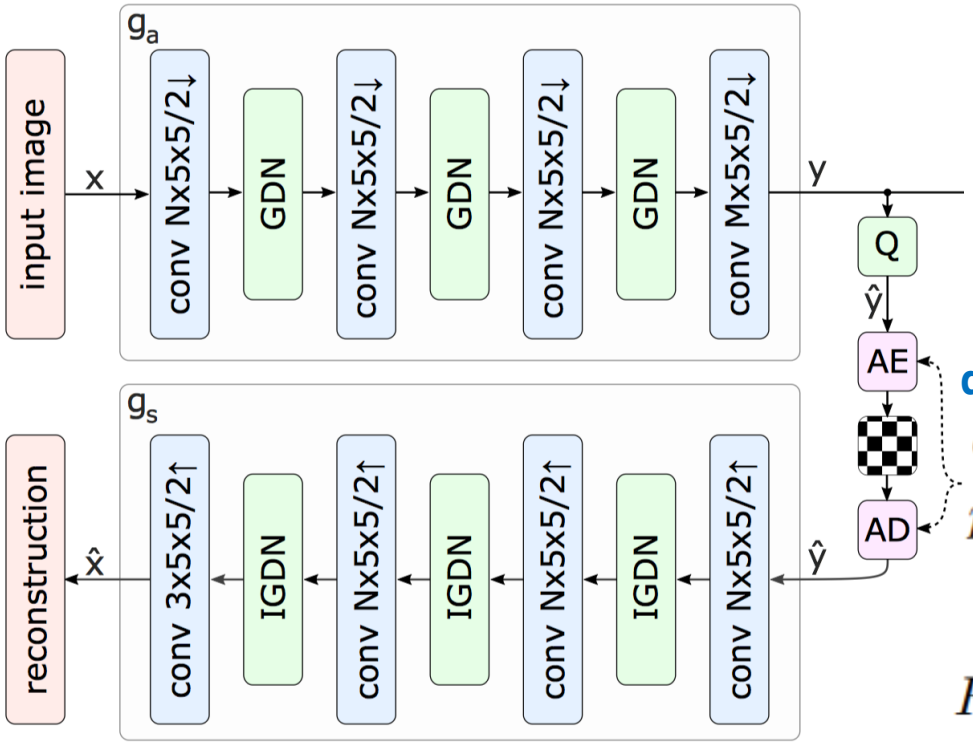
$$L[g_a, g_s, P_{\boldsymbol{q}}] = \underbrace{-\mathbb{E}\left[\log_2 P_{\boldsymbol{q}}\right]}_{R} + \lambda\, \mathbb{E}\left[d(\boldsymbol{x}, \hat{\boldsymbol{x}})\right]$$

[15] Ballé, Johannes, et al. "End-to-end optimized image compression." in ICLR. 2017.

Learned Image Compression

- CNN transformer + **factorized** entropy model [15]





$$\tilde{\boldsymbol{y}} = \boldsymbol{y} + \Delta\boldsymbol{y} \quad \Delta\boldsymbol{y} \sim \mathcal{U}(0,1)$$

Training: differentiable quantization

$$\hat{\boldsymbol{y}} = \mathrm{round}(\boldsymbol{y})$$

Inference: quantization (not differentiable)

differentiable entropy

$$c = f_K \circ f_{K-1} \cdots f_1$$

$$p = f'_K \cdot f'_{K-1} \cdots f'_1$$

$\tilde{\boldsymbol{y}}$ or $\hat{\boldsymbol{y}}$

$$f_k(\underline{\boldsymbol{x}}) = g_k\left(\boldsymbol{H}^{(k)}\underline{\boldsymbol{x}} + \boldsymbol{b}^{(k)}\right) \qquad 1 \le k < K$$

$$f_K(\underline{\boldsymbol{x}}) = \mathrm{sigmoid}\left(\boldsymbol{H}^{(K)}\underline{\boldsymbol{x}} + \boldsymbol{b}^{(K)}\right)$$

$$g_k(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{a}^{(k)} \odot \tanh(\boldsymbol{x})$$

$$g'_k(\boldsymbol{x}) = 1 + \boldsymbol{a}^{(k)} \odot \tanh'(\boldsymbol{x})$$

$$\boldsymbol{H}^{(k)} = \mathrm{softplus}\left(\hat{\boldsymbol{H}}^{(k)}\right)$$

$$\boldsymbol{a}^{(k)} = \tanh\left(\hat{\boldsymbol{a}}^{(k)}\right)$$

$$R = \mathbb{E}_{\boldsymbol{x}\sim p_{\boldsymbol{x}}}\left[-\log_2 p_{\hat{\boldsymbol{y}}}\left(Q(g_a(\boldsymbol{x};\boldsymbol{\phi}_g))\right)\right]$$

estimated bit-rate

$$L(\boldsymbol{\theta},\boldsymbol{\phi}) = \mathbb{E}_{\boldsymbol{x},\Delta\boldsymbol{y}}\left[-\sum_i \log_2 p_{\tilde{y}_i}(g_a(\boldsymbol{x};\boldsymbol{\phi}) + \Delta\boldsymbol{y};\boldsymbol{\psi}^{(i)}) + \lambda d\left(g_p(g_s(g_a(\boldsymbol{x};\boldsymbol{\phi}) + \Delta\boldsymbol{y};\boldsymbol{\theta})), g_p(\boldsymbol{x})\right)\right]$$

bit-rate **trade-off** **distortion**

Optimized in an end-to-end manner

[15] Ballé, Johannes, et al. "End-to-end optimized image compression." in ICLR. 2017.

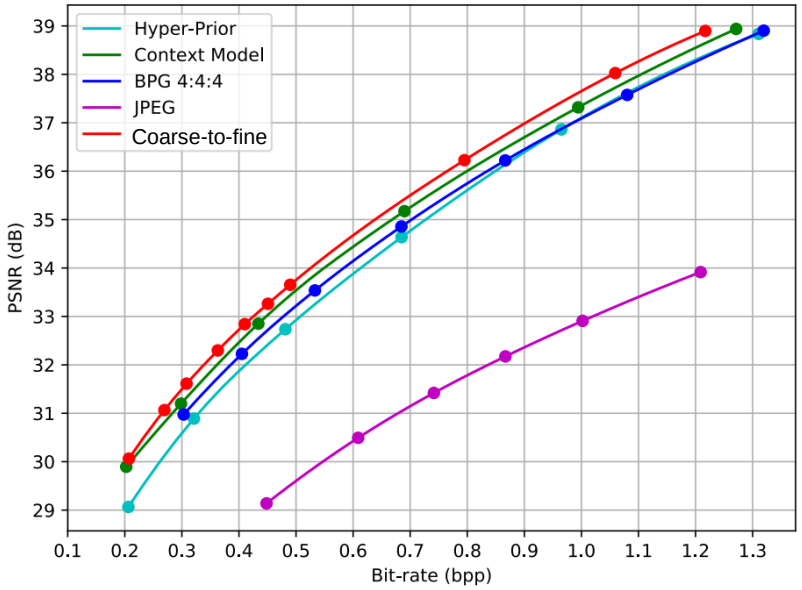
Learned Image Compression

Tutorial at IEEE VCIP 2020

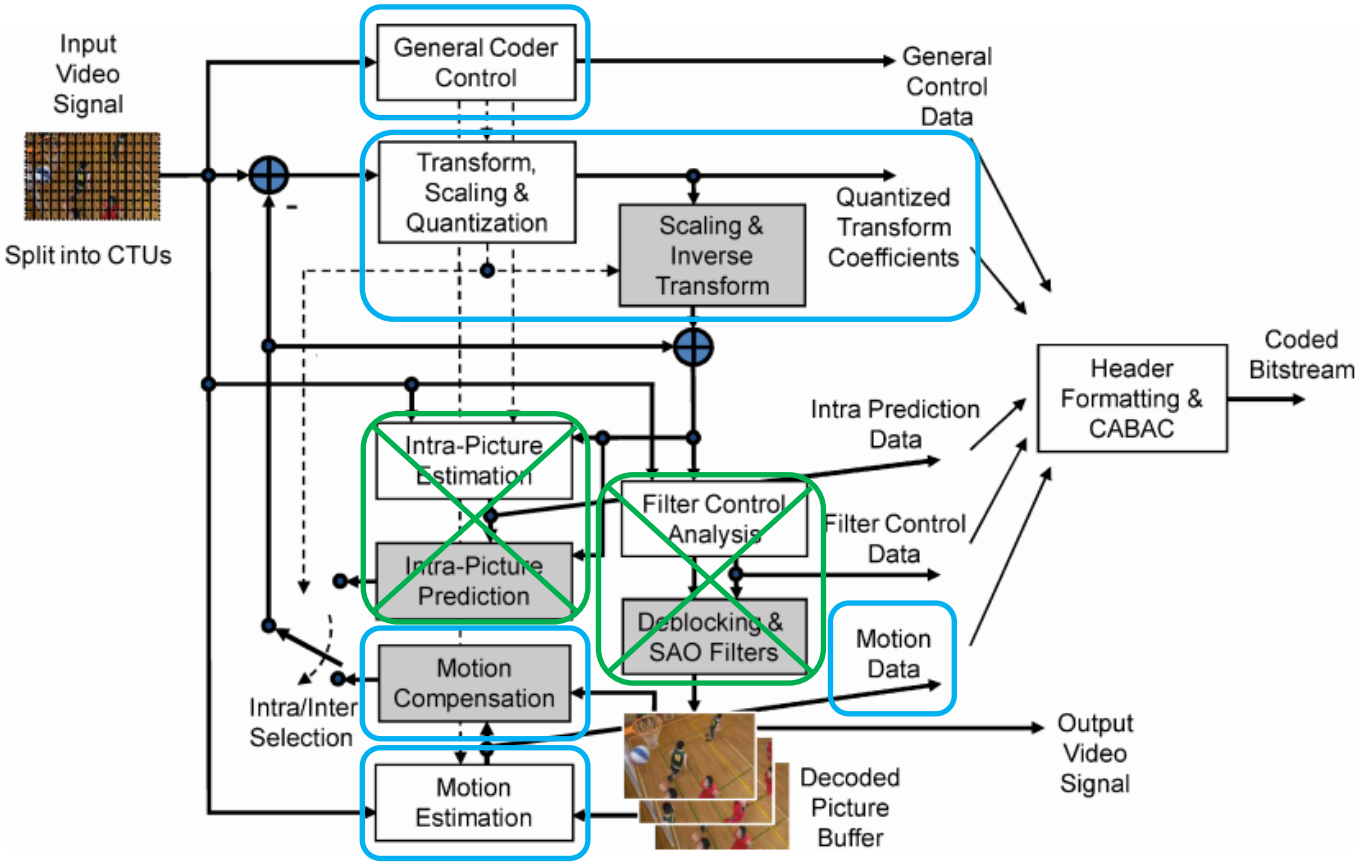
https://renyang-home.github.io/#tutorials

- CNN-based methods
  - Factorized entropy model
  - Hyperprior entropy model
  - Autoregressive entropy model
  - Coarse-to-fine entropy model
  - Conditional auto-encoder (variable bit-rates)
  - Invertible auto-encoder (lossy and lossless by one framework)
- RNN-based methods
  - Variable bit-rate
- GAN-based methods
  - Photo-realistic compressed image with low bit-rate

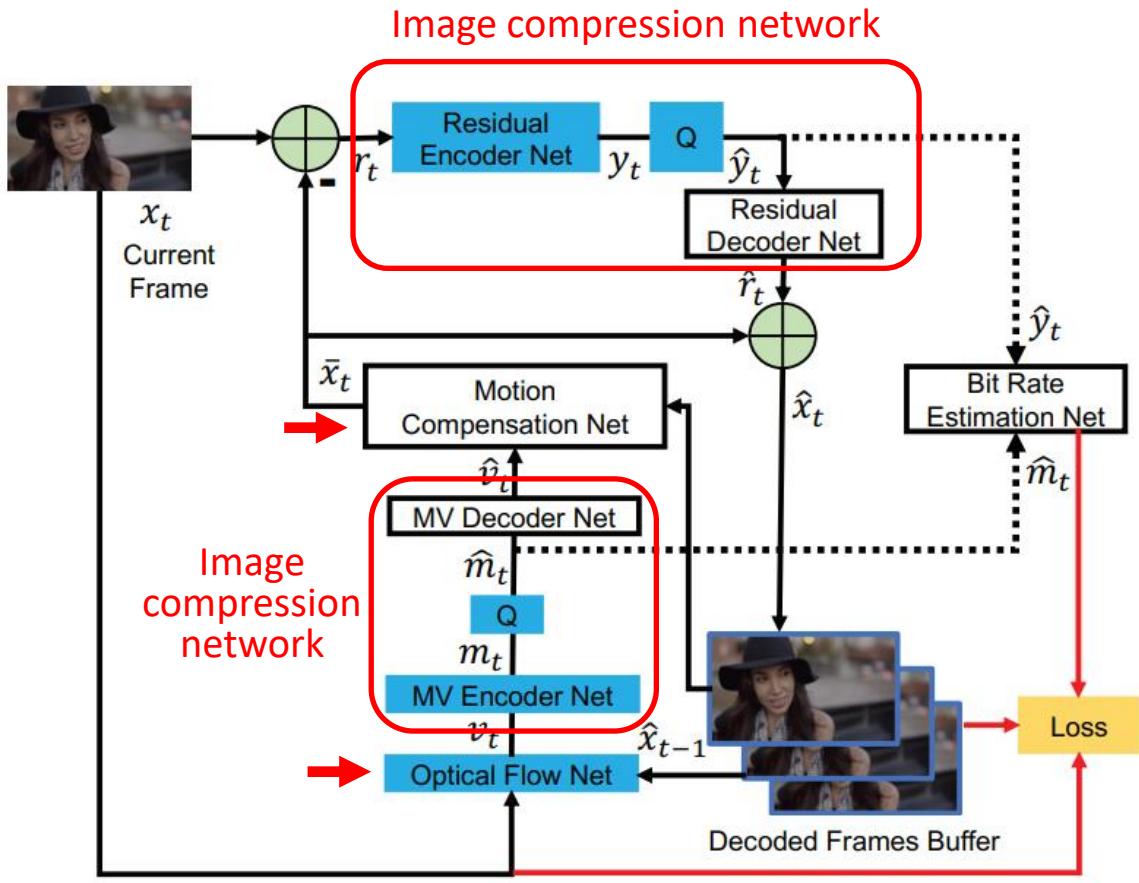
The state-of-the-art learned image compression methods successfully outperform the latest traditional compression standard BPG 4:4:4



3. End-to-end deep video compression network



Traditional video compression framework (HEVC) [1]



End-to-end deep video compression framework (DVC) [14]

Loss: $\lambda D + R = \lambda d(x_t, \hat{x}_t) + (H(\hat{m}_t) + H(\hat{y}_t))$

[1] Sullivan, Gary J., et al. "Overview of the high efficiency video coding (HEVC) standard." IEEE T-CSVT. 2012.
[14] Lu, Guo, et al. "DVC: An end-to-end deep video compression framework." in CVPR. 2019.

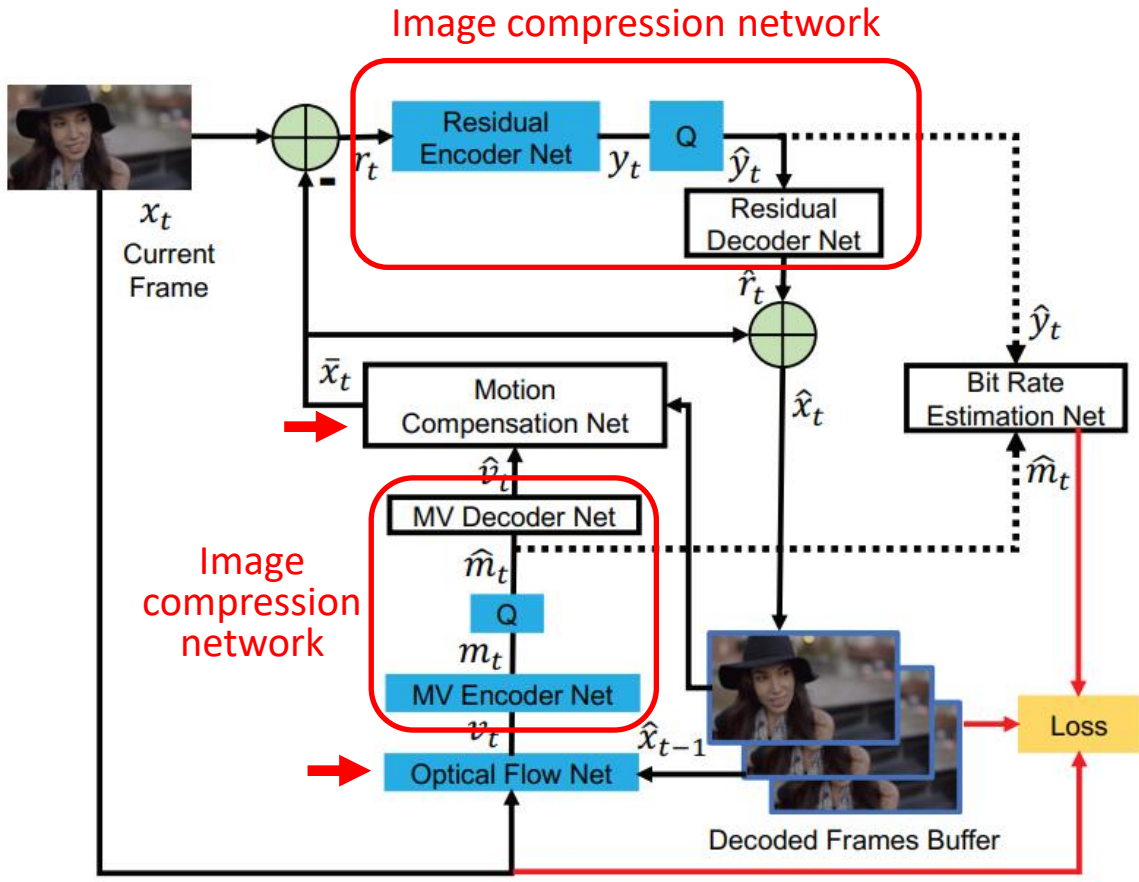
3. End-to-end deep video compression network

Official test code (no open source codes) [14]

https://github.com/GuoLusjtu/DVC

Open source implementation: OpenDVC [16]

https://github.com/RenYang-home/OpenDVC



End-to-end deep video compression framework (DVC) [14]

Loss: $\lambda D + R = \lambda d(x_t, \hat{x}_t) + (H(\hat{m}_t) + H(\hat{y}_t))$

Citation:

Yang, Ren, Luc Van Gool, and Radu Timofte. "OpenDVC: An Open Source Implementation of the DVC Video Compression Method." *arXiv preprint arXiv:2006.15862* (2020).

[14] Lu, Guo, et al. "DVC: An end-to-end deep video compression framework." in CVPR. 2019.

[16] Yang, Ren, et al. "OpenDVC: An Open Source Implementation of the DVC Video Compression Method." arXiv preprint arXiv:2006.15862.

3. End-to-end deep video compression network

Official test code (no open source codes) [14]

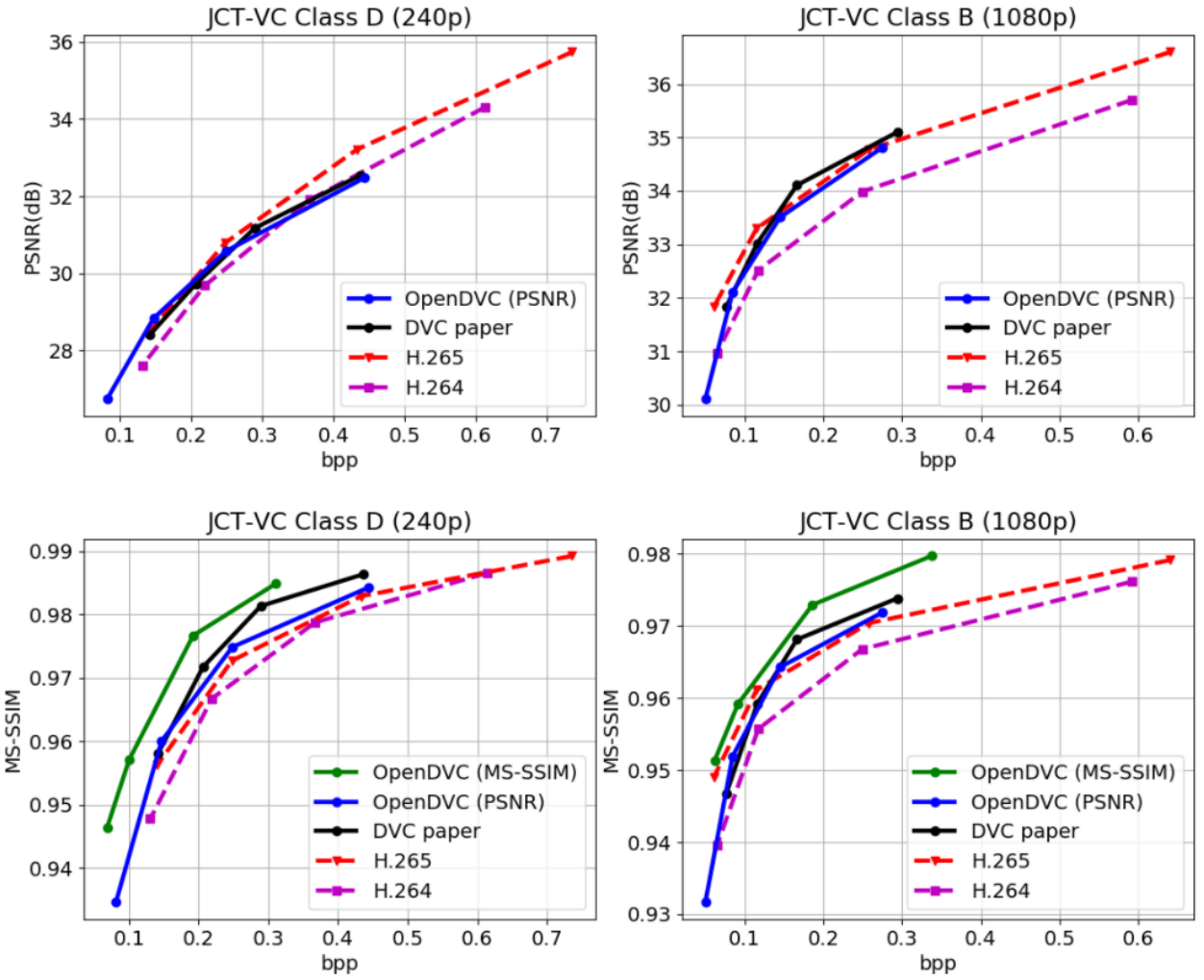
https://github.com/GuoLusjtu/DVC

Open source implementation: OpenDVC [16]

https://github.com/RenYang-home/OpenDVC

Citation:

Yang, Ren, Luc Van Gool, and Radu Timofte. "OpenDVC: An Open Source Implementation of the DVC Video Compression Method." *arXiv preprint arXiv:2006.15862* (2020).



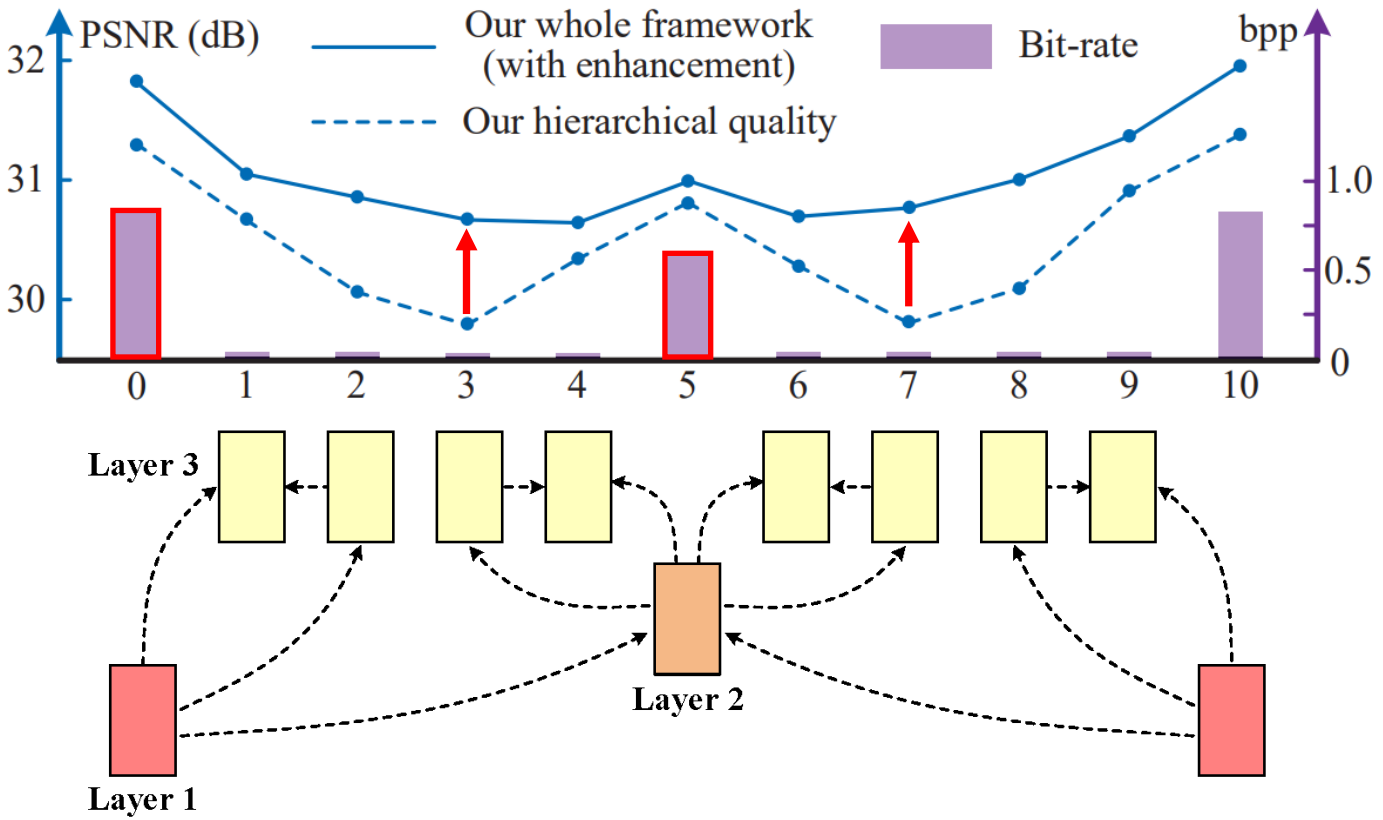
H.264/5: x264/5 LDP veryfast

[14] Lu, Guo, et al. "DVC: An end-to-end deep video compression framework." in CVPR. 2019.

[16] Yang, Ren, et al. "OpenDVC: An Open Source Implementation of the DVC Video Compression Method." arXiv preprint arXiv:2006.15862.

3. End-to-end deep video compression network

- Hierarchical Learned Video Compression (HLVC) with recurrent enhancement [17]



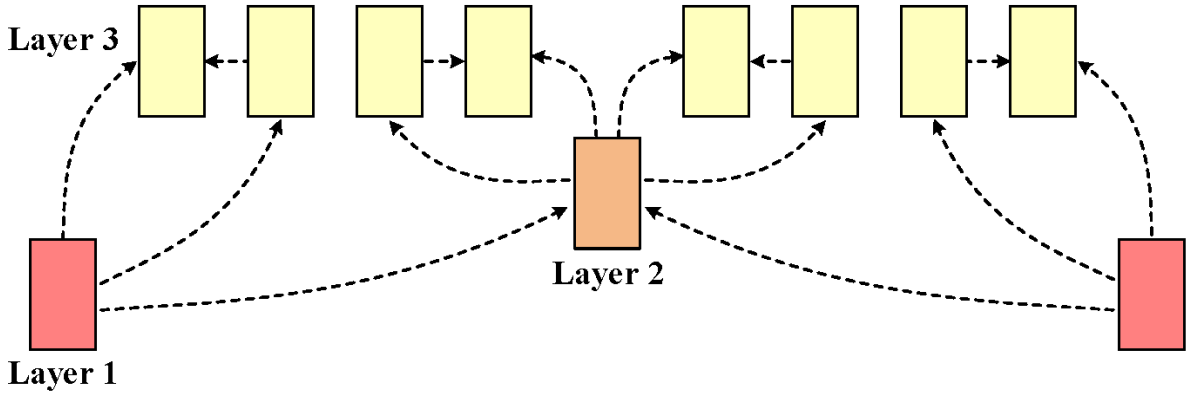
The benefits of hierarchical quality are two-fold:

- **At encoder side**, the high quality frames provide high quality references to improve the compression performance of other frames.
- **At decoder side**, the low quality frames can be enhanced by taking advantage of high quality frames without bit-rate overhead. It is equivalent to reducing bit-rate on low quality frames.

[17] Yang, Ren, et al. "Learning for Video Compression with Hierarchical Quality and Recurrent Enhancement." in CVPR. 2020.

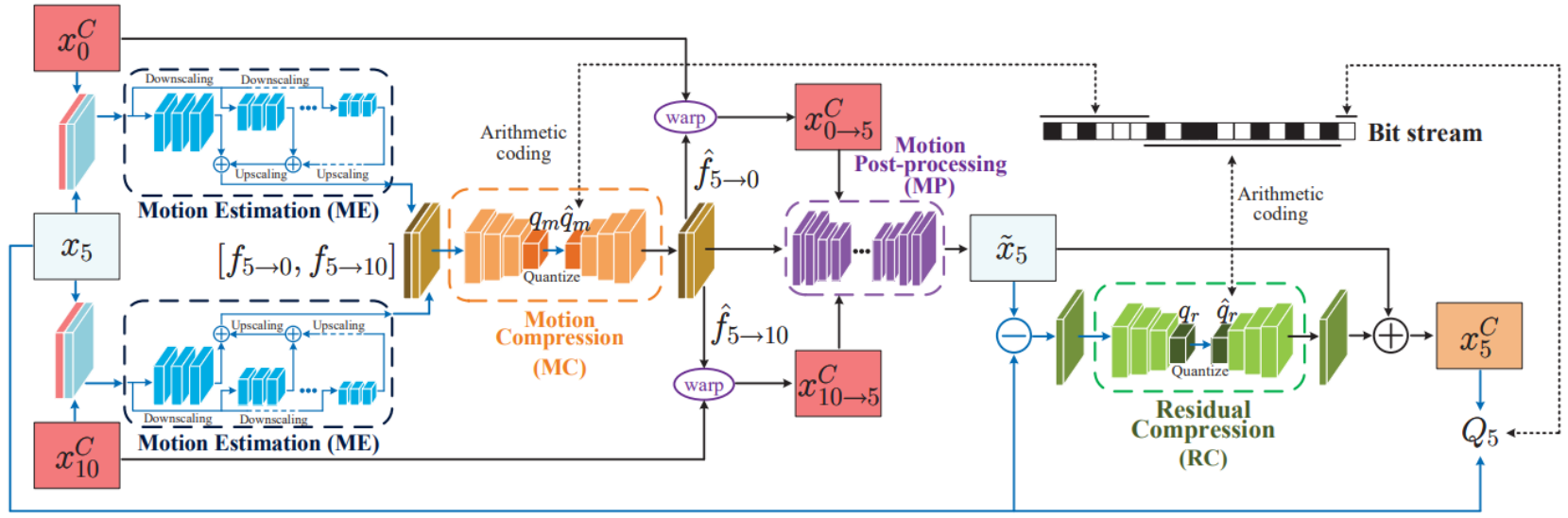
3. End-to-end deep video compression network

- Hierarchical Learned Video Compression (HLVC) with recurrent enhancement [17]



Layer 1:
Compressed by BPG for PSNR model, and by Lee *et al.* ICLR 2019 for MS-SSIM model.

Layer 2:
Compressed by the proposed Bi-Directional Deep Compression (BDDC) network

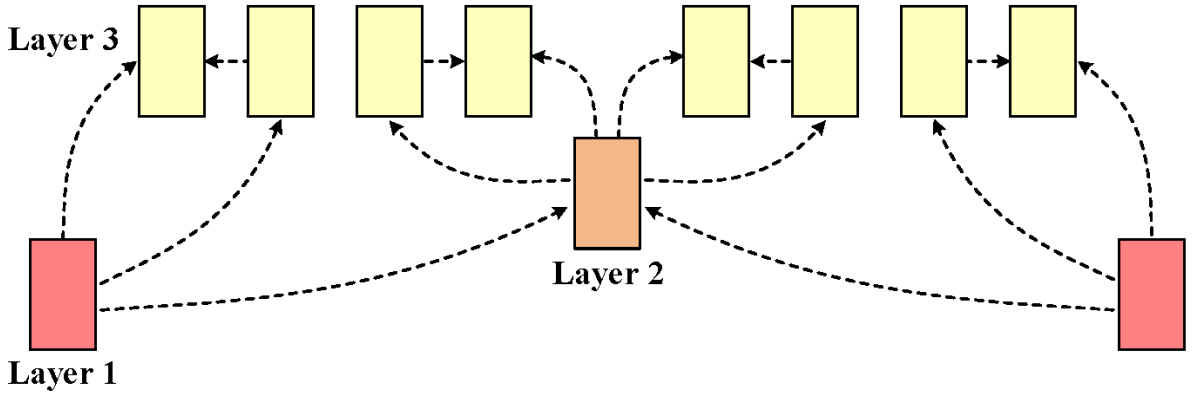


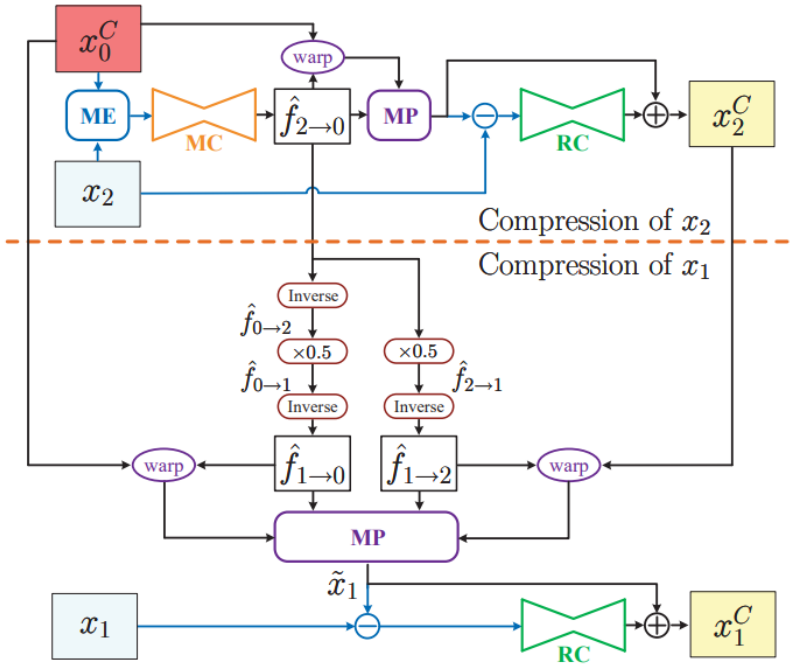
[17] Yang, Ren, et al. "Learning for Video Compression with Hierarchical Quality and Recurrent Enhancement." in CVPR. 2020.

3. End-to-end deep video compression network

- Hierarchical Learned Video Compression (HLVC) with recurrent enhancement [17]

Layer 3:
Compressed by the proposed Single Motion Deep Compression (SMDC) network





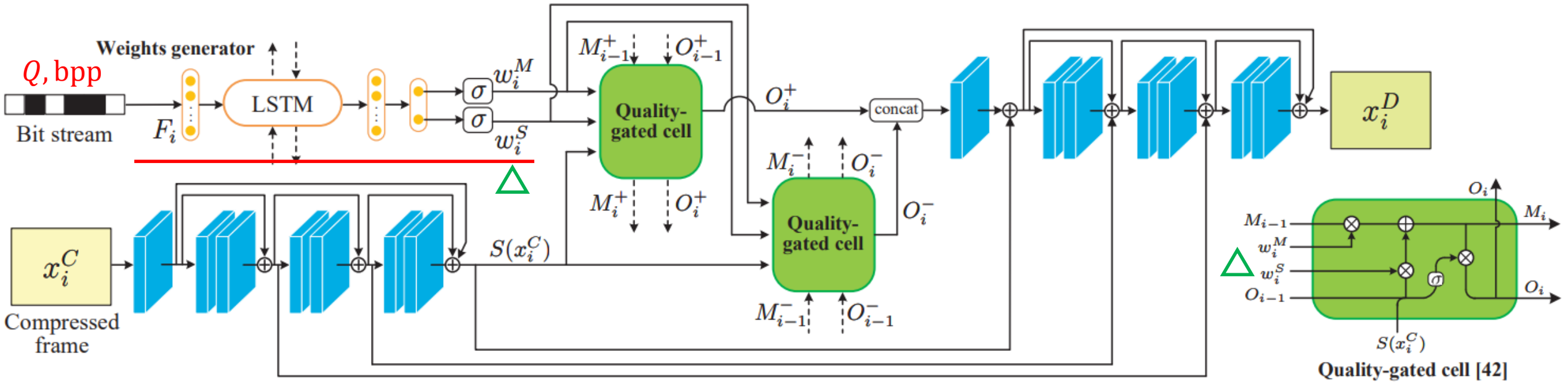
Due to the correlation of motions among multiple neighboring frames, we propose using the motion between x_0^C and x_2 to predict the motions between x_1 and x_0^C or x_2. That is,

$$\hat{f}_{1\to 0} = \text{Inverse}(0.5 \times \underbrace{\text{Inverse}(\hat{f}_{2\to 0})}_{\hat{f}_{0\to 2}}).$$
$$\hat{f}_{0\to 1}$$

As such, x_1 can be compressed with the reference frames of x_0^C and x_2, **without bits consumed for motion map**, thus improving the rate-distortion performance.

3. End-to-end deep video compression network

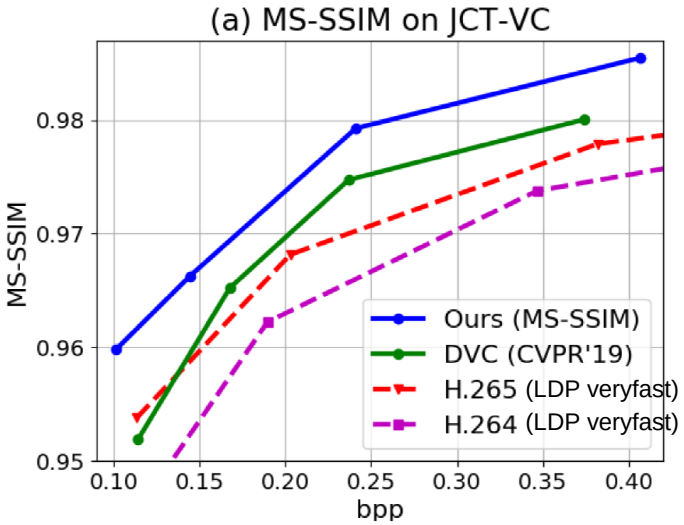
- Hierarchical Learned Video Compression (HLVC) with recurrent enhancement [17]

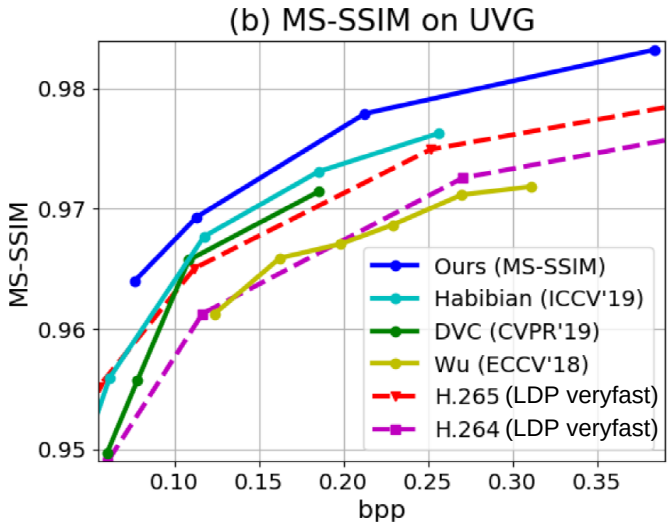


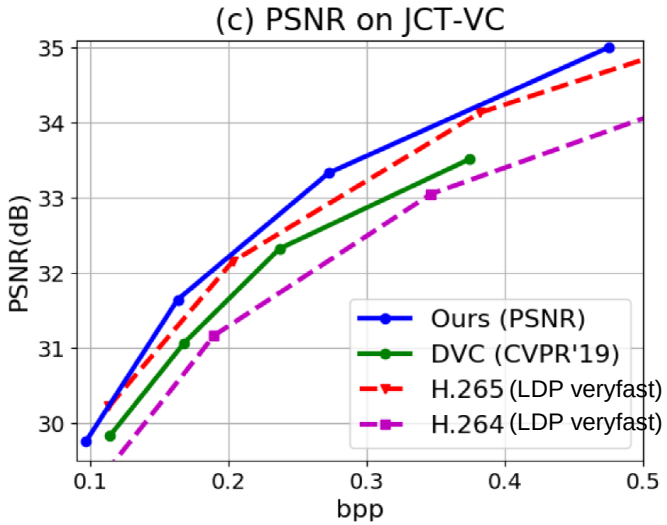
[17] Yang, Ren, et al. "Learning for Video Compression with Hierarchical Quality and Recurrent Enhancement." in CVPR. 2020.

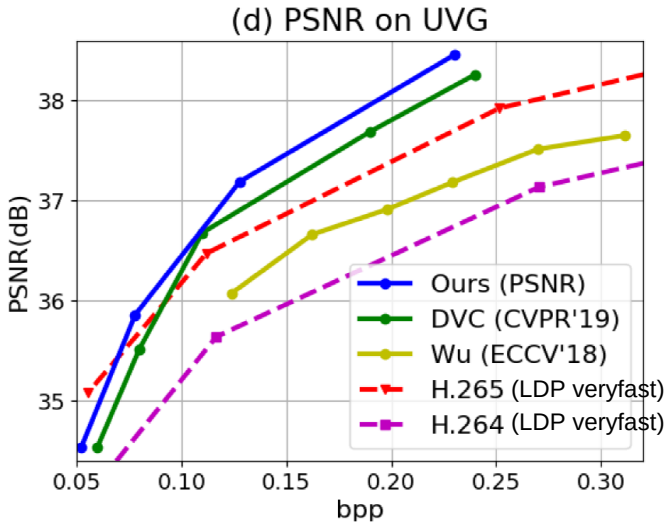
3. End-to-end deep video compression network

- Hierarchical Learned Video Compression (HLVC) with recurrent enhancement [17]









https://github.com/RenYang-home/HLVC

Open source models:

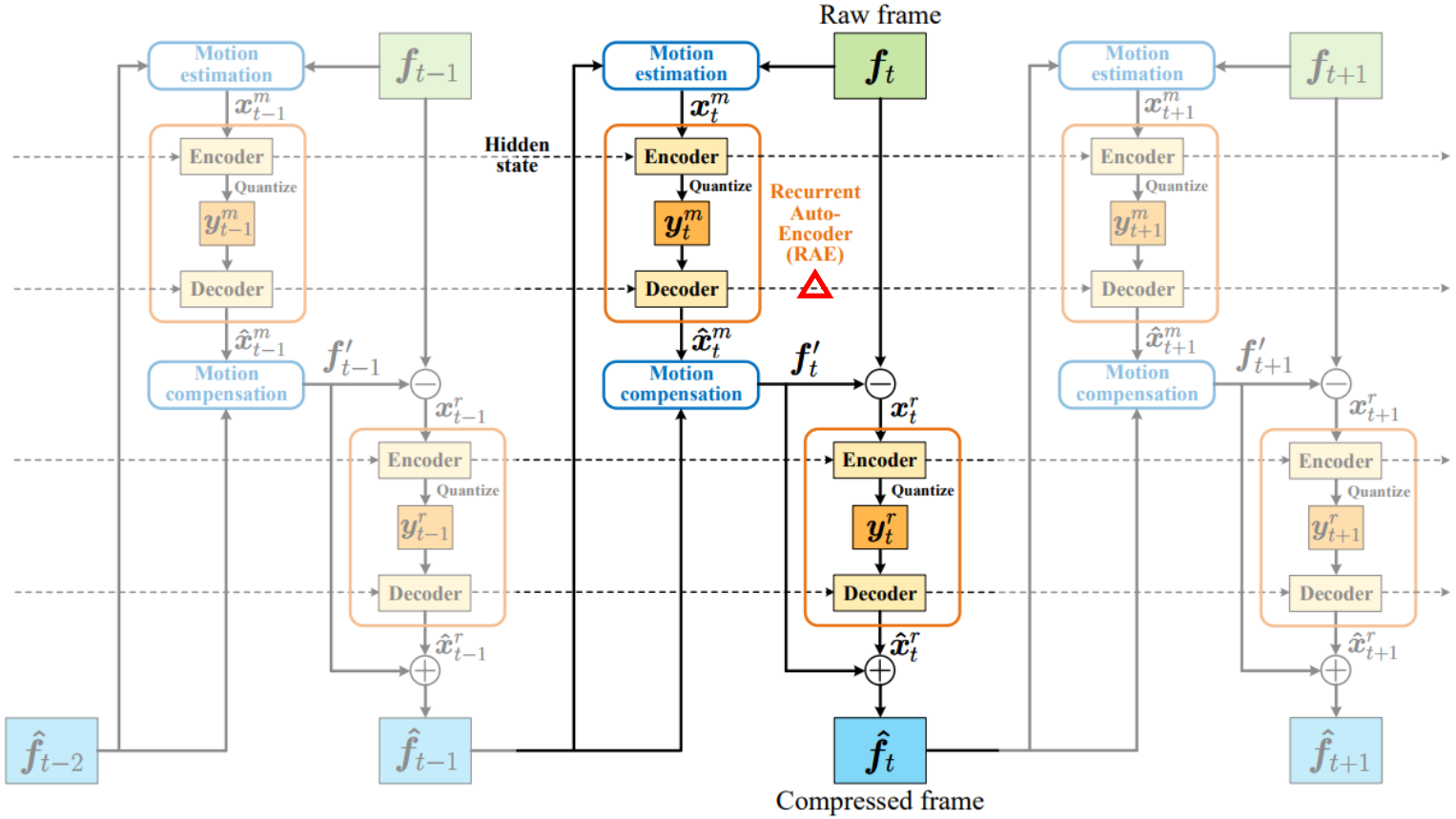
1. Long-distance P-frame compression
2. Long-distance B-frame compression
3. Short-distance P-frame compression
4. Combination of B- and P-frames:
   medium-distance P-frame
   + short-distance B-frame.

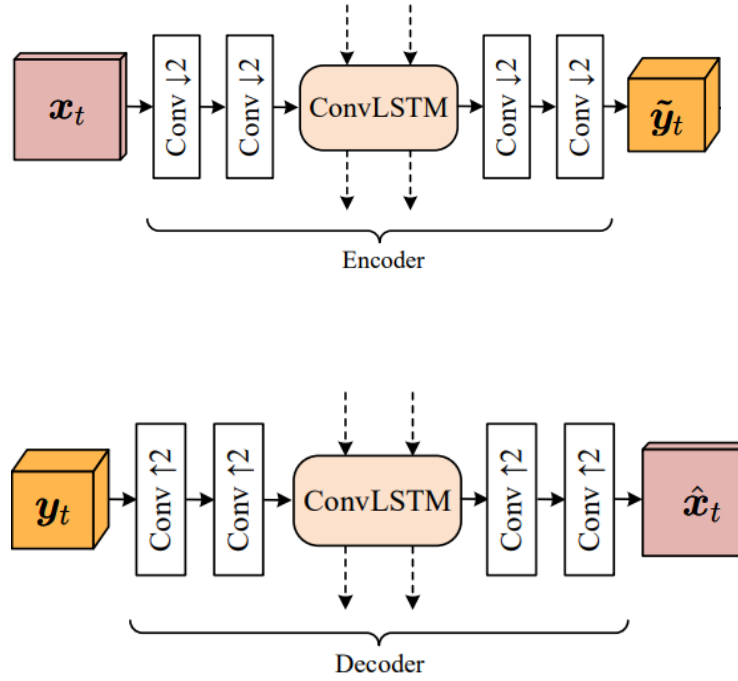
5. Demos for compressing video sequence:
   "fast" and "slow" modes

[17] Yang, Ren, et al. "Learning for Video Compression with Hierarchical Quality and Recurrent Enhancement." in CVPR. 2020.

Learned Video Compression

- Recurrent Learned Video Compression (HLVC) [18]

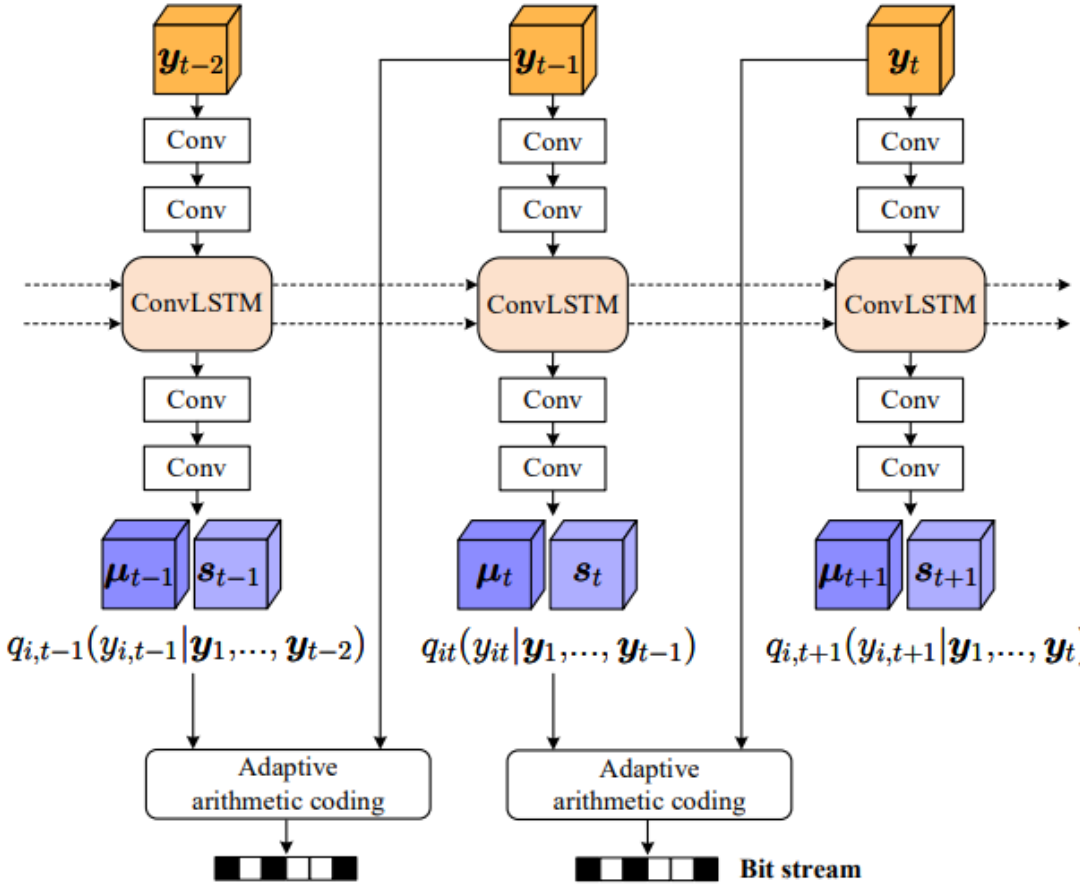




[18] Yang, Ren, et al. "Learning for Video Compression with Recurrent Auto-Encoder and Recurrent Probability Model." IEEE J-STSP. 2021

3. End-to-end deep video compression network

- Recurrent Learned Video Compression (HLVC) [18]



Recurrent Probability Model (RPM)

Codes (to be released)

$$H(p_t, q_t) = \mathbb{E}_{\boldsymbol{y}_t \sim p_t}[-\log_2 q_t(\boldsymbol{y}_t \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1})]$$

$$q_t(\boldsymbol{y}_t \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}) = \prod_{i=1}^{N} q_{it}(y_{it} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1})$$

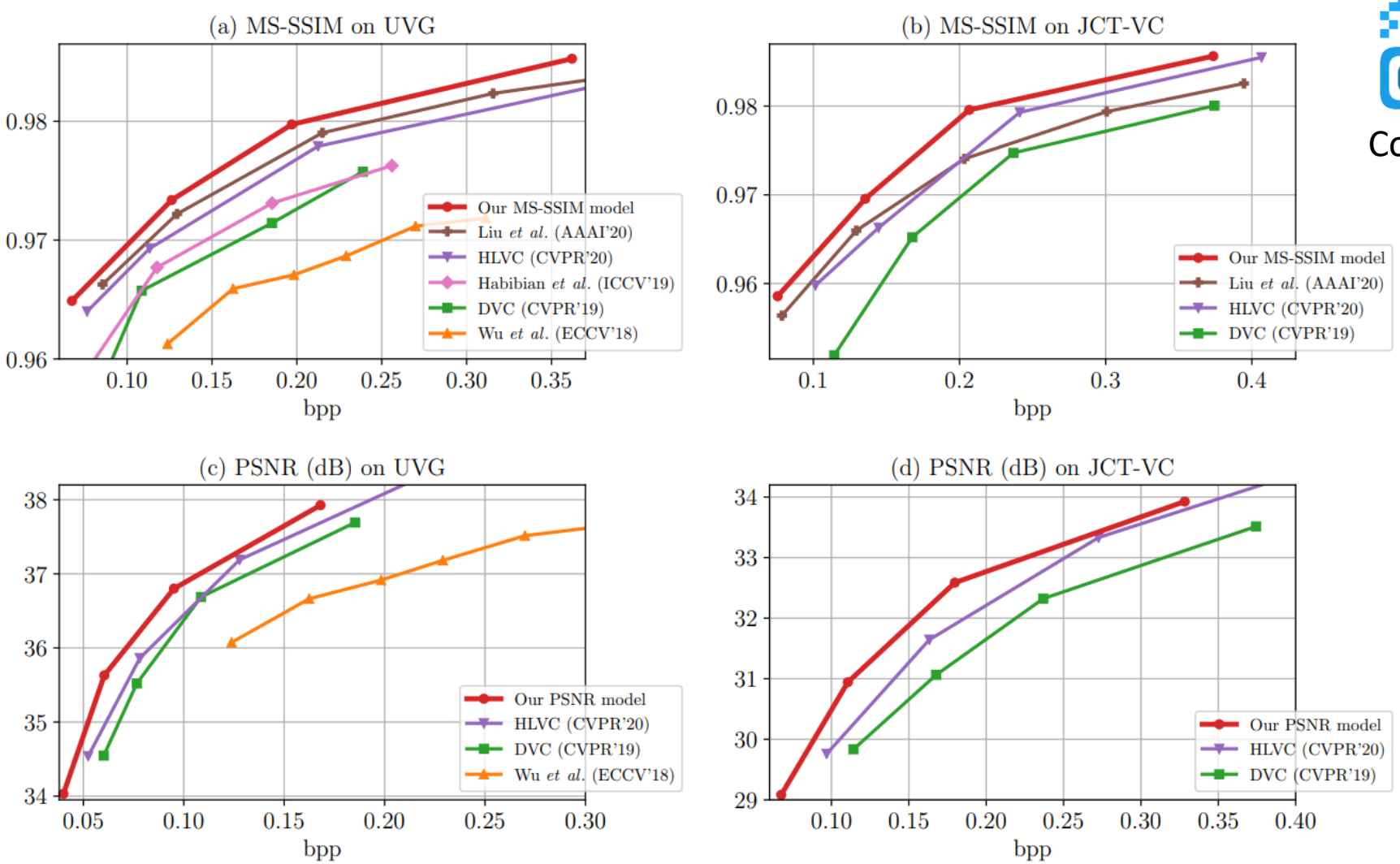
$$q_{it}(y_{it} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}) = \int_{y_{it}-0.5}^{y_{it}+0.5} \text{Logistic}(y; \mu_{it}, s_{it}) dy$$

[18] Yang, Ren, et al. "Learning for Video Compression with Recurrent Auto-Encoder and Recurrent Probability Model." IEEE J-STSP. 2021

3. End-to-end deep video compression network

- Recurrent Learned Video Compression (HLVC) [18]

Codes (to be released)

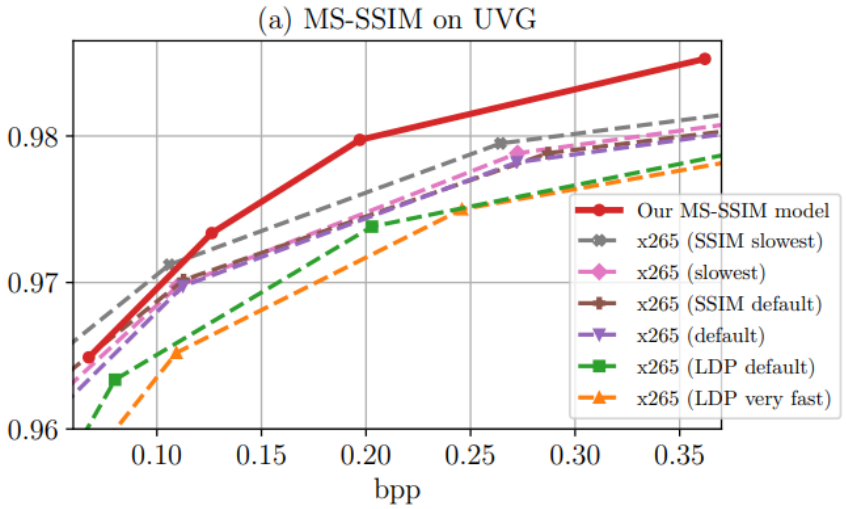


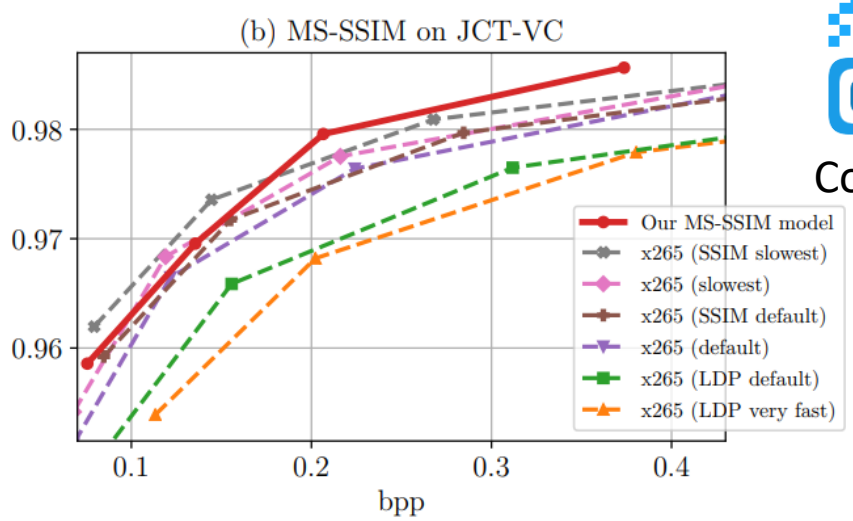
[18] Yang, Ren, et al. "Learning for Video Compression with Recurrent Auto-Encoder and Recurrent Probability Model." IEEE J-STSP. 2021

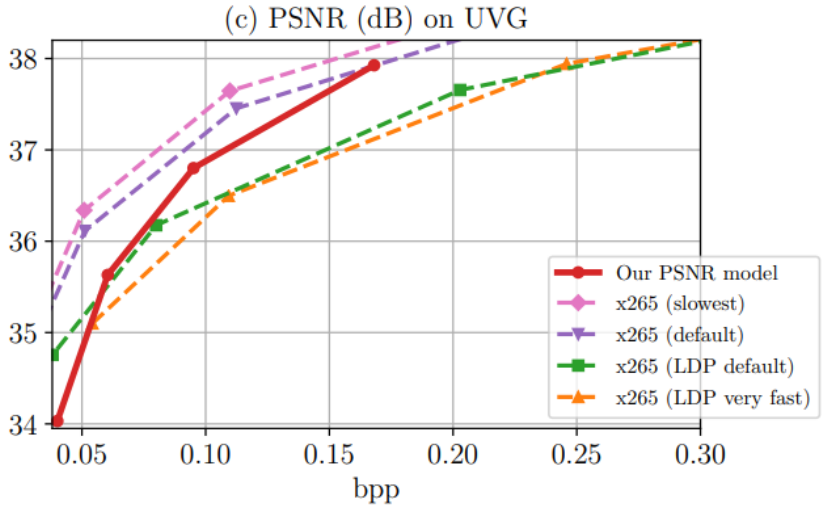
3. End-to-end deep video compression network

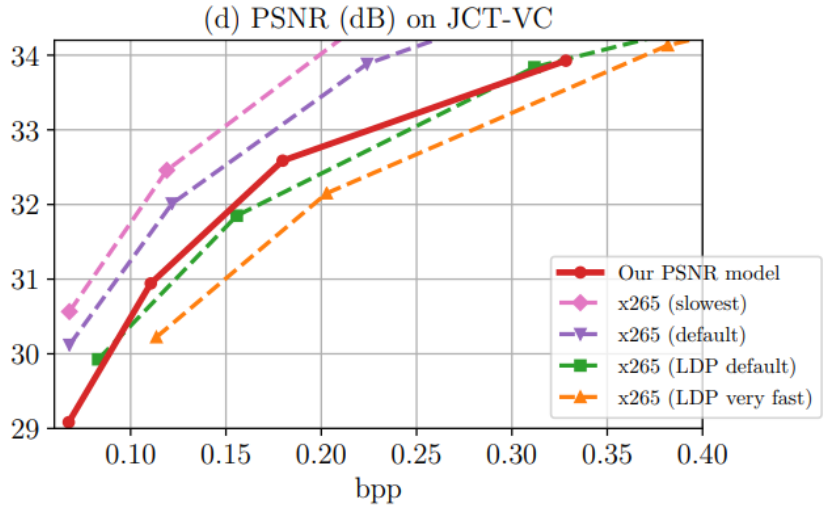
- Recurrent Learned Video Compression (HLVC) [18]

Codes (to be released)









[18] Yang, Ren, et al. "Learning for Video Compression with Recurrent Auto-Encoder and Recurrent Probability Model." IEEE J-STSP. 2021

3. End-to-end deep video compression network

- Other recent works:

[19] Djelouah, Abdelaziz, et al. "Neural inter-frame compression for video coding." Proceedings of the IEEE International Conference on Computer Vision. 2019.

[20] Habibian, Amirhossein, et al. "Video compression with rate-distortion autoencoders." Proceedings of the IEEE International Conference on Computer Vision. 2019.

[21] Lin, Jianping, et al. "M-LVC: Multiple Frames Prediction for Learned Video Compression." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[22] Agustsson, Eirikur, et al. "Scale-Space Flow for End-to-End Optimized Video Compression." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[23] Lu, Guo, et al. "Content Adaptive and Error Propagation Aware Deep Video Compression." Proceedings of the European Conference on Computer Vision. 2020.

[24] Hu, Zhihao, et al. "Improving deep video compression by resolution-adaptive flow coding." Proceedings of the European Conference on Computer Vision. 2020.

[25] Golinski, Adam, et al. "Feedback Recurrent Autoencoder for Video Compression." Proceedings of the Asian Conference on Computer Vision. 2020.

3. End-to-end deep video compression network

- Will learning-based compression be standardized?
- Can learning-based method be compatible with traditional standards (e.g., HEVC and JPEG)?

JPEG initiates standardisation of image compression based on AI

The 89th JPEG meeting was held online from 5 to 9 October 2020.

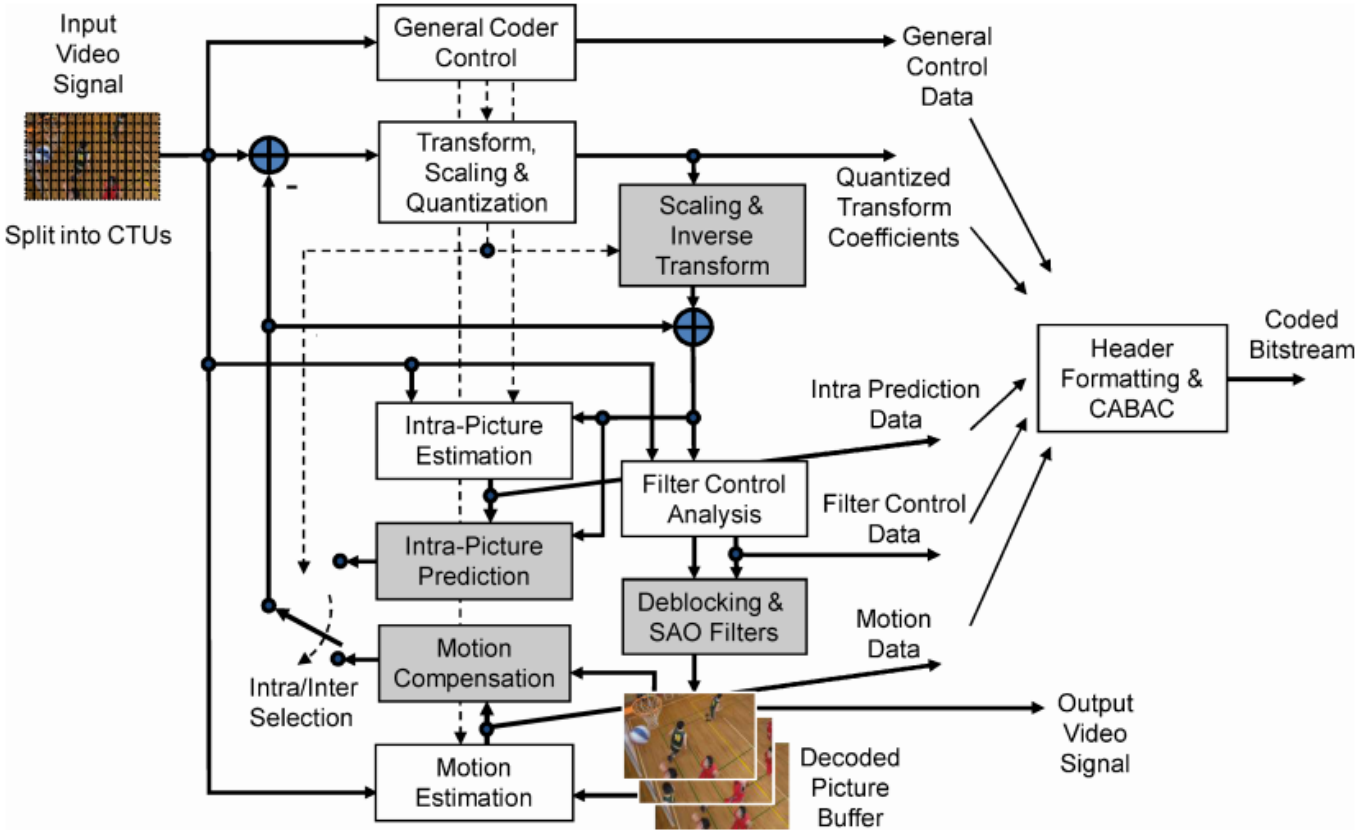
During this meeting multiple JPEG standardisation activities and explorations were discussed and progressed. Notably, the call for evidence on learning-based image coding was successfully completed and evidence was found that this technology promises several new functionalities while offering at the same time superior compression efficiency, beyond the state of the art.

JPEG AI

At the 89th meeting the submissions to the Call for Evidence on learning-based image coding were presented and discussed. Four submissions were received in response to the Call for Evidence. The results of the subjective evaluation of the submissions to the Call for Evidence were reported and discussed in detail by experts. It was agreed that there is strong evidence that learning-based image coding solutions can outperform the already defined anchors in terms of compression efficiency, when compared to state-of- the-art conventional image coding architecture. Thus, it was decided to create a new standardisation activity for a JPEG AI on learning-based image coding system, that applies machine learning tools to achieve substantially better compression efficiency compared to current image coding systems, while offering unique features desirable for an efficient distribution and consumption of images. This type of approach should allow to obtain an efficient compressed domain representation not only for visualisation, but also for machine learning based image processing and computer vision. JPEG AI releases to the public the results of the objective and subjective evaluations as well as a first version of common test conditions for assessing the performance of leaning-based image coding systems.

3. End-to-end deep video compression network

- Will learning-based compression be standardized?
- Can learning-based method be compatible with traditional standards (e.g., HEVC and JPEG)?



Traditional video compression framework (HEVC) [1]

Gray: modules in both encoder and decoder; White: modules only in encoder

4. Learning for compression with standard decoder (compatible with common image and video viewer)

Accelerating HEVC encoding without changing the decoder [26]

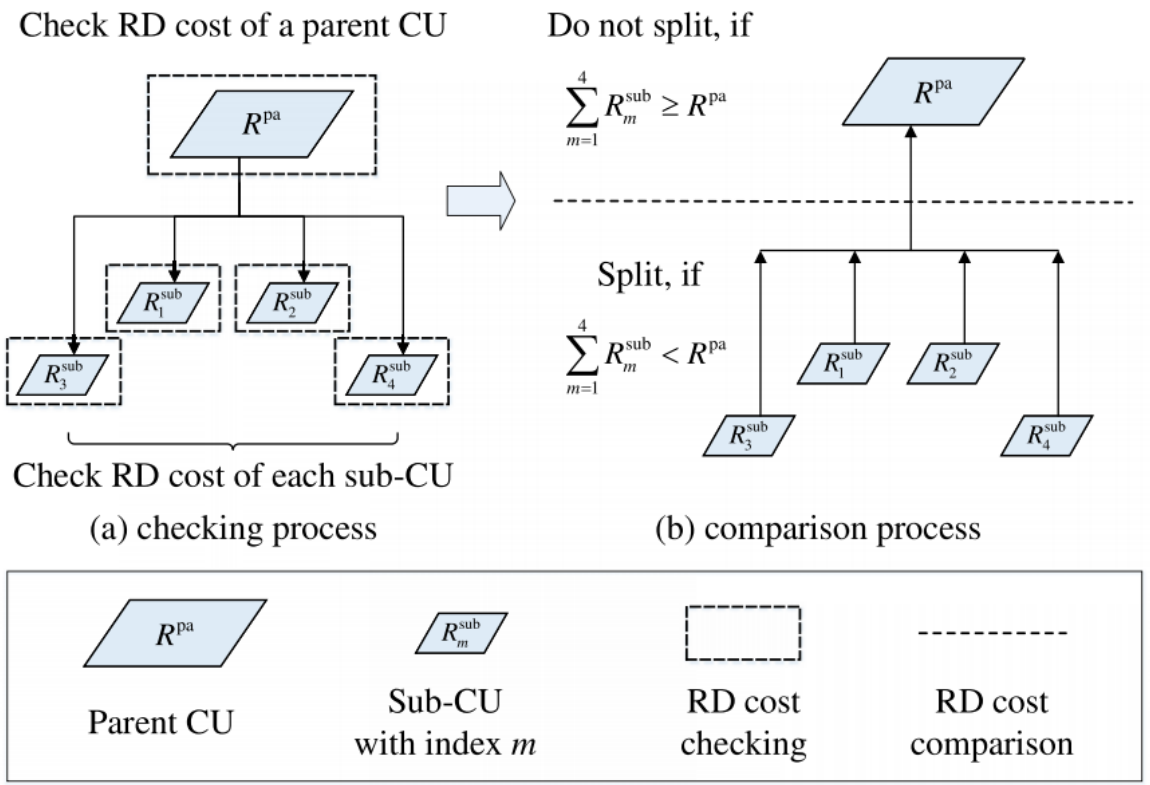


Fig. 1. Illustration of checking and comparing RD cost between a parent CU and its sub-CUs. Note that this illustration can be applied to the splitting of $64 \times 64 \rightarrow 32 \times 32$, $32 \times 32 \rightarrow 16 \times 16$ or $16 \times 16 \rightarrow 8 \times 8$.

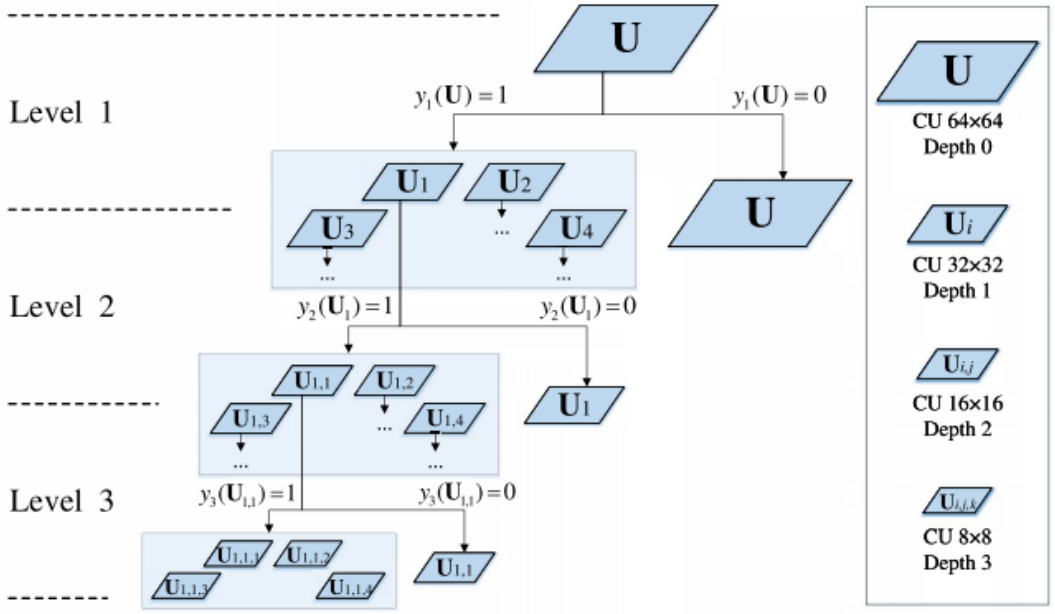
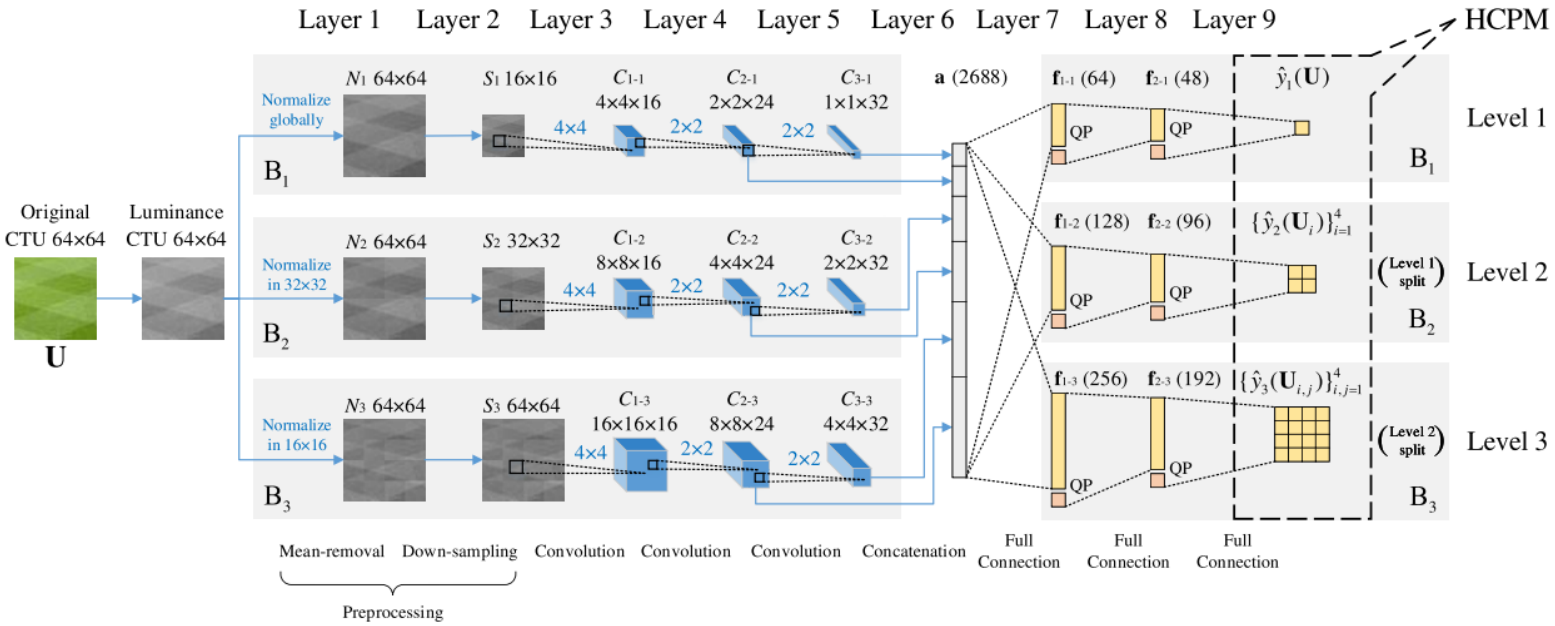


Fig. 2. Illustration of three-level CU classifier.

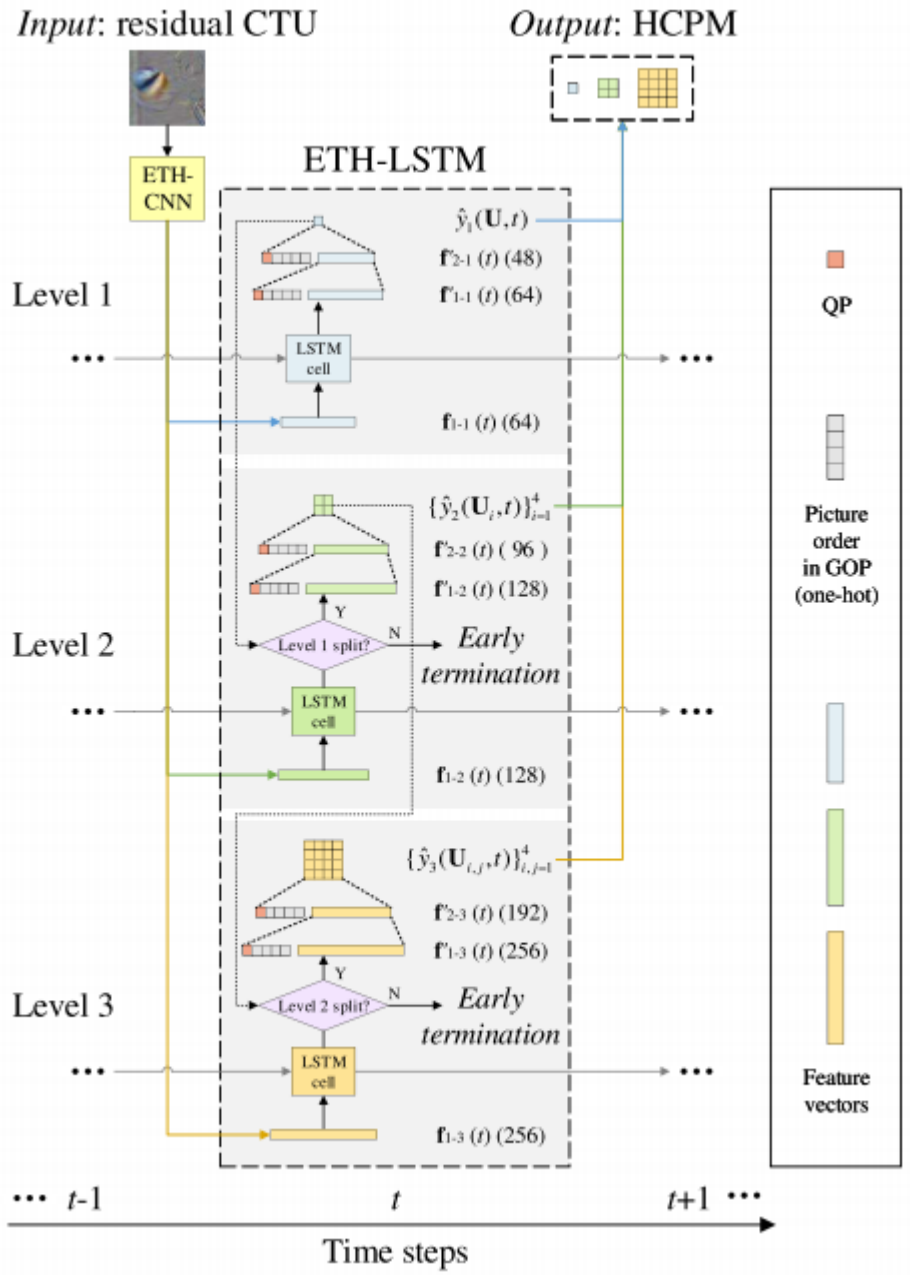
[26] Li, Tianyi, et al. "Reducing complexity of HEVC: A deep learning approach." IEEE T-IP. 2018.

4. Learning for compression with standard decoder (compatible with common image and video viewer)

Accelerating HEVC encoding without changing the decoder



Intra-mode: CNN-based network for predicting CU partition



Inter-mode: LSTM-based network

[26] Li, Tianyi, et al. "Reducing complexity of HEVC: A deep learning approach." IEEE T-IP. 2018.

4. Learning for compression with standard decoder (compatible with common image and video viewer)

Accelerating HEVC encoding without changing the decoder [26]

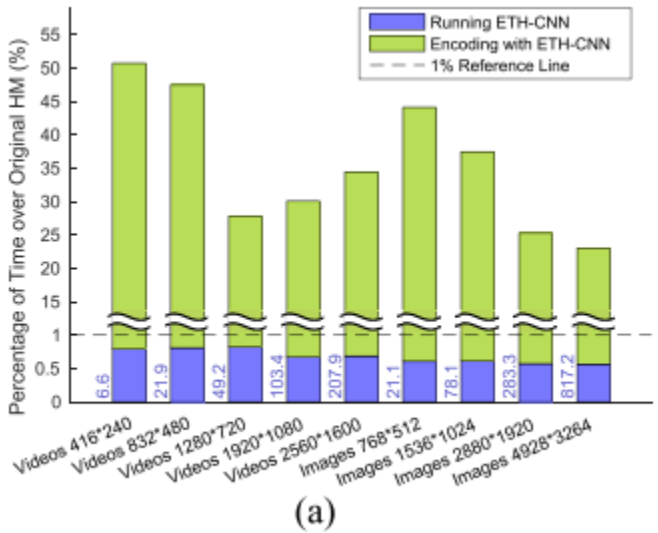
Codes

RESULTS FOR SEQUENCES OF THE JCT-VC TEST SET (AI)

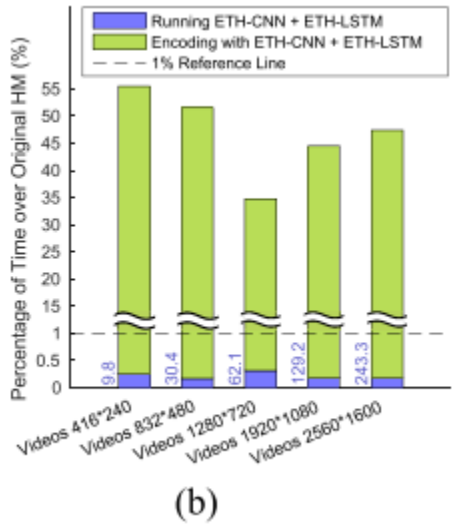
Class	Sequence	Appr.	BD-BR (%)	BD-PSNR (dB)	ΔT (%) QP=22	QP=27	QP=32	QP=37
Std. dev.		[25]	2.553	0.175	9.42	10.16	13.04	15.01
		[9]	2.603	0.158	**7.10**	**6.24**	**6.09**	**5.63**
		Our	**1.020**	**0.039**	12.67	12.96	11.96	10.75
Best		[25]	3.630	-0.149	-67.32	-63.09	-63.98	-70.70
		[9]	2.382	-0.082	-70.66	-72.75	-73.62	-73.86
		Our	**0.622**	**-0.039**	**-80.40**	**-84.32**	**-84.45**	**-84.97**
Average		[25]	8.559	-0.419	-50.05	-46.72	-43.09	-45.01
		[9]	6.189	-0.316	-56.41	-60.23	-62.62	-65.04
		Our	**2.247**	**-0.104**	**-56.92**	**-60.38**	**-63.61**	**-66.47**

RESULTS FOR SEQUENCES OF THE JCT-VT TEST SET (LDP)

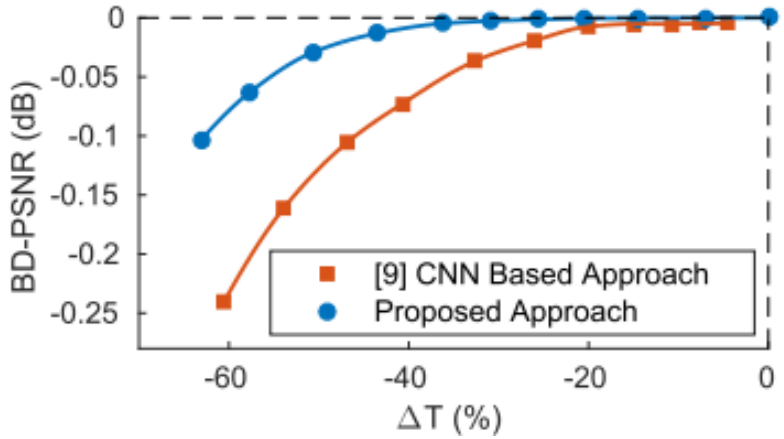
Class	Sequence	Appr.	BD-BR (%)	BD-PSNR (dB)	ΔT (%) QP=22	QP=27	QP=32	QP=37
Std. dev.		[8]	0.534	**0.013**	10.84	13.44	14.82	14.88
		[7]	2.146	0.100	16.62	18.13	17.18	16.08
		[10]	1.282	0.037	6.17	9.34	12.04	13.30
		Our	**0.448**	0.014	**5.60**	**9.27**	**8.84**	**8.45**
Best		[8]	0.974	-0.029	-59.90	-72.00	-74.29	-73.91
		[7]	1.934	-0.028	**-65.45**	**-75.72**	**-81.45**	**-84.83**
		[10]	1.949	-0.042	-55.45	-58.49	-59.50	-60.03
		Our	**0.770**	**-0.017**	-53.36	-68.77	-72.19	-75.27
Average		[8]	1.799	-0.054	-38.31	-46.10	-48.95	-46.50
		[7]	5.051	-0.163	-32.16	-39.59	-49.31	-57.20
		[10]	3.616	-0.108	-43.67	-43.10	-43.54	-43.98
		Our	**1.495**	**-0.046**	**-43.84**	**-52.13**	**-57.89**	**-62.94**



(a)



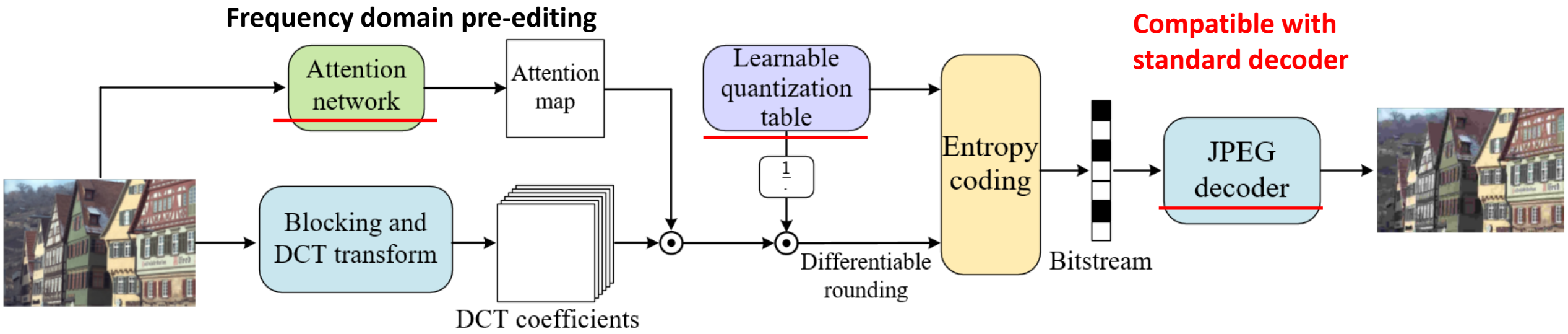
(b)



[26] Li, Tianyi, et al. "Reducing complexity of HEVC: A deep learning approach." IEEE T-IP. 2018.

4. Learning for compression with standard decoder (compatible with common image and video viewer)

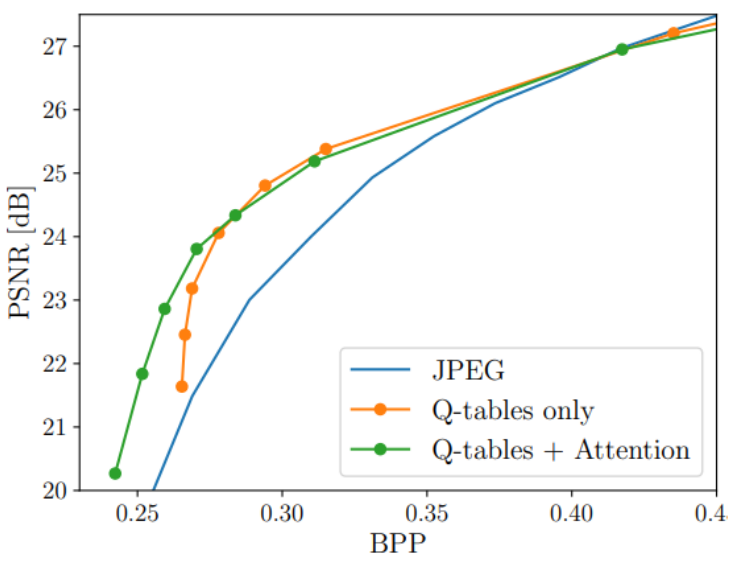
Learning to improve image compression with standard JPEG decoder [27]



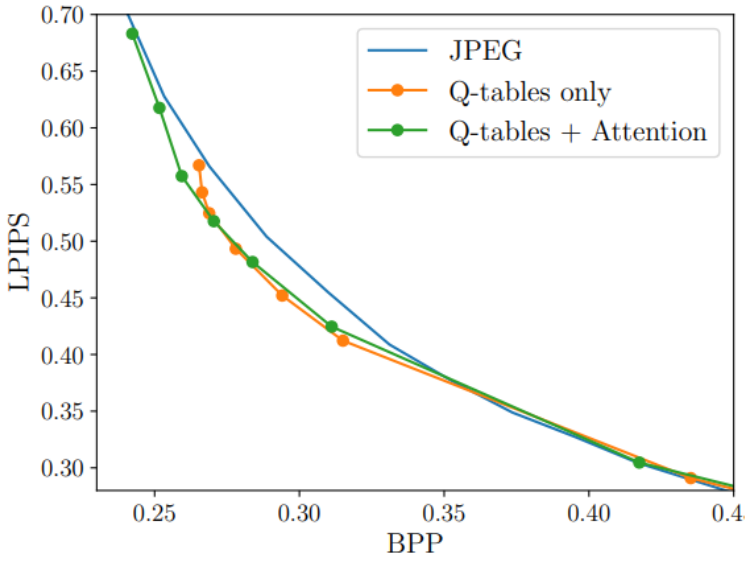
[27] Strümpler, Yannick, Ren Yang, Radu Timofte. "Learning to Improve Image Compression without Changing the Standard Decoder." in ECCVW.

4. Learning for compression with standard decoder (compatible with common image and video viewer)

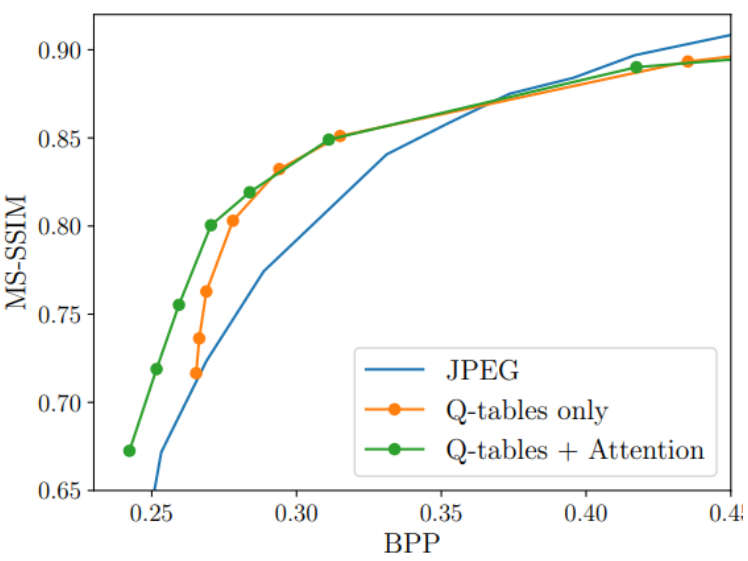
Learning to improve image compression with standard JPEG decoder [27]



PSNR on Kodak



LPIPS on Kodak

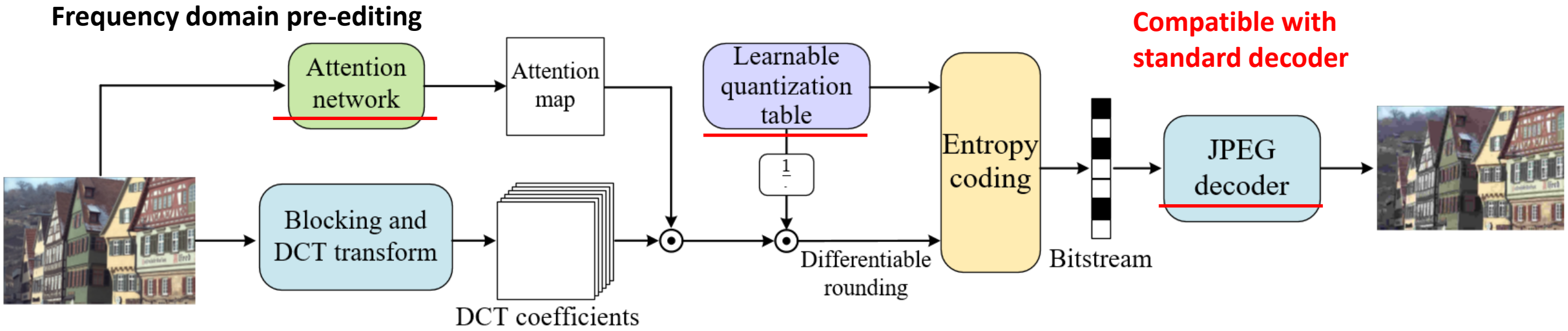


MS-SSIM on Kodak

[27] Strümpler, Yannick, Ren Yang, Radu Timofte. "Learning to Improve Image Compression without Changing the Standard Decoder." in ECCVW.

4. Learning for compression with standard decoder (compatible with common image and video viewer)

Learning to improve image compression with standard JPEG decoder [27]



- We achieve better rate-distortion performance without changing the standard decoder
- The compressed image can be decoded (viewed) on any common device, e.g., mobile, ipad, PC, etc.

[27] Strümpler, Yannick, Ren Yang, Radu Timofte. "Learning to Improve Image Compression without Changing the Standard Decoder." in ECCVW.

ETH zürich

Thanks for your attention

Q & A

Ren Yang (杨韧)

Doctoral Student | Scientific Assistant
Computer Vision Laboratory | ETH Zürich
Sternwartstrasse 7 | 8092 Zürich, Switzerland

E-mail: ren.yang@vision.ee.ethz.ch

https://renyang-home.github.io/

This slide is available at my homepage