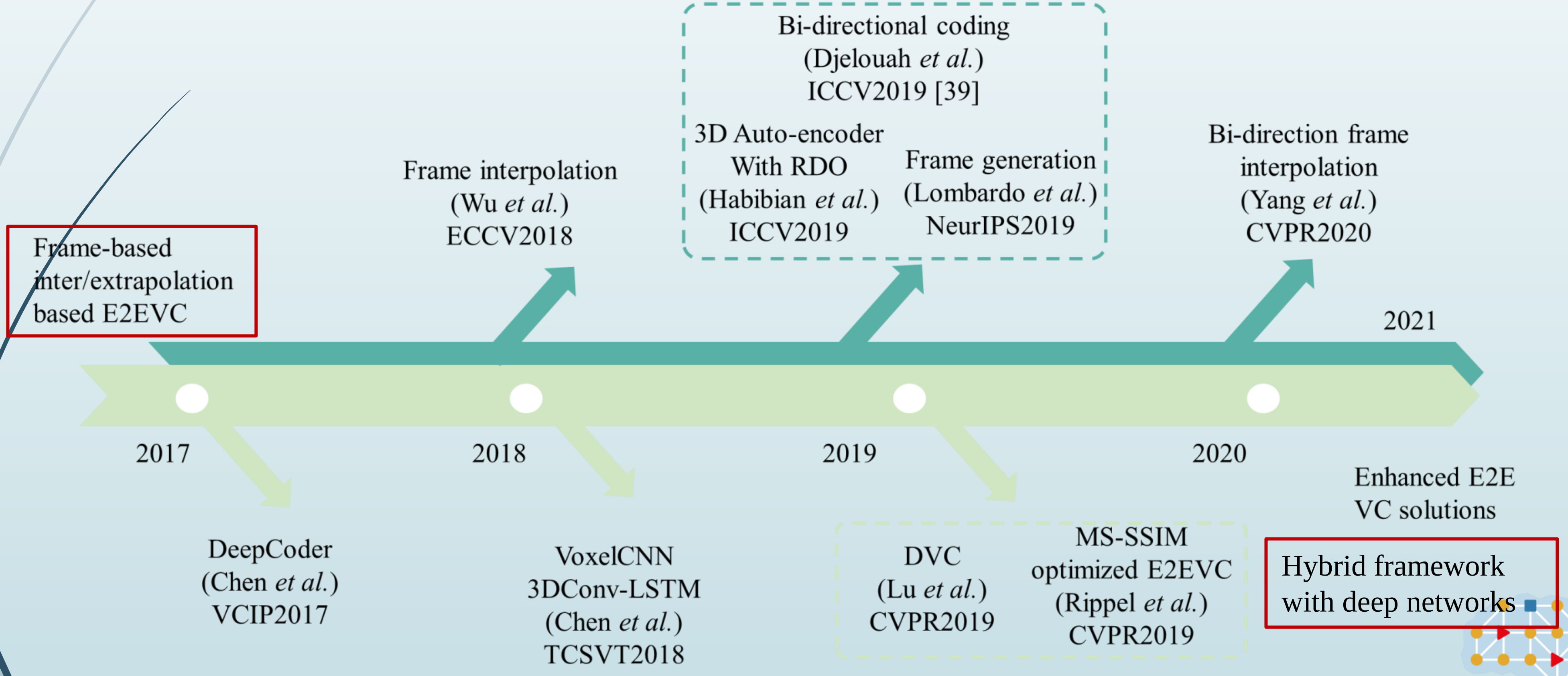
MPAI – End-to-end Video (EEV) Project and Activities

2023/03/01

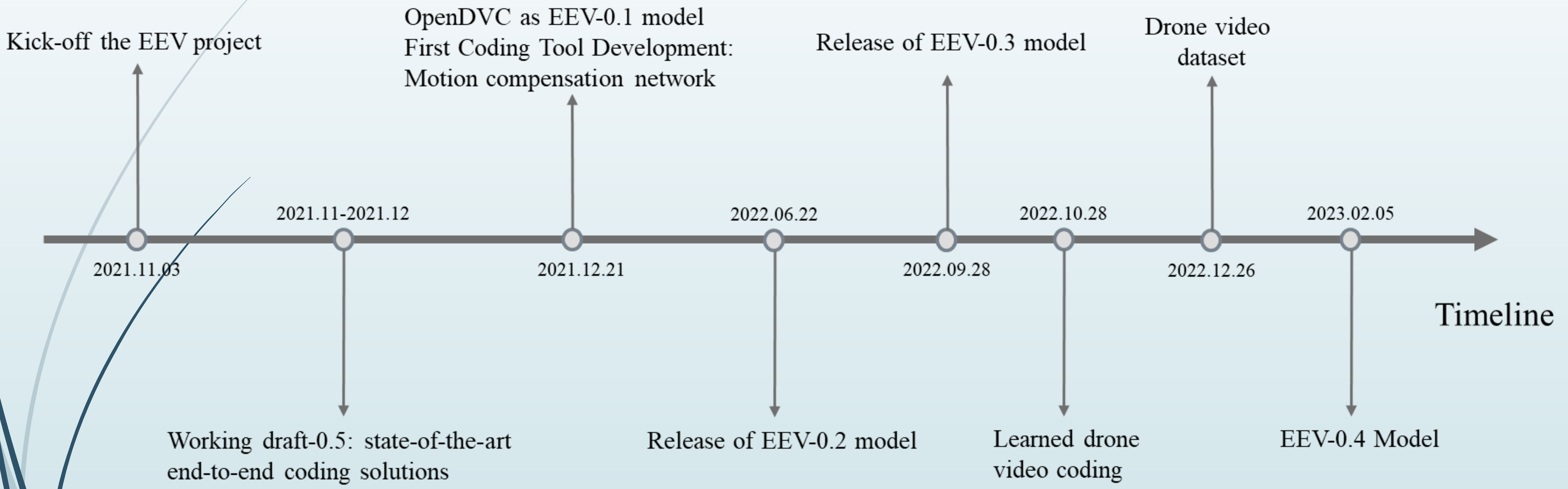
Chuanmin Jia
Peking University / Peng Cheng Laboratory
cmjia@pku.edu.cn

Video Coding Schemes using Fully Neural Models

- Frame generation based End-to-end video coding (EEV)
  - Neural intra codec plus frame interpolation
- Hybrid framework EEV
  - Prediction plus transform based codec using fully networks



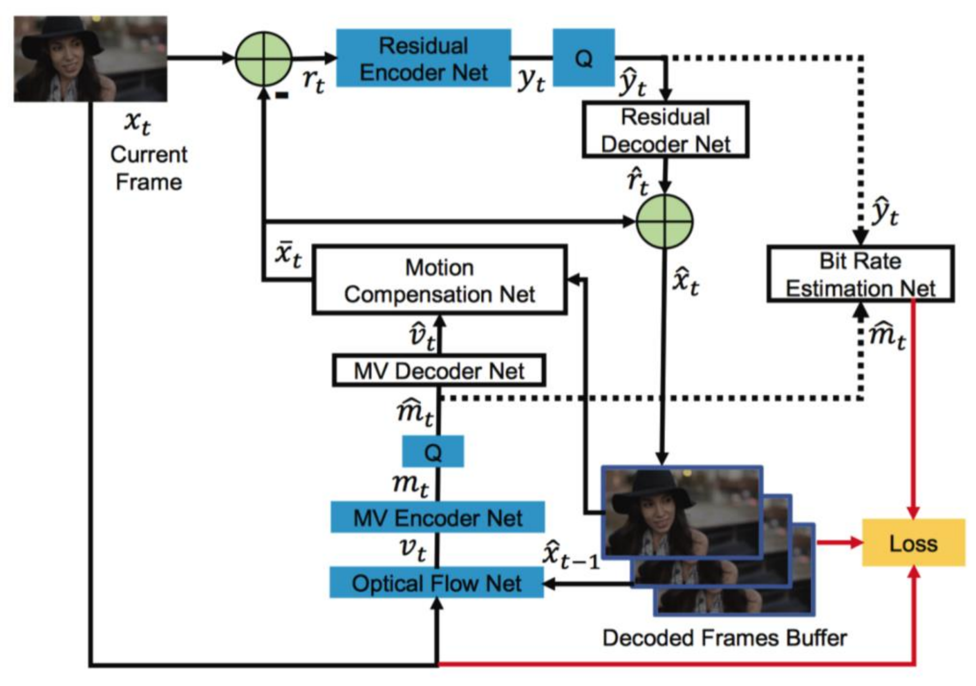
MPAI EEV Milestone



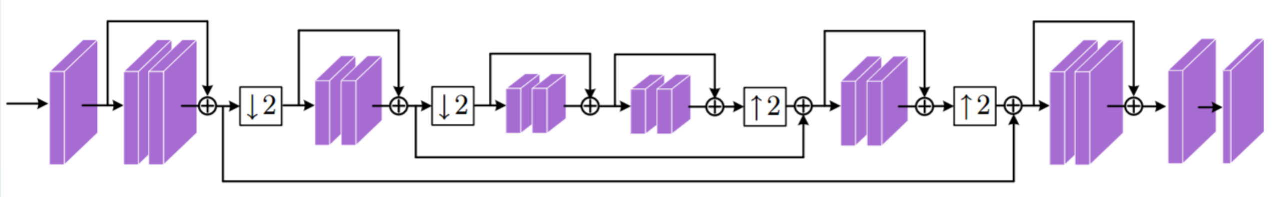
AI-based End-to-End Video Coding & OpenDVC

Reference model of EEV-0.1: OpenDVC

- Neural Motion Estimation
  - Optical Flow (motion field)
- Neural Motion Compensation
  - Predicted Frame
- Residual Compression
  - Autoencoder
- Bit-rate estimation
  - Motion field
  - Residual coding



Framework



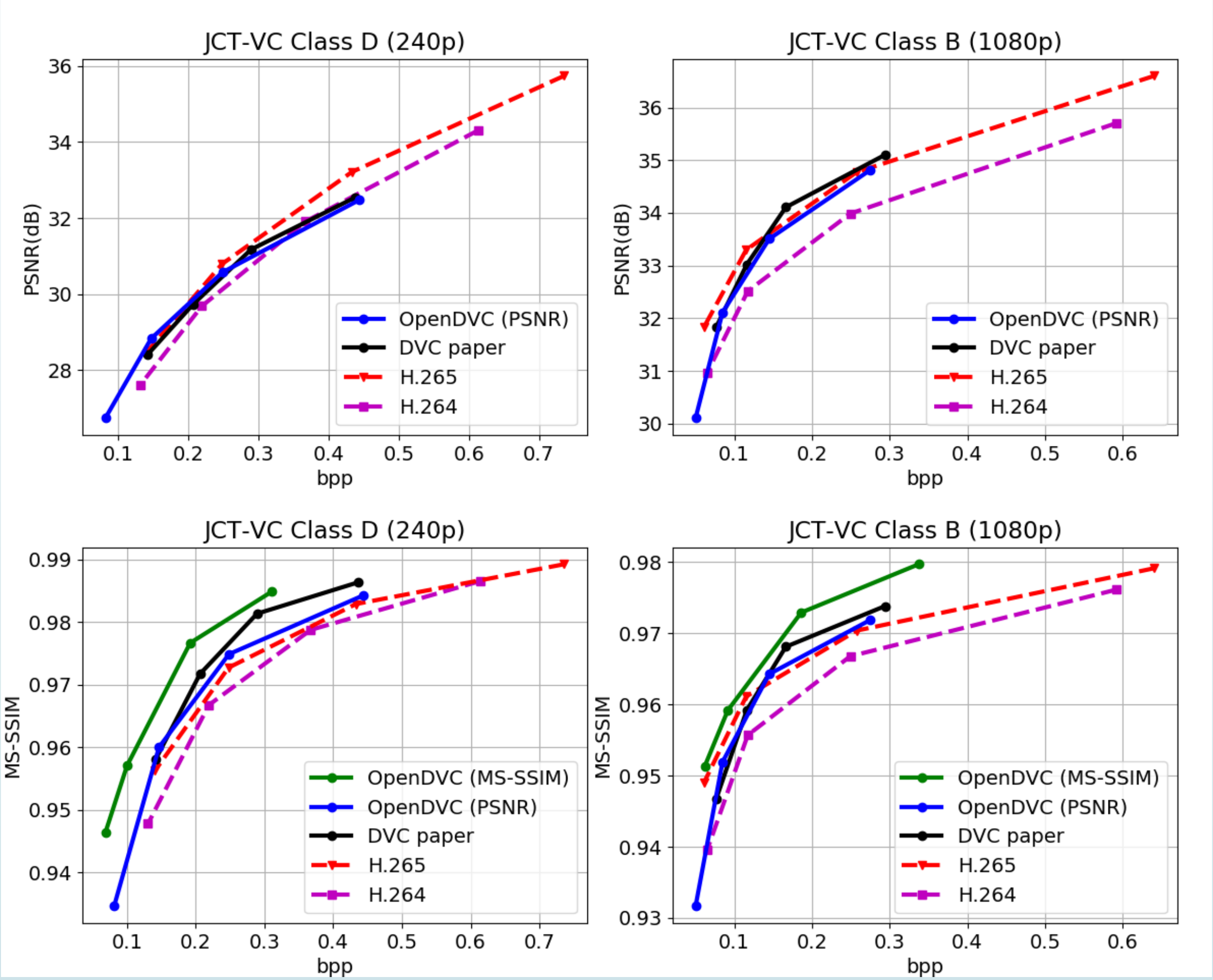
Motion Estimation Net

OpenDVC: An Open Source Implementation of the DVC Video Compression Method, https://arxiv.org/abs/2006.15862

MPAI. community

Performance

- Test data
  - JCT-VC sequences
- Performance of OpenDVC
  - Better than x264 and x265



OpenDVC: An Open Source Implementation of the DVC Video Compression Method, https://arxiv.org/abs/2006.15862

EEV Model Development

- Drone videos benchmark including the following sequences

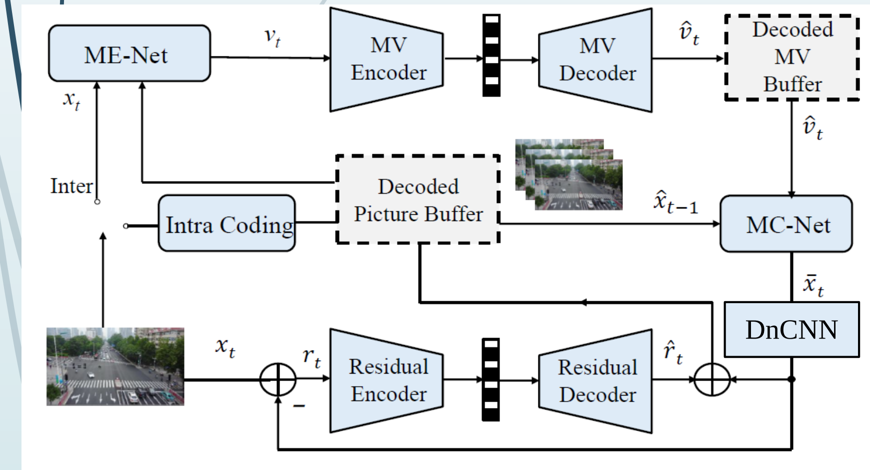
EEV Model Development

- Drone videos benchmark including the following sequences

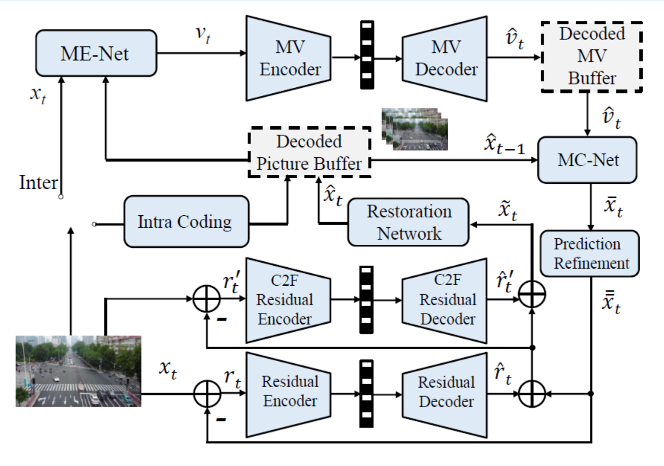
Source	Sequence Name	Spatial Resolution	Frame Count	Frame Rate	Bit Depth	Scene Feature
Class A VisDrone-SOT TPAMI2021 [1]	BasketballGround	960x528	100	24	8	Outdoor
	GrassLand	1344x752	100	24	8	Outdoor
	Intersection	1360x752	100	24	8	Outdoor
	NightMall	1920x1072	100	30	8	Outdoor
	SoccerGround	1904x1056	100	30	8	Outdoor
Class B VisDrone-MOT TPAMI2021 [1]	Circle	1360x752	100	24	8	Outdoor
	CrossBridge	2720x1520	100	30	8	Outdoor
	Highway	1344x752	100	24	8	Outdoor
Class C Corridor IROS2018 [9]	Classroom	640x352	100	24	8	Indoor
	Elevator	640x352	100	24	8	Indoor
	Hall	640x352	100	24	8	Indoor
Class D UAVDT_S ECCV2018 [10]	Campus	1024x528	100	24	8	Outdoor
	RoadByTheSea	1024x528	100	24	8	Outdoor
	Theater	1024x528	100	24	8	Outdoor

EEV Model Development

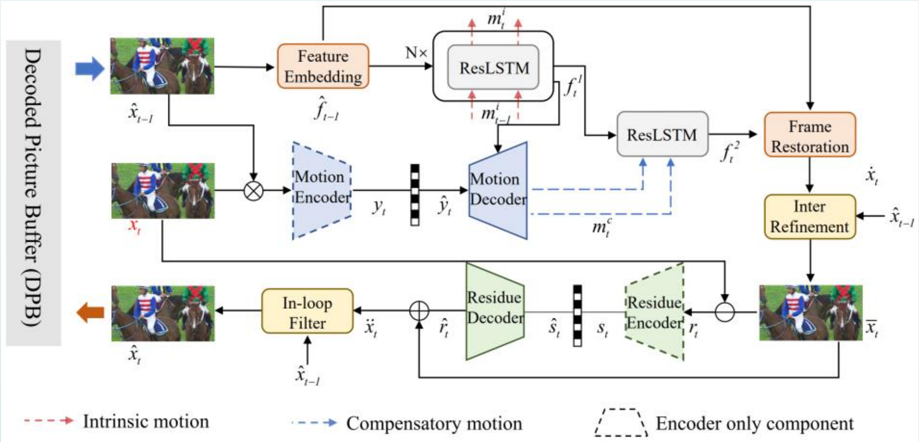
EEV-0.2



EEV-0.3



EEV-0.4



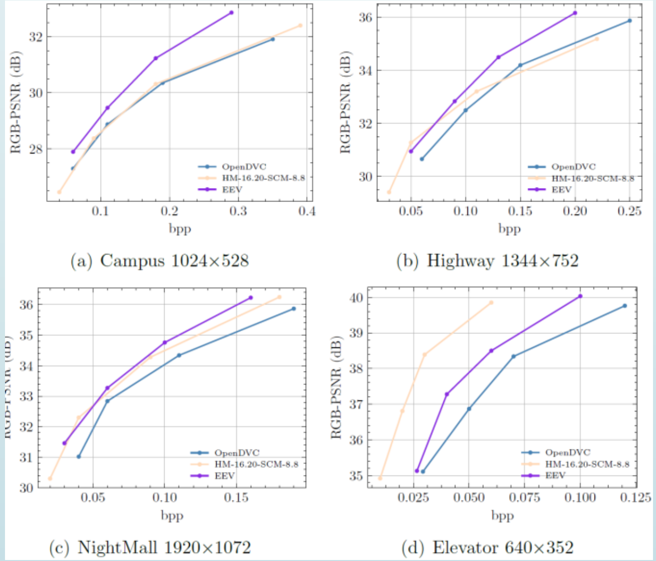
K. Lin, C. Jia*, X. Zhang, S. Ma and W. Gao, "DMVC: Decomposed Motion Modeling for Learned Video Compression." *IEEE Transactions on Circuits and Systems for Video Technology* (accepted).

MPAI. community

EEV-0.3 Performance

- Test conditions
  - TAppEncoder -c encoder LDP.cfg -InputBitDepth 8 -InputChromaFormat 444 -Level 6.2 -wdt seq wid -hgt seq hgt -f 100 -fr fps -q QP -IntraPeriod 16 -InputColourSpaceConvert RGBtoGBR -SNRInternalColourSpace 1 -OutputColourSpaceConvert GBRtoRGB
  - python test opendvc.py -path seqname -mode PSNR -IntraPeriod 16 –metric PSNR -l λ
  - python test eev.py -path seqname -mode PSNR -IntraPeriod 16 -metric PSNR -l λ

Category	Sequence Name	BD-Rate Reduction EEV vs OpenDVC	BD-Rate Reduction EEV vs HEVC
Class A VisDrone-SOT	BasketballGround	-23.84%	9.57%
	GrassLand	-16.42%	-38.64%
	Intersection	-18.62%	-28.52%
	NightMall	-21.94%	-6.51%
	SoccerGround	-21.61%	-10.76%
Class B VisDrone-MOT	Circle	-20.17%	-25.67%
	CrossBridge	-23.96%	26.66%
	Highway	-20.30%	-12.57%
Class C Corridor	Classroom	-8.39%	178.49%
	Elevator	-19.47%	109.54%
	Hall	-15.37%	58.66%
Class D UAVDT_S	Campus	-26.94%	-25.68%
	RoadByTheSea	-20.98%	-24.40%
	Theater	-19.79%	2.98%
Class A		**-20.49%**	**-14.97%**
Class B		**-21.48%**	**-3.86%**
Class C		**-14.41%**	**115.56%**
Class D		**-22.57%**	**-15.70%**
Average		**-19.84%**	**15.23%**



(a) Campus 1024×528
(b) Highway 1344×752
(c) NightMall 1920×1072
(d) Elevator 640×352

EEV-0.4 Performance

- MS-SSIM based Rate-distortion curves
  - VTM is using LDB config

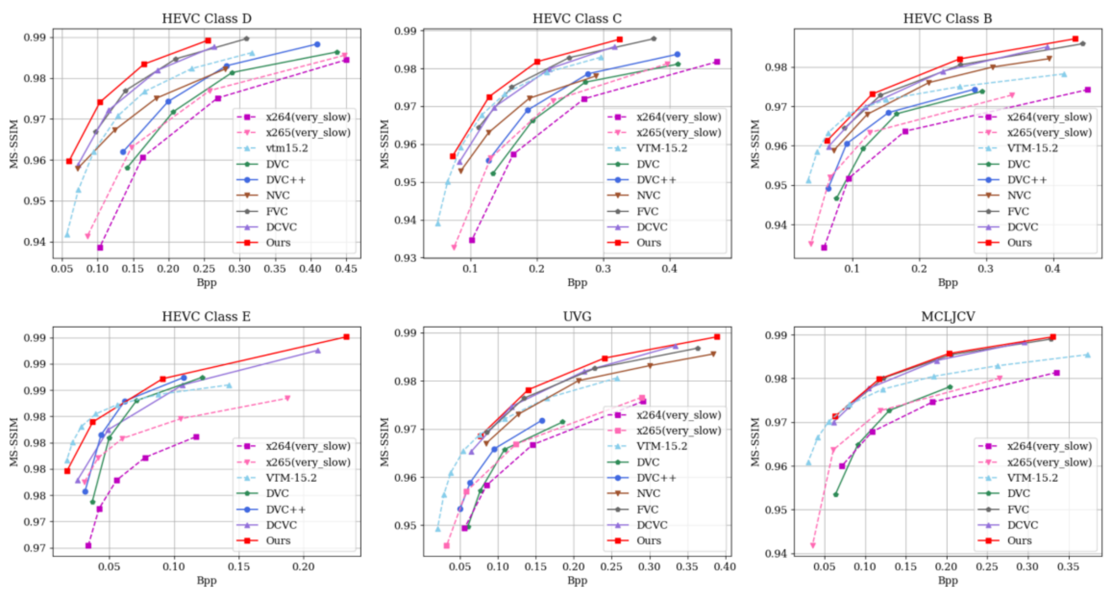


TABLE I: The coding performance of the proposed method, where x265 (*very_slow*) is used as anchor. The best and the second best neural video coding methods are respectively marked as red and blue.

	MS-SSIM (%)						PSNR (%)					
	ClassB	ClassC	ClassD	ClassE	UVG	MCLJCV	ClassB	ClassC	ClassD	ClassE	UVG	MCLJCV
x264	39.75	22.31	17.91	-	26.20	31.38	54.95	31.13	27.96	82.12	49.10	47.49
VTM-15.2 [14]	-49.53	-37.77	-15.26	-61.73	-47.83	-41.32	-58.73	-46.00	-42.32	-66.70	-61.86	-50.18
DVC [15]	13.68	7.67	1.27	6.37	17.29	31.29	14.60	39.42	29.95	4.59	8.45	15.18
DVC++ [27]	-12.52	-7.27	-12.70	-7.32	-	-	-10.86	10.46	4.00	-15.94	-17.80	-
NVC [28]	-33.01	-20.02	-12.24	-	-	-	-9.52	19.00	15.75	-	-	-
FVC [29]	-46.95	-38.39	-45.76	-	-49.12	-46.80	-15.22	-4.76	-8.26	-	-28.71	-21.08
DCVC [32]	-43.64	-35.24	-44.75	-17.88	-48.32	-43.79	-33.33	-7.27	-16.55	-21.75	-35.00	-23.08
Ours	-50.79	-44.95	-54.48	-48.26	-51.95	-47.18	-33.40	-12.05	-24.62	-35.75	-31.45	-17.00

Sequence Name	BD-Rate vs VVC (MS-SSIM)
BasketballGround	11.87%
GrassLand	3.71%
Intersection	-14.25%
NightMall	-12.16%
SoccerGround	-7.00%
Circle	-8.92%
CrossBridge	-5.57%
Highway	-13.56%
Classroom	-22.53%
Elevator	-34.83%
Hall	-29.98%
Campus	6.71%
RoadByTheSea	4.64%
Theater	34.22%
Average	**-6.26%**

EEV-0.4 Performance

- Visual quality

Fig. 10: The subjective quality comparison of the proposed method with VVC. The quantization parameter (QP) for VVC is set as 35. To align the consumed bits, λ equals to 8 for the proposed method.

Discussion

- For future use case
  - Support compress domain analysis without re-training
  - Hardware encoding/decoding
- Operation/Complexity
  - EEV-0.3: 7.7T, 50M Params
  - EEV-0.4: 5.4T FLOPs, 23M Params

Thank you!

https://eev.mpai.community

MPAI.
community