北京大学
PEKING UNIVERSITY

MPAI End-to-end (EEV) Project

Chuanmin Jia

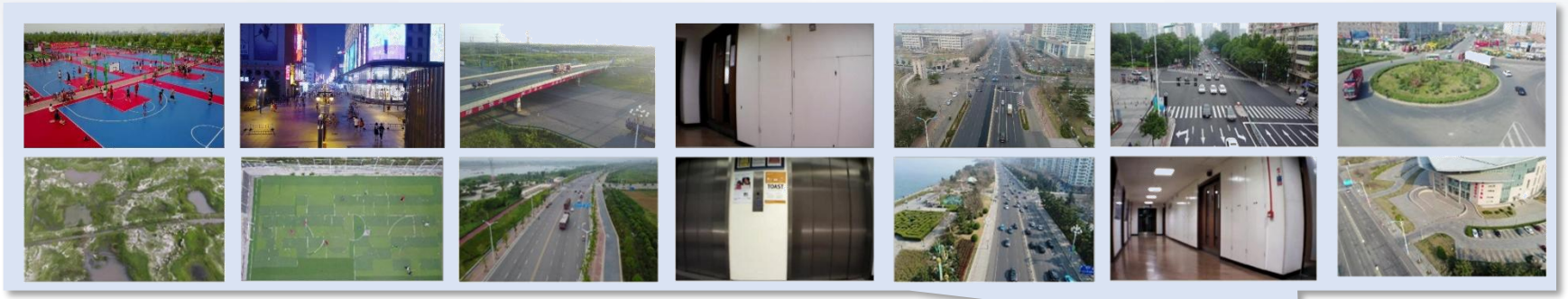
Peking University/Pengcheng Lab

2026.07.15

Outline

MPAI-EEV

- **MPAI-EEV: The first video coding standard for fully neural network solutions**



Switching to Random Access Configuration: surpasses VTM

2022.01 EEV established

2022.6 Release a work draft

2022.9 EEV-0.1

2022.10 EEV-0.2

2022.12 EEV-0.3

2023.2 Drone video benchmark (DCC 23)

2024.7 EEV-0.4 (TCSVT 24)

2025.8 EEV-0.5 (MM 24)

2026.2 Overview release (IJCAI 25)

EEV-0.6 (CVPR 26)

Goal: Build 1st E2E video coding standard

Low-delay P configuration: surpasses VTM

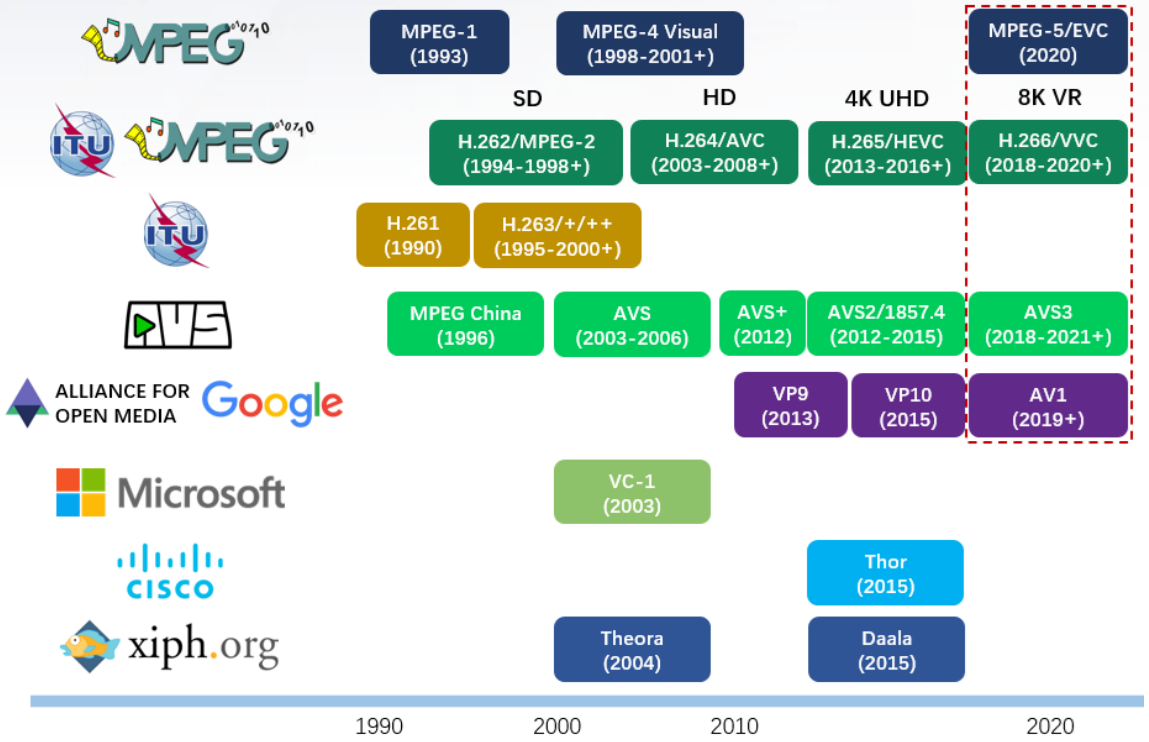
01

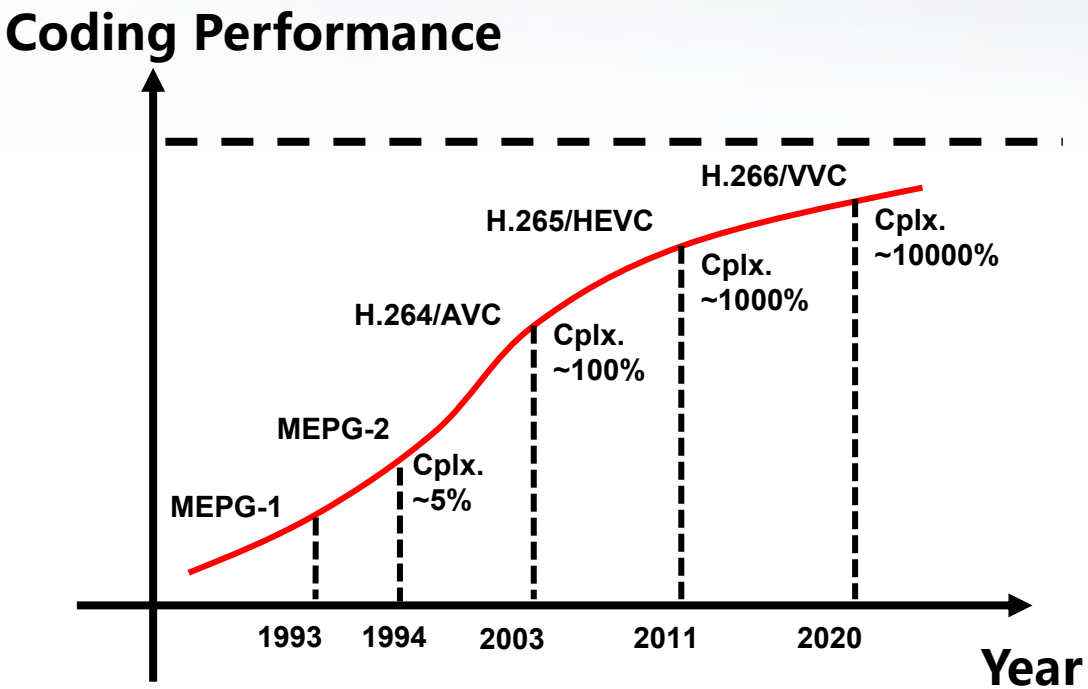
Background & Motivation

Why end-to-end neural video coding? Why MPAI-EEV?

Background

- **Traditional coding (H.26x, AVS, AOM) faces bottlenecks:**





Paradigm shift: hand-crafted traditional coding → end-to-end neural video coding

Motivation

- **Established Time**
  - **Early 2022**
- **Goal**
  - **Build a leading end-to-end video coding standard**
- **Key Characteristics**
  - **Fully neural network solutions — unconstrained by IP legacy**
  - **End-to-end optimization for bitrate and quality metrics**
  - **Content-adaptive learning — adapted to specific scenarios like drone videos**
  - **Easy upgrade — continuous software updates**

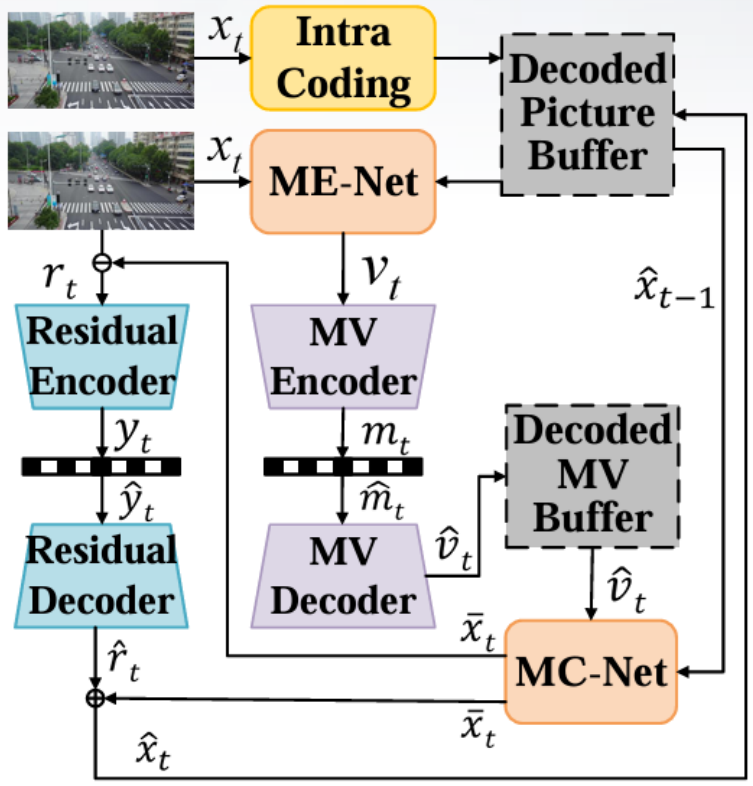
02

EEV Verification Model Evolution

EEV-0.1 to EEV-0.4: Building the foundation

EEV-0.1: The Baseline (OpenDVC)

- **EEV-0.1 adopted OpenDVC as the starting point**
  - **A publicly available implementation of the DVC framework.**
- **Key Components:**
  - **ME-Net: Motion estimation network**
  - **MC-Net: Motion compensation network**
  - **Residual Coding: Encode prediction residual**
  - **Training: Vimeo-90K dataset**



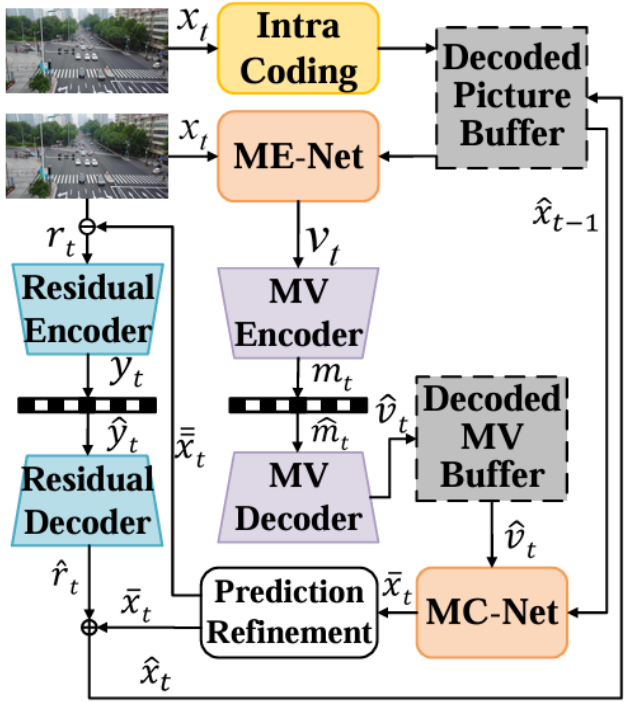
EEV-0.1

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

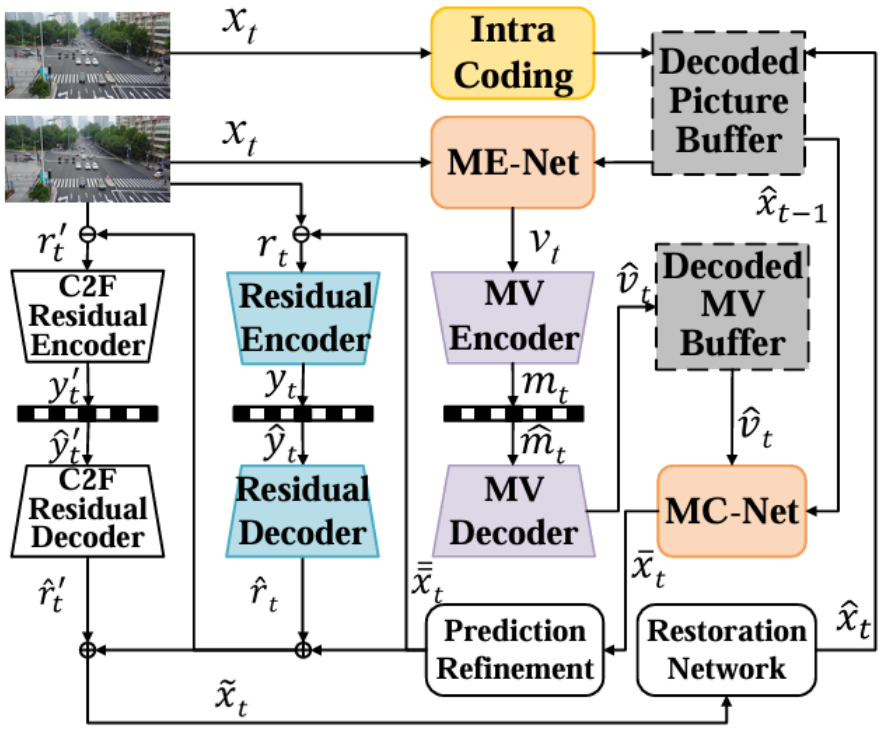
EEV-0.2 & EEV-0.3: Enhanced Tools

- **EEV-0.2 : Jun. 2022**
  - **Enhanced Motion Compensation Prediction (MCP) using Prediction Refinement**
- **EEV-0.3: Sep. 2022**
  - **Coarse-to-Fine (C2F) Residual Modeling & In-Loop Restoration (ILR) Network**

EEV-0.2:

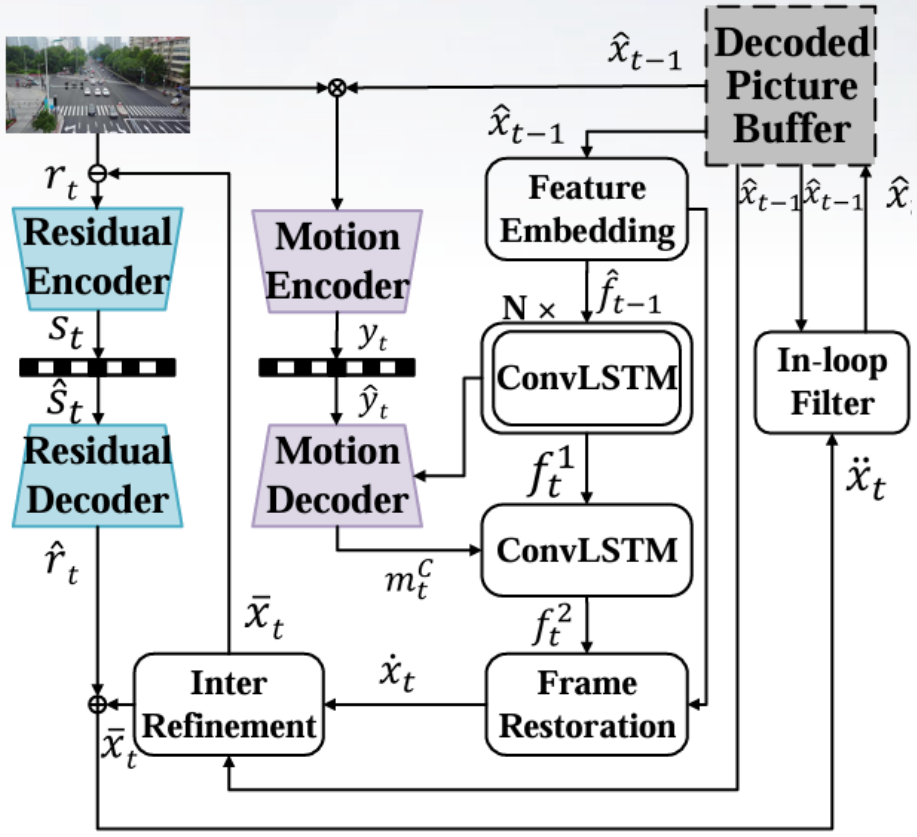


EEV-0.3:



EEV-0.4: Motion Decoupling

- **EEV-0.4 introduces advanced motion representation via motion decoupling**
- **Two-stage Motion Representation:**
  - **MV Prediction : from temporal context via ConvLSTM, coarse motion**
  - **MV Difference : fine motion correction**
- **Additional Modules:**
  - **Feature embedding & restoration for compact latent representation**
  - **Spatio-temporal inter-prediction enhancement**



EEV-0.4

Chuanmin Jia, Feng Ye, Fanke Dong, Kai Lin, Leonardo Chiariglione, Siwei Ma, Huifang Sun, Wen Gao: MPAI-EEV: Standardization Efforts of Artificial Intelligence Based End-to-End Video Coding. IEEE Trans. Circuits Syst. Video Technol. 34(5): 3096-3110, 2024

R-D Performance

- **Each generation of EEV has a significant performance boost, and EEV-0.4 performs on par with H.266/VVC-LDP (RGB-PSNR).**

Metric / Class	EEV-0.4	EEV-0.3	H.266/VVC	EEV-0.1
JCT-VC Class A	**-43.04**	-12.12	-39.89	10.84
JCT-VC Class B	-34.38	9.63	**-40.64**	39.48
JCT-VC Class C	-34.13	6.47	**-38.26**	38.14
JCT-VC Class D	**-49.29**	-24.38	-36.10	-1.77
JCT-VC Class E	-37.54	26.45	**-49.05**	62.97
Average	**-39.68**	**1.21**	**-40.79**	**29.93**

Chuanmin Jia, Feng Ye, Fanke Dong, Kai Lin, Leonardo Chiariglione, Siwei Ma, Huifang Sun, Wen Gao: MPAI-EEV: Standardization Efforts of Artificial Intelligence Based End-to-End Video Coding. IEEE Trans. Circuits Syst. Video Technol. 34(5): 3096-3110, 2024

03

Specialized Applications

Drone Video Coding Benchmark

DCC 2023

UAV Video Coding Benchmark

- **First comprehensive benchmark for learned UAV video coding, accomplished as a milestone of the MPAI-EEV project.**

- **Dataset Characteristics**
  - **14 video clips from 4 sources:**
  - **Resolutions: 640×352 to 2720×1520**
  - **Frame rate: 24-30 fps, 100 frames each**

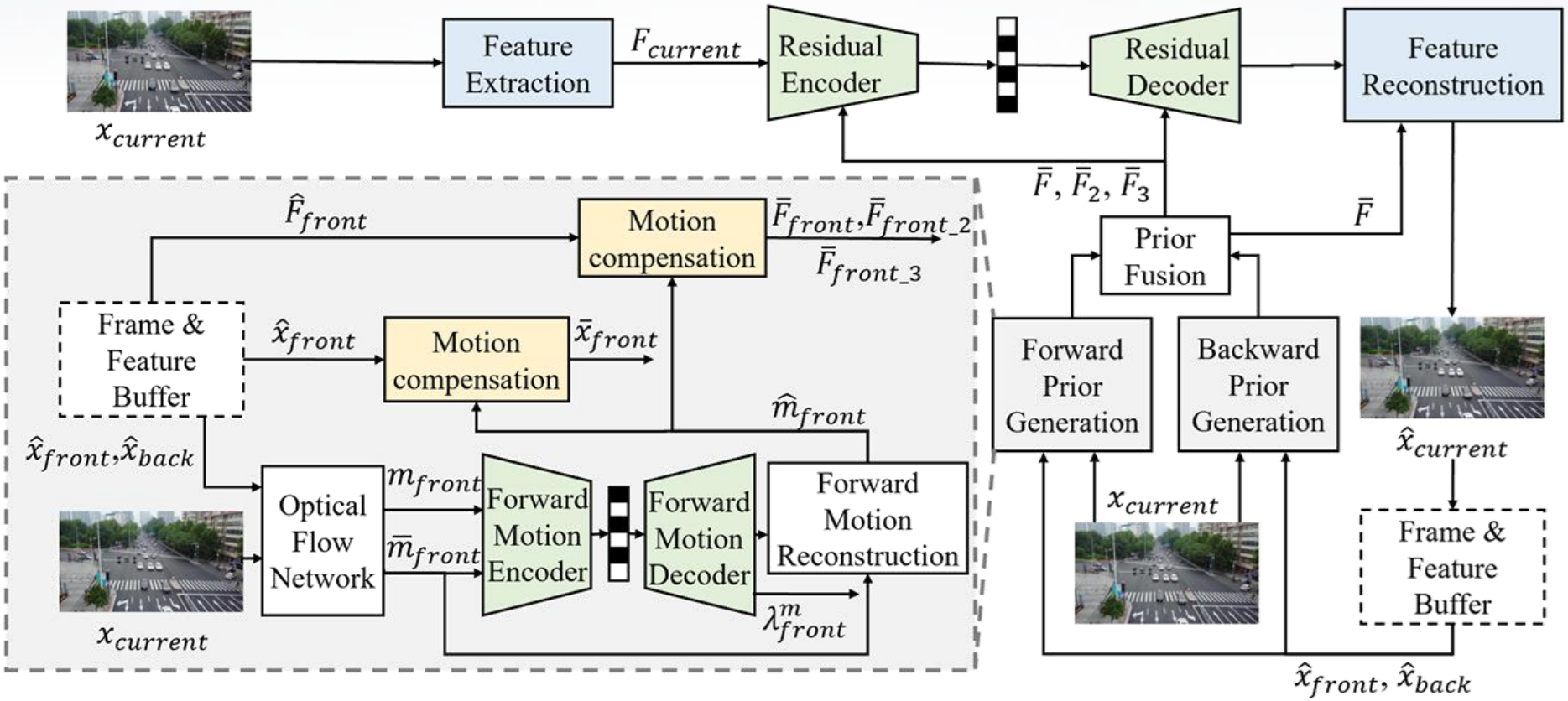
04

EEV-0.5 & 0.6: Bi-directional Breakthrough

High Resolution Scaled Hierarchical Bi-directional Motion Model

EEV-0.5

- Since EEV-0.5, research focus is shifted from P-frame to B-frame based coding
- Propose a scaled hierarchical bi-directional coding structure, using **Parameter-Shared Motion Codecs**, **Scaled Motion Priors**, and **Trustworthy Motion Modeling**



EEV-0.5

- **Revisited the emerging advances in learned video compression in EEV**
- **Point out EEV needs to explore high-resolution, random access configuration coding**

1. Architectures

Uni-directional & Bi-directional prediction models

2. Optimization

Online learning, content adaptation, variable bitrate encoding, rate control, perceptual quality optimization, model quantization

3. System Design

Edge computing, mobile, semantic communication

4. Standardization

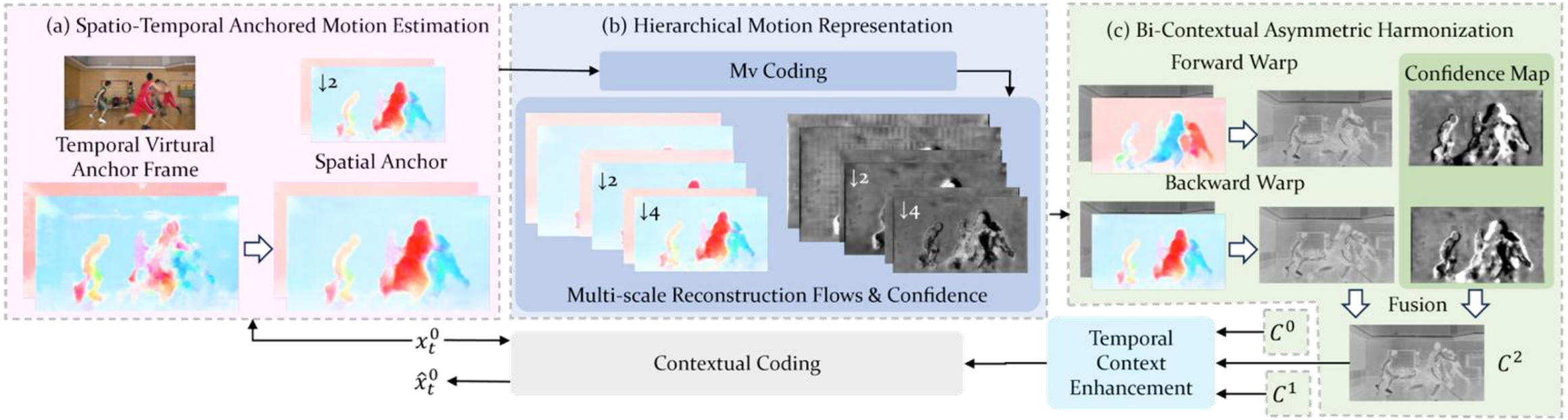
MPAI-EEV, JVET NNVC, AVS EEM, IEEE 1857.11

5. Experiments

Comprehensive performance evaluation

EEV-0.6

- Takes a systematic approach from the perspective of reference information modeling, optimizing the representation, translation, and harmonization of reference information.
- Proposes a hierarchical B-frame coding framework for high resolution videos.

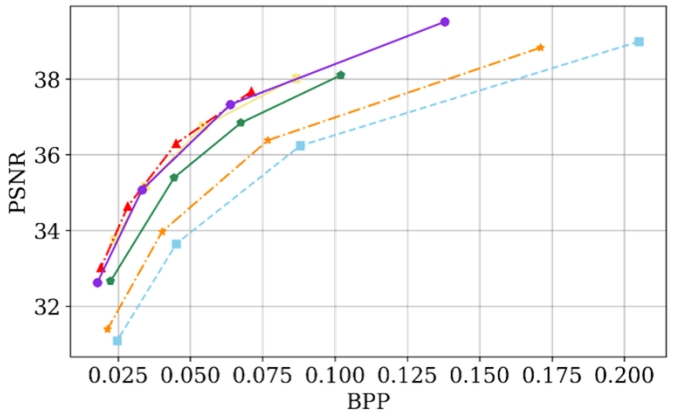


EEV-0.6: R-D Performance

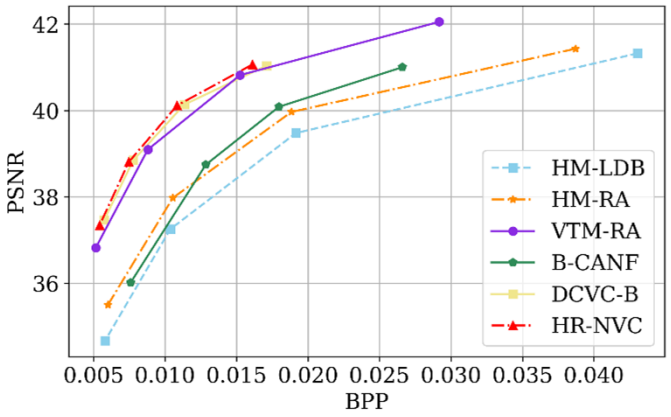
- **EEV-0.6 is the first end-to-end video coding method tested on 4K videos and achieves good compression performance. *97 frames, RGB-PSNR**

	JVET Class A1	JVET Class A2	Average
HM-16.20-LDB	0.00	0.00	0.00
HM-16.20-RA	-12.44	-17.49	-14.97
DCVC-B	-22.91	-32.16	-27.53
HR-NVC	**-29.18**	**-34.49**	**-31.84**

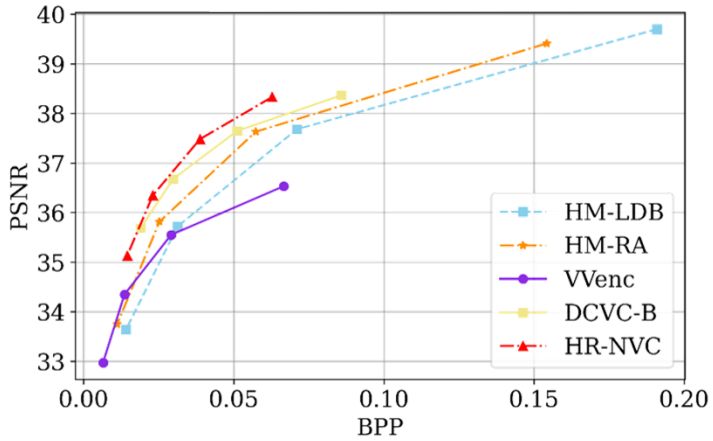
JCT-VC Class A



JCT-VC Class E



JVET Class A2



Feng Ye, Kai Zhang, Li Zhang, Chuanmin Jia: High Resolution Neural Video Coding with Bi-directional Confidence-Guided Reference Information Modeling. CVPR 2026. (highlight)

EEV-0.7: Future Directions

- **Semantic Confidence Modeling:**
  - **Extend confidence estimation from pixel-level motion cues to object-level scene understanding for better spatio-temporal consistency**
- **Advanced Rate-distortion Optimization:**
  - **Broaden adaptive fusion to GOP-level to achieve optimized bit allocation**
- **Real-Time Edge Deployment:**
  - **Prune EEV-0.6 on demand to satisfy computational constraints of edge devices.**
- **Task-Aware Coding:**
  - **Incorporate downstream vision semantics into confidence modeling and bit allocation, enabling joint compression-analysis optimization**

MPAI-EEV

Key Achievements

Better Than VVC/H.266

Both Low-delay P and RA configs

Drone Video Benchmark

First learned UAV coding benchmark

Top Publications

TCSVT, DCC, MM, IJCAI, CVPR

Thank You !