MMVC: Learned Multi-Mode Video Compression with Block-based Prediction Mode Selection and Density-Adaptive Entropy Coding

Bowen Liu, Yu Chen*, Rakesh Chowdary Machineni*, Shiyu Liu, Hun-Seok Kim

University of Michigan, Ann Arbor

Paper Tag: THU-AM-192

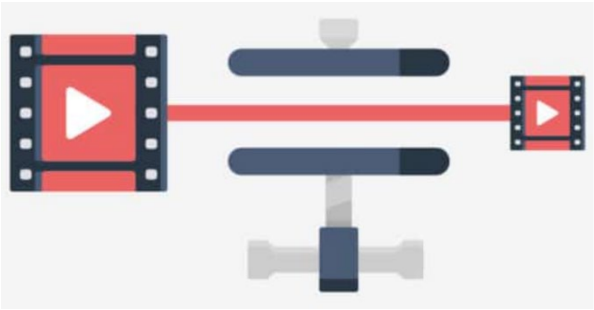
Preview

- **Overview: learning-based multi-mode video compression**
  - Proposed a multi-mode feature domain prediction framework, which adapts to different motion and contextual patterns in the unit of block
  - Applied a block wise channel removal scheme to improve residual sparsity under similar quality
  - Implemented a density adaptive entropy coding scheme to encode dense and sparse residual blocks separately
- **Performance**
  - Superior or comparable to existing approaches on benchmarking datasets
- **Ablation study**
  - Multi-mode prediction performance and the utilizaton of different modes for various scenarios

Video Compression: Intro and Goals

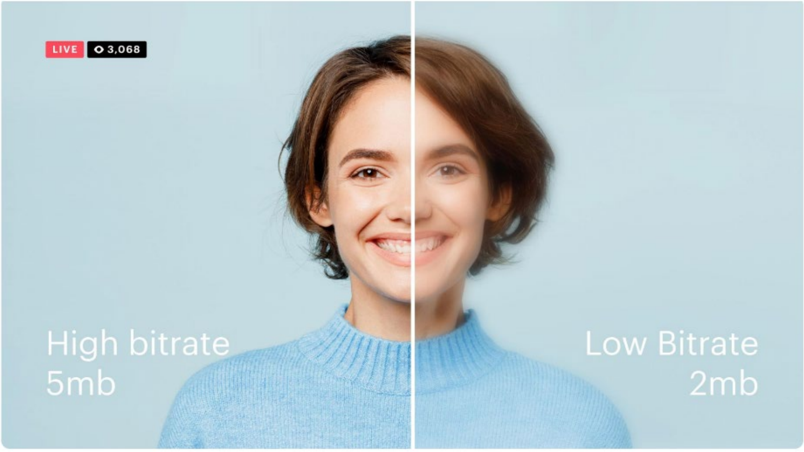
- **Why do we need video compression?**
  - To reduce storage requirements
  - To allow faster transmission over netwroks
  - To improve the streaming user experience

[Credit: cordis.europa.eu]



[Credit: Resi.io]

- **Goal of video compression**
  - Minimize the file size while maintaining acceptable video quality
  - Balance the trade-off between video quality and bitrate



[Credit: Restream.io]

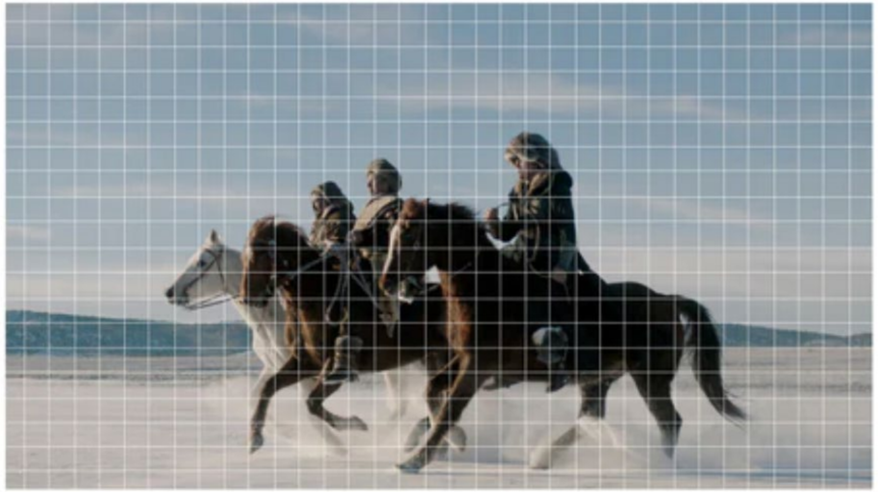
Video Compression: General Methodologies

- **How to compress the video?**
  - Remove the redundancy in the video sequence
    - Spatial Redundancy: Removing redundant information within a single frame.
    - Temporal Redundancy: Exploiting similarities between consecutive frames.
- **Video Prediction**
  - In video codecs, prediction helps to temporally decorrelate the stored information, so as to reduce the code size under a certain quality
- **Video Compression methods**
  - Standard codec
    - AVC, HEVC, VVC, VP9, etc.
  - Learning-based video compression algorithm
    - DVC, DCVC, RLVC, VCT, etc.

[Credit: S. Oprea, 2020]

Block-wise Compression

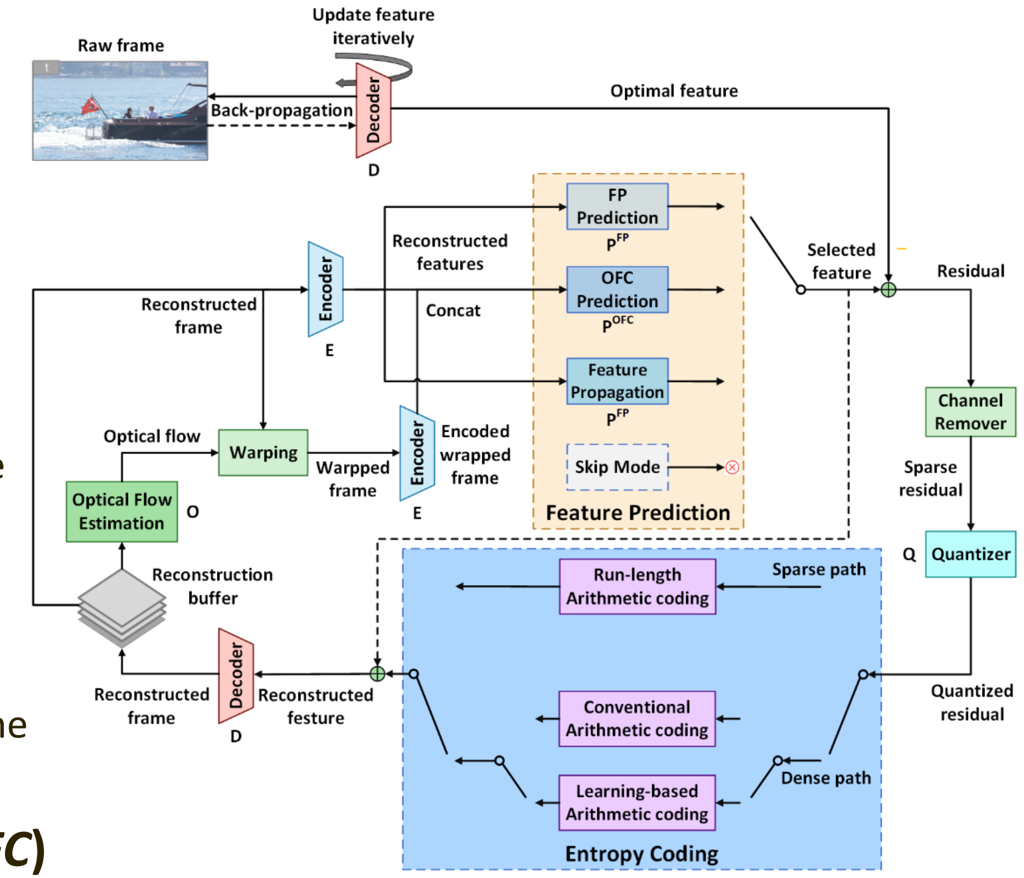
- **Motivation**
  - Inspired by conventional video codec
    - Typically address various types of motions in the unit of macroblocks
  - Presented a learning-based, block wise video compression scheme
    - Applied content-driven mode selection on the fly



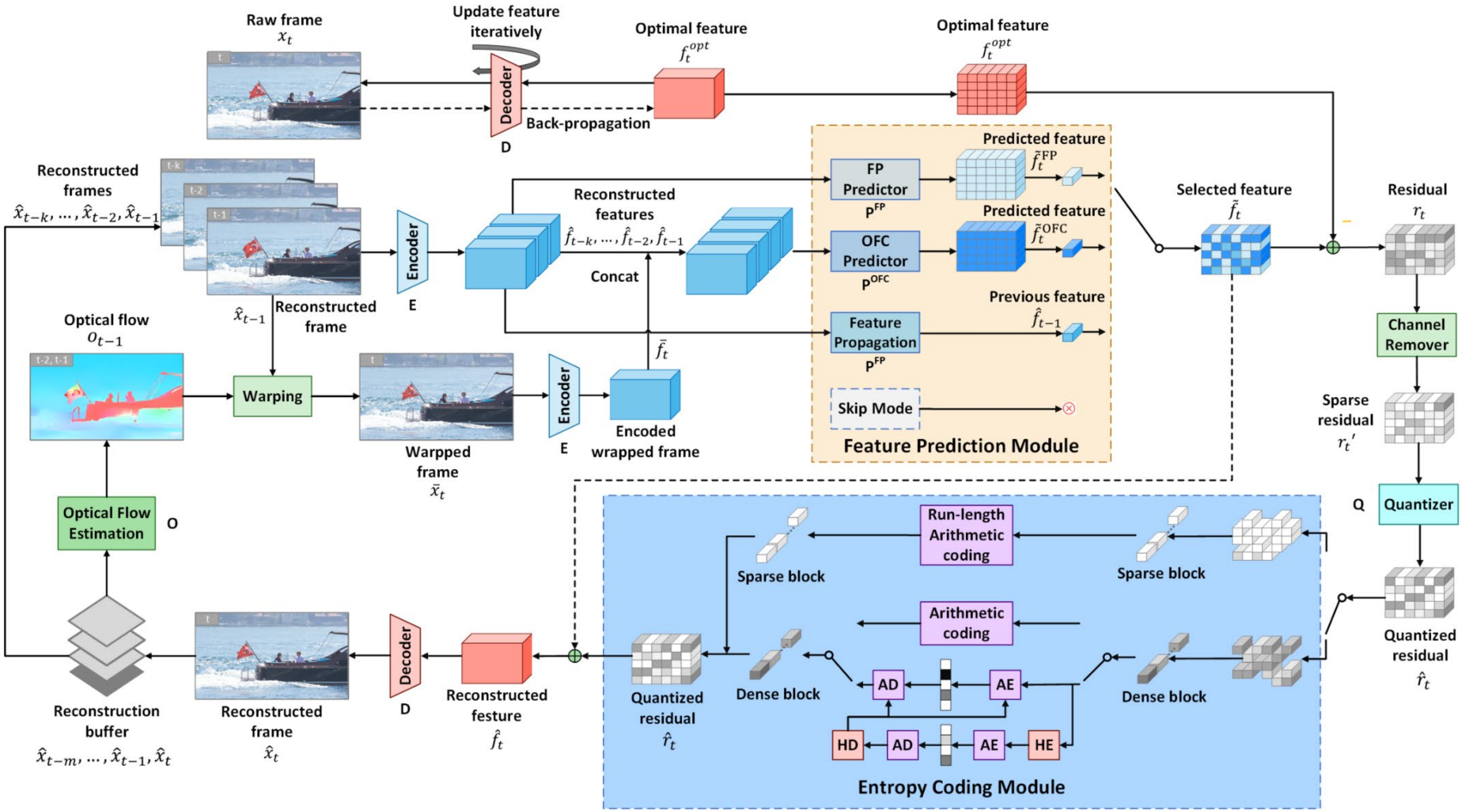
[Credit: Teradek.com]

Multi-Modes for Video Prediction

- **Our proposed modes**
  - **Skip mode (*S*)**
    - This mode aims to find the most condensed representation to transmit unchanged blocks
    - *S* mode is particularly useful for static scenes where same backgrounds are captured by a fixed view camera
  - **Feature Prediction mode (*FP*)**
    - This mode is designed for the generic cases in the video sequence
  - **Feature Propagation mode (*FPG*)**
    - This mode is applied to some changed blocks, where no better prediction mode is available
    - *FPG* mode copies the previously reconstructed feature block as the predicted result
  - **Optical Flow Conditioned Feature Prediction mode (*OFC*)**
    - *OFC* mode leverages the temporal locality of motions
    - The frame warped with optical flow information is treated as a preliminary prediction of the current frame, providing guidance to this prediction model

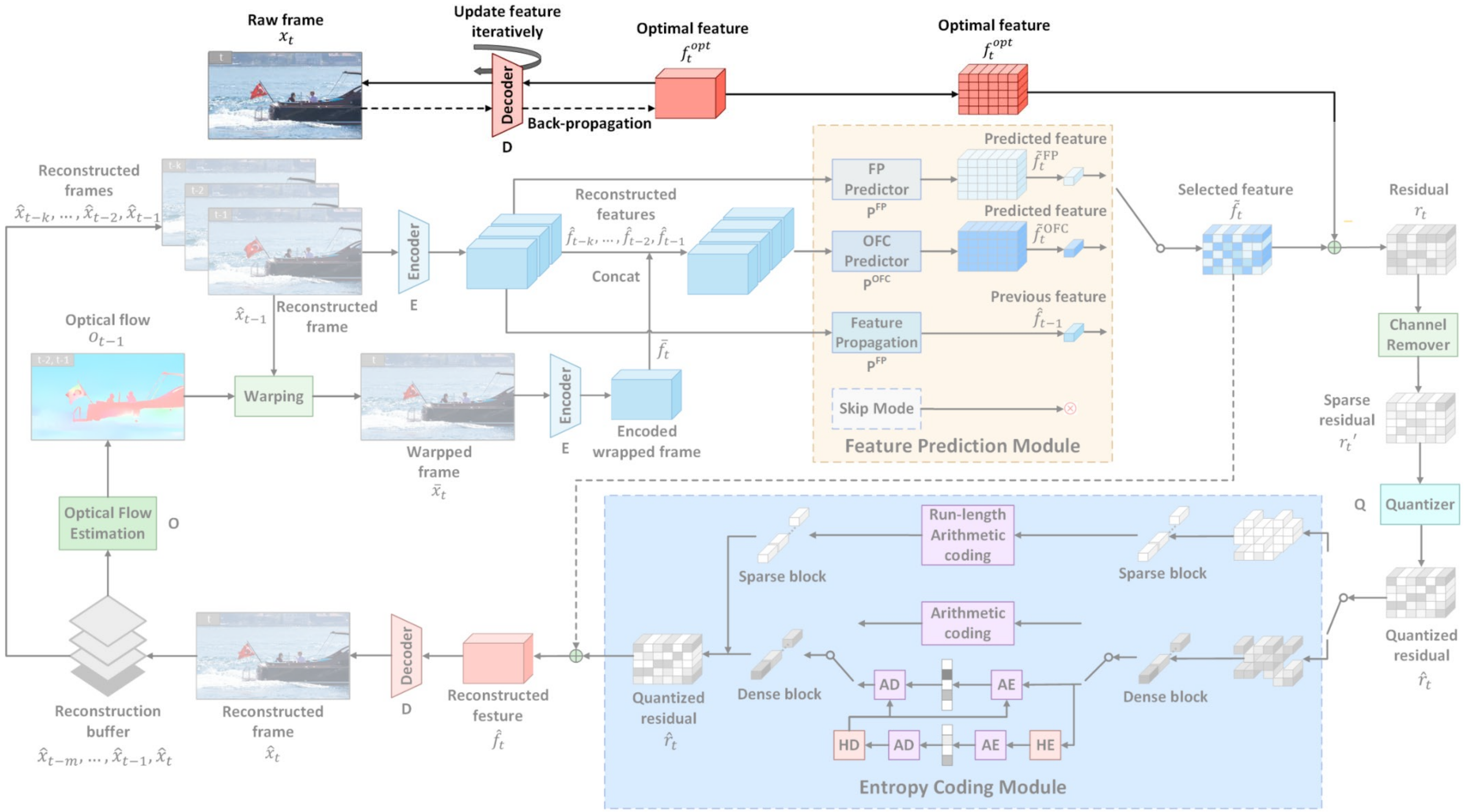


Video Coding Framework



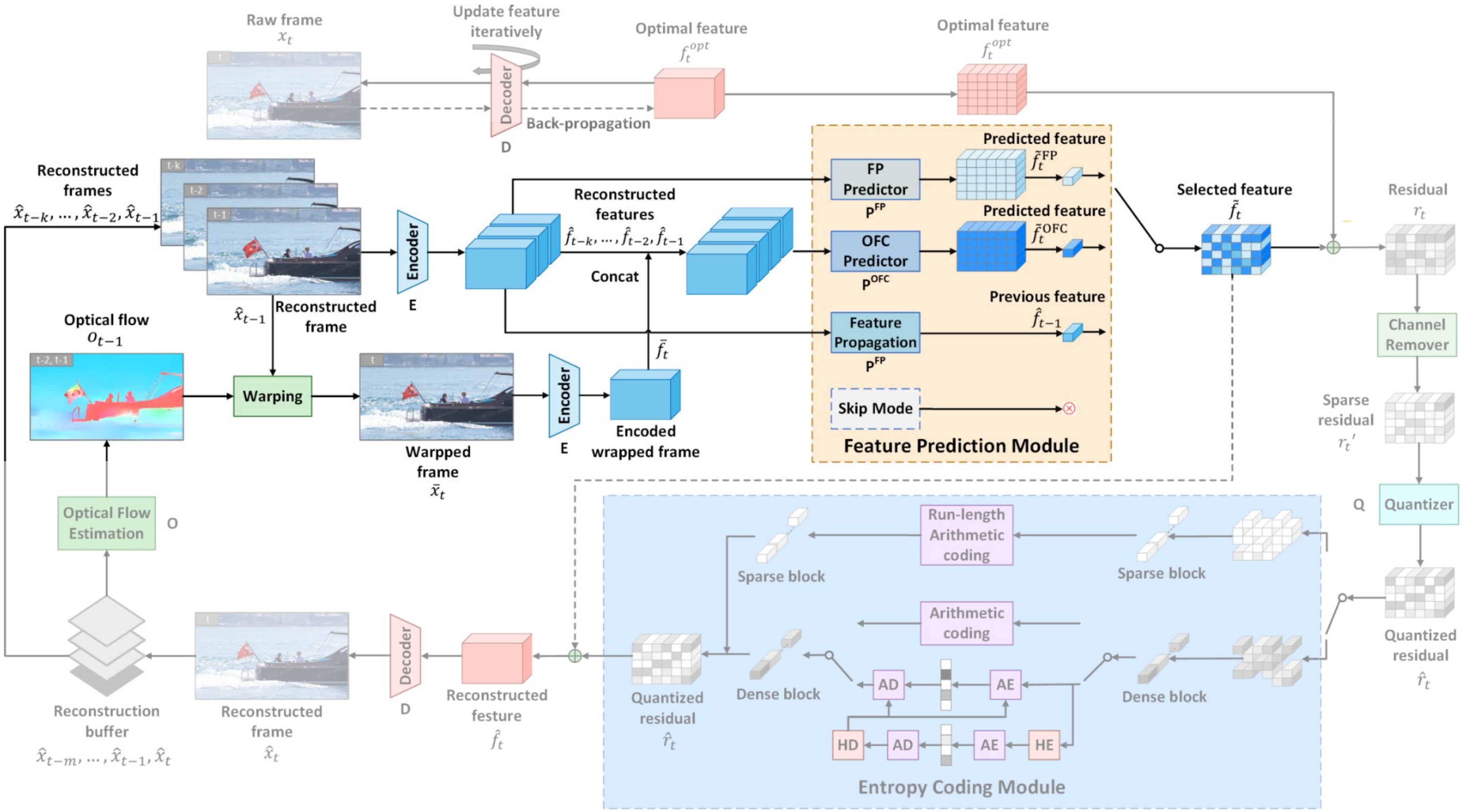
- Video coding pipeline

Step 1. Acquisition of Feature Representation



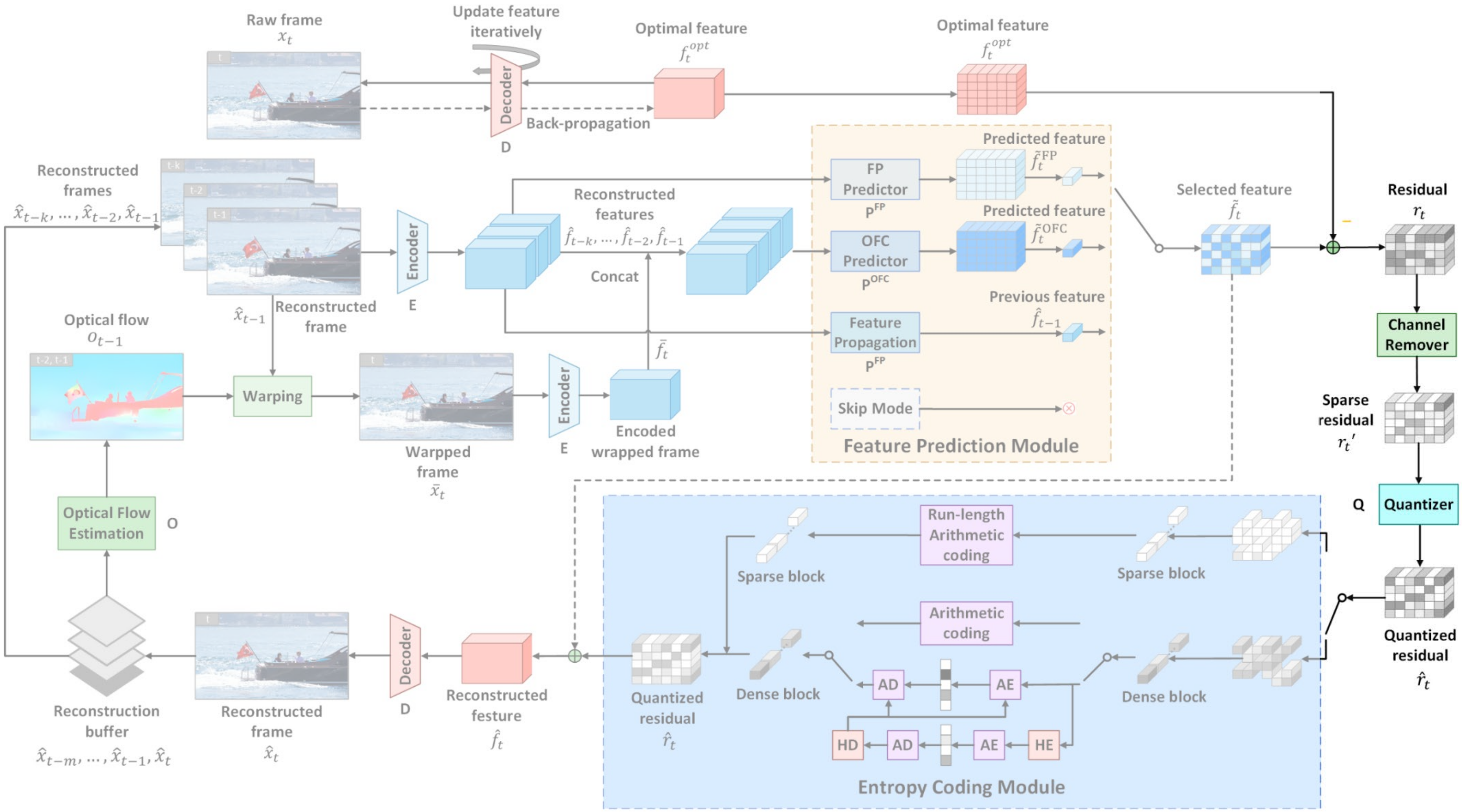
- Each frame of the video sequence is first individually compressed to an optimal feature representation

Step 2. Feature Domain Video Prediction



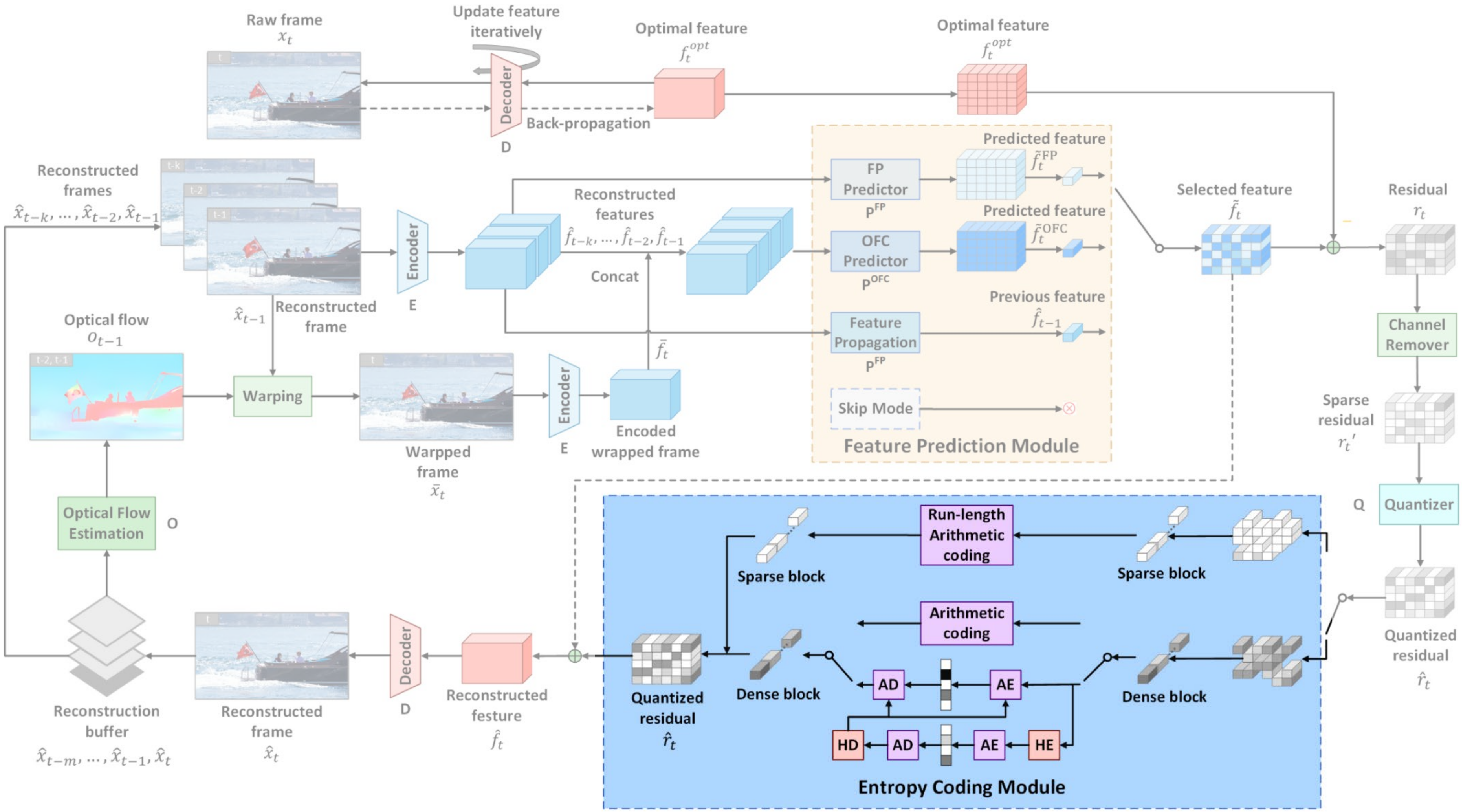
- The proposed multi-mode prediction framework selects the best prediction for each block with the smallest entropy coded residual

Step 3. Channel Removal



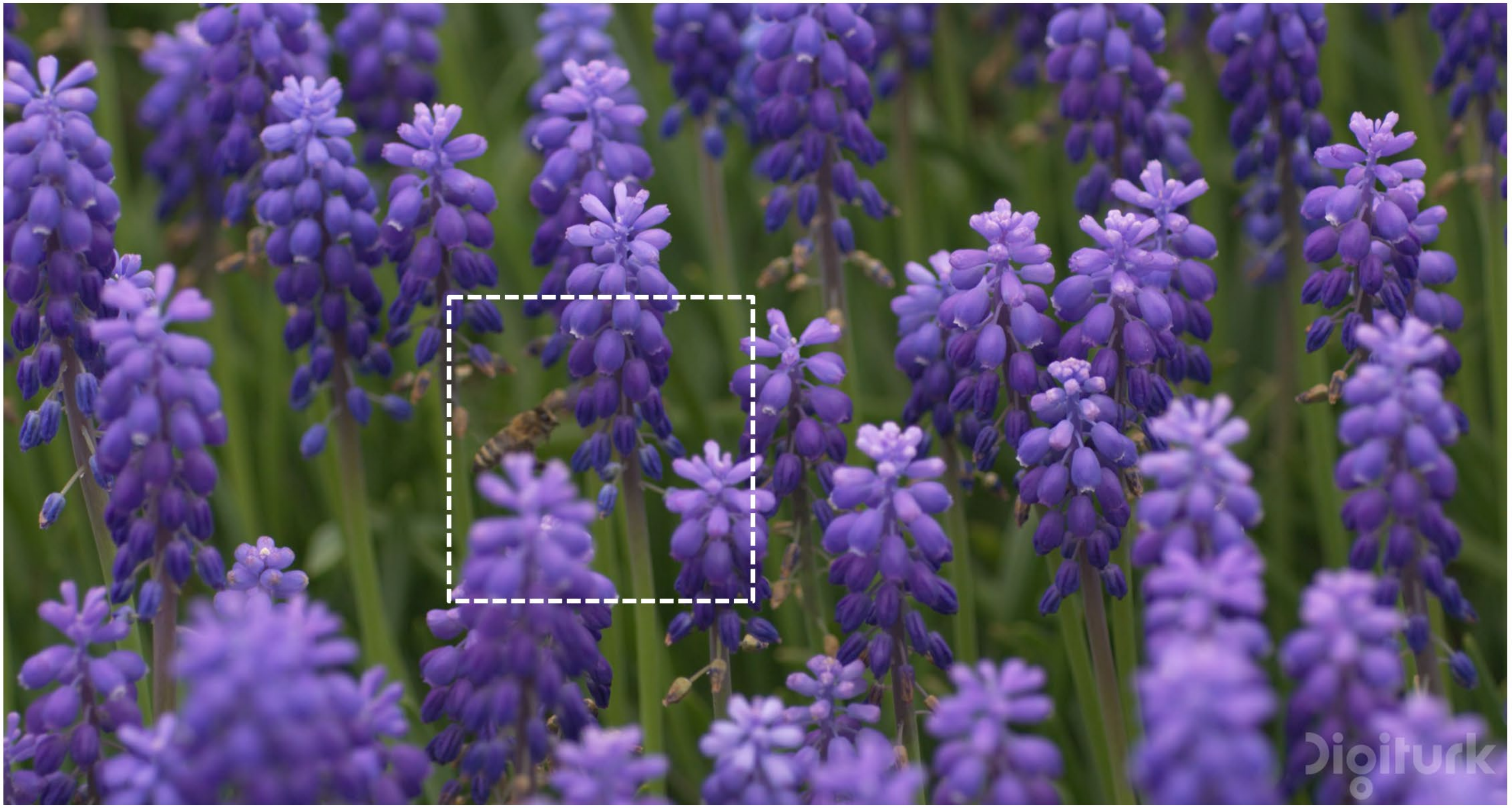
- The adaptive residual channel removal guarantees the reconstruction quality while reducing the number of bits consumed by unimportant residual feature channels

Step 4. Density Adaptive Entropy Coding

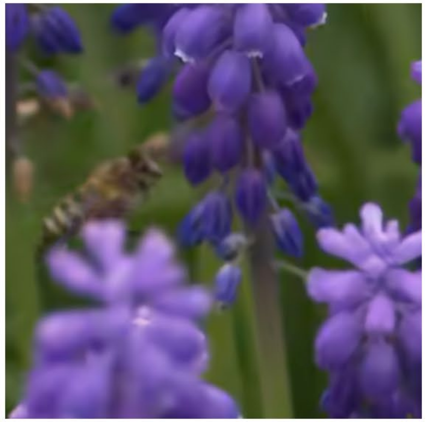


- To further reduce the bitrate, the adaptive entropy coding encodes the dense residual blocks and sparse residual blocks separately with different coding schemes

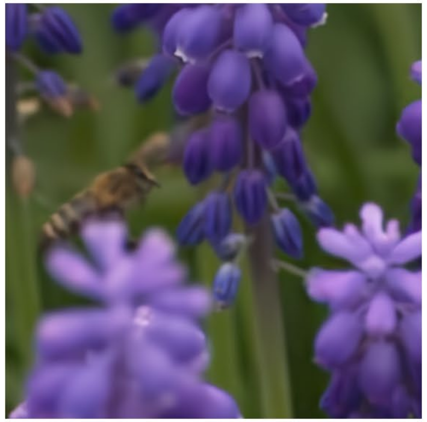
Visualization



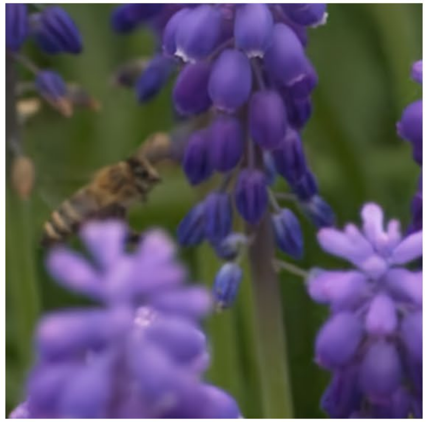
Ours 0.06bpp



HEVC 0.08bpp



Ours Prediction

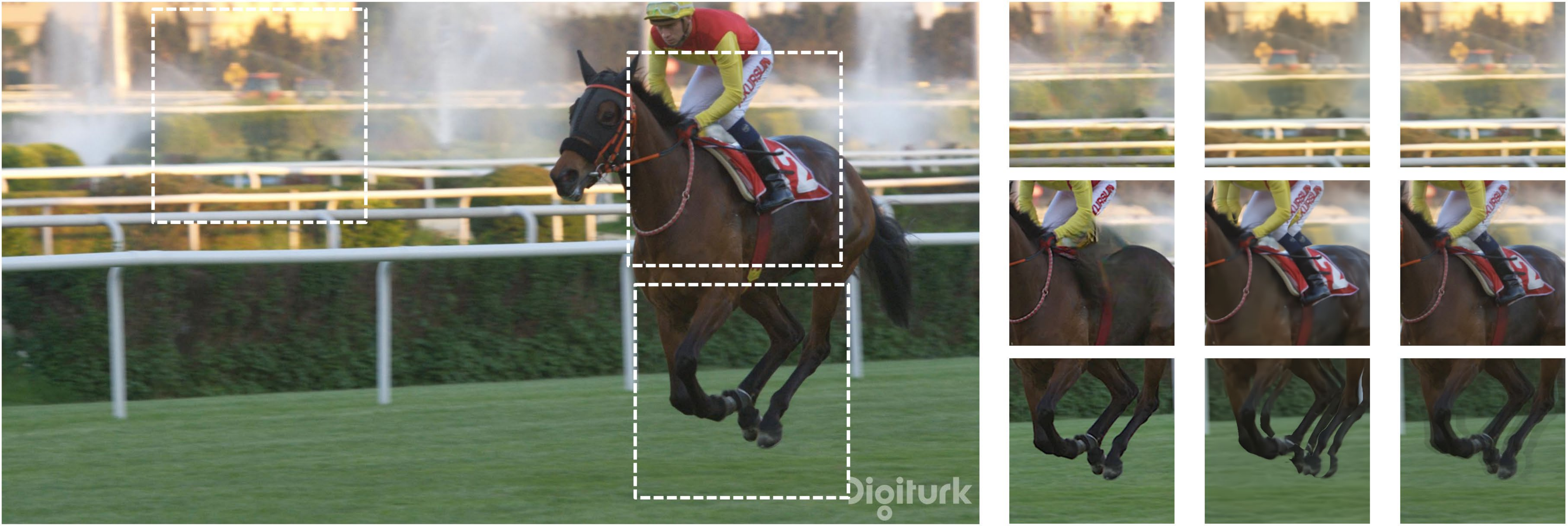


VVC 0.06bpp

Raw

- Details of the static background and dynamic objects are well preserved
- Compared with HEVC, our result yields less block artifacts preserving finer details

Visualization of Multiple Predictions



Raw FP OFC FP + OFC

- The decoded scenes are obtained from the predicted features without residual
- By adopting multiple prediction modes that complement each other, our prediction is able to cover content variety in the original frame with a shorter bitstream

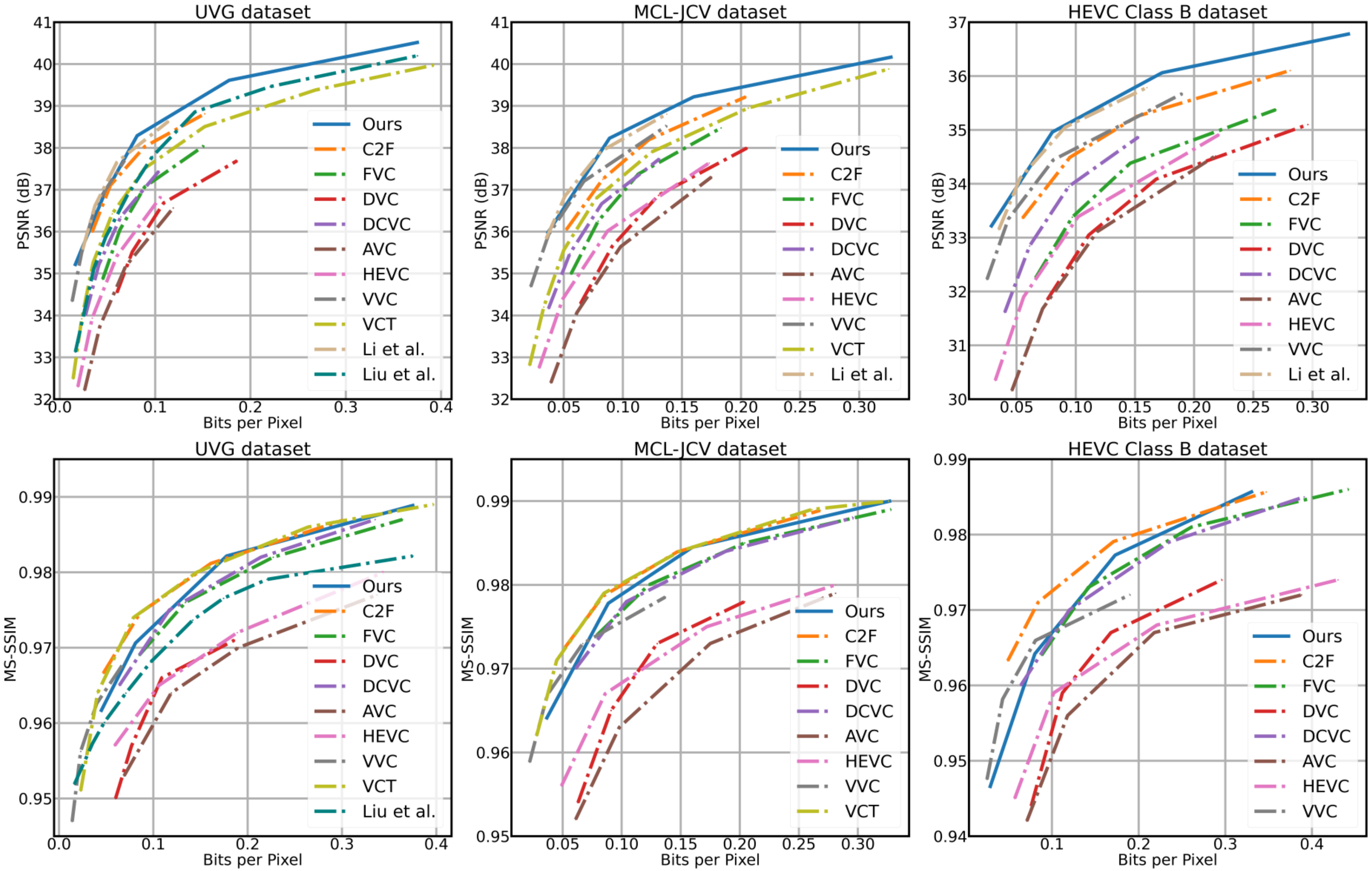
Ablation Study: Mode Utilization

Dataset	UVG				Kinetics			
Prediction mode	PSNR (dB)	Removed channels	Bpp	Bitrate saving	PSNR (dB)	Removed channels	Bpp	Bitrate saving
FP	38.0	23%	0.146	0%	37.7	29.8%	0.136	0%
OFC	36.9	47%	0.118	19.2%	37.4	49.3%	0.106	22.1%
FP+OFC	38.1	27%	0.096	34.3%	37.7	41.6%	0.099	27.2%
FP+OFC+FPG	38.2	27%	0.084	42.5%	37.7	43.5%	0.096	29.4%
FP+OFC+FPG+S	38.2	44%	0.081	44.5%	37.8	50.8%	0.088	35.3%
Mode utilization	FP	OFC	FPG	S	FP	OFC	FPG	S
	78.1%	10.6%	6%	5.3%	37.6%	38.3%	12%	11.6%

- By adopting the ensemble of both modes (FP + OFC), the quality is preserved with a lower bitrate, indicating that two prediction modes can complement each other by capturing different motion
- Including the Feature Propagation mode as an alternative prediction path further reduces the bitrate without degrading the quality
- The compression ratio improves noticeably by introducing the Skip mode as the final additional mode
- The results collected from the Kinetics dataset, where the scenes are captured mostly by a fixed-view camera, showcased higher utilization of skip mode brings a significant bitrate reduction

Evaluation: Performance Comparison

- Achieves state-of-the-art performance on benchmark datasets



Thank you!