2021 IEEE International Conference on Image Processing

Deep Video Compression for Interframe Coding

David Alexandre, Hsueh-Ming Hang, Wen-Hsiao Peng, Marek Domański

National Yang Ming Chiao Tung University, Taiwan
Poznań University of Technology, Poland

國立陽明交通大學
NATIONAL YANG MING CHIAO TUNG UNIVERSITY

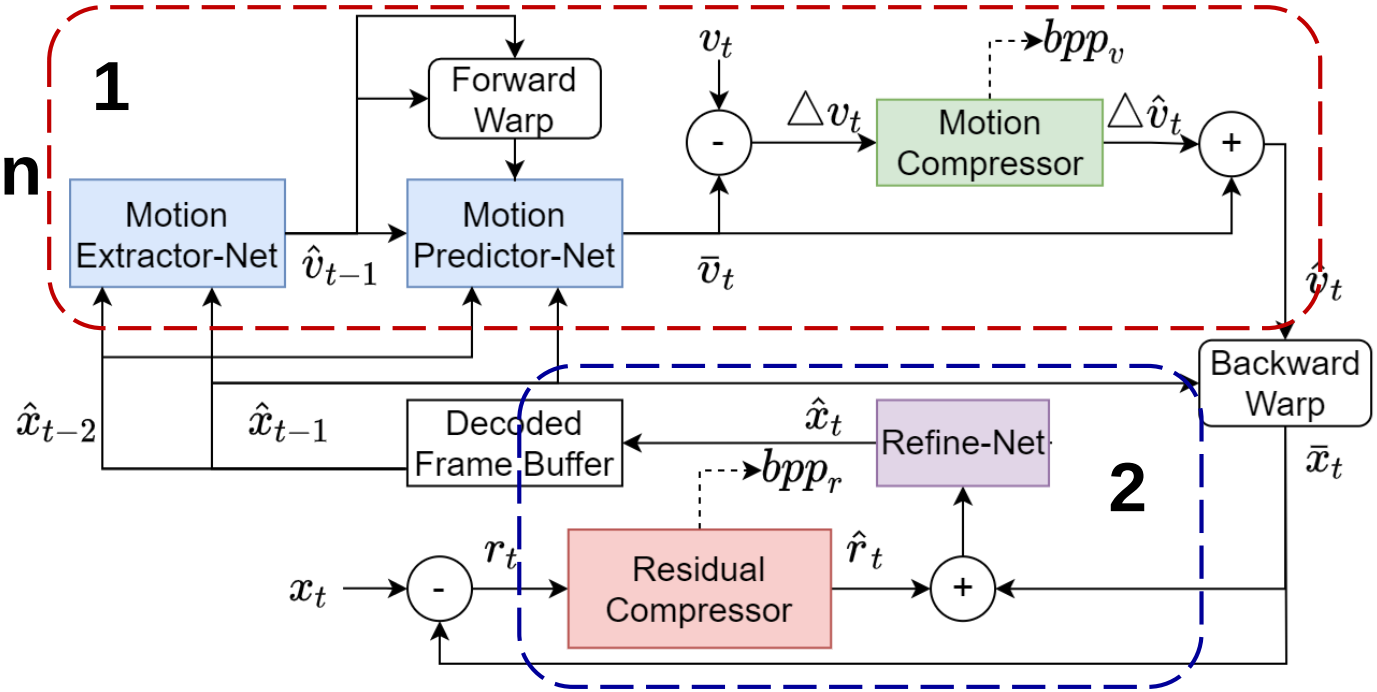
POZNAN UNIVERSITY OF TECHNOLOGY

Contributions

1. We design an **inter-frame video coding** scheme that includes I- and P-frames within a GOP.
2. We propose a **local motion predictor,** which predicts the current-frame motion vectors using the previous two coded frames. It enables our scheme to transmit only the differential motion vectors (or optical flow) to the decoder to save bits.
3. Our **refine-net** is paired with **the residual codec** to improve the reconstructed image quality.

Proposed Method

1. Motion Information Coding



2. Image Residual Coding

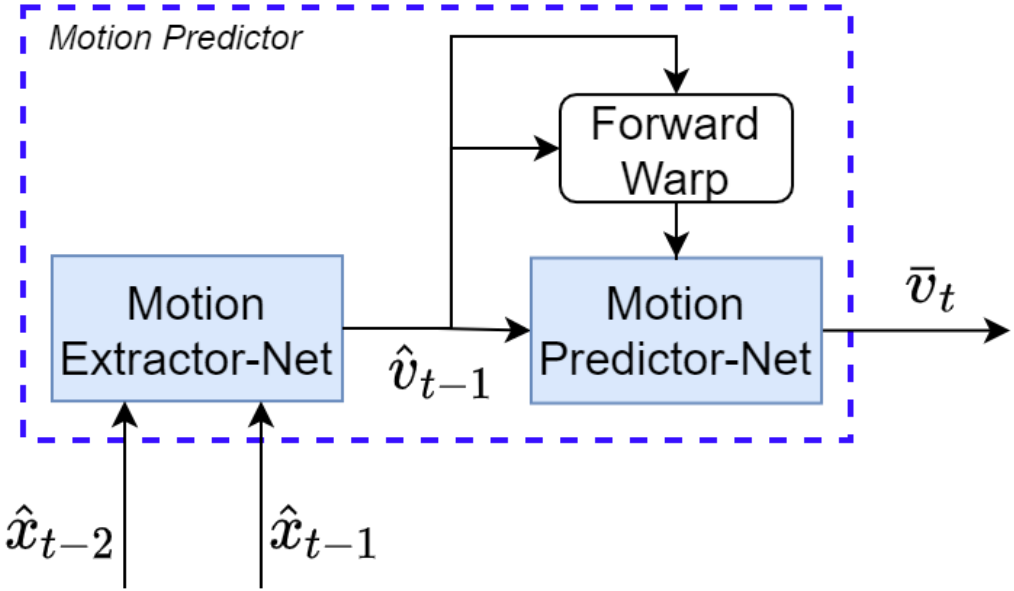
Proposed Method (1) - Motion Extractor

Motion Extractor-Net

- Use PWC-Net
- Produce optical flow as motion information

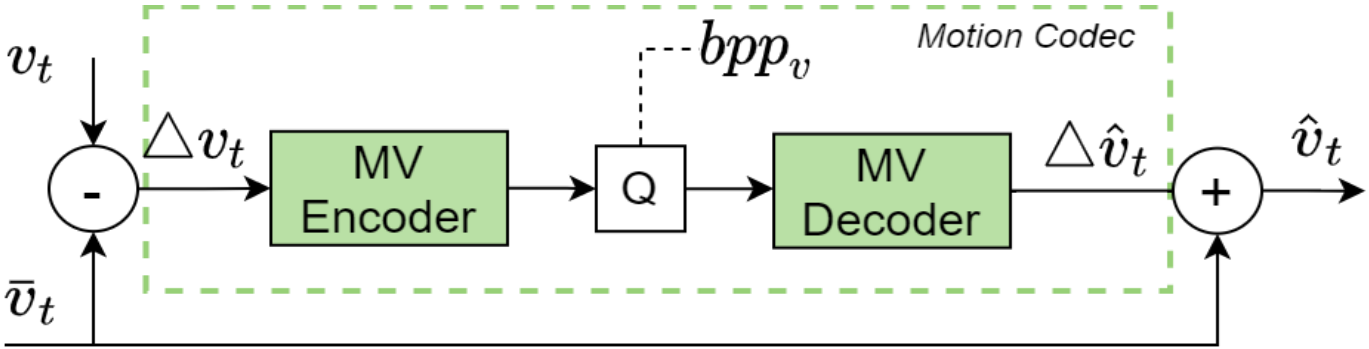
Motion Predictor-Net

- Extrapolate the optical flow based on two previous decoded frames



$\hat{x}_{t-2}$, $\hat{x}_{t-1}$: decoded frame *t-2* and *t-1*.
$\hat{v}_{t-1}$: estimated opt flow based on $\hat{x}_{t-2}$, $\hat{x}_{t-1}$.
$\bar{v}_t$: extrapolated opt flow for frame *t*.

Proposed Method (2) - Motion Compression



$\bar{v}_t$: extrapolated opt. flow for frame t.
v_t: estimated opt. flow based on current frame t.
$\triangle \hat{v}_t$: decoded differential opt. flow

Motion Compressor

- Transmit the differences between the estimated current opt. flow and the extrapolated opt. flow.
- Adopt the learning-based compressor with hyperprior

(Minnen, et al., “Joint Autoregressive and Hierarchical Priors for Learned Image Compression,” NIPS 2018)

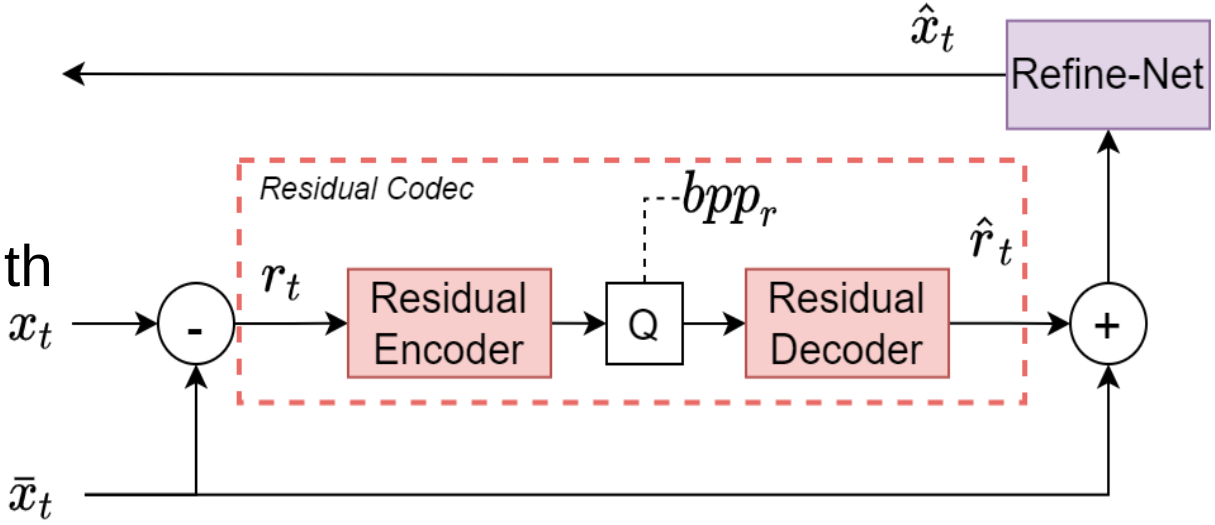
Proposed Method (3) - Residual Compressor

Residual Compressor

- Adopt the learning-based compressor with hyperprior

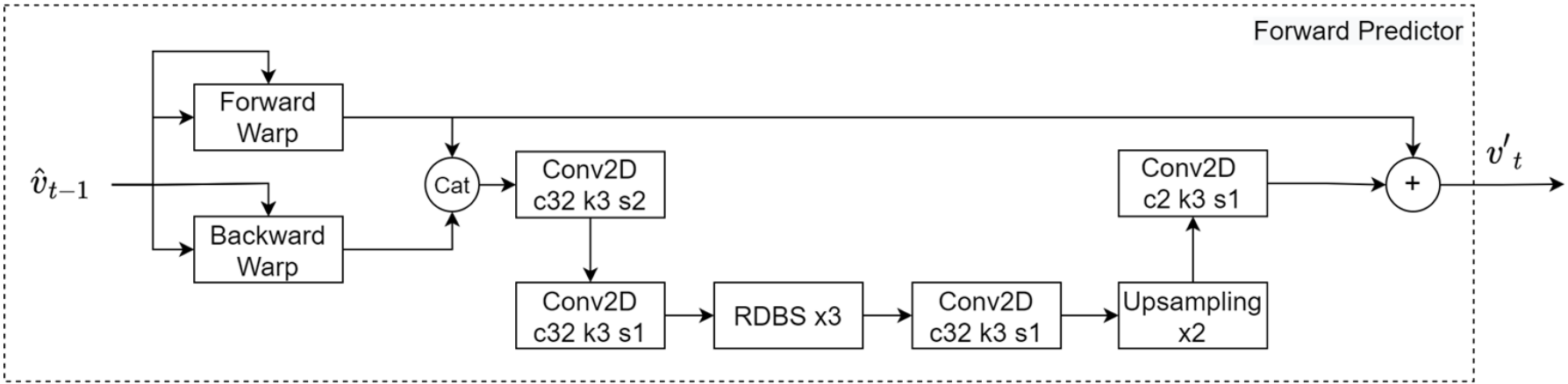
Refine-Net

- Design a multi-scale refinement network



x_t: original frame t.
$\bar{x}_t$: motion-compensated frame t.
r_t, $\hat{r}_t$: original and decoded residuals.

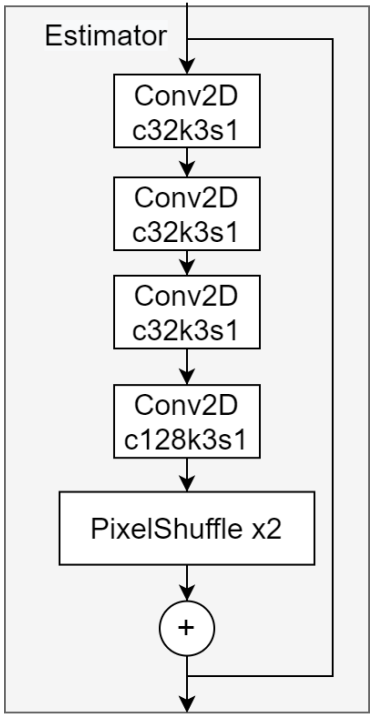
Architecture (1) - Motion Extractor/Prediction



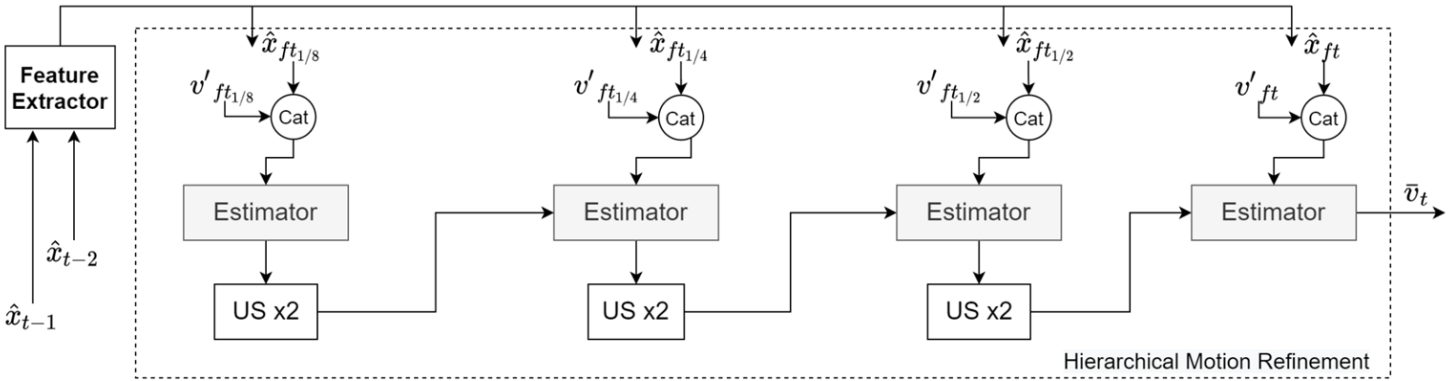
Motion Extractor-Net -- optical flow estimation based on two previous decoded frames

Motion Predictor-Net (motion extrapolation) -- forward warping to predict future optical flow

Architecture (2) - Motion Extractor/Prediction

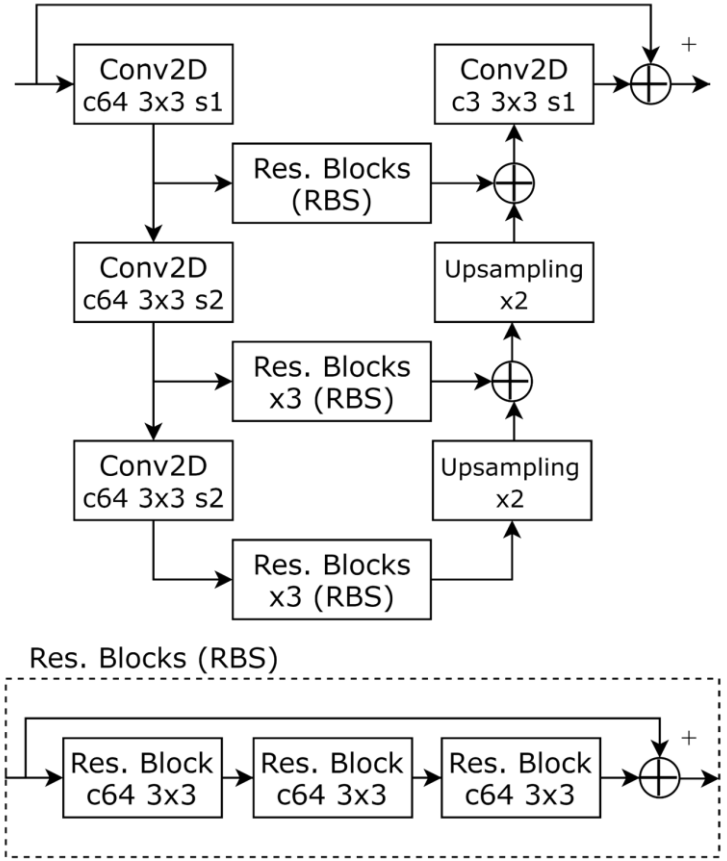


Hierarchical Motion Refinement -- enhance the quality of estimated optical flow



Architecture (3) - Refine-Net

- **The multi-scale refinement network** uses residual blocks.
- It uses the motion-compensated frame and the decoded residuals.
- Our compressor sends extra signaling to *hint* the Refine-Net to reconstruct a higher quality target frame.



Experiments - Training Setup – 3 Phases

Loss Functions for three phase of training

1. Motion Predictor

$$D_E = MSE(x'_t, x_t)$$

2. Motion Coding

$$D_v = MSE(\bar{x}_t, x_t)$$

3. Residual Coding / End-to-End

$$D_r = MSE(\hat{x}_t, x_t) \,/\, MSSSIM(\hat{x}_t, x_t)$$

$$L = \lambda_v * D_v + \lambda_r * D_r + R_v + R_r$$

$$\lambda_v = 0.2 * \lambda_r$$

x_t: original frame t

x'_t: motion compensated frame using predicted flow

$\bar{x}_t$: motion-compensated frame with coded motion information

$\hat{x}_t$: decoded frame t

R_v: bit rate for motion info.

R_r: bit rate for residuals

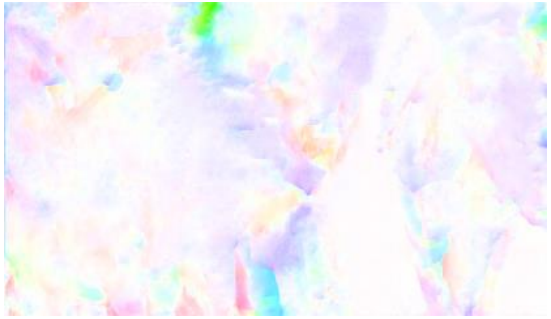
Experiments - Extrapolated Motion (1)

Reference Frame

Using Extrapolated Flow

Target Frame

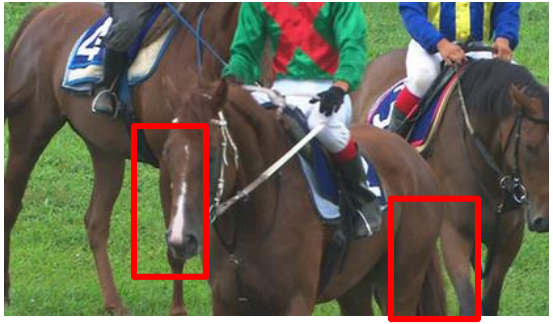
Extrapolated Flow



Flow Residual

Ground Truth Flow

Experiments - Extrapolated Motion (2)

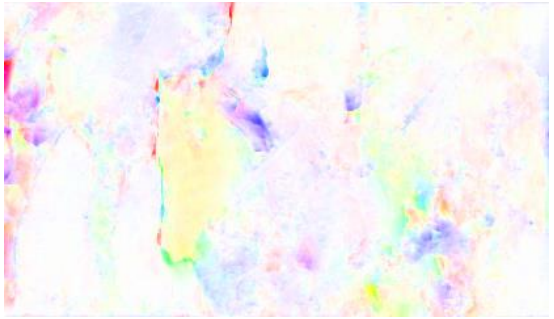


Reference Frame

Using Extrapolated Flow

Target Frame

Extrapolated Flow



Flow Residual

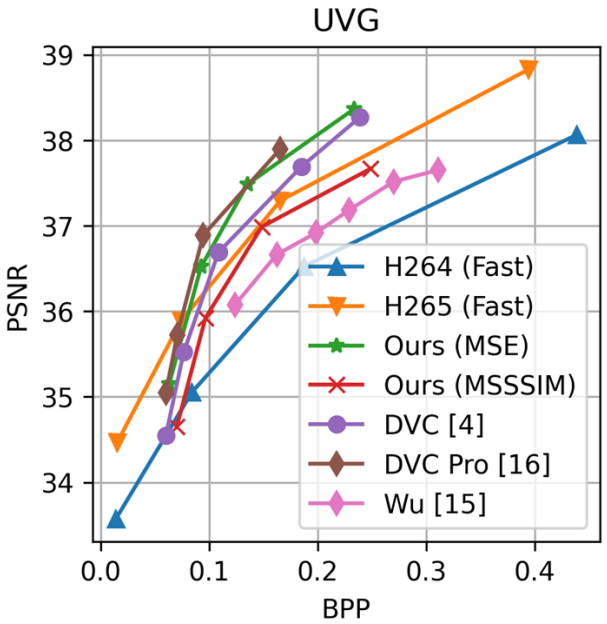
Ground Truth Flow

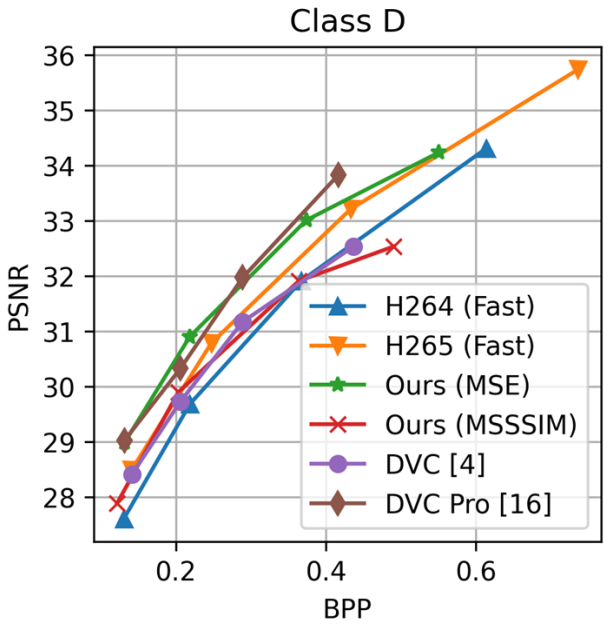
Experiments - Video Coding Sample

HEVC Class D – BasketballPass
416x240
Avg. PSNR: 31.15
Avg. BPP: ~0.11

HEVC Class D – HorseRace
416x240
Avg. PSNR: 29.49
Avg. BPP: ~0.2

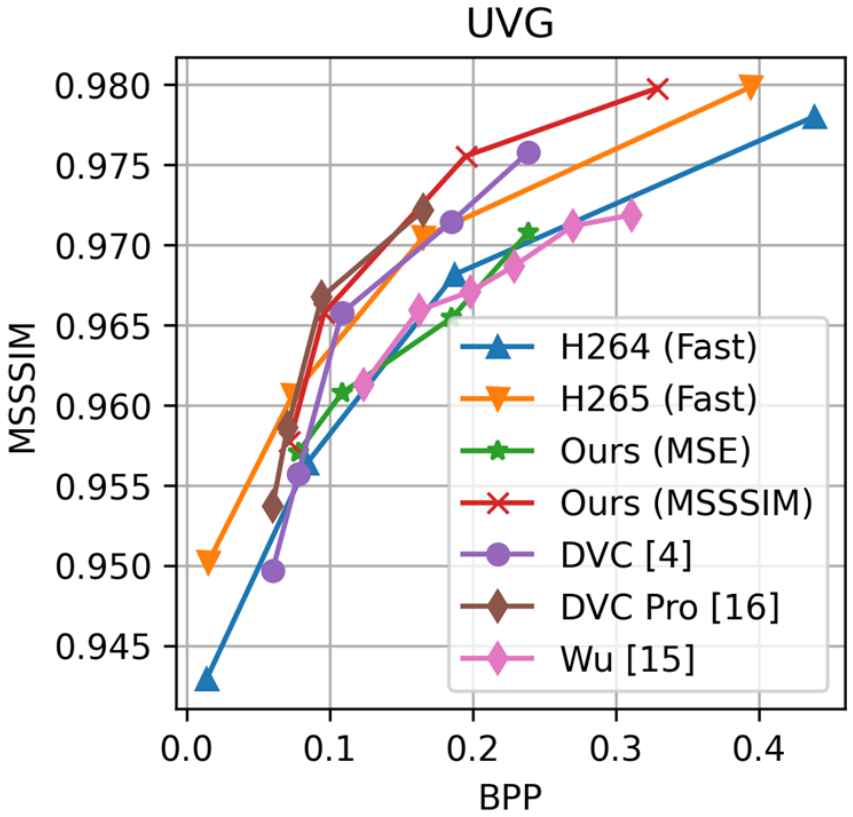
Experiments - R-D Performance-1

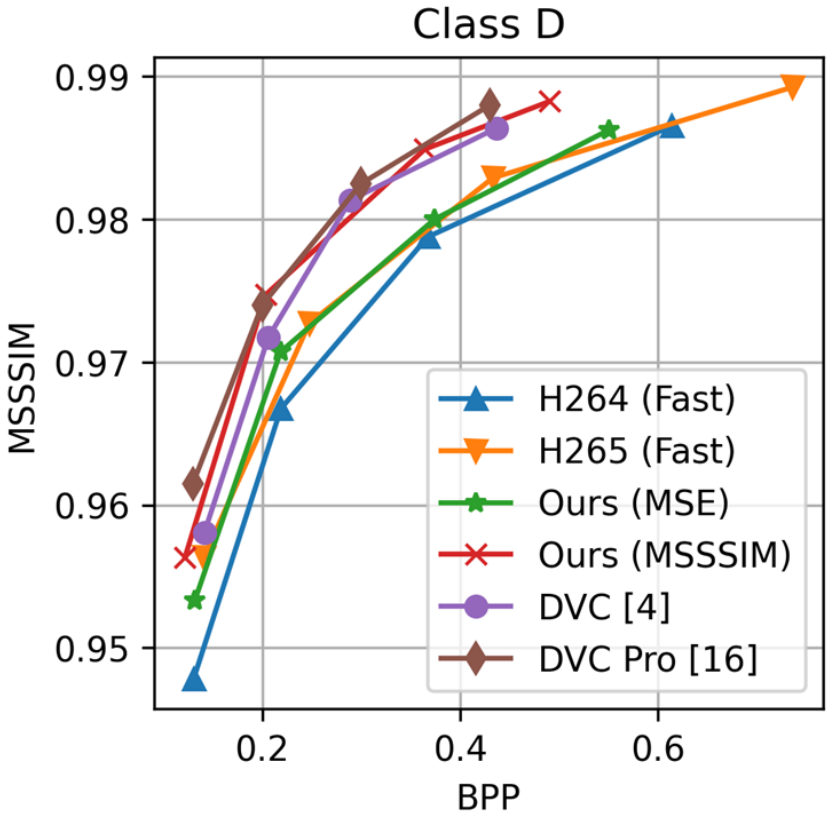




- Training dataset: Vimeo-90k Septuplet
- Validation dataset: UVG, HEVC Class D
- Testing scenario: 10 frames per GOP, using first 100 frames from each dataset.

Experiments - R-D Performance-2





Conclusions

1. A **learning-based inter-frame video compression** system is presented for interframe coding.
2. We propose a motion predictor-net, which predicts the motion vectors for the target frame based on two previously coded frames.
3. Our residual compressor generates side information embedded in the coded residuals to assist Refine-Net for better image reconstruction.
4. Its RD performance is comparable with the other SOTA learning-based video codecs.

Thank you for your attention

Additional Information

Network Parameters

The detail for our network parameters is shown in Table 1. The intra coding has the largest parameter number around 7.2M. The motion coding uses 6.3M, and the residual coding.

Model	Params
Intra coding	7.2 M
Motion extractor-net	700 K
Motion estimator-net	640 K
Motion compressor-net	5.0 M
Residual compressor-net	5.0 M
Residual refine-net	501 K

Refine-net in Residual Information

Portion of the reference frame (left) to compare with (center), which contains “extra signaling” (dotted pattern) next to the rider arms.
The extra signaling is used the refine-net to produce the correct reconstruction (right).

$\bar{x}_{t-1}$

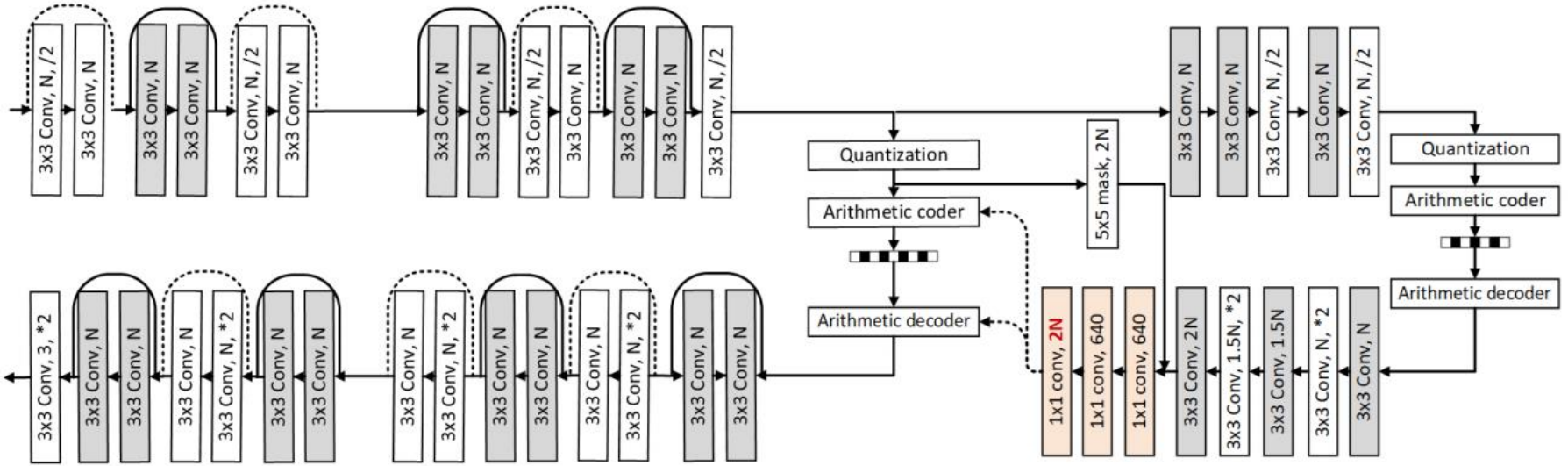
$\bar{x}_{t-1} + \hat{r}_t^*$

$\hat{x}_t$

Decoded Residual (an example)

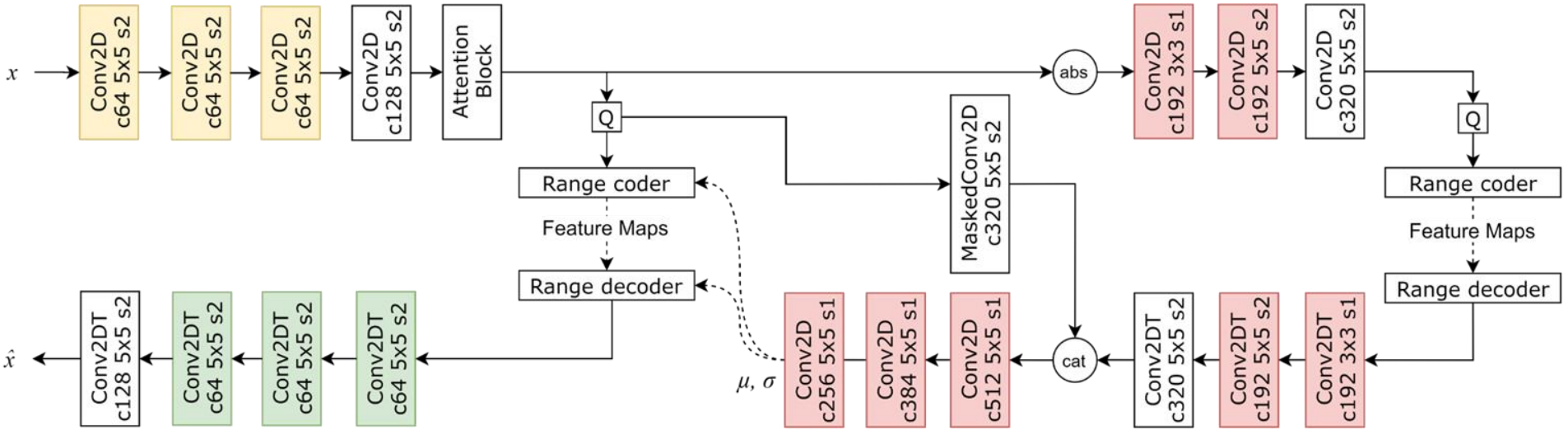
Amplified with a factor of 5

Intra Coding



Our intra-coding used image compression network from Cheng, *et al*. (Learned Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules, 2020)

Motion / Residual Compressor



The motion / residual compressors are taken from the design of Minnen, et al. (Joint Autoregressive and Hierarchical Priors for Learned Image Compression, 2019)

Experiments - More Results