WED-AM-193

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

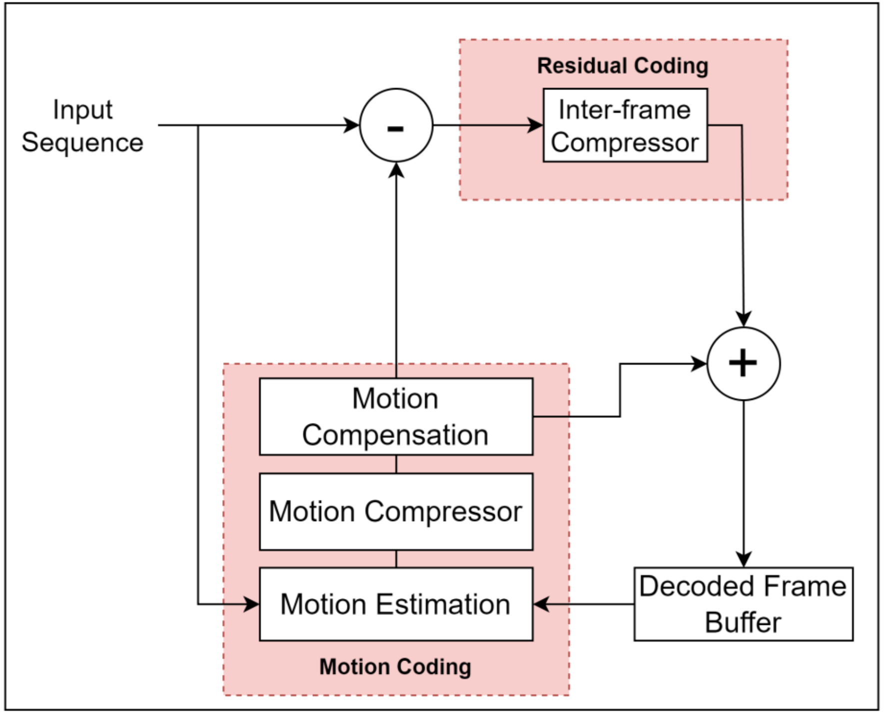
Hierarchical B-Frame Video Coding Using Two-Layer CANF Without Motion Coding

David Alexandre, Hsueh-Ming Hang, and Wen-Hsiao Peng

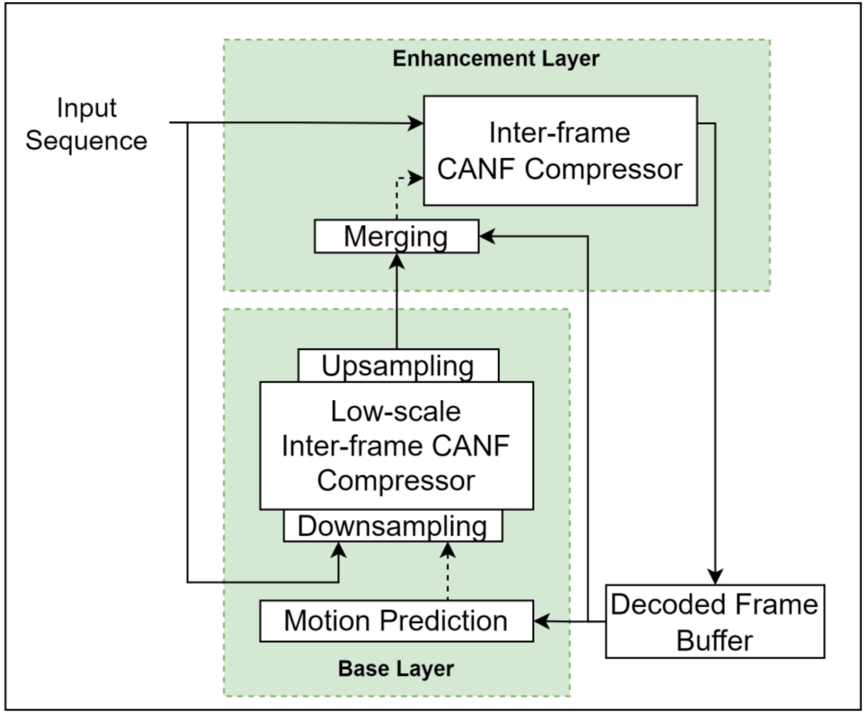
National Yang Ming Chiao Tung University (NYCU), Taiwan

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

Hybrid Coding vs. This Work

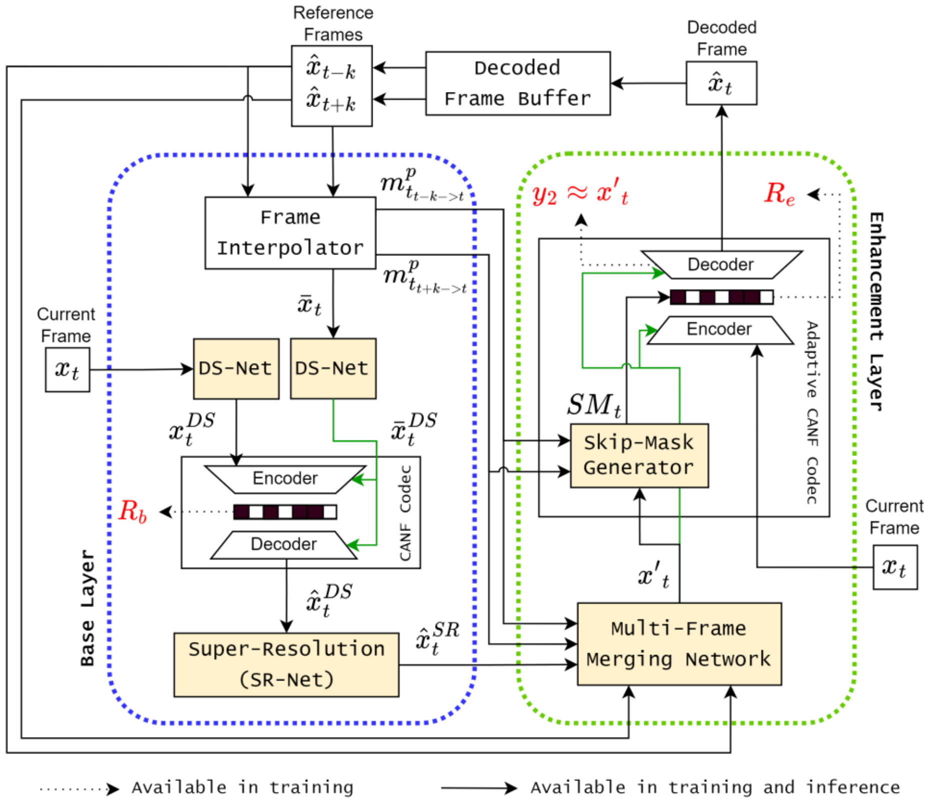


Common Generic Hybrid Video Coding System



Proposed Video Coding System
(Two-Layer without Motion Coding)

TLZMC Framework



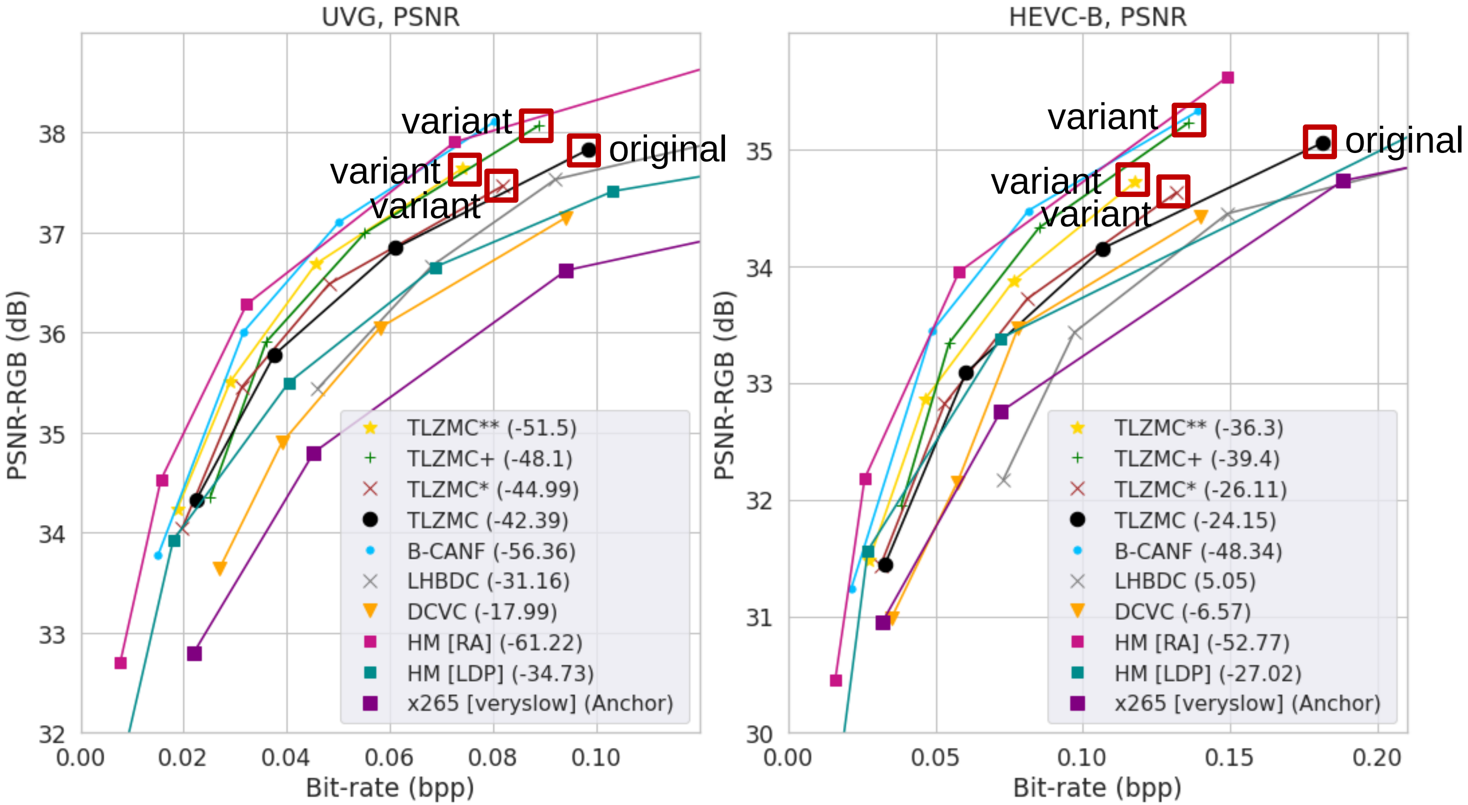
Base-Layer

- Frame Interpolator
- Downsampler (DS-Net)
- CANF Codec
- Super-Resolution (SR-Net)

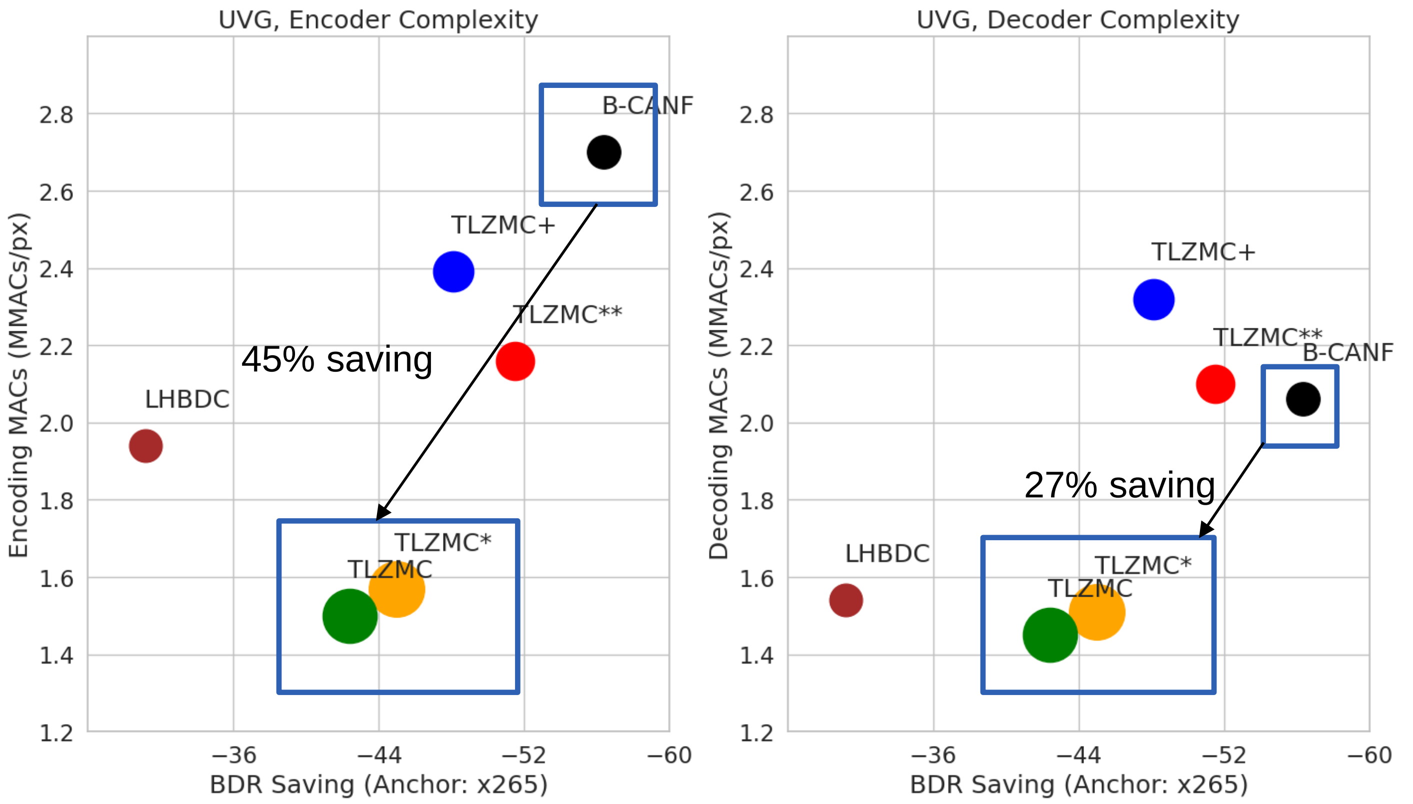
Enhancement Layer

- Multi-Frame Merging Network
- Skip-Mask Generator
- Adaptive CANF Codec

RD-Curve and BD-Rate (UVG, HEVC-B)



Complexity Comparison



Contributions

01 **THE FIRST ATTEMPT**

The first attempt at two-layer coding structure without motion coding

02 **COMPLEXITY REDUCTION**

Shows complexity reduction over the state-of-the-art methods

03 **CONTEXT ADAPTIVE FRAMEWORK**

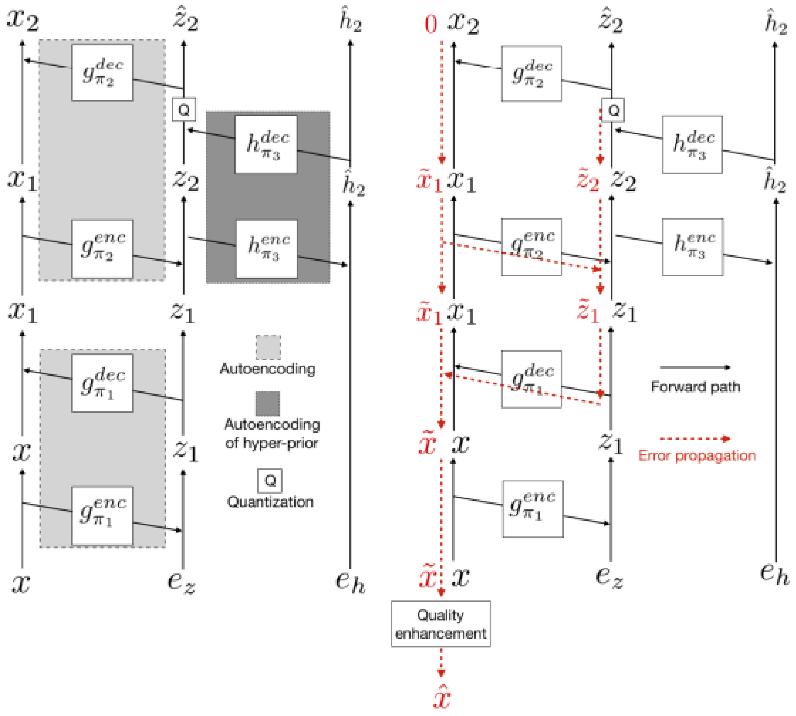
Demonstrates context adaptive framework via rate control

04 **GOOD RD PERFORMANCE**

Comparable RD performance to SOTA for certain video scenes

ANFIC

ANFIC: Image Compression Using Augmented Normalizing Flows

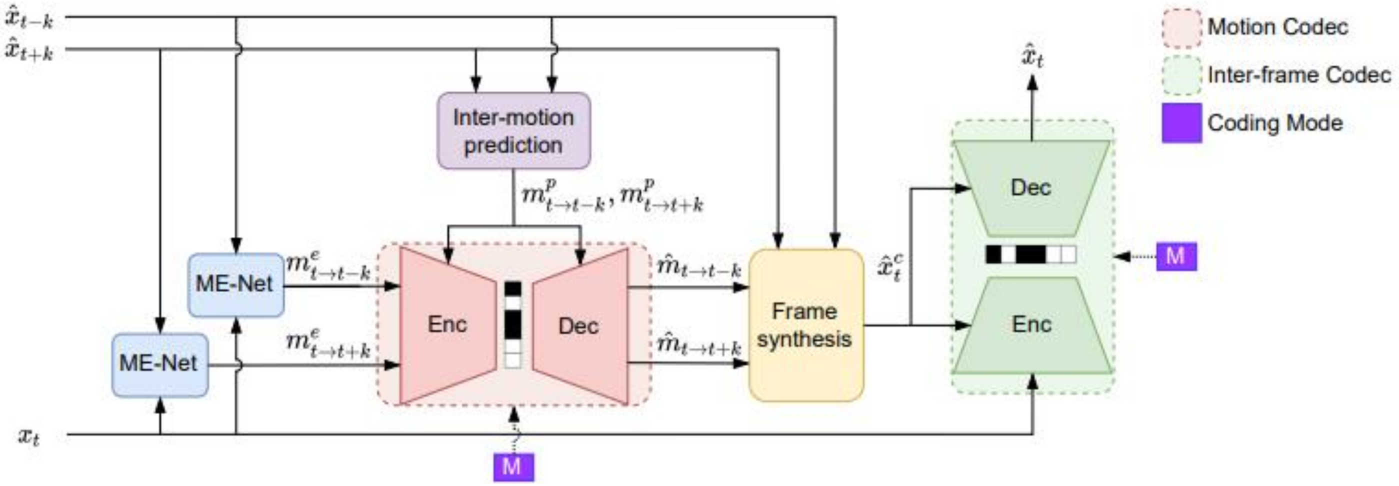


- Based on Augmented Normalizing Flows (ANF)
- Uses a flow-based framework that hierarchically stacks and extends multiple VAEs
- Achieves state-of-the-art coding performance in terms of PSNR-RGB

*Figure from ANFIC: Image Compression Using Augmented Normalizing Flows, (Ho, et al. 2021)

B-CANF: Deep B-frame Coding

B-CANF: Adaptive B-frame Coding with Conditional Augmented Normalizing Flows (Chen, et al. 2022)



- Utilizes conditional augmented normalizing flows
- Incorporates frame-type adaptive coding

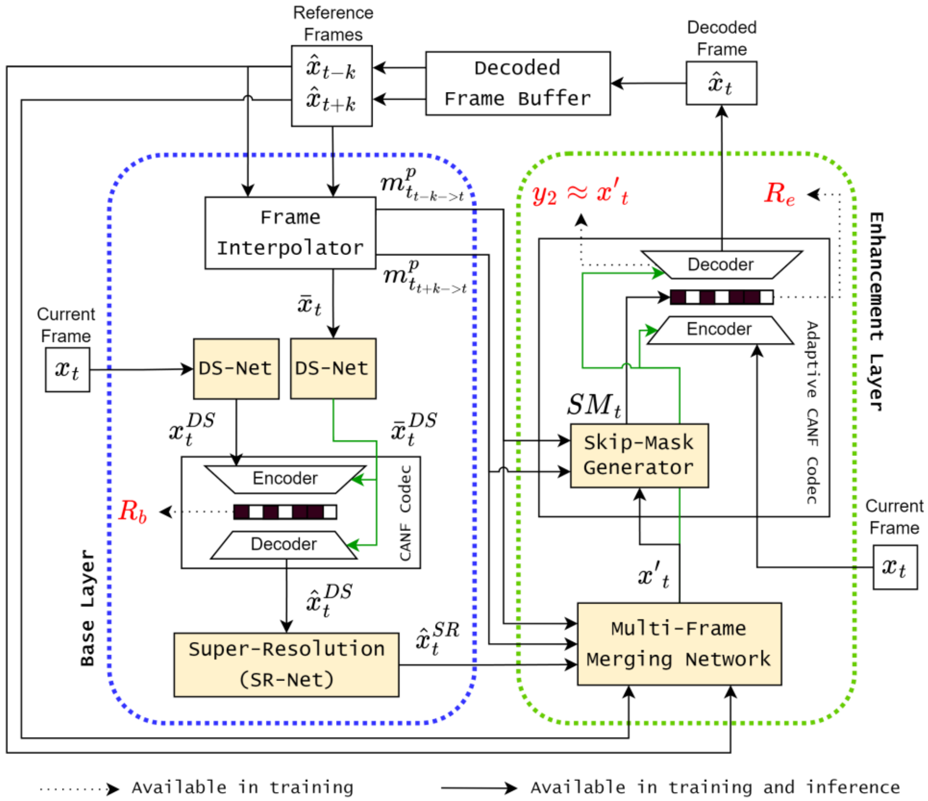
TLZMC

Two-**L**ayer **Z**ero **M**otion **C**oding for Hierarchical B-frames

Key features:

- Lower complexity than the state-of-the-art, B-CANF
- Comparable results to B-CANF in some scenarios
- Outperform other deep hierarchical B-frame coding
- Context adaptive coding and bit rate control

TLZMC Framework



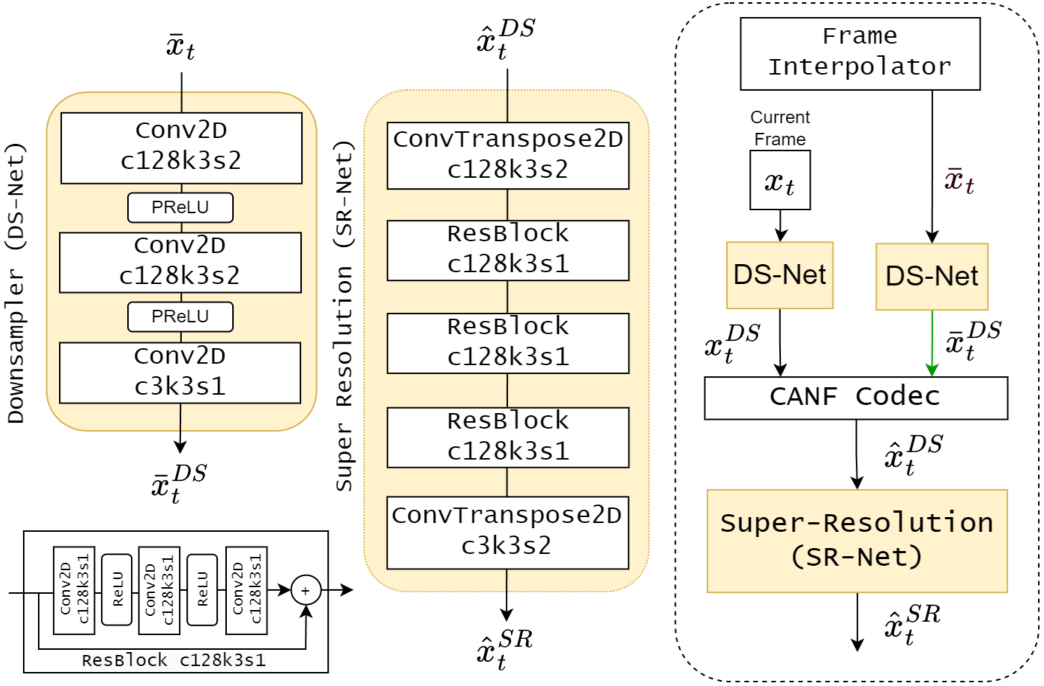
Base Layer

- Frame Interpolator
- Downsampler (DS-Net)
- CANF Codec
- Super-Resolution (SR-Net)

Enhancement Layer

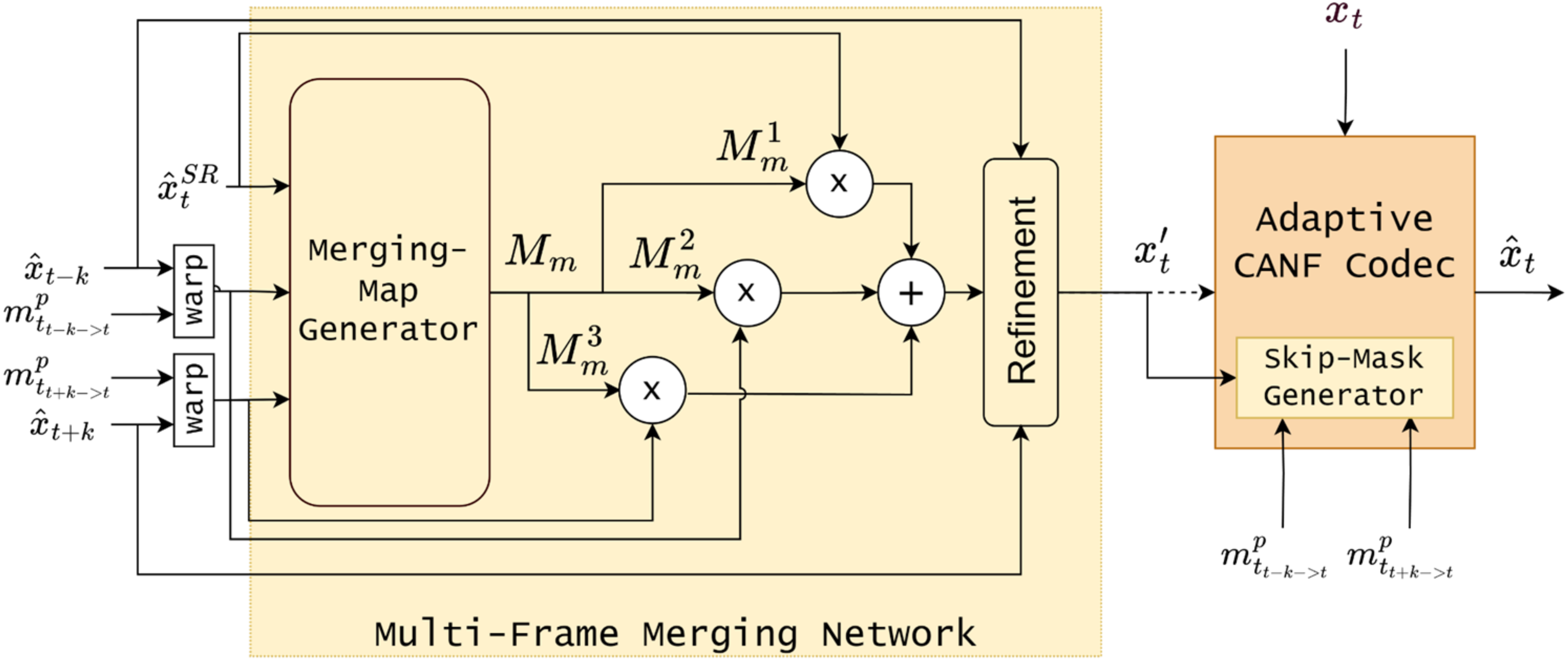
- Multi-Frame Merging Net
- Skip-Mask Generator
- Adaptive CANF Codec

DS-Net and SR-Net



- Uses simple neural networks to perform downsampling (1/4x) and super-resolution
- Supports end-to-end training

Enhancement Layer (EL)



Merging Maps Examples (UVG ShakeNDry)

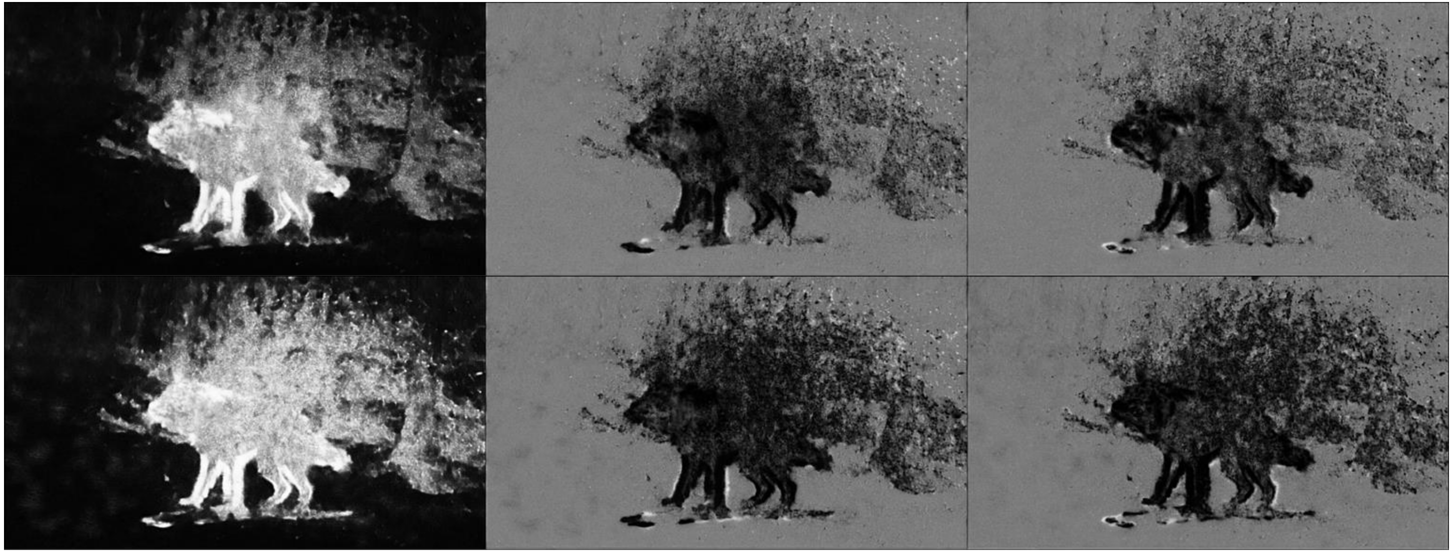
M_m^1
$(*\hat{x}_t^{SR})$

M_m^2
$(*warp(\hat{x}_{t-k}, m^p_{t_{t-k->t}}))$

M_m^3
$(*warp(\hat{x}_{t+k}, m^p_{t_{t+k->t}}))$

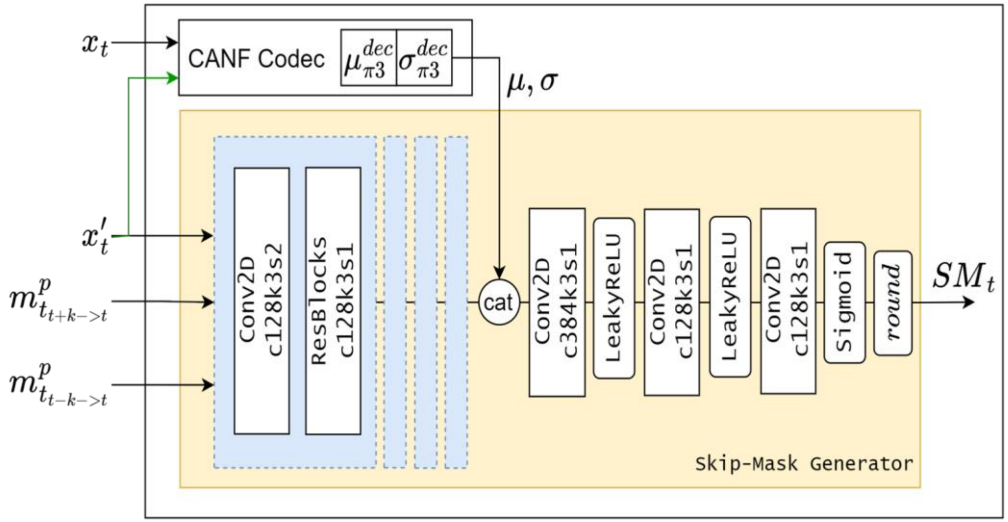
$\lambda = 256$

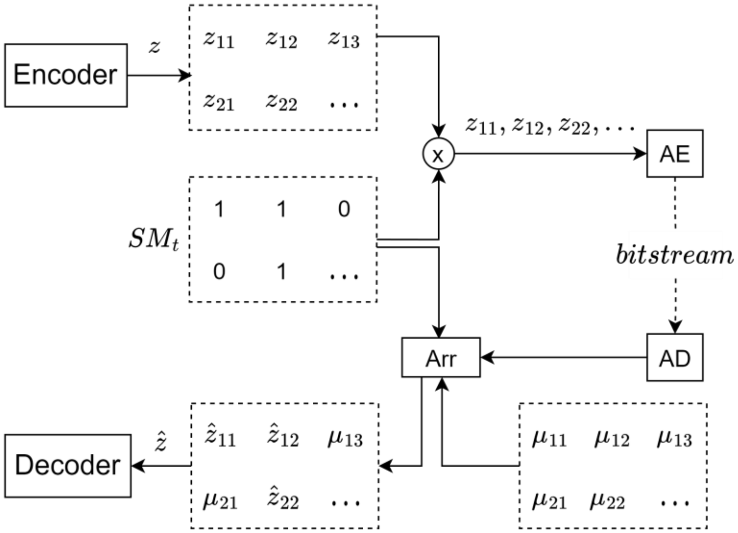
$\lambda = 2048$



Adaptive CANF Codec

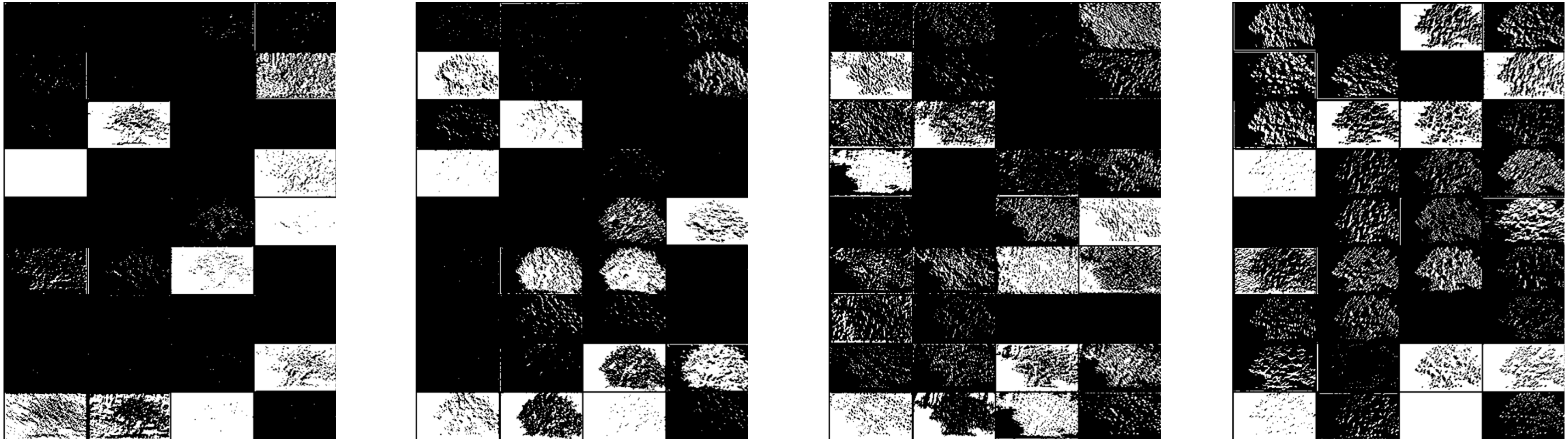
- Skips latent samples by using their mean values for reconstruction
- Predicts a “skip mask” on both the encoder and decoder sides





Skip Mask Examples (UVG ShakeNDry)

Black = skipped



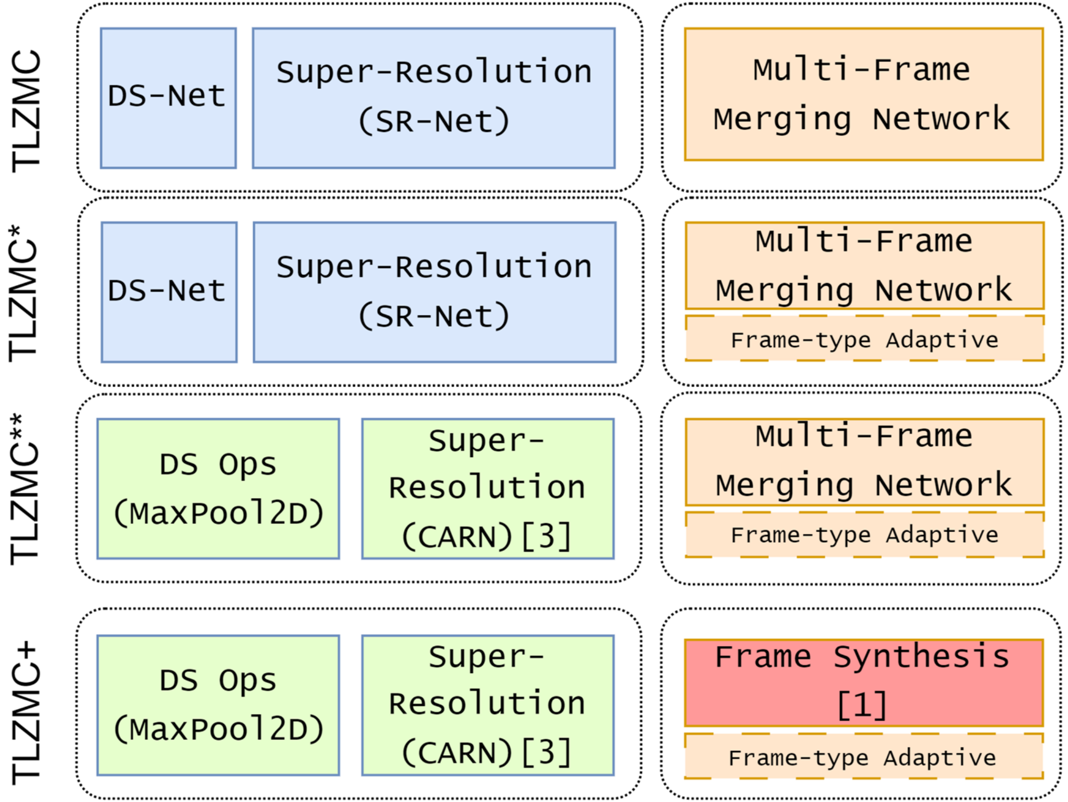
λ=256

λ=512

λ=1024

λ=2048

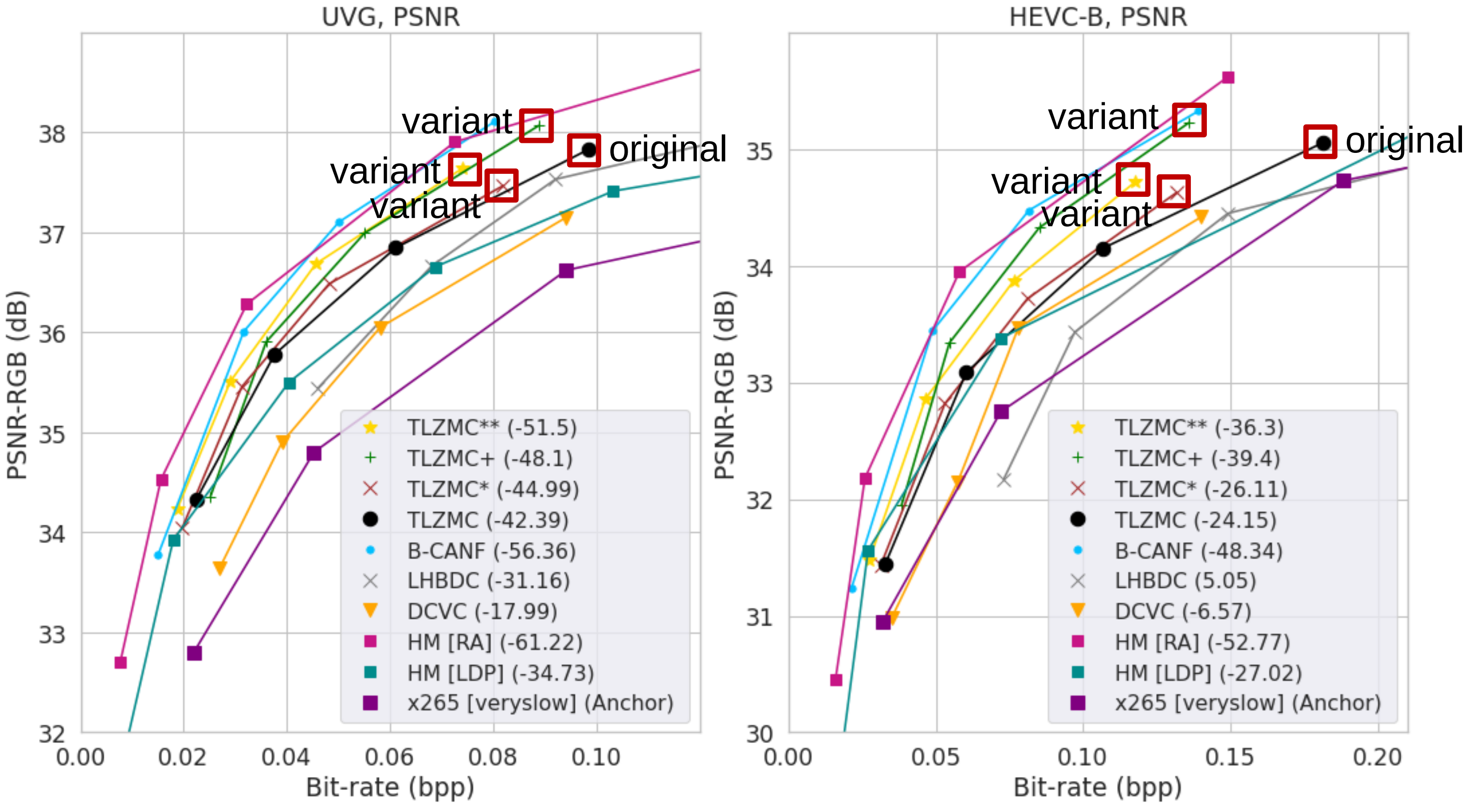
TLZMC Variants



[1] B-CANF (Chen, et al. 2022)

[3] CARN (Ahn, 2018)

RD-Curve and BD-Rate (UVG, HEVC-B) Results



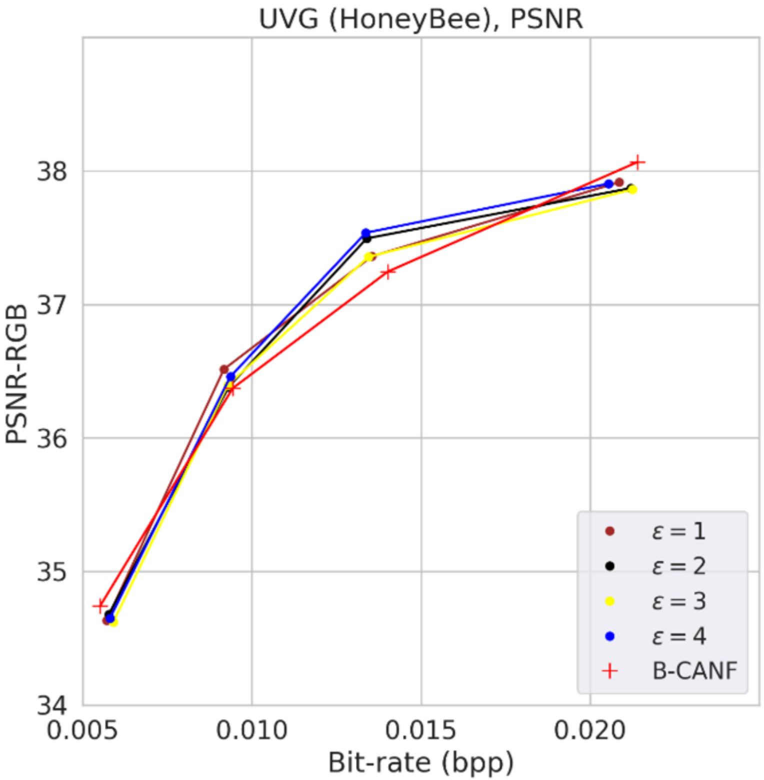
Training Procedure

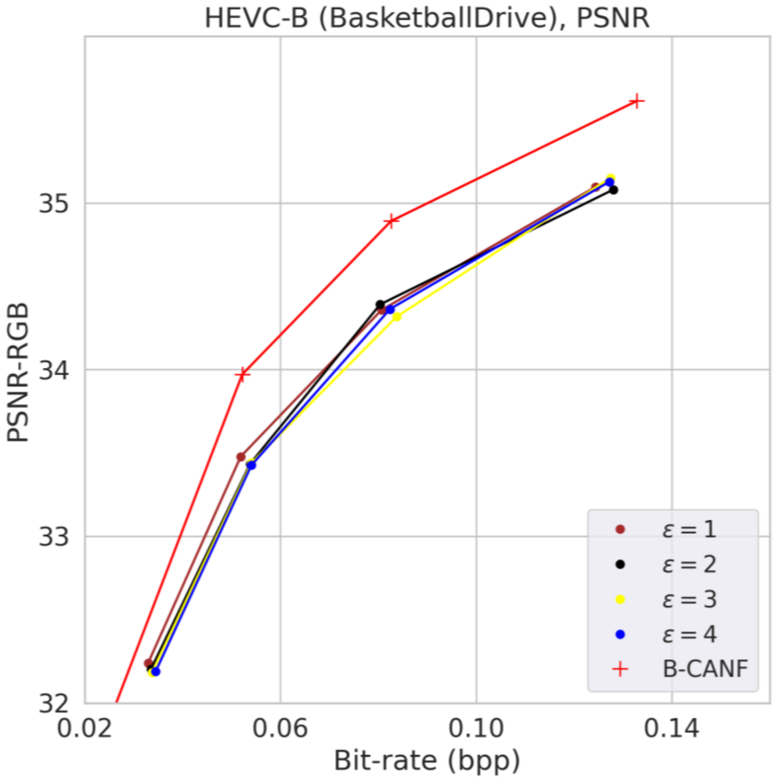
Training Sequence

	TLZMC	Loss Function
1	Frame Interpolator	$D(\bar{x}_t, x_t)$
2	DS + CANF Compressor	$D(\hat{x}_t^{DS}, x_t^{DS}) + R_b$
3	Super Resolution	$D(\hat{x}_t^{SR}, x_t) + R_b$
4	Multi-Frame Merging Network	$D(x'_t, x_t)$
5	Adaptive CANF Compressor	$D(\hat{x}_t, x_t) * \lambda + R_b + R_e$
6	End-to-end	$D(\hat{x}_t, x_t) * \lambda + \varepsilon * R_b + R_e + Aux$

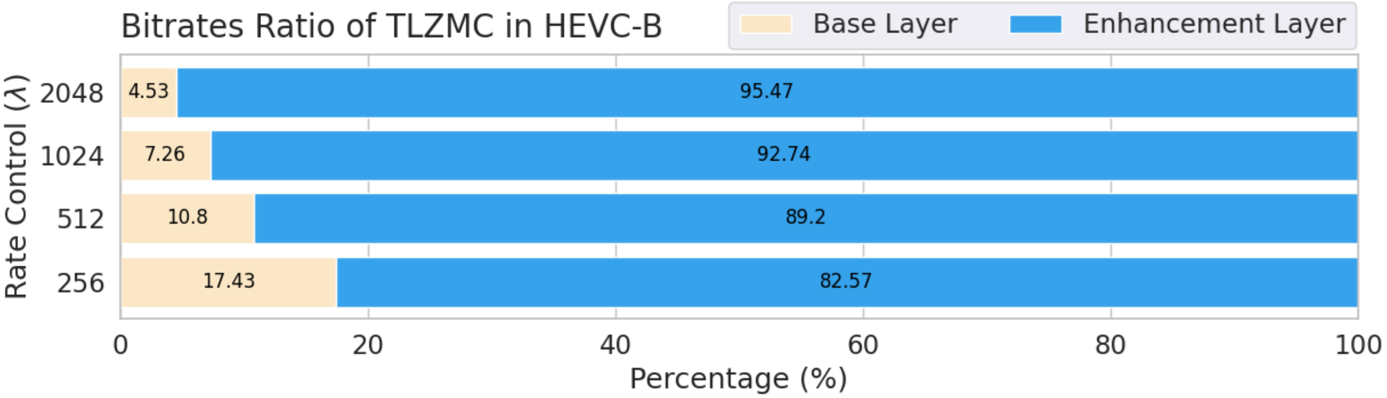
$$Aux = (D(y_2, x'_t) + D(x'_t, x_t) + D(\hat{x}_t^{SR}, x_t)) * 0.01 * \lambda$$

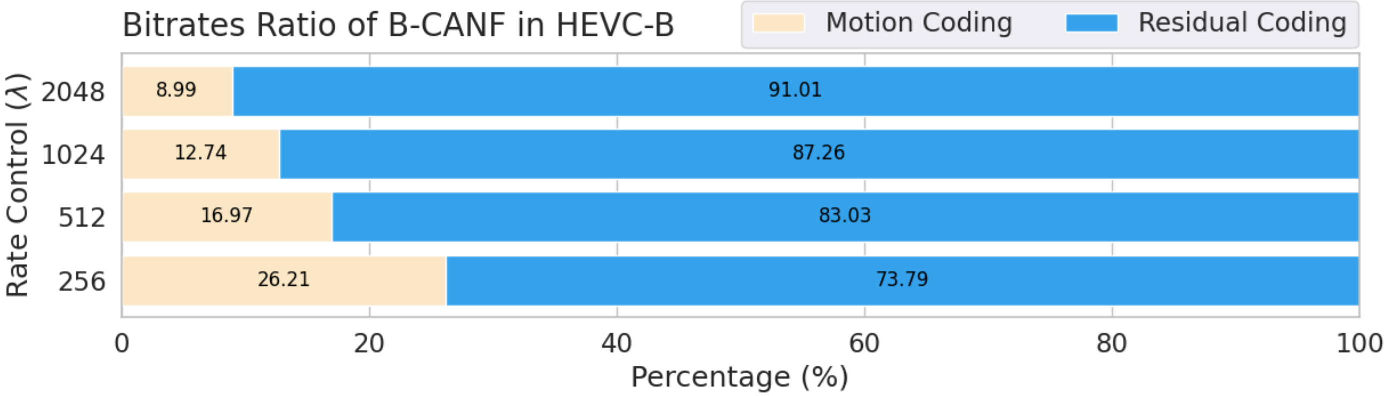
Bit Allocation for BL and EL



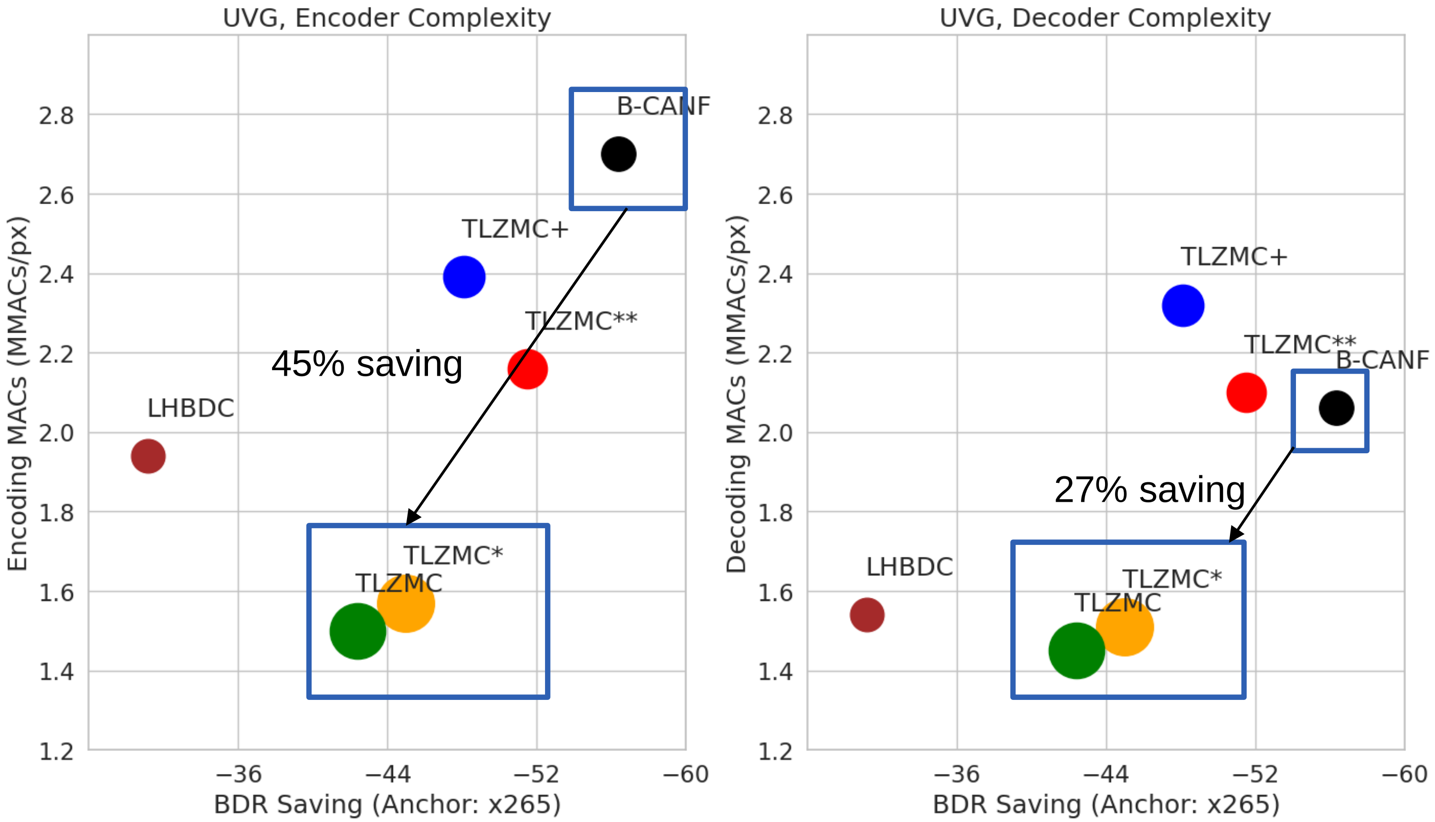


Bit Rate Ratio of BL to EL





Complexity Comparison



45% saving

27% saving

Conclusions

- Motion coding replaced by “neural prediction + base-layer coding”
- Less computational complexity
- RD performance close to the SOTA
- Content adaptivity is crucial to coding gain

Thank you for your attention