DCC2023 Poster Session

Temporal Down-sampling based Video Coding with Frame-Recurrent Enhancement

Presenter: Keren He
Co-authors: Keren He, Chen Fu, Chi Do -Kim Pham, Lu Zhang, and Jinjia Zhou
HOSEI University
2023/3/23

法政大学
HOSEI University

Guideline

- Introduction
- Proposals
- Experimental results
- Coclusion

Introduction

I Introduction

Traditional video coding **VS** Downsampling-based coding method

Traditional video codig method

Compression → **Transmission** → **Decompression**

Down-samlping

Comparing with traditional video coding framework:

- Receive low bitrate for narrow bandwidth transmission.
- Small computational cost.

Compression → **Transmission** → **Decompression**

SR Network

Up-samlping

Downsampling-based coding method

I Introduction

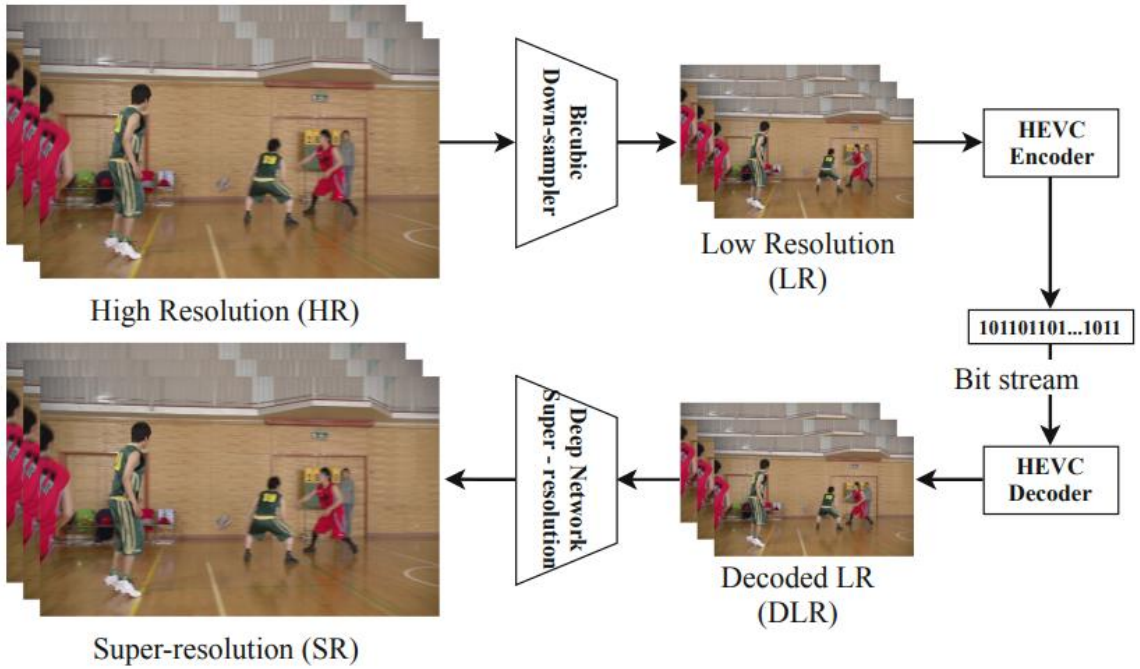
Problem: In many digital systems, the transmission bandwidth, as well as storage capacity, are usually very limited.

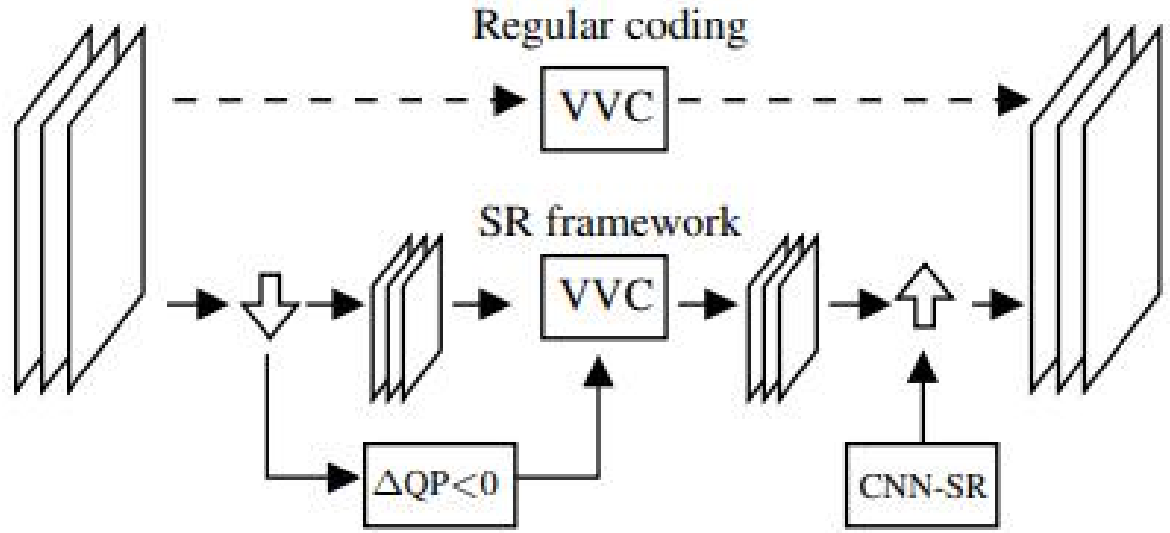
Downsampling -based coding method + **deep learning techniques**

Aim: To seek lower bit rates and further obtain high-quality up-sampled videos.

Introduction

Related works—Downsampling-based coding method



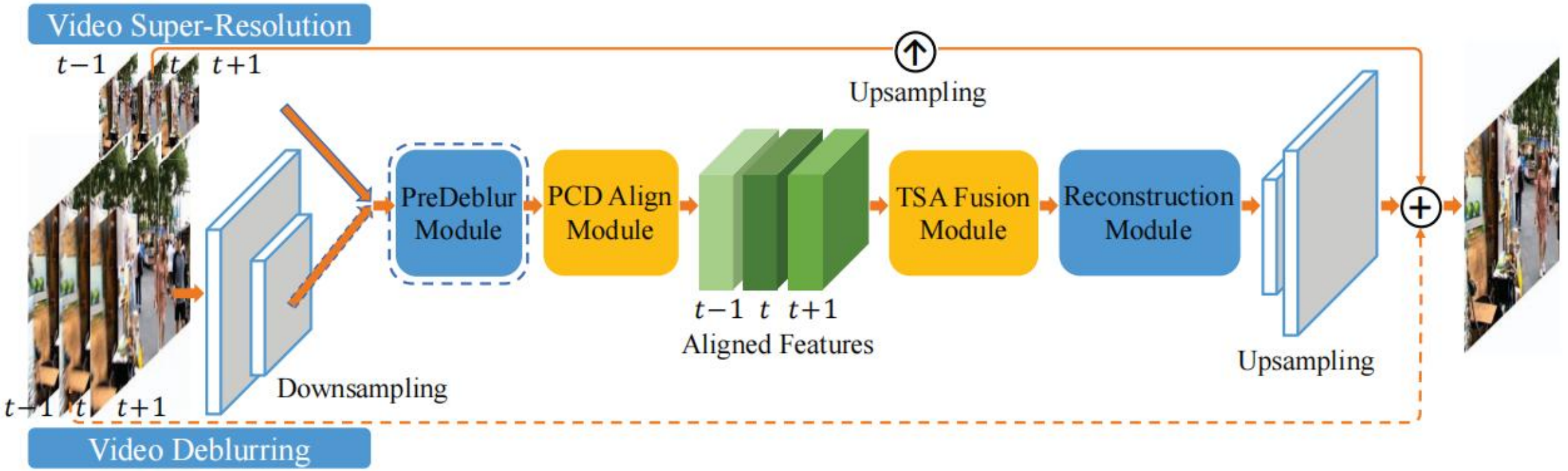


Unlike these works, we propose a temporal down sampling method to select down-sampled frames and use inter-frame information for enhancement.

- *Ho, M. M., He, G., Wang, Z., & Zhou, J. (2020, January). Down-sampling based video coding with degradation-aware restoration-reconstruction deep neural network. In International Conference on Multimedia Modeling (pp. 99-110). Springer, Cham.*
- *Fatemeh Nasiri, Wassim Hamidouche, Luce Morin, Gildas Cocherel, and Nicolas Dhollande, "A study on the impact of training data in cnn based super-resolution for low bitrate end-to-end video coding," in 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, 2020, pp. 1–5.*

Introduction

Related work—EDVR: Video Restoration with Enhanced Deformable Convolutional Networks

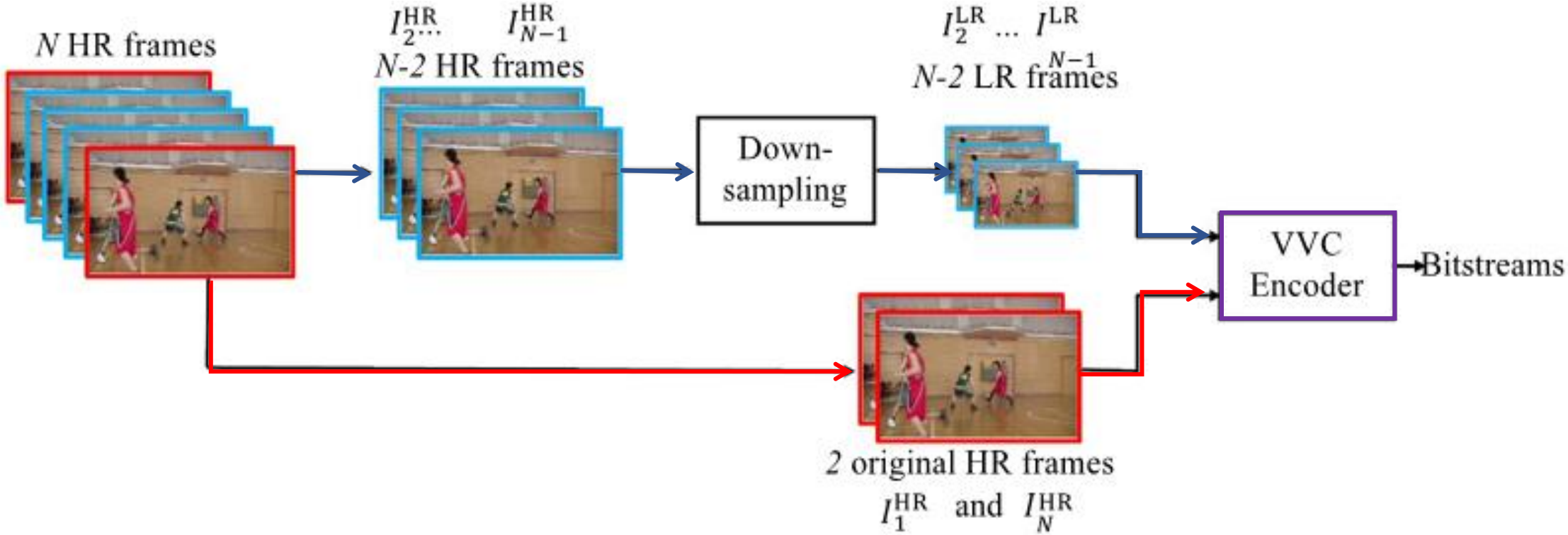


Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy, "Edvr: Video restoration with enhanced deformable convolutional networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 0–0.

Proposals

The Proposal 1—TDS Framework

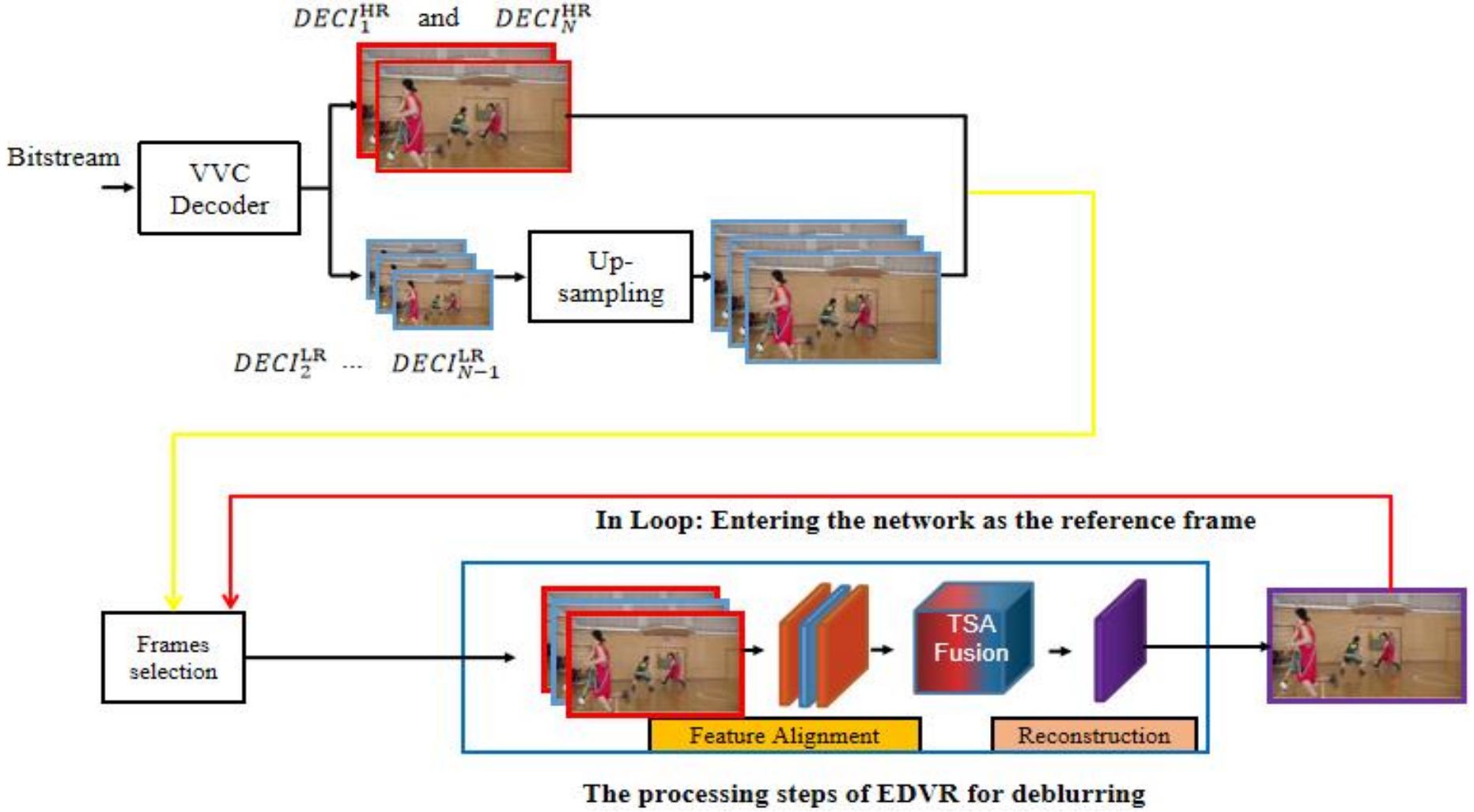
Temporal Down-sampling based Video Coding



(a) Encoder structure

The Proposal 1—TDS Framework

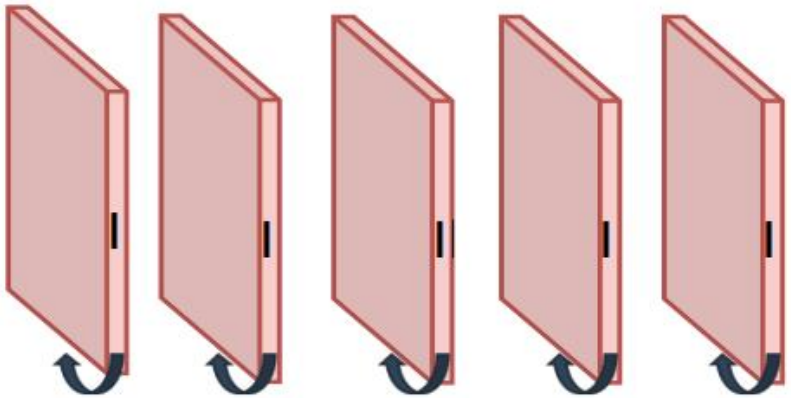
Temporal Down-sampling based Video Coding



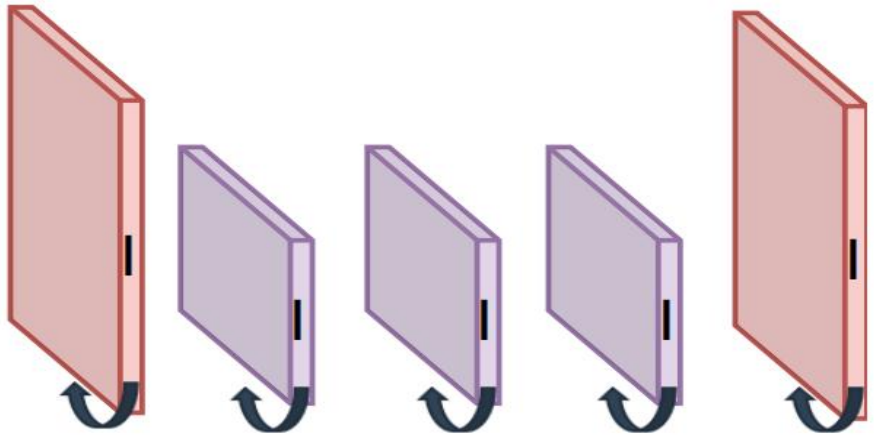
(b) Decoder structure

The Proposal 1—TDS Framework

Original All intra(AI) configuration **VS** Our proposal modules based on the AI configuration



All-intra(Original)



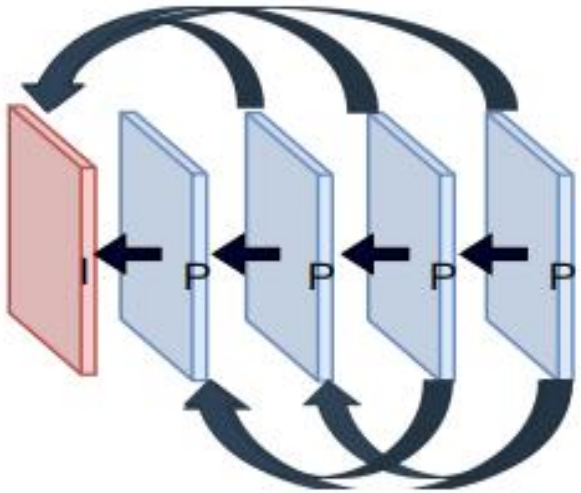
ours

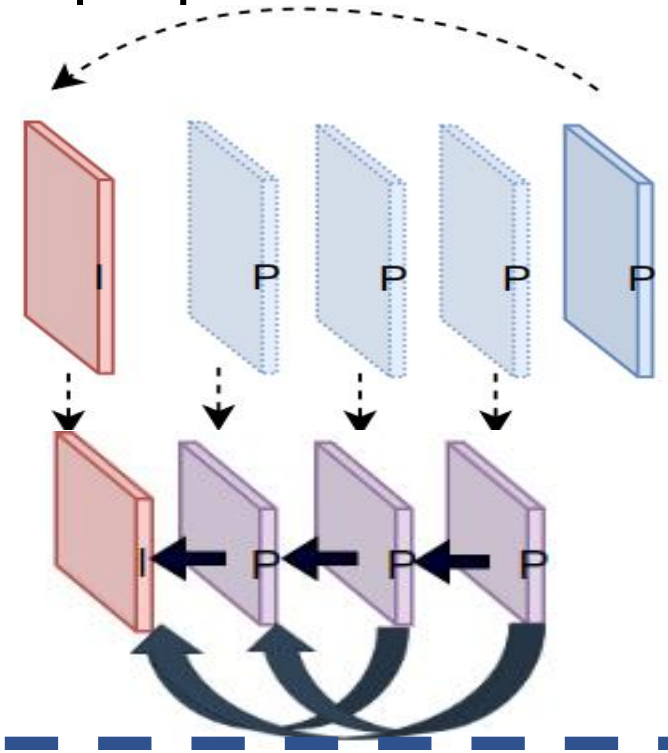
Our method: Under All intra configuration, we only downsample the middle frames into low resolution images, the first and last two frames remain the original size.

The Proposal 1—TDS Framework

Original (Low-Delay-P)LDP configuration **VS** Our proposal modules based on the LDP configuration

Low-Delay-P(Original)





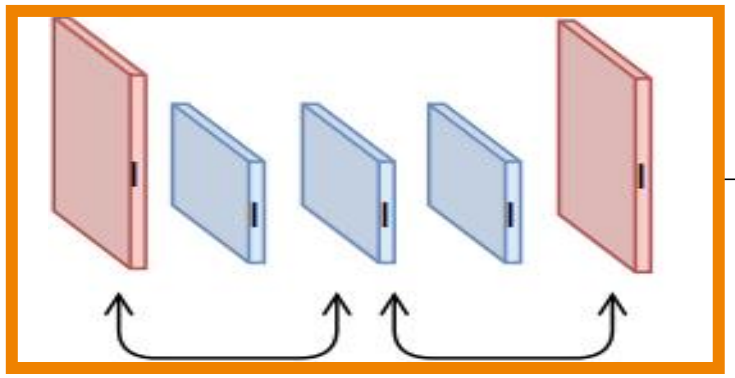
ours

Problem: Down-sampled milddle frames need to refer to I frame which has same size for decoding.

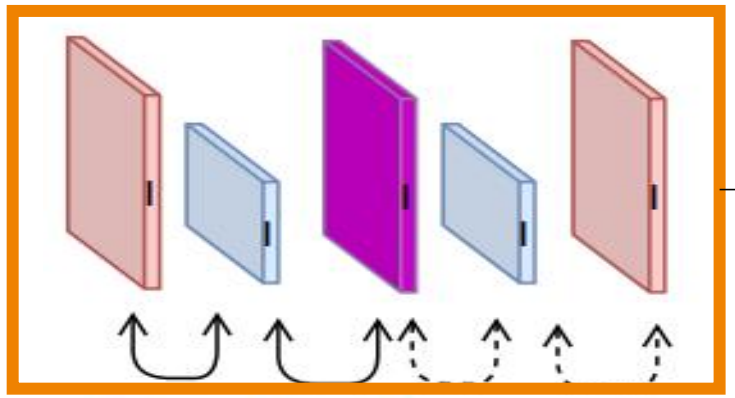
Our method: We only downsample the middle N frames(N can be 3,5,7....)into low resolution images, the first and last two frames remain the original size. Then encode an extra down-sampled I frame, use this frame to decode down-sampled middle frames.

The Proposal 2—SRFR Framework

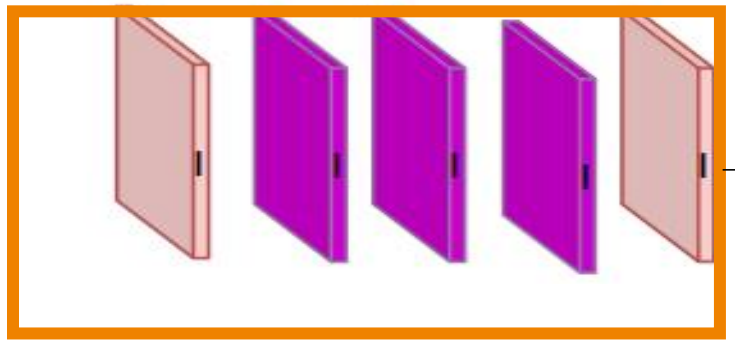
Up-sampling with frame-Recurrent enhancement



Enhancing the middle frame by referring to the first and last frames



The second frame is enhanced by referring to the first and third frames, and the fourth frame is enhanced by referring to the third and fifth frames.



The essence of enhancement is to refine the current frame by referring to nearby frames.

Experimental Results

Experimental Results

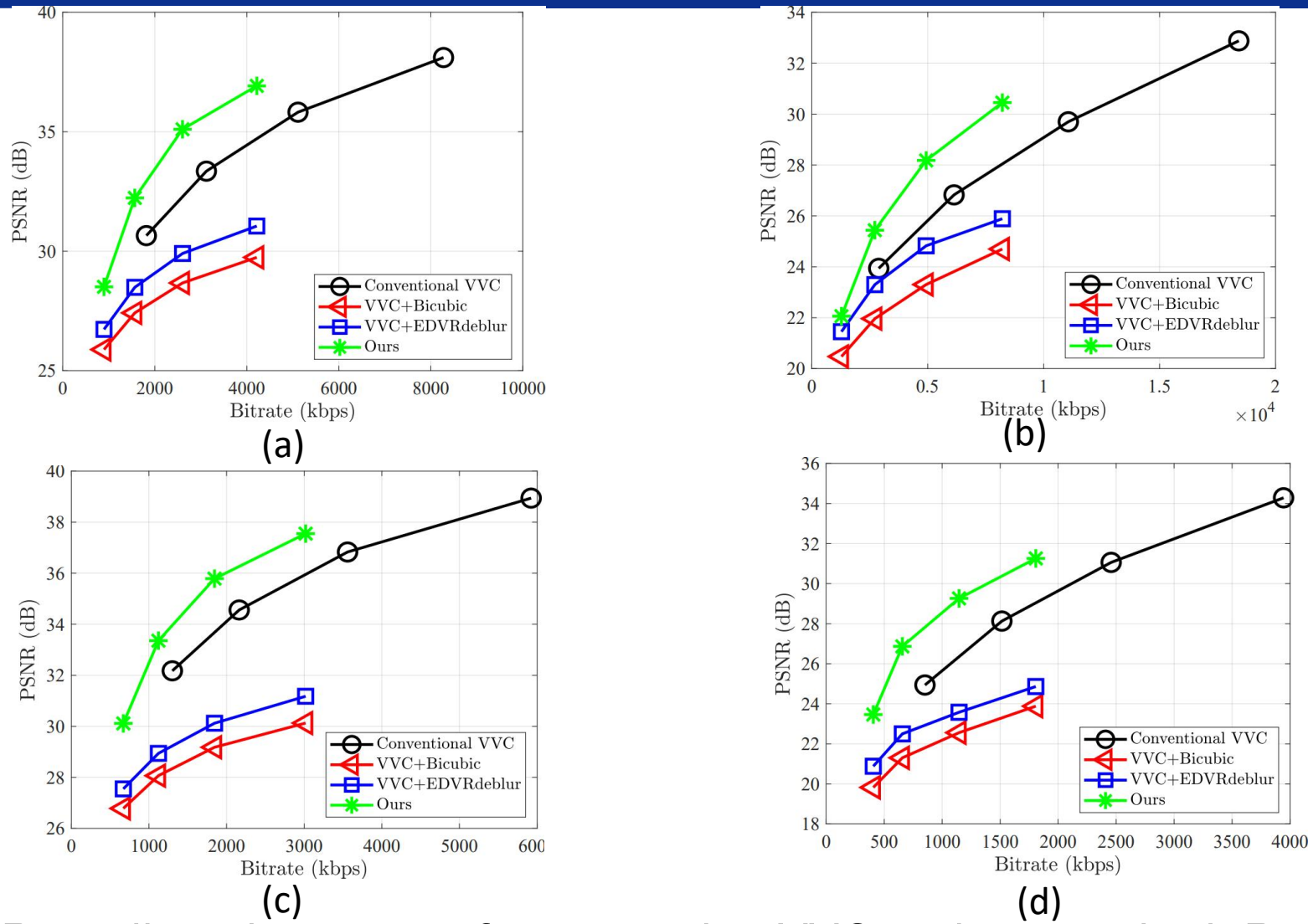


Figure 2. Rate-distortion curves for comparing VVC and our method. Results for sequences compressed by All Intra: (a) Fourpeople; (b) Partyscene; (c) KristenAndSara; (d) BQSquare.

Experimental Results

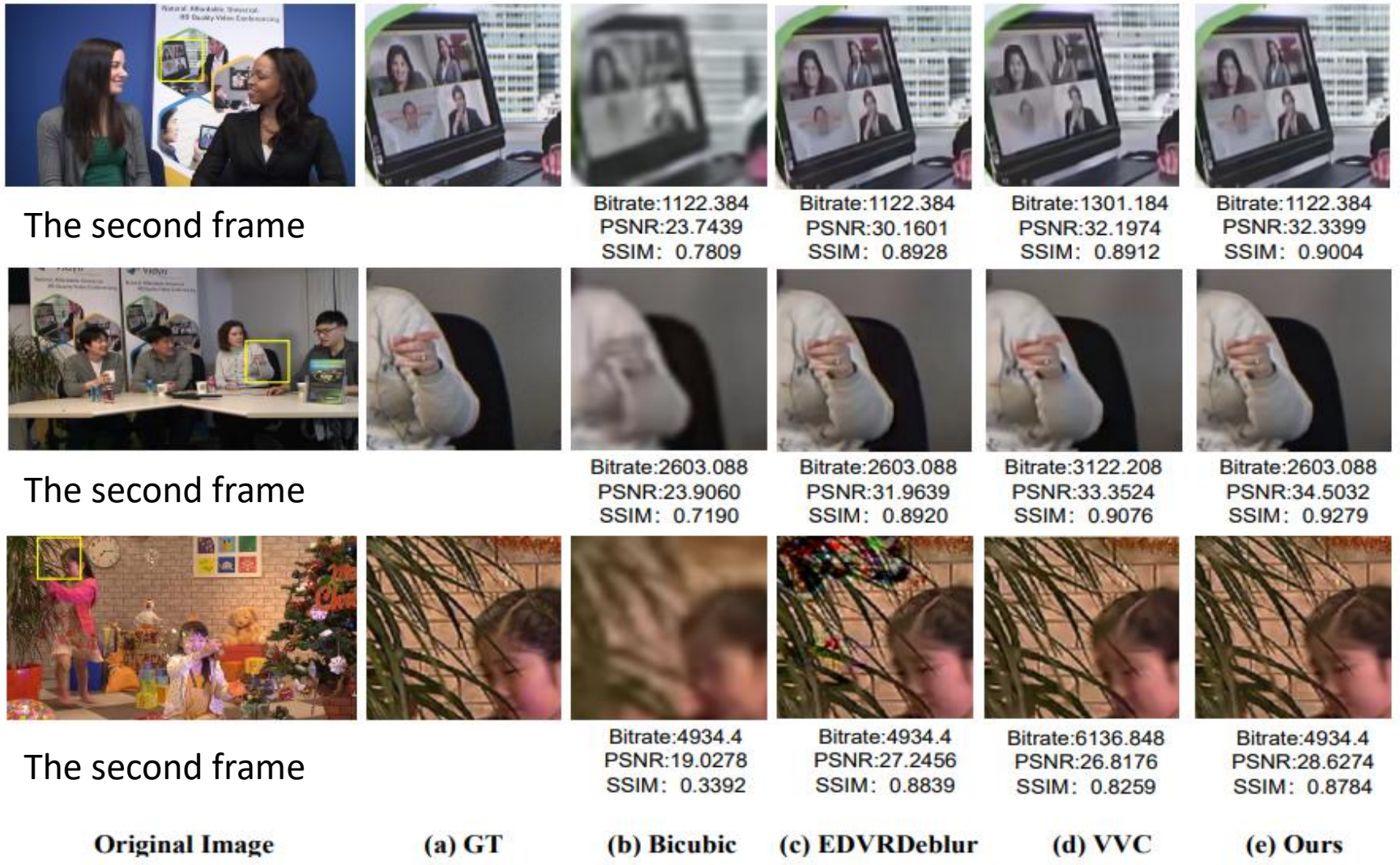


Figure 3. Subjective results comparisons between the proposed method and the other three methods in PSNR (dB)/SSIM. Results for sequences: (1) KristenAndSara; (2) Fourpeople; (c) PartyScene.

Experimental Results

Resolution	Sequence	Low Delay P		All Intra	
		BD-rate	BD-psnr	BD-rate	BD-psnr
1920x1080	Cactus	-60.47	0.468	-59.803	1.933
	BQTerrance	-39.261	0.543	-77.574	3.776
832x480	BQMall	-61.248	2.115	-75.61	5.102
	partyScene	-82.764	4.705	-85.455	7.238
416x240	BQSquare	-58.396	2.332	-80.975	7.677
1280x720	KristenAndSara	-65.751	4.517	-68.7	5.296
	Fourpeople	-58.582	0.679	-73.889	6.673
	Johnny	-61.863	0.679	-63.591	3.956
Average		-61.041875	2.00475	-73.199625	5.206375

Table 1. Objective comparison between our proposed method using AI configuration and the standard VVC in QPs = {22,27,32,37,42,47}. Ours outperforms VVC in the average BD-rate, BD-psnr. The table shows the results of the calculations for middle down-sampled frames.

Conclusion

Ⅳ Conclusion

- We design a **temporal down-sampling based video coding framework** (TDS).It can be combined with all the existing coding standards.

- A method of **super-resolution with frame recurrent image enhancement** (SRFR) is applied to improve the downsampled frames. The temporal information from high resolution frames can be fully used to improve the video quality through frame recurrent.

- When compared to VVC, the BD-rate of the down-sampled frames can be reduced from **39.261% ~ 85.455 %** in AL and LDP configuration.

Thank you

For your listening

法政大学
HOSEI University