3D-CVQE：An Effective 3D-CNN Quality Enhancement for Compressed Video Using Limited Coding Information

Xuan Sun, Pengyu Liu, Kebin Jia and Congcong Wang

Reporter：Xuan Sun

catalog

01 Introduction

02 Research Contents

03 Experiments

01 Introduction

1 Introduction

Background

The current encoding framework is still block-based. Due to crude quantification and inaccurate motion compensation techniques, some high-frequency information is lost in the encoding process. The blurring of content edges and significant compression distortion will bring great negative effects on subjective quality with limited coding resources.

Therefore, it is very necessary to research a quality enhancement method under the same coding resources.

Contribution

This paper proposes the 3D-CNN Compressed Video Quality Enhancement (3D-CVQE) network for enhancing the quality of the compressed video. The main **contributions** of this paper are:

1. A dataset is established with 148 pairs of YUV videos. According to the dataset, the “quality fluctuation”, “pixel missing”, and “position fluctuation” characteristics of compressed video are found by rethinking the video coding process.

Contribution

2. Based on the characteristics of “**quality fluctuation**” and “**position fluctuation**”, this paper proposes the use of **non-aligned networks** with multi-frame input to solve the reconstruction problem of compressed video.

3. Based on the characteristic of “**pixel missing**”, this paper proposes to treat the compressed video quality enhancement task **as a special video super-resolution task** with no increase in resolution. It is demonstrated that the average PSNR of 18 HEVC standard sequences is enhanced **0.4652 dB** and the number of parameters remains at **1.98 million**.

02 Research Contents

Dataset preparation

In order to analyze the characteristics of compressed video systematically, this paper prepares **148** pairs of YUV video sequences to establish the dataset.

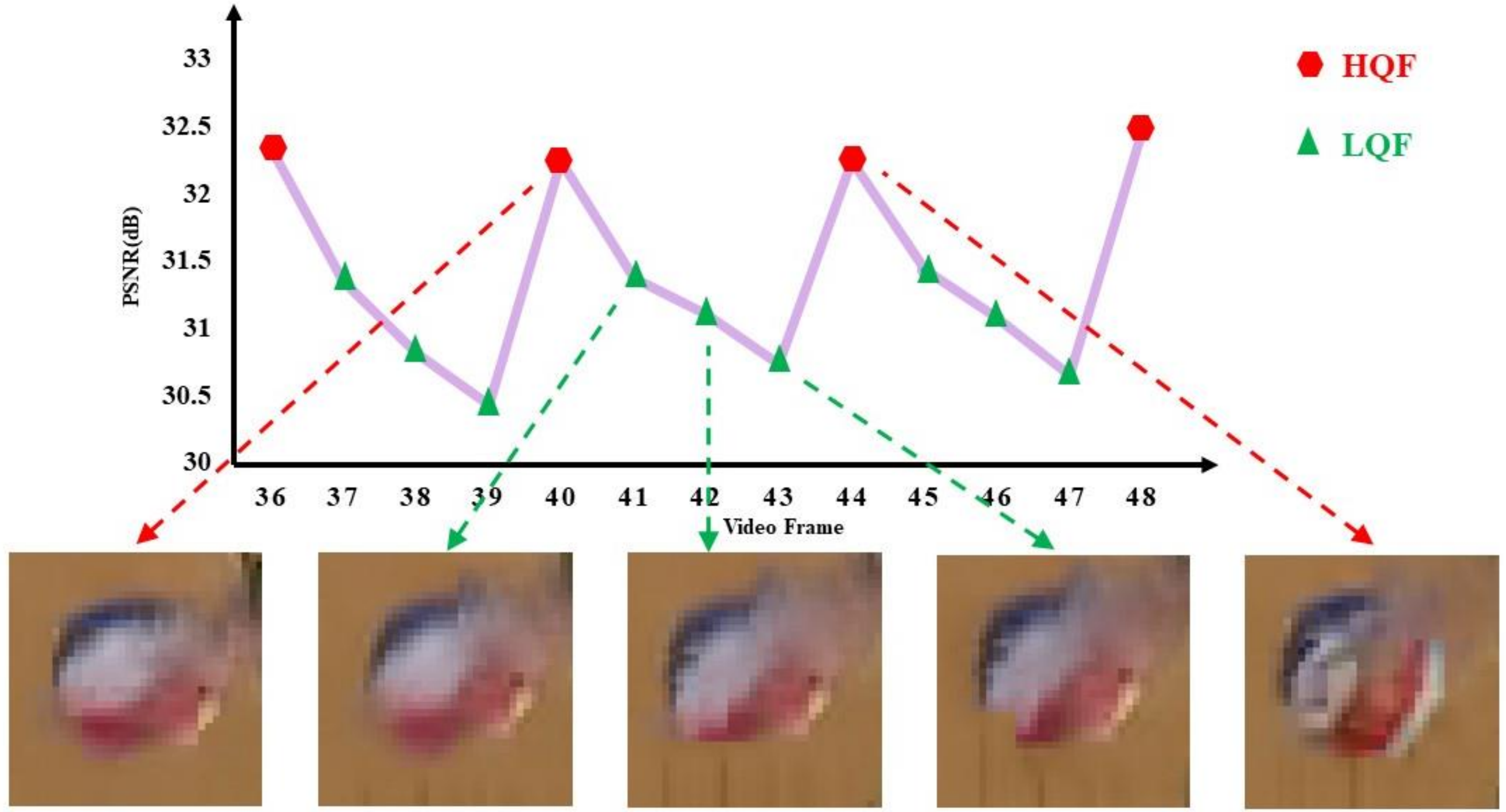
The video sequences contained in the dataset are at large range of **resolutions**: CIF (352 × 288), 4CIF (704 × 576), 240p (416 × 240), 360p (640 × 360), 480p (832 × 480), 720p (1280 × 720), 1080p (1920 × 1080), and WQXGA (2560 × 1600).

First characteristic: quality fluctuation

For Low-Delay (LD) configuration, although the video is compressed as QP is set to 37, actually every frame of compressed video has a QP offset. There are high-quality compressed frames (HQFs) at QP 38 and low-quality compressed frames (LQFs) at QP 39 and 40 in the compressed video. Different compression configurations determine real QP of each frame which is correlated with the quality of the frame.

Therefore, it is feasible to improve the quality of LQFs with the large amount of valid information in HQFs.

First characteristic: quality fluctuation



Frame 40: HQF
QP: 38
PSNR: 32.24 dB

Frame 41: LQF
QP: 40
PSNR: 31.37 dB

Frame 42: LQF
QP: 39
PSNR: 31.10 dB

Frame 43: LQF
QP: 40
PSNR: 30.73 dB

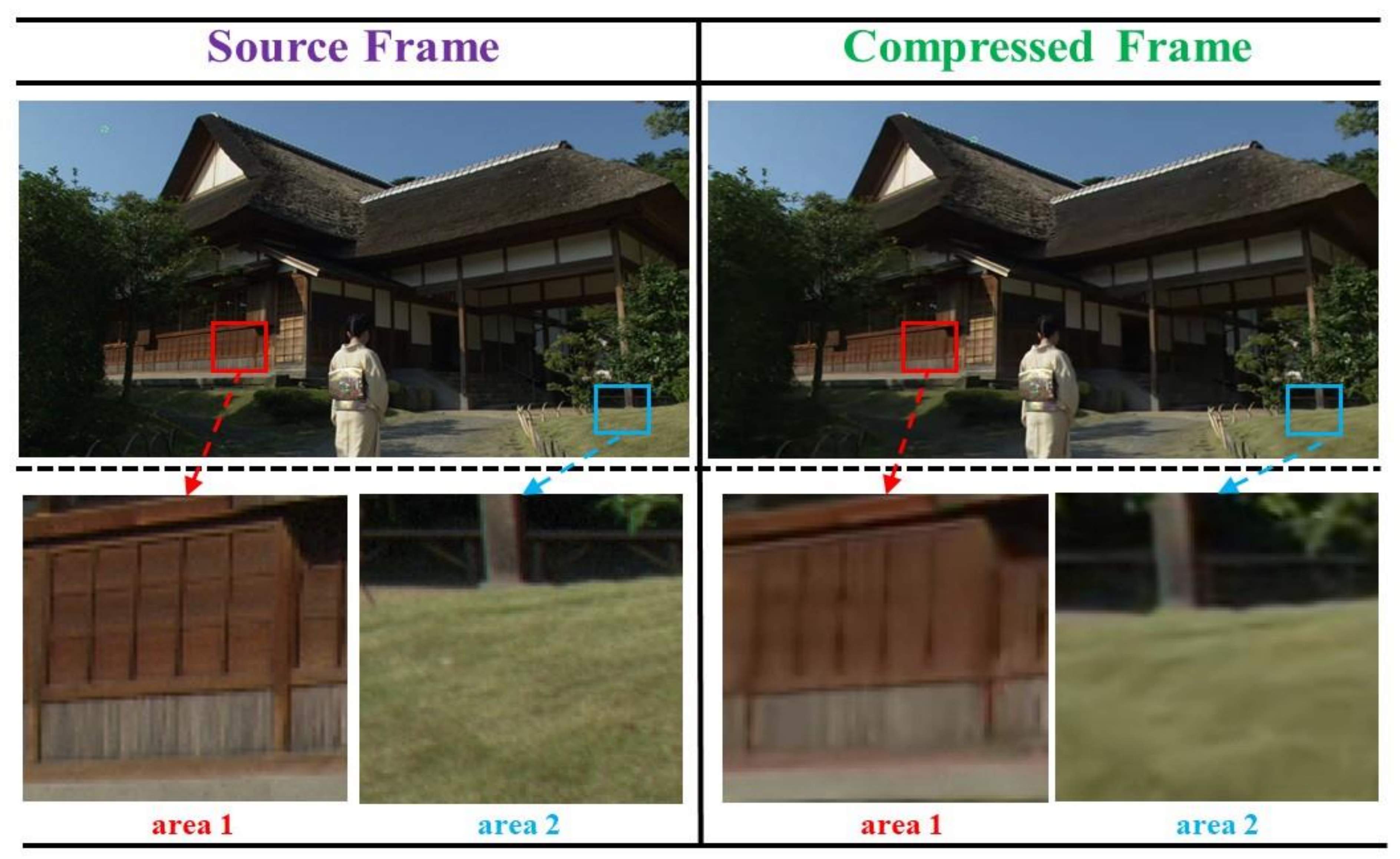
Frame 44: HQF
QP: 38
PSNR: 32.24 dB

Second characteristic: pixel missing

An important part of video coding is to find the optimal pixel block which has the highest correlation with the source pixel block in the current and nearby frames. With limited coding resources, the final selection of the optimal block may not be particularly similar to the source block.

The lack of valid pixel information makes it difficult to reconstruct the video, and a large number of blurred regions and artifacts makes it difficult for the network to learn features. For this reason, the new network has to focus on how to efficiently retain and utilize the limited source information.

Second characteristic: pixel missing

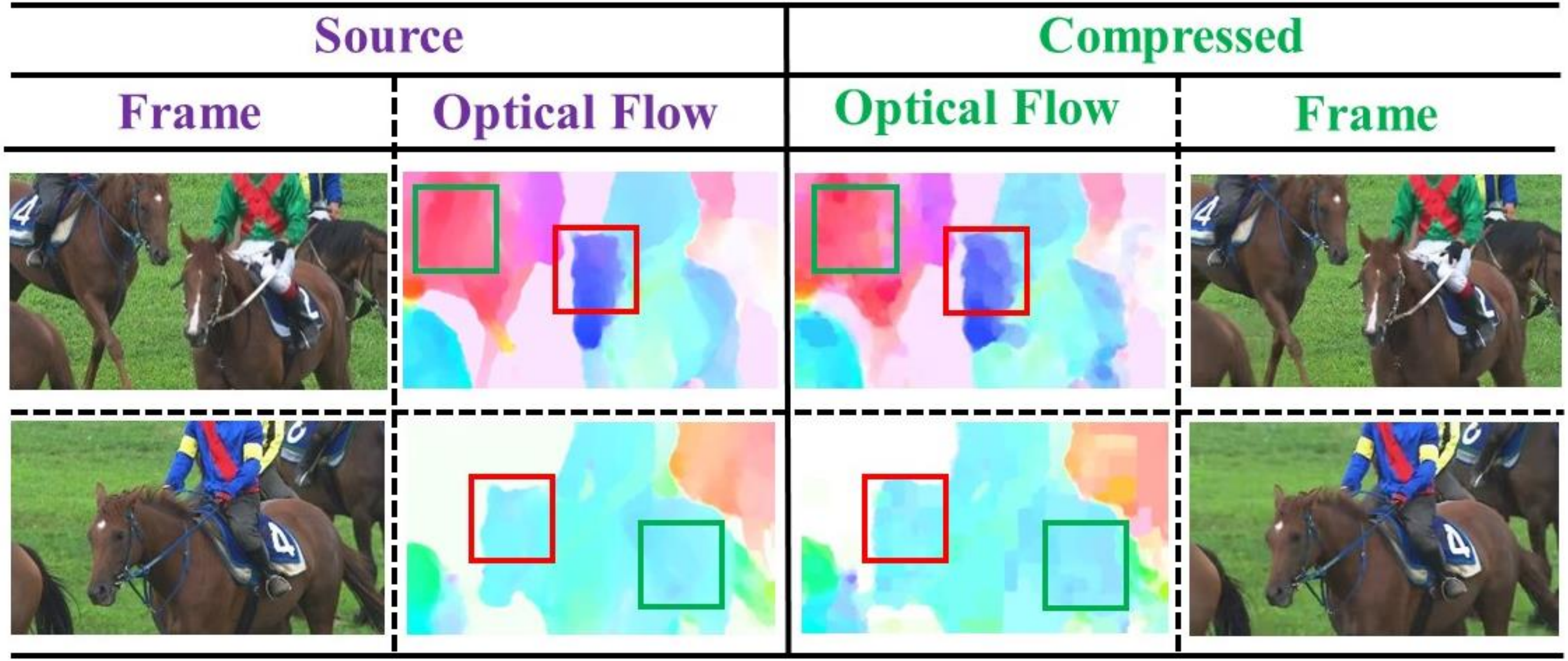


Third characteristic: position fluctuation

Though the coding has tried to find the optimal block, there is a certain difference in texture between the optimal block and the source block. There are the optical flow frames and its corresponding.

Therefore, frame alignment is difficult to be carried out with such drastic position fluctuation in compressed video, so it is proposed to solve the alignment problem of compressed video implicitly by using non-aligned networks.

Third characteristic: position fluctuation



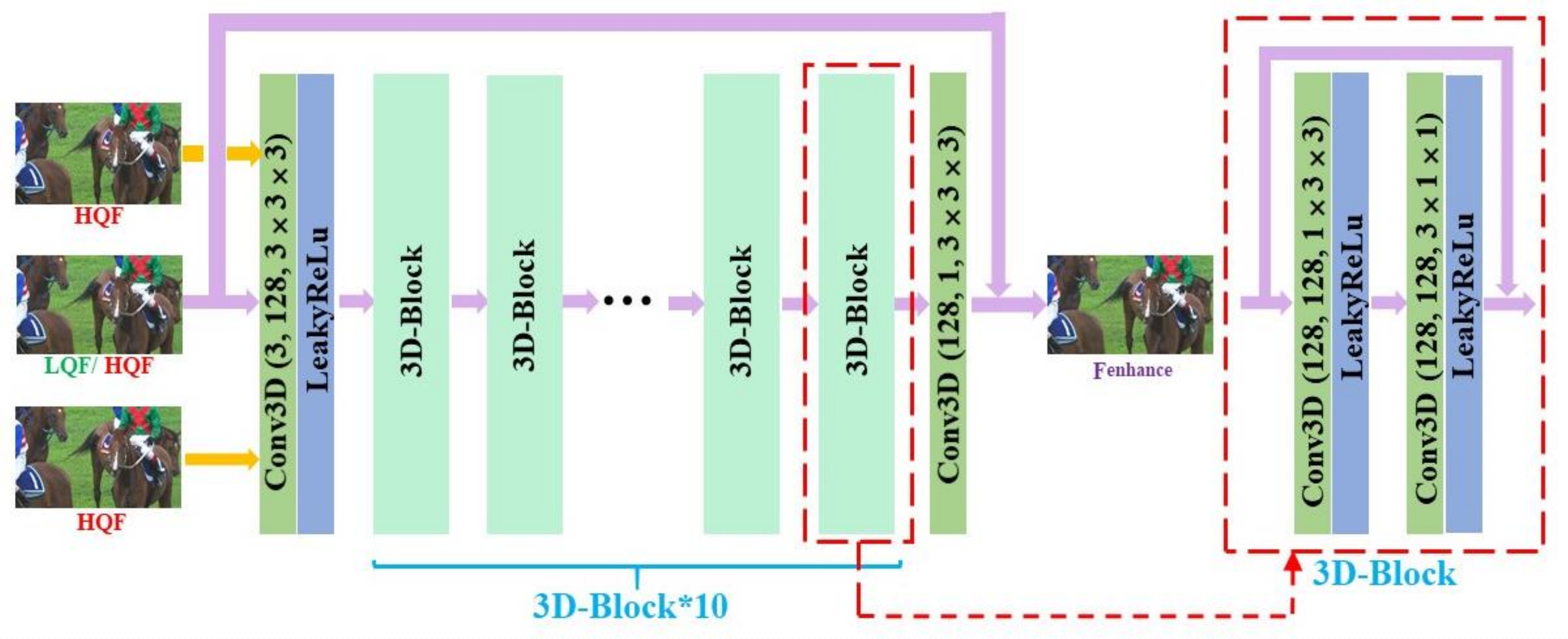
Network Architecture

3D-CNNs have been successfully implemented in high level vision tasks for videos such as action recognition and event classification. They are also effectively applicable to a low-level vision task for videos such as compressed video quality enhancement.

In this paper, 3D-CVQE approach is designed with the 3D-CNN which makes motion alignment not necessary thanks to its spatial-temporal feature representation ability.

2 Research Contents

The whole framework


HQF
LQF/ HQF
HQF
Conv3D (3, 128, 3 × 3 × 3)
LeakyReLu
3D-Block
3D-Block
• • •
3D-Block
3D-Block
Conv3D (128, 1, 3 × 3 × 3)
Fenhance
Conv3D (128, 128, 1 × 3 × 3)
LeakyReLu
Conv3D (128, 128, 3 × 1 × 1)
LeakyReLu
3D-Block
3D-Block*10

The design of 3D-Block

The input of the framework is three compressed frames: one LQF and two HQFs closed to it. Similarly, each HQF is also enhanced with the help of its nearest HQFs.

Since 3D-CNNs have an inherent flaw that is its huge number of parameters, the 3D-Block is proposed to solve this problem. Each 3D-Block consists of one 1×3×3 convolutional filter on spatial domain and one 3×1×1 convolutional on temporal domain.

03 Experiments

3 Experiments

configuration

The framework is implemented on PyTorch. For fairness, all experiments are conducted on the same dataset with the same training configuration. The experiments are conducted on a PC with Intel Xeon E5 CPU and Nvidia GeForce GTX 1080Ti GPU.

18 standard test sequences of JCT-VC are test set. Other 130 sequences are randomly divided into training set (100 sequences) and validation set (30 sequences).

Performance of 3D-CVQE approach

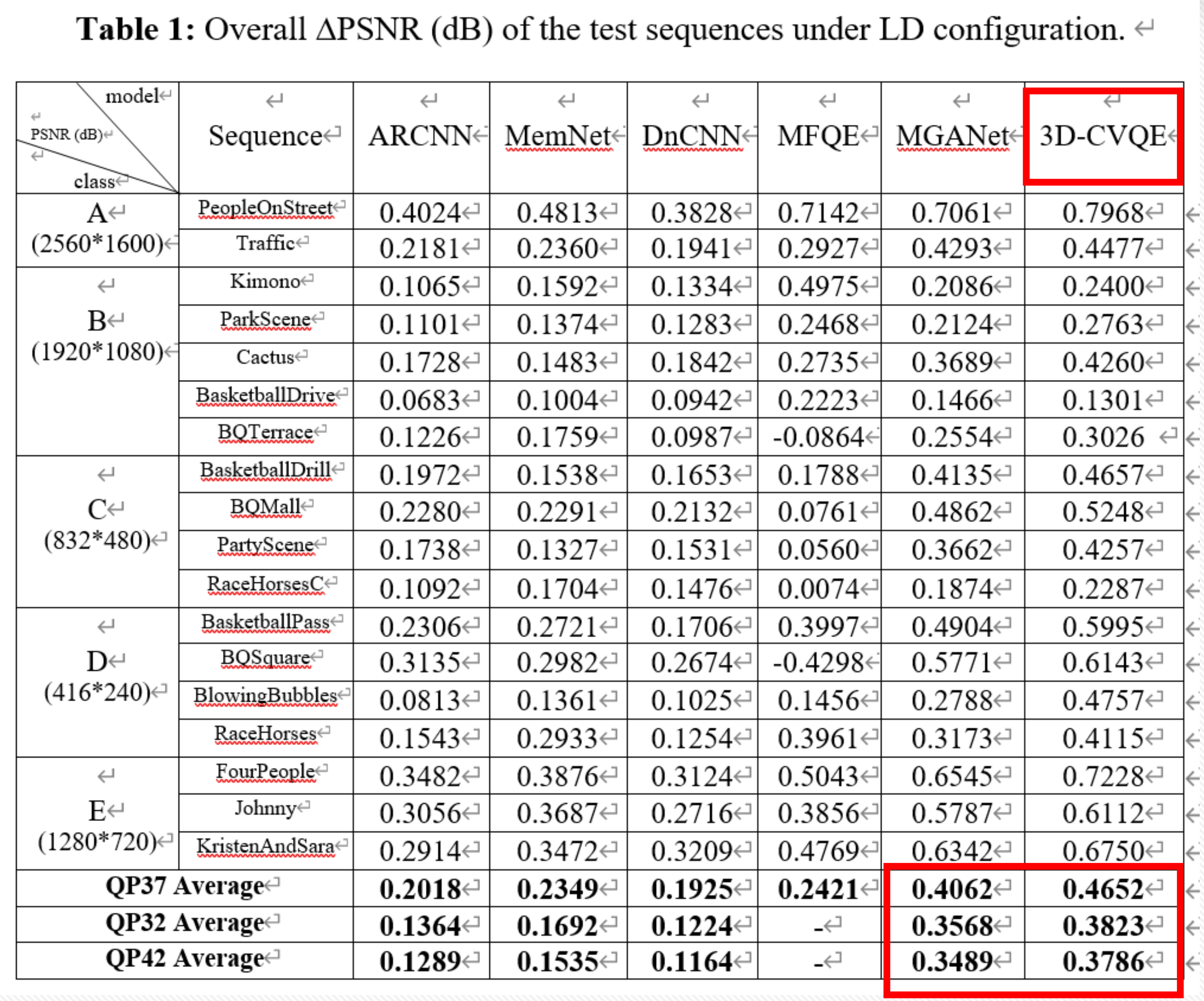
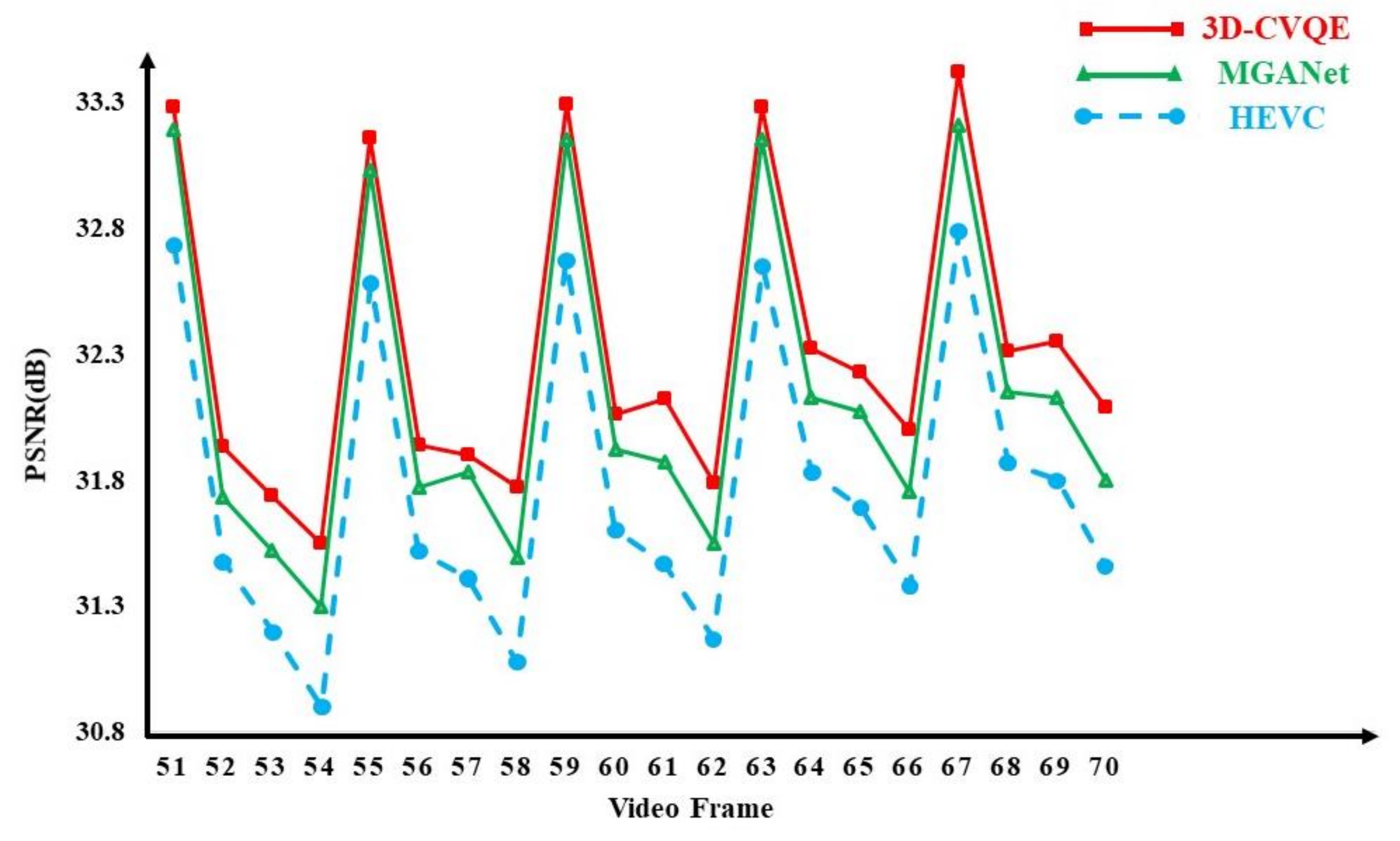


Table 1: Overall ΔPSNR (dB) of the test sequences under LD configuration.

model PSNR (dB) class	Sequence	ARCNN	MemNet	DnCNN	MFQE	MGANet	3D-CVQE
A (2560*1600)	PeopleOnStreet	0.4024	0.4813	0.3828	0.7142	0.7061	0.7968
	Traffic	0.2181	0.2360	0.1941	0.2927	0.4293	0.4477
B (1920*1080)	Kimono	0.1065	0.1592	0.1334	0.4975	0.2086	0.2400
	ParkScene	0.1101	0.1374	0.1283	0.2468	0.2124	0.2763
	Cactus	0.1728	0.1483	0.1842	0.2735	0.3689	0.4260
	BasketballDrive	0.0683	0.1004	0.0942	0.2223	0.1466	0.1301
	BQTerrace	0.1226	0.1759	0.0987	-0.0864	0.2554	0.3026
C (832*480)	BasketballDrill	0.1972	0.1538	0.1653	0.1788	0.4135	0.4657
	BQMall	0.2280	0.2291	0.2132	0.0761	0.4862	0.5248
	PartyScene	0.1738	0.1327	0.1531	0.0560	0.3662	0.4257
	RaceHorsesC	0.1092	0.1704	0.1476	0.0074	0.1874	0.2287
D (416*240)	BasketballPass	0.2306	0.2721	0.1706	0.3997	0.4904	0.5995
	BQSquare	0.3135	0.2982	0.2674	-0.4298	0.5771	0.6143
	BlowingBubbles	0.0813	0.1361	0.1025	0.1456	0.2788	0.4757
	RaceHorses	0.1543	0.2933	0.1254	0.3961	0.3173	0.4115
E (1280*720)	FourPeople	0.3482	0.3876	0.3124	0.5043	0.6545	0.7228
	Johnny	0.3056	0.3687	0.2716	0.3856	0.5787	0.6112
	KristenAndSara	0.2914	0.3472	0.3209	0.4769	0.6342	0.6750
QP37 Average		**0.2018**	**0.2349**	**0.1925**	**0.2421**	**0.4062**	**0.4652**
QP32 Average		**0.1364**	**0.1692**	**0.1224**	-	**0.3568**	**0.3823**
QP42 Average		**0.1289**	**0.1535**	**0.1164**	-	**0.3489**	**0.3786**

3 Experiments

Improvement of quality fluctuation



It can be seen that the PSNR fluctuation of 3D-CVQE approach is significantly smaller than the HEVC baseline. In summary, 3D-CVQE approach is also capable of reducing the quality fluctuation of video compression.

Conclusions

This paper proposes 3D-CVQE approach with multi-frame input to enhance the quality of compressed video by reducing compression artifacts. The proposed method focuses on the essential relationship between video coding and compressed video quality enhancement tasks, and finds the unique features of compressed video by reflecting on the video coding process.

Based on these characteristics, a quality enhancement network applicable to compressed video is proposed. This paper opens up new space for future exploration to use temporal and spatial information for quality enhancement of compressed video.

Thank you!