Long-distance Information Filtering Network for Compressed Video Quality Enhancement

Xuan Sun*, Ziyue Zhang*, Guannan Chen* and Dan Zhu*
*BOE Technology Group Co., LTD.
sunxuan@boe.com.cn

Introduction

A Frame of Kimono Sequence

Raw ***Compressed*** ***Raw*** ***Compressed***

- Restoring high-quality videos from low-quality compressed ones is an important topic in video coding.

- Most existing methods do not exploit the information in the long-distance compressed frames. Even when they do, these methods ignore the effect of interference information.

- Hence, we need to study compressed video quality enhancement (VQE) tasks.

Contribution

- In this paper, we propose a novel Long-distance Information Filtering (LIF) scheme to accomplish the VQE task. The main idea of LIF is to fully exploit the filtered and valid clues from long-distance frames to guide video reconstruction.

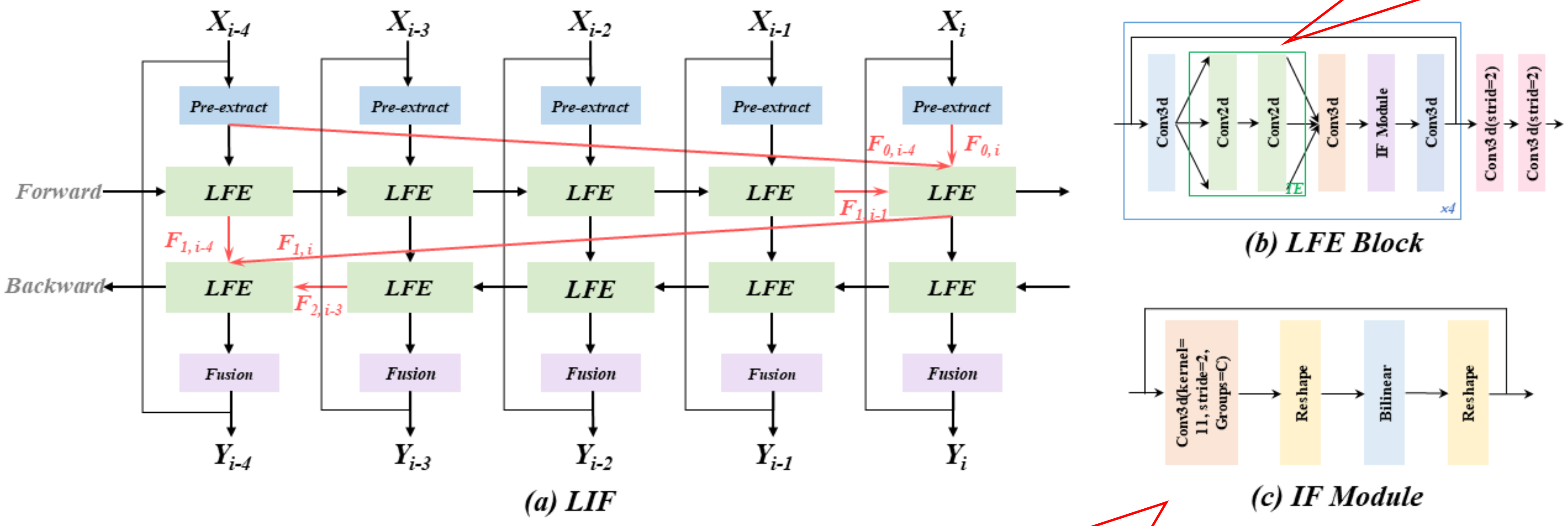
- Specifically, we first develop a Long-distance Feature Extraction (LFE) block to capture the most valid context. This block mainly consists of concise 3D-convolutional layers without utilizing an optical flow prediction module and deformable convolutional layers, making it friendly to implement on hardware devices.

- Furthermore, we design an Information Filtering (IF) module to filter the spatio-temporal information. This module uses large convolutional kernels to model effective local spatial relationships.

Methodology

Two propagation approach:
Extract long-distance temporal information

TE:
Split the feature maps into three parts in the temporal dimension and feed them into 2D convolutional layers, which helps enhance the texture



(a) LIF

(b) LFE Block

(c) IF Module

IF:
Model a local spatial relationship via the large 11×11 kernel to remove interferential information

Experiments

Table 1: Quantitative results of ΔPSNR (dB) / ΔSSIM ($\times 10^{-2}$) on test videos at 4 QPs

Class	Sequence	MRDN	STDF-R3	STDF-R3L	RFDA	Ours
A (2560*1600)	PeopleOnStreet	1.23/1.19	1.18/1.82	1.25/1.96	1.44/2.22	1.33/2.16
	Traffic	0.72/1.16	0.65/1.04	0.73/1.15	0.80/1.28	**0.84/1.33**
B (1920*1080)	Kimono	0.82/1.65	0.77/1.47	0.85/1.61	1.02/1.86	0.96/1.80
	ParkScene	0.60/1.54	0.54/1.32	0.59/1.47	0.64/1.58	**0.68/1.70**
	Cactus	0.67/1.30	0.70/1.23	0.77/1.38	0.83/1.49	0.78/1.45
	BQTerrace	0.55/0.97	0.58/0.93	0.63/1.06	0.65/1.06	0.61/ 0.99
	BasketballDrive	0.71/1.25	0.66/1.07	0.75/1.23	0.87/1.40	0.81/1.35
C (832*480)	RaceHorses	0.60/1.48	0.48/1.09	0.55/1.35	0.48/1.23	0.52/**1.35**
	BQMall	0.90/1.73	0.90/1.61	0.99/1.80	1.09/1.97	1.04/1.96
	PartyScene	0.55/1.66	0.60/1.60	0.68/1.94	0.66/1.88	**0.79/2.22**
	BasketballDrill	0.73/1.54	0.70/1.26	0.79/1.49	0.88/1.67	**0.89**/1.65
D (416*240)	RaceHorses	0.83/2.09	0.73/1.75	0.83/2.08	0.85/2.11	0.81/2.10
	BQSquare	0.75/1.07	0.91/1.13	0.94/1.25	1.05/1.39	**1.20/1.63**
	BlowingBubbles	0.66/2.08	0.68/1.96	0.74/2.26	0.78/2.40	**0.93/2.87**
	BasketballPass	0.98/2.03	0.95/1.82	1.08/2.12	1.12/2.23	**1.15/2.32**
E (1280*720)	FourPeople	0.94/1.18	0.92/1.07	0.94/1.17	1.13/1.36	**1.15/1.38**
	Johnny	0.75/0.78	0.69/0.73	0.81/0.88	0.90/0.94	**0.96/1.08**
	KristenAndSara	0.95/1.01	0.94/0.89	0.97/0.96	1.19/1.15	1.14/**1.18**
LD	QP37 Average	0.78/1.47	0.75/1.32	0.83/1.51	0.91/1.62	**0.92/1.70**
	QP32 Average	0.81/1.02	0.73/0.87	0.86/1.04	0.87/1.07	**0.93/1.20**
	QP27 Average	0.83/0.72	0.67/0.53	0.72/0.57	0.82/0.68	**0.91/0.78**
	QP22 Average	0.70/0.40	0.57/0.30	0.63/0.34	0.76/0.42	**0.87/0.49**
LD	Total	0.78/0.90	0.68/0.76	0.76/0.87	0.84/0.95	**0.91/1.04**

- Training dataset: MFQE 2.0 dataset
- Testing dataset: 18 standard test sequences of JCT-VC
- Coding configuration: HM 16.2, LDP
- QP: 22, 27, 32, 37
- Training iterations: 150K

LIF outperforms the existing state of the arts on MFQE2.0 datasets at four QPs

Experiments

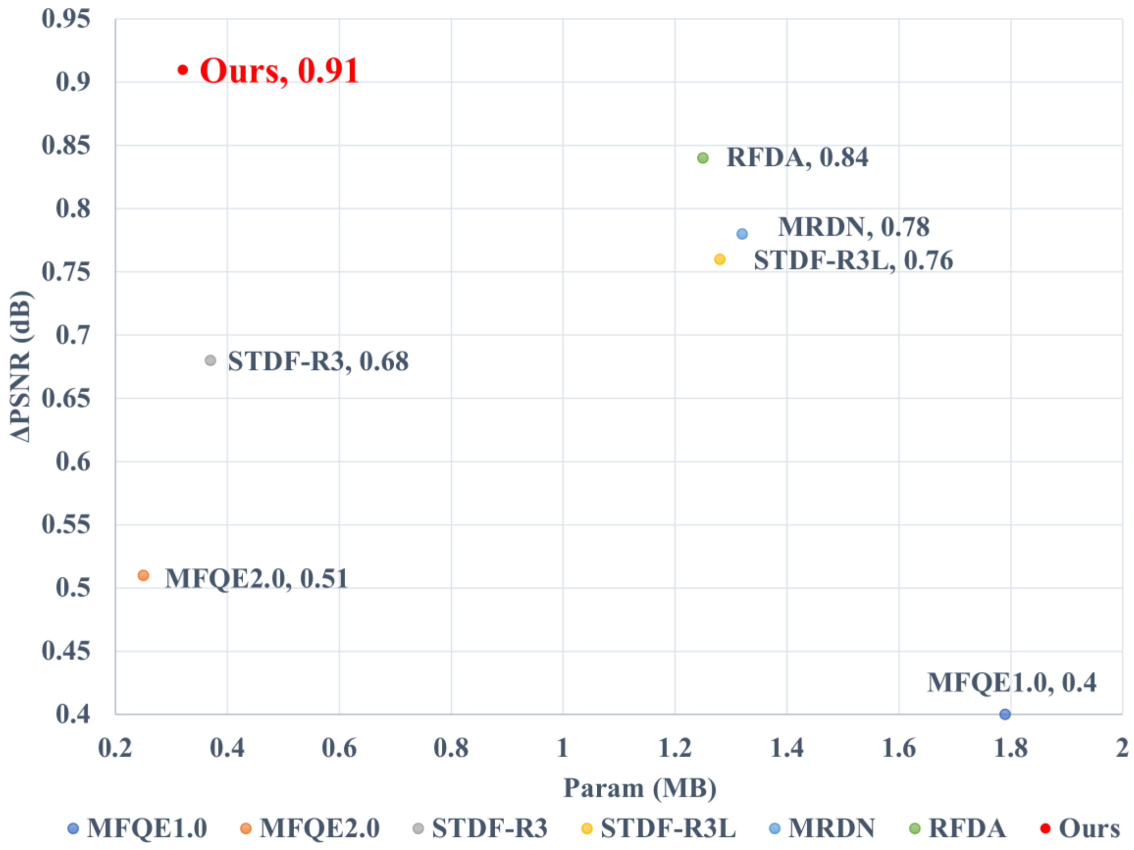


Figure 3: The parametres and performance comparison at 4 QPs.

- The parameters and performance comparison are provided in Figure 3. The specific value of each sequence in the table is measured when the QP is 37.
- In particular, LIF gets +33.82%/+0.23 dB higher ΔPSNR than STDF-R3 with fewer parameters.
- Compared with STDF-R3L, LIF achieves +19.74%/+0.15 dB higher ΔPSNR with 25.1% of the parameters.
- Moreover, LIF obtains +8.33%/+0.07dB higher ΔPSNR than RFDA with 25.6% of the parameters.

It is important to note that our training volume is nearly 50% of RFDA and STDF-R3L.

Experiments

Table 2: Impact of components

Method	Base	IF	TE	ΔPSNR(dB)	ΔSSIM($\times 10^{-2}$)
1	✓		✓	0.74	1.41
2	✓	✓		0.79	1.51
LIF	✓	✓	✓	0.92	1.70

We conduct several ablation studies to validate the effectiveness of LIF and the necessity of every proposed module.

We validate the necessity of the IF module and the TE module.

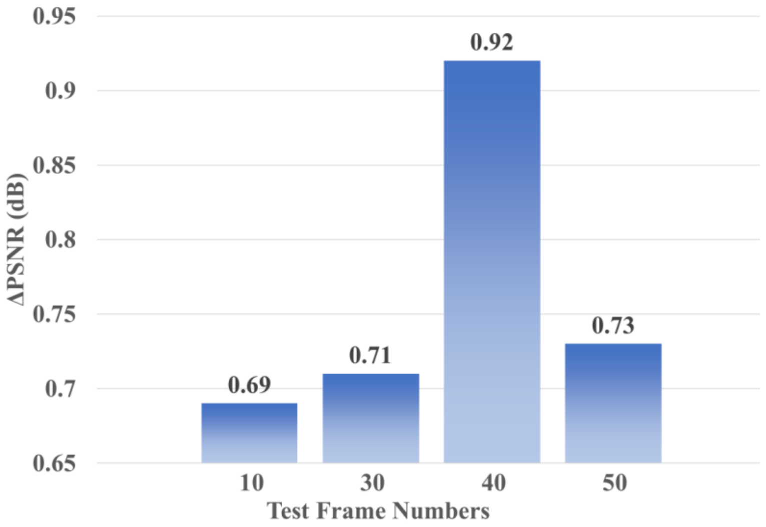


Figure 6: PSNR of Different Inferential Frame Numbers.

We use different numbers of consecutive frames as input to demonstrate the feature utilization ability of LIF for long-distance frames in Figure 6.

As we can see, with more frames used, the model can get better performance. Our LIF can take up to 40 frames.

Experiments

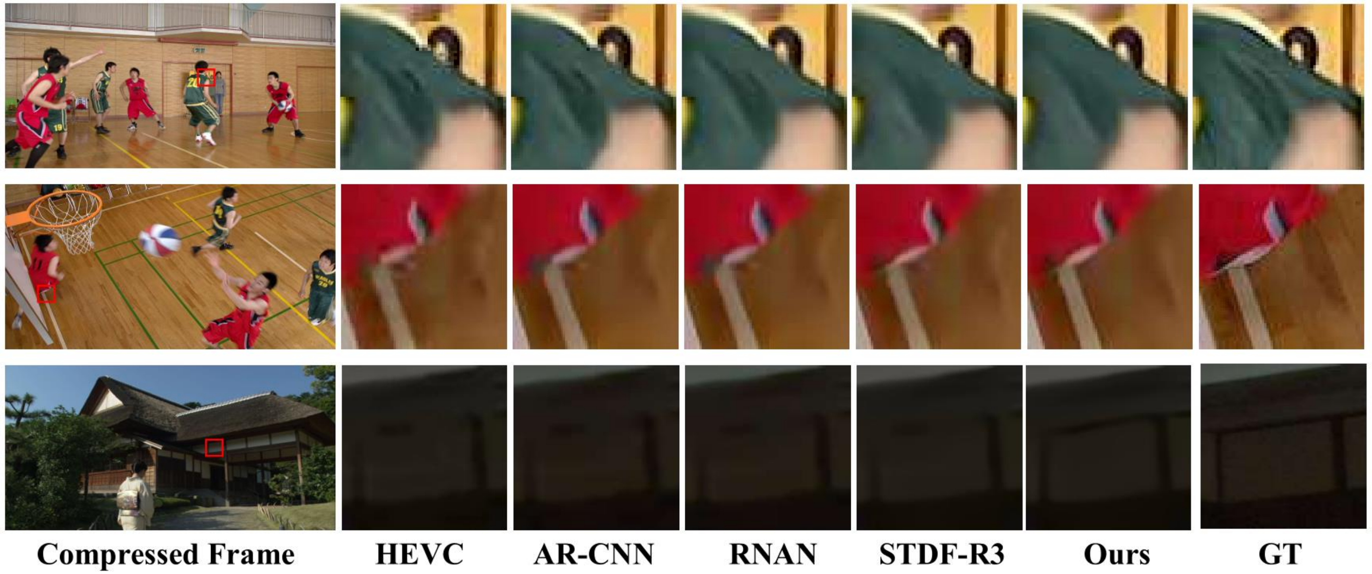


Figure 5: The Subjective performance.

- For instance, as shown in the last row in the figure, LIF successfully recovers more explicit window borders than others, which proves the effectiveness of using long-distance information.
- As shown in the first two rows in the figure, only LIF removes the additional artifacts on the shoulders and the floor, demonstrating the IF module's filtering capability.

Qualitative comparisons are shown in Figure 5. Our LIF can recover finer details and sharper edges than AR-CNN, RNAN, and STDF-R3.

Conclution

To summarize, our contributions are as follows:

- We propose a novel Long-distance Information Filtering (LIF) scheme to model the spatio-temporal dependency for the VQE task.
- We exhibit the experiments to understand the long-distance information flow between compressed frames and the effect of the large kernel on VQE.
- LIF demonstrates new state-of-the-art performance on the VQE benchmark dataset.

Thank you!