Video Enhancement Network Based on Max-pooling and Hierarchical Feature Fusion

Background

Video Enhancement

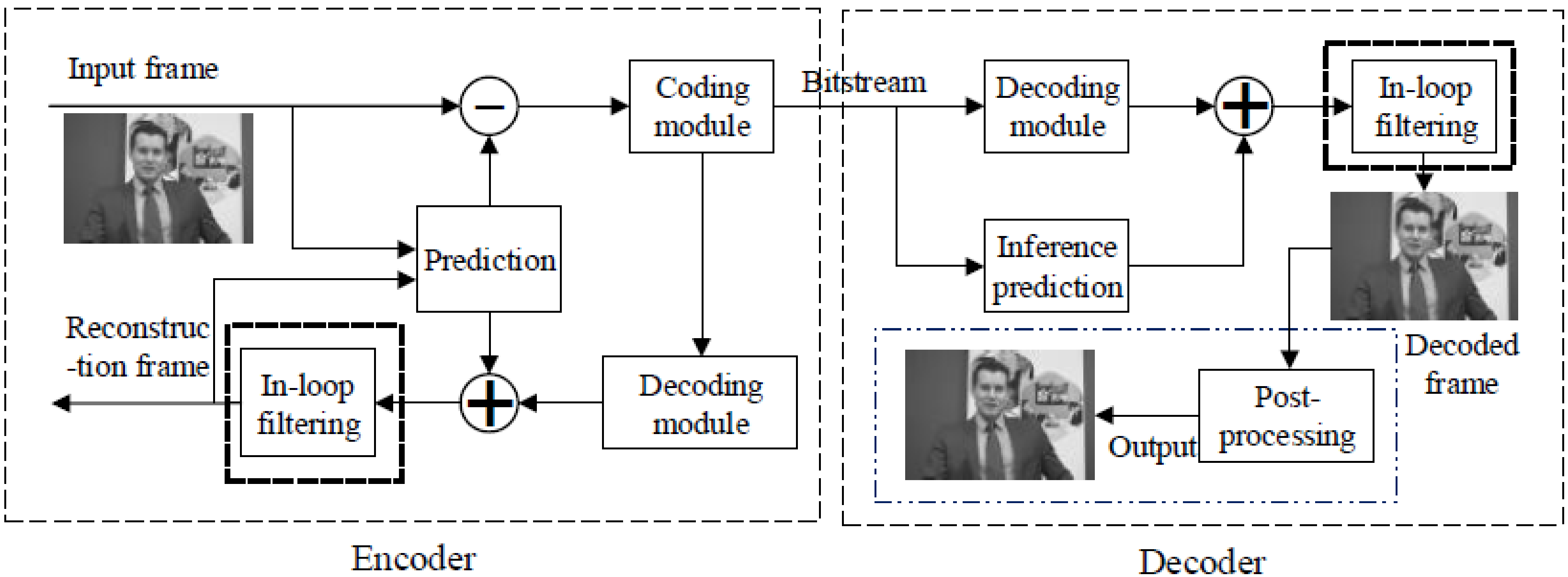


Figure 1: In-loop filtering and post-processing in Codec

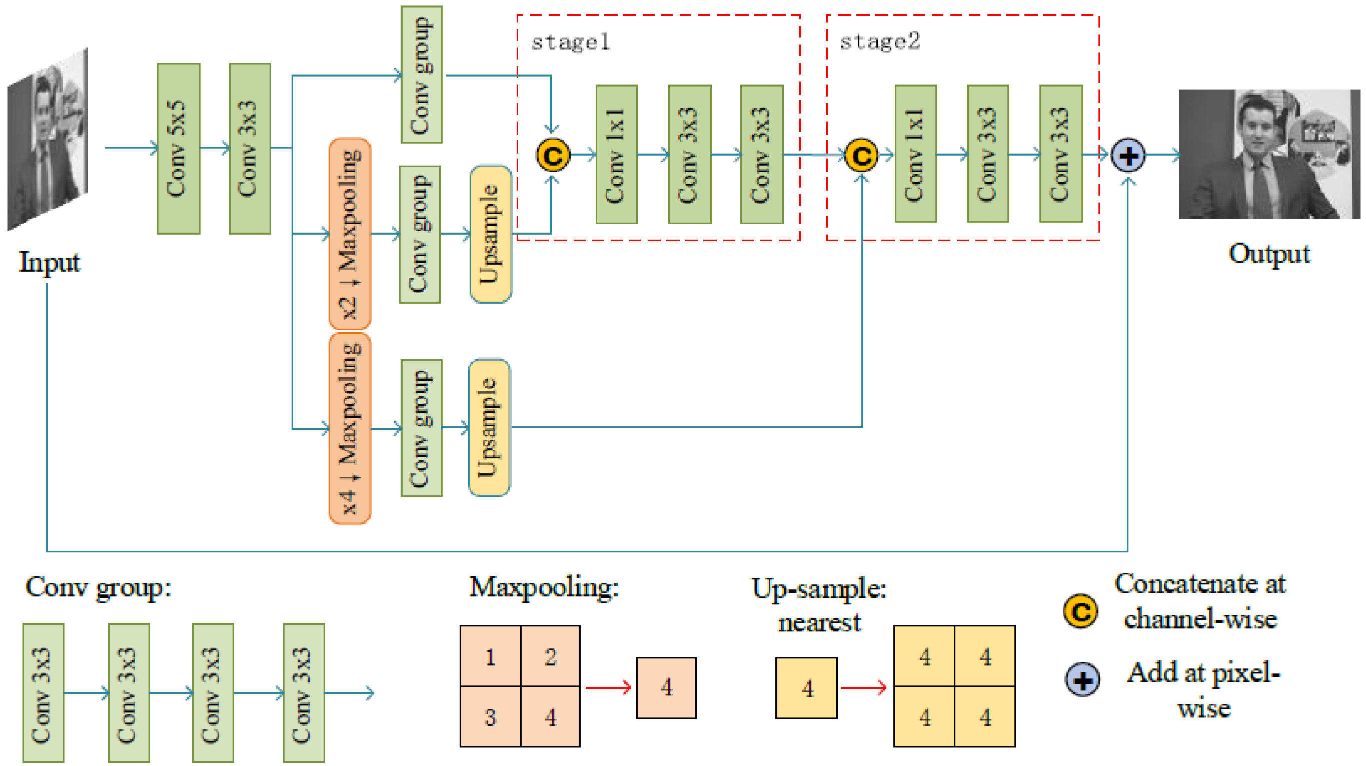
Background

Analysis(Learning-base Method)

Name	In-loop filtering	Post-processing
Related Work	IFCNN[1] RHCNN[2]	DCAD[3] DSCNN[4]
Analysis	It is stable and can be integrated into the encoder, but it will increase the complexity of the encoder and the encoding time.	It does not increase the complexity of the encoder, but does not make full use of the image feature information

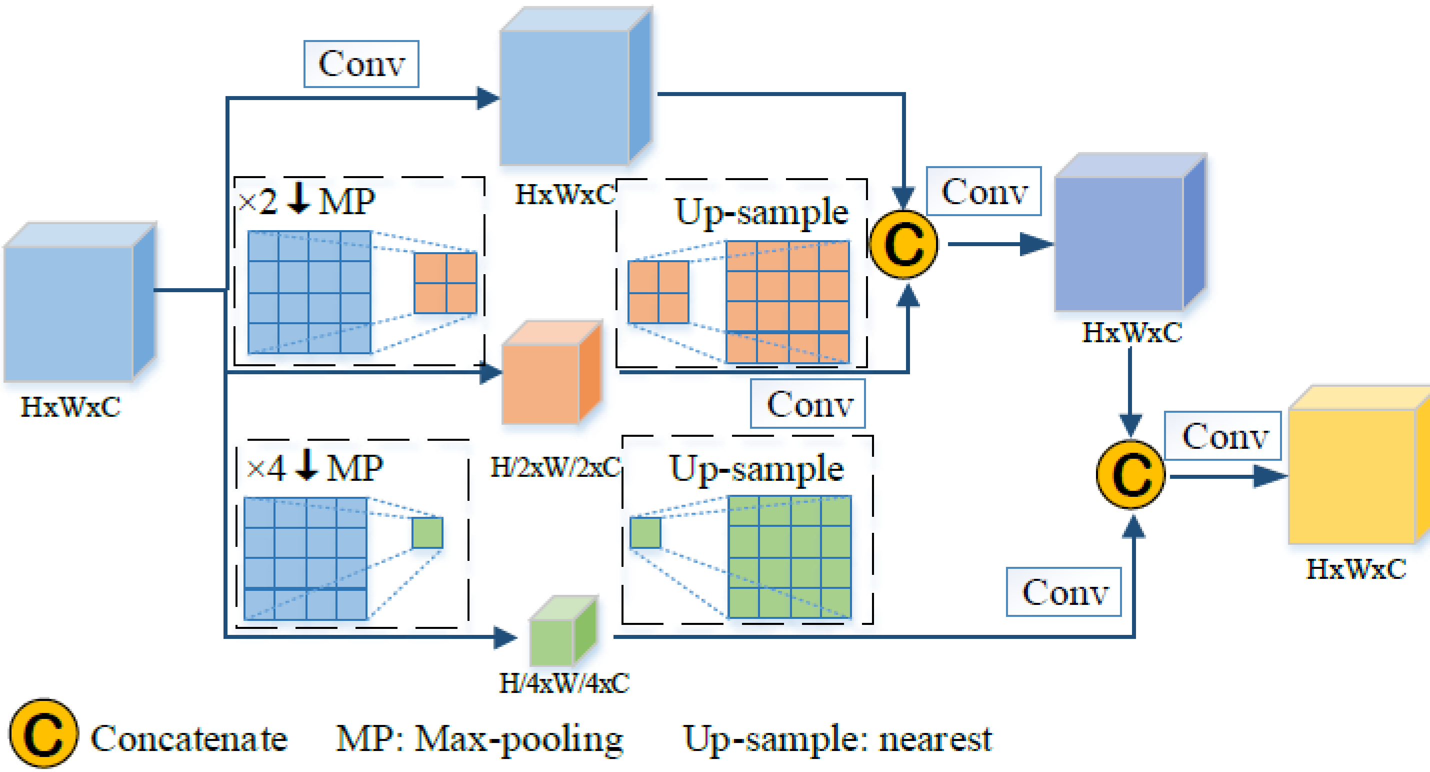
Proposed Method

Overview of the Network



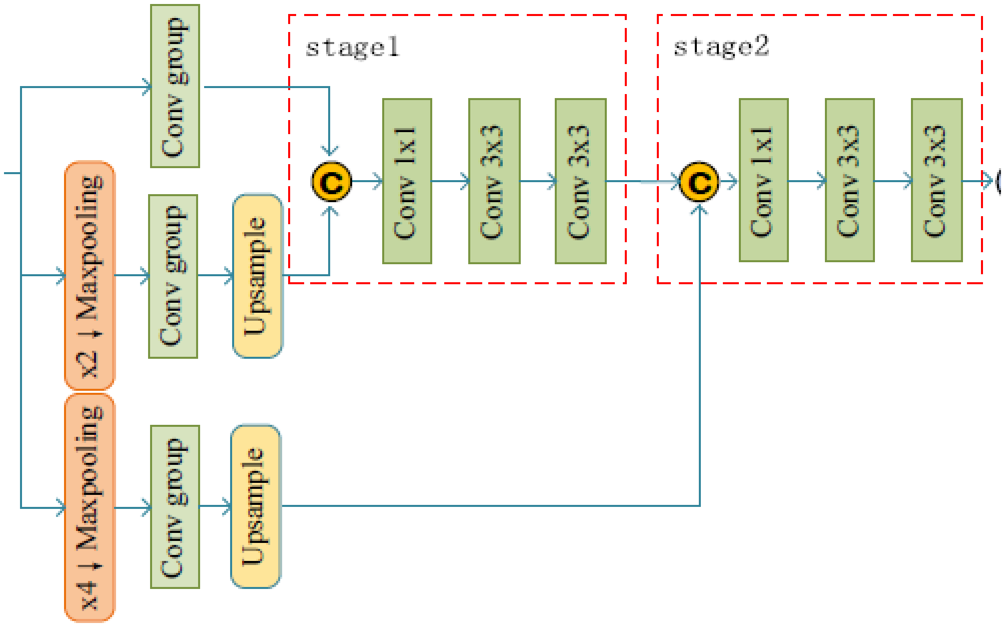
Proposed Method

Feature Extraction



Proposed Method

Hierarchical Feature Fusion



In order to make full use of different scale features, we design a hierarchical feature fusion module to fuse multi-scale features.(Stage 1 and stage 2 in the figure)

Proposed Method

Training

- The dataset in this paper comes from 16 HEVC test sequences. All the data are compressed by HEVC reference software (HM) with the configuration QP = 42, All Intra mode. We only use luminance component Y for training and testing.
- In training stage, we decomposed each compressed frame into 64x64 patches. In the testing stage, we take the whole frame as input. The purpose of this design of dataset is to obtain a large number of samples with dierent content features. At the same time, accelerating the convergence of the model in the training stage.
- We train the model with Adam optimizer, learning rate 1e-4, batch size 64, and mean square error as loss function. The deep learning framework is Pytorch 1.1.0.

Result

Compared with HEVC(PSNR)

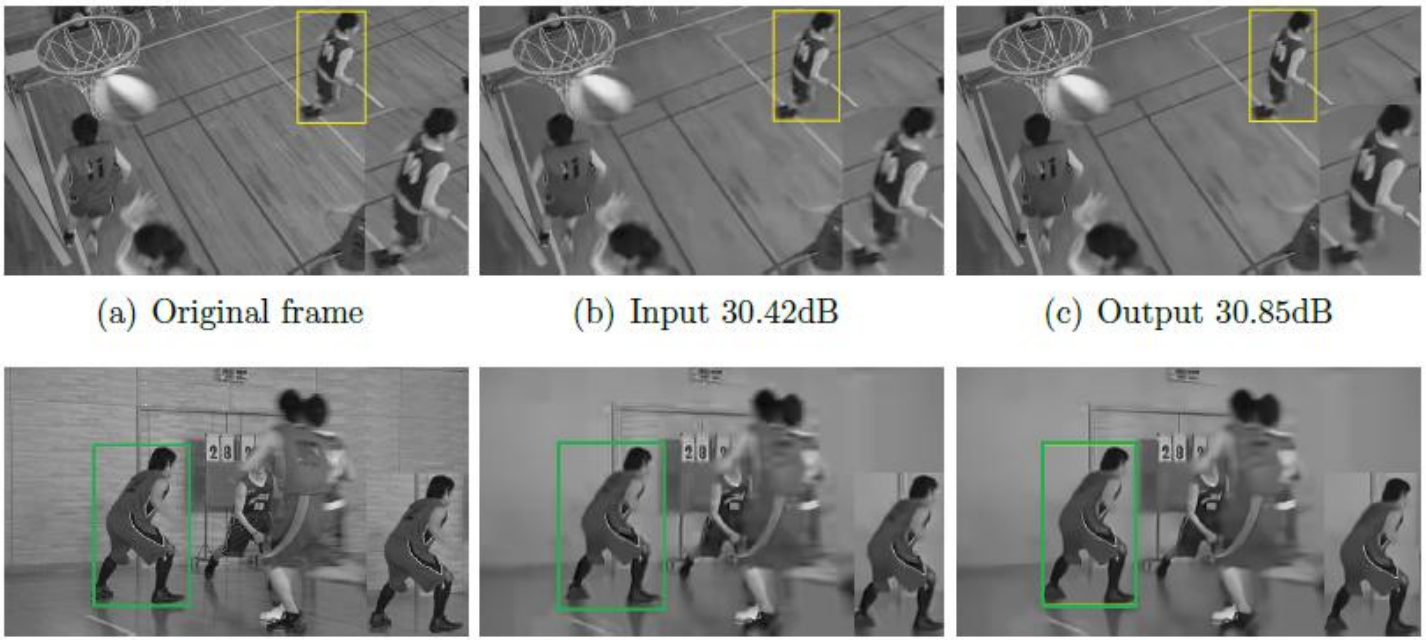
Table 2: Generalization Test Results

	BasketballDrill ΔPSNR/dB	BasketballPass ΔPSNR/dB	Johnny ΔPSNR/dB
VRCNN[20]	0.4104	0.2911	0.4734
DSCNN[13]	0.4033	0.2621	0.4988
DCAD[12]	0.4486	0.3208	0.5573
Ours	**0.4861**	**0.3415**	**0.5613**

Results of test 2. ΔPSNR denotes the difference between output and HEVC baseline. Positive indicates performance improvement. The black font is the best.

Result

Visual Quality



(a) Original frame (b) Input 30.42dB (c) Output 30.85dB

(d) Original frame (e) Input 30.03dB (f) Output 30.41dB

Figure 5: Visual Quality Comparisons in test 2. The left column is original frame. The middle column is frame compressed by HEVC. The right column is output of our model.

Reference

1. W. Park and M. Kim, “Cnn-based in-loop fltering for coding efficiency improvement,” in 2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), 2016, pp. 1-5.
2. Y. Zhang, T. Shen, X. Ji, Y. Zhang, R. Xiong, and Q. Dai, “Residual highway convolutional neural networks for in-loop fltering in hevc," IEEE Transactions on Image Processing, vol. 27, no. 8, pp. 3827-3841, 2018.
3. T. Wang, M. Chen, and H. Chao, \A novel deep learning-based method of improving coding efficiency from the decoder-end for hevc," in 2017 Data Compression Conference (DCC), 2017, pp. 410-419.
4. Yang R , Xu M , Wang Z . Decoder-side HEVC quality enhancement with scalable convolutional neural network[C]. 2017 IEEE International Conference on Multimedia and Expo (ICME). IEEE Computer Society, 2017.

Thanks !