Hybrid Video Coding for CLIC 2021

David Alexandre, Wen-Hsiao Peng, Hsueh-Ming Hang
Dept. of Electronics Engineering
National Yang Ming Chiao Tung University, Taiwan
davidalexandre.eed05g@nctu.edu.tw, wpeng@cs.nctu.edu.tw, hmhang@nctu.edu.tw

Abstract

We propose a hybrid video coding scheme, which includes a VVC/H.266 codec and a learning-based postprocessing module. The CNN-based postprocessing network enhances the quality of the VVC-decoded frames. The postprocessor takes the previously reconstructed image and the current decoded image to produce the current reconstructed image. Our experimental results show that for typical camera-captured videos, the proposed postprocessor improves both PSNR and MS-SSIM.

1. Introduction

We propose a hybrid video coding scheme, which includes a VVC/H.266 codec and a learning-based postprocessing module. The CNN-based postprocessing network enhances the quality of the VVC-decoded frames. The postprocessor takes the previous reconstructed image and the current decoded image to produce the current reconstructed image. Our experimental results show that for typical camera-captured videos, the proposed postprocessor improves both PSNR and MS-SSIM. Fig. 1 shows our postprocessing scheme.

2. Proposed Method

We use the VVenC/VVdeC software of the VVC codec [1]. By setting appropriate bit rates to compress the video frames, we obtained the decoded images. Often, the first I-frame is set to a higher quality to achieve an over-all better average quality.

Generally, the previous reconstructed frame (or the intra decoded frame for $t = 1$) is used as the reference frame (x_{t-1}) and the current frame as the target frame (x_t), the proposed CNN module produces the reconstructed target frame. The input frames are normalized in the YUV444 format between range of 0 to 1. The first stage of the proposed system is a 2D convolutional network used to extract the features. Next, we use multi-scale residual dense networks [2] to manipulate/process the information combined from

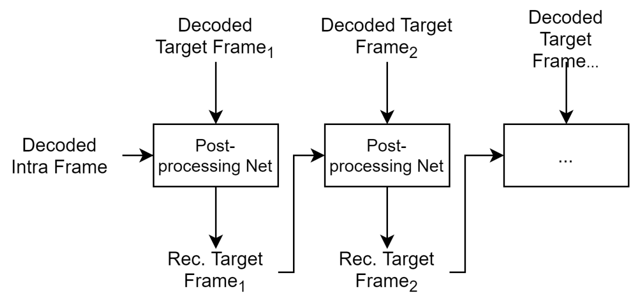

Figure 1: Our postprocessing scheme after VVC codec. The first target (P-)frame is processed using the decoded intra frame and the current decoded frame. It produces the first reconstructed target frame. Then, starting from frame 2, the postprocessor takes in the previous reconstructed frame and the current decode frame to produce the current reconstructed frame.

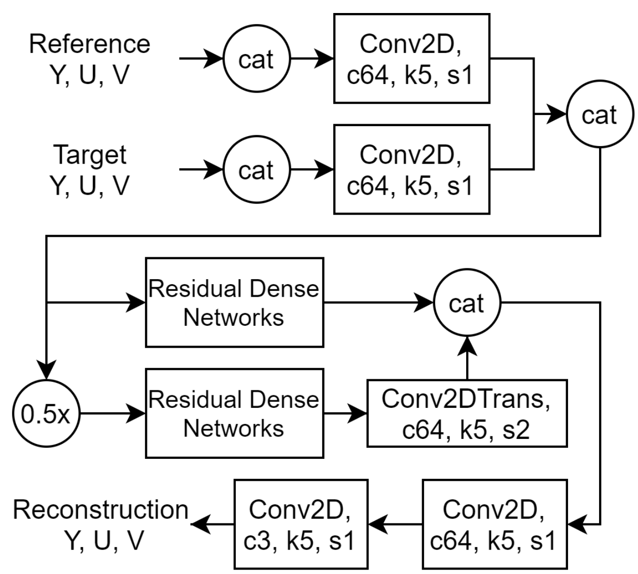

Figure 2: The postprocessing network. c64, k5, s1 refers to 64 channel output, kernel 5x5 and stride 1.

(a) Postprocessed frame (PSNR:40.208, MS-SSIM:0.998) from decoded frame (PSNR:38.144, MS-SSIM:0.997)

(b) Postprocessed frame (PSNR:36.565, MS-SSIM:0.992) from decoded frame (PSNR:36.618, MS-SSIM:0.991)

Figure 3: Examples for postprocessed frames taken from the CLIC2021 validation dataset.

the reference frame and the target frame. At the end, it generates the reconstructed frame $\hat{x}_t$. The network structure is shown in Fig.2. During the training, the postprocessing network is optimized for MS-SSIM loss between target frame and reconstructed frame, $L = 1 - MSSSIM(x, \hat{x}_t)$.

3. Experiments

We use the UGC Dataset as the training dataset. First, we generate the reconstructed frame using VVC codec at a certain QP value. As shown in Fig. 3, for the CLIC 2021 validation set we have quality gain in terms of MS-SSIM or PSNR. To give the best performance, we put a mechanism to skip the postprocessing network in the case that the postprocessor produces a lower reconstruction quality due to the particularity of certain videos. For the submission, our team name is NYCU_C and achieves 0.97106 using the preset fast mode of VVenC.

4. Acknowledgment

This work is partially supported by the Ministry of Science and Technology, Taiwan under Grant MOST 109-2634-F-009-020 through Pervasive AI Research (PAIR) Labs, National Yang Ming Chiao Tung University, Taiwan. Also, we would like to thank National Center for High-performance Computing (NCHC) for providing computational and storage resources.

References

[1] Fraunhofer Versatile Video Decoder (VVdeC). https://github.com/fraunhoferhhi/vvdec.

[2] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution, 2018.