# Learned Inter-frame Compression for CLIC 2021

Chih-Peng Chang[1] Chih-Hsuan Lin[1] Wen-Hsiao Peng[1,3] Hsueh-Ming Hang[2,3]

cpchang.cs08g@nctu.edu.tw meerkat10.cs09g@nctu.edu.tw wpeng@cs.nctu.edu.tw hmhang@nctu.edu.tw

[1]Computer Science Dept., [2]Electronics Engineering Dept.,
[3]Pervasive AI Research (PAIR) Labs, National Chiao Tung University, Taiwan

## Abstract

*We propose a hybrid-based video coding framework and submit it to the video track of Challenge on Learned Image Compression (CLIC) 2021. Our system consist of a conventional codec as intra-frame coder and a learned network for inter-frame coding. The residual coder additionally introduce an adjustable quantization step size to fit the target bitrate. The team name for the proposing method is NYCU.*

## 1. Introduction

In 2021, Challenge on Learned Image Compression (CLIC) launches a video coding track where participants need to submit a codec that can encode a 60-frames video sequence into bitstream. Before validation phase is over, each need to submit a decoder and a compressed bitstream that is limit by a specified file size.

To participate in this challenge, we design and submit a hybrid coding framework that include a rate adjustable residual coder. Our team name is NYCU. In this factsheet, we briefly describe the design in section 2 and the experimental results in section 3.

## 2. Proposed Method

The proposed scheme is a hybrid coding method, where intra-frame coding adopts conventional codec and inter-frame coding is a learned codec. In Fig. 1, we show the diagram to better illustrate the overall architecture. $x_1$ represents I-frames which is compressed with intra-frame codec. And $x_2$ indicates the following frames that are going to be coded by the learned inter-frame codec. There are four learning-based components included in the inter-frame module, which include motion estimation, motion coding, motion compensation, and residual coding.

The intra-coding tool and the four components in inter-frame part will be further describe in the following sections.

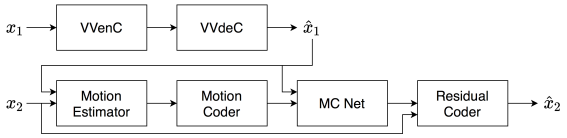


Figure 1. The overall architecture of our proposed framework.

### 2.1. Intra Codec

We use VVC as intra-coding module for the proposing scheme. The implementation of VVC is using VVenC and VVdeC software. In the submitted version of our codec, the QP value is set to 38.

### 2.2. Motion Estimation

We adopt PWC-Net as motion estimation network in our framework. This network accepts reference frame and coding frame as input and expected to output the optical flow map for the corresponding input.

In order to improve the convergence speed of the whole framework, we pre-train the PWC-Net before including it in the joint training scenario.

### 2.3. Motion Coding

In order to transmit the motion generated by motion estimation network, we need a motion coder to encode the optical flow into bitstream. The architecture of motion coding network follows the design described in Balle et al. [1].

### 2.4. Motion Compensation

Motion compensation is needed in order to generate the predicted frame based on the estimated flow map, and it is done in two steps. First, we perform warping with reference frame and decoded flow map. Second, we input the warped frame with the reference frame into a motion compensation network. The architecture of this network follows the design in [2]. The expected output of this network is a compensated frame.

### 2.5. Residual Coding

We need a residual coder to encode the prediction error produced by motion prediction. The residual frame is the difference of the target frame and the compensated frame. Same as motion coding module, residual coding also follows Balle et al. [1] for the network architecture. In addition to the original design in Balle et al. [1], we include a scaling factor to adjust the quantization step size for the latent code. This modification make sure that we can perform adjustment to the compressed file size in order to fit the target bit-rate at test phase.

## 3. Experiments

The entire inter-frame coding framework is trained on vimeo dataset and tested on CLIC 2021 validation dataset. The GoP size is set to 10 and the scaling factor is set to 1 for validation phase submission. The MS-SSIM for validation phase submission is 0.962.

## 4. Acknowledgement

This work is partially supported by the MOST, Taiwan under Grant MOST 108-2634-F-009-007 through Pervasive AI Research (PAIR) Labs, National Chiao Tung University, Taiwan.

## References

[1] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. Variational image compression with a scale hyperprior. In *International Conference on Learning Representations*, 2018. 4321, 4322

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 4321