|  |  |
| --- | --- |
|  | Moving Picture, Audio and Data Coding  by Artificial Intelligence  www.mpai.community |

|  |  |
| --- | --- |
| **N409** | 2021/10/27 |
| **Source** | Requirements (EVC) |
| **Title** | MPAI-EVC Evidence Project report and plan |
| **Target** | MPAI Members |

The goal of the group is to enhance EVC (Essential Video Coding) using AI-tools to reach at least 25% improvement over the baseline profile. The group is currently working on three coding tool: Intra prediction, Super Resolution, and in-loop filtering. For each tool, in the following we describe the proposed approach and the steps of database building, learning phase and inference.

* **Intra prediction tool:** 
  + The group put together a dataset of 32x32 intra prediction blocks with 1.5M patches and another for 16x16 prediction blocks with 5.5M patches from the *AROD* dataset.
  + Another dataset was put together for testing purposes extracting patches from the first frame of the JVET Class B sequences *BasketballDrive*, *BQTerrace*, *Cactus*, *Kimono1*, *ParkScene*

The approach consists in feeding the EVC intra predictor to an autoencoder inspired by the *Contex Encoders* architecture together with the relative causal context, recasting the problem as an inpainting task via masked convolutions.  
The autoencoder is trained offline over patches extracted from the AROD dataset to minimize the MSE between its output and the original image block to predict.

In the inference phase the EVC encoder sends, for each 32x32 and 16x16 CU, the 64x64 decoded context to the server for the mode 0 EVC intra predictor (DC). The server feeds the received 64x64 context into the trained autoencoder and returns to the EVC encoder the new 32x32 or 16x16 predictor, depending on the case. The EVC encoder finally replaces EVC DC intra predictor with the autoencoder-generated predictor and this predictor is then put into competition with the other 4 EVC intra predictors (modes 1-4) and encoding proceeds as usual. The generated bitstream remains fully decodable under the assumption that the autoencoder network is available at the decoder side.

Table 1 shows the BD-rate gains of the DC-enhanced EVC encoder over the reference EVC encoder, with gains in the 1% to 2% range depending on the considered QP range.  
Future plans include extending the proposed approach also to 8x8 and 4x4 CUs, and experimenting with other network architectures than convolutional and training over different datasets than AROD.

|  |  |  |
| --- | --- | --- |
|  | **BD-rate variation [%]** | |
| **Sequence** | **QPs 22-47** | **QPs 32-47** |
| BasketballDrive | -2.54 | -4.19 |
| BQTerrace | -0.16 | -0.51 |
| Cactus | -1.31 | -1.90 |
| Kimono | -1.13 | -1.40 |
| ParkScene | -1.06 | -1.51 |
| AVG | -1.24 | -1.90 |

Table 1: BD-rate gains over the EVC baseline encoder where the 32x32 and 16x16 mode 0 (DC) Intra predictor is replaced by that generated by a convolutional autoencoder.

* **Super resolution tool:**

The group has worked on the test phase.

We built a dataset to train the super resolution network: 2000 pictures (KAGGLE DATASET 4K standard resolution images (2057 files) <https://www.kaggle.com/evgeniumakov/images4k>).

We have experimented the performances of the trained network on 8 sequences of 500 frames each for the super resolution SD to HD, and on 3 sequences of 500 frames each for the super resolution HD 2 4K

The results so far obtained are described in the two subsections below.

1. **SD to HD**

The SD to HD testing phase has been finalized on all QPs (15,32,37 and 45), with activated

and not activated the in-loop filter, which is a deblocking filtering in EVC codec.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sequence** | **BD-rate variation [%] - QP15** | **BD-rate variation [%] - QP30** | **BD-rate variation [%] - QP37** | **BD-rate variation [%] - QP45** |
| Rome\_1 | -83.70 | -72.51 | -66.13 | -73.79 |
| Rome\_2 | -80.16 | -53.25 | -60.82 | -67.55 |
| Talk\_show | -64.63 | -51.82 | -61.68 | -70.72 |
| Rush\_hours | -76.84 | -60.21 | -62.87 | -67.05 |
| Duck \_take\_off | -80.54 | -73.68 | -72.84 | -77.01 |
| Diego\_and\_the\_owl | -77.05 | -57.60 | -61.60 | -66.63 |
| Crowd\_run | -74.70 | -64.95 | -68.93 | -75.72 |
| Parkjoy | -76.45 | -73.22 | -74.61 | -79.21 |
| AVG | -76.76 | -63.42 | -66.19 | -72.21 |

Table 3: BD-rate variation for the EVC baseline encoder where the super resolution block is replaced by the deep-learning based super resolution (QPs 15-45) with inloop filter activated.

|  |  |
| --- | --- |
| **Sequence** | **BD-rate QP Averaged [%]** |
| Rome\_1 | -30.99 |
| Rome\_2 | -54.76 |
| Talk\_show | -59.39 |
| Rush\_hours | -1.32 |
| Duck \_take\_off | -7.72 |
| Diego\_and\_the\_owl | 7.43 |
| Crowd\_run | -13.94 |
| Parkjoy | -23.31 |

Table 4: BD-rate variation averaged over the QPs of Table 3, for the EVC baseline encoder where the super resolution block is replaced by the deep-learning based super resolution with inloop filter activated.

1. **HD to 4K**

To solve existing issues with the memory management of the on board GPU memory, we have first decompose the input frame, for each sequence used in the test, into 4 of identical size blocks, applied on each of them the super resolution step and then merged the results to re-compose the original size of the frame.

Experiments HD to 4K, referred to the quantizer parameters (QP) 15, 32, 37, 45, with activated and not activated the in-loop filter, which is a deblocking filtering in EVC codec, are ongoing.

The next steps are:

|  |  |  |  |
| --- | --- | --- | --- |
| **Tool** | **Date** | **Topic** | **Who** |
| Intra prediction | 1 meeting cycle | More experiments to improve the BD-rate | Attilio, Davide, Roberto |
| 1 meeting cycle | Experiments on 8x8 block size | Attilio, Davide |
| 2 meeting cycles | Measure the performances after training (BD-Rate) | Attilio |
| Super Resolution | 2 meeting cycles | Solve issue of the memory management for HD24K | Alessandro, Mattia, Francesco, Antonio Kebula |
| 2 meeting cycles | More experiments to improve the BD-rate | Francesco, Antonio, Mattia and Alessandro |
| In-loop | 2 meeting cycle | Dataset Building for In-loop filter | Roberto |

**Future plan**

* motion compensation: improve the motion compensation using NN architecture
* inter prediction: use NN architectures to refine the quality of inter-predicted blocks; introduce new inter prediction mode which tries to predict a frame directly without the use of side information; leverage on Optical Flow algorithm for the motion estimation.
* quantization: uniform scalar quantization used in classical video codec standard does not conform to the characteristics of human visual system. It is possible to use a quantization strategy based on neural networks.
* arithmetic encoder: improve the CABAC performance by leveraging NN to directly predict the probability distribution of intra modes instead of the handcraft context models