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

|  |  |
| --- | --- |
| N374 | 2021/09/30 |
| Source | Video group |
| Title | MPAI-EVC Evidence Project report and plan |
| Target | MPAI Members |

The goal of the group is to enhance EVC using AI-tools to reach at least 25% improvement over EVC baseline profile.

The group is working on three tools (Intra prediction, Super Resolution, in-loop filtering) and for each tool there are three phases: database building, learning phase and inference phase.

The state of the art is:

* Intra prediction track:
	+ Dataset building
		- Training dataset of 32x32 intra prediction block created, 1.5M pictures
		- Training dataset 16x16: 5.5M pictures.
	+ Neural network trained for 32x32 intra prediction
	+ Test on 1st frame of Class B sequences *BasketballDrive*, *BQTerrace*, *Cactus*, *Kimono1*, *ParkScene*
	+ Computed BD rate over QPs = {22, 32, 37, 42, 47}.

The group has worked on the signalling issue: the EVC encoder sends, for each 32x32 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 enhanced 32x32 predictor. The EVC encoder finally replaces the standard DC intra predictor with the autoencoder-enhanced DC predictor and this predictor is then put into competition with the other 4 EVC intra predictors (modes 1-4) and encoding proceeds as usual.

 Table 1 shows the BD-rate gains of the DC-enhanced EVC encoder over the reference EVC encoder (we recall that 32x32 CUs intra predictors only are enhanced). The proposed method achieves consistent gains over almost all tested sequences, with an average gain of about 0.8% and over 1.5% gains with the sequence BasketballDrive (gains stem from high QPs, above 32: about 1%, Table 2). The rightmost column of the Table 1 shows the fraction of times our enhanced DC predictor is preferred by the EVC encoder RDO algorithm to the other 4 intra predictors: our predictor achieves comparatively lower RD cost in over 50% of the cases.

|  |  |  |
| --- | --- | --- |
| **Sequence** | **BD-rate variation [%]** | **Selection rate** |
| BasketballDrive | -1.6037 | 56.15 |
| BQTerrace | 0.0020 | 73.59 |
|  Cactus | -0.4680 | 82.83 |
| Kimono | -0.9291 | 50.62 |
| ParkScene | -0.5807 | 79.16 |
|  AVG | -0.7159 | 68.47 |

 Table 1: BD-rate gains for the EVC baseline encoder where the 32x32 mode 0 (DC) Intra predictor is replaced by the predictor enhanced by a convolutional autoencoder (QPs 22-47)

|  |  |
| --- | --- |
| **Sequence** | **BD-rate variation [%]** |
| BasketballDrive | -2.5180 |
| BQTerrace | -0.1287 |
| Cactus | -0.7256 |
| Kimono | -1.1523 |
| ParkScene | -0.8006 |
| AVG | -1.06504 |

Table 2: BD-rate gains for the EVC baseline encoder where the 32x32 mode 0 (DC) Intra predictor is replaced by the predictor enhanced by a convolutional autoencoder (QPs 32-47).

We recall that in these preliminary experiments, only 32x32 CUs intra predictors were enhanced with the proposed method, so further gains can be expected when the proposed method is extended to smaller CUs (16x16 and below), which represent the large bulk of CUs actually chosen by the encoder RDO mechanism.

In the following there are the PSNR-bitrate graphs for each sequence:



Figure 1: sequence Kimono, QPs(22-47)



Figure 2: sequence BasketballDrive, QPs(22-47)

 

Figure 3: sequence BQTerrace, QPs(22-47)



Figure 4: sequence Cactus, QPs(22-47)



Figure 5: sequence Parkscene, QPs(22-47)

* Super resolution track: the group has worked on the test phase.

All the experiments have been conducted on two uses cases, e.g., SD to HD and HD to 4K, with in mind to extend them in the future to SD to 4K. The chosen quantizer parameter (QP) values have been 15, 30 and 45, with activated and not activated the in-loop filter, which in our case is a deblocking filtering. Based on the above possible configurations we have 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, and in the following the graphs with the results.

Figure 6: sequence park joy, QPs(15,30,45)

Figure 7: sequence Crowd run, QPs(15,30,45)

Figure 8: sequence diego and the owl, QPs(15,30,45)

Figure 8: sequence ducks take off, QPs(15,30,45)

Figure 9: sequence Rome 1, QPs(15,30,45)

Figure 10: sequence Rome 2, QPs(15,30,45)

Figure 11: sequence Rush hour, QPs(15,30,45)

Figure 12: sequence Talk show, QPs(15,30,45)

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 | Train on database of 16x16 block size | Attilio, Davide |
| 1 meeting cycle | Random crop on the picture buffer | Roberto, Attilio |
| 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, Francesco Kebula |
| 2 meeting cycles | More experiments to improve the BD-rate | Francesco and Alessandro |
| In-loop filter | 1 meeting cycle | More experiments to improve the BD-rate | Alessandro, Mattia |

**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