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

|  |  |
| --- | --- |
| N335 | 2021/08/25 |
| Source | Video group |
| Title | Status report on MPAI-EVC Evidence Project |
| Target | MPAI-11 |

MPAI has created a collaborative environment that will allow testing independently developed AI tools on a common EVC (Essential Video Coding) code base. The goal is to enhance EVC to reach at least 25% improvement over EVC baseline profile.

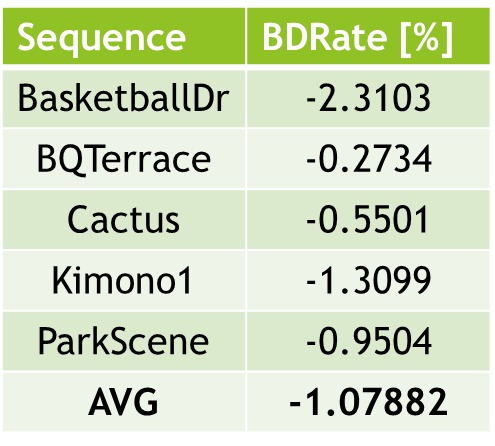
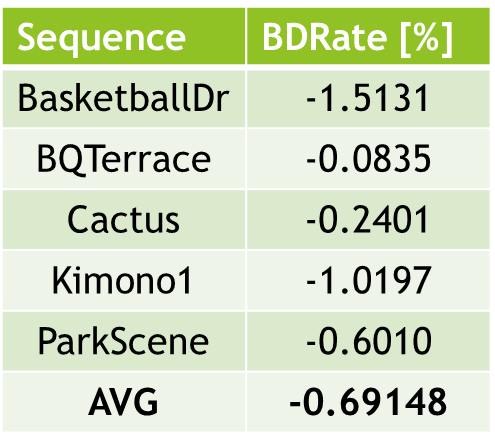
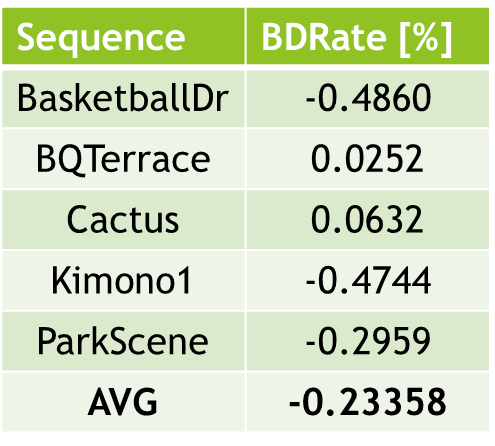
The group is working on two tools (Intra prediction and Super Resolution) and for each tool there are three phases: database building, learning phase and inference phase. All these phases will be developed in parallel.

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}.

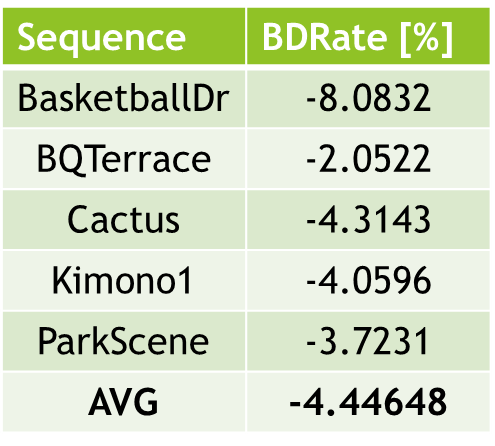
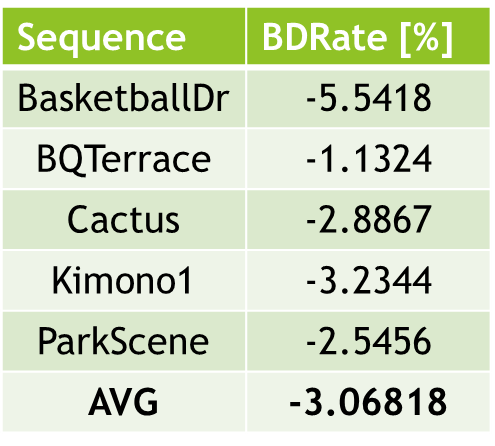
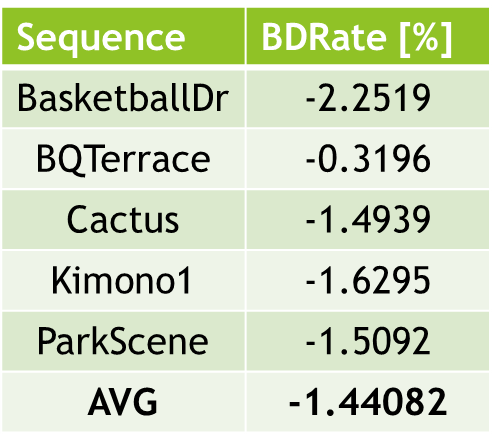
In the following tables are shown the current results for the Unconditioned mode (the NN is always ON) and the Oracle mode (NN is switchable ON/OFF according to some metrics).

There is no signalling cost calculation in the Oracle mode results.



QP 22-37 QP 22-47 QP 32-47

Table 1: BD-rate over 5 test sequences (Unconditioned mode)



QP 22-37 QP 22-47 QP 32-47

Table 2: BD-rate over 5 test sequences (Oracle mode)

* + We are computing the signalling cost:

Encoder

If (NN\_RDO<EVC\_RDO)

Use NN 🡪 Best predictor = IPD\_NN

Else

Use EVC predictor 🡪 best predictor = IPD\_DC (for example)

Decoder

If (best predictor==IPD\_NN)

## Use NN

* Super resolution track: we have tested few neural networks available on Internet and selected the Densely Residual Laplacian Super Resolution (DRLN) as a candidate one.
  + We have built a dataset to **test** the super resolution network:

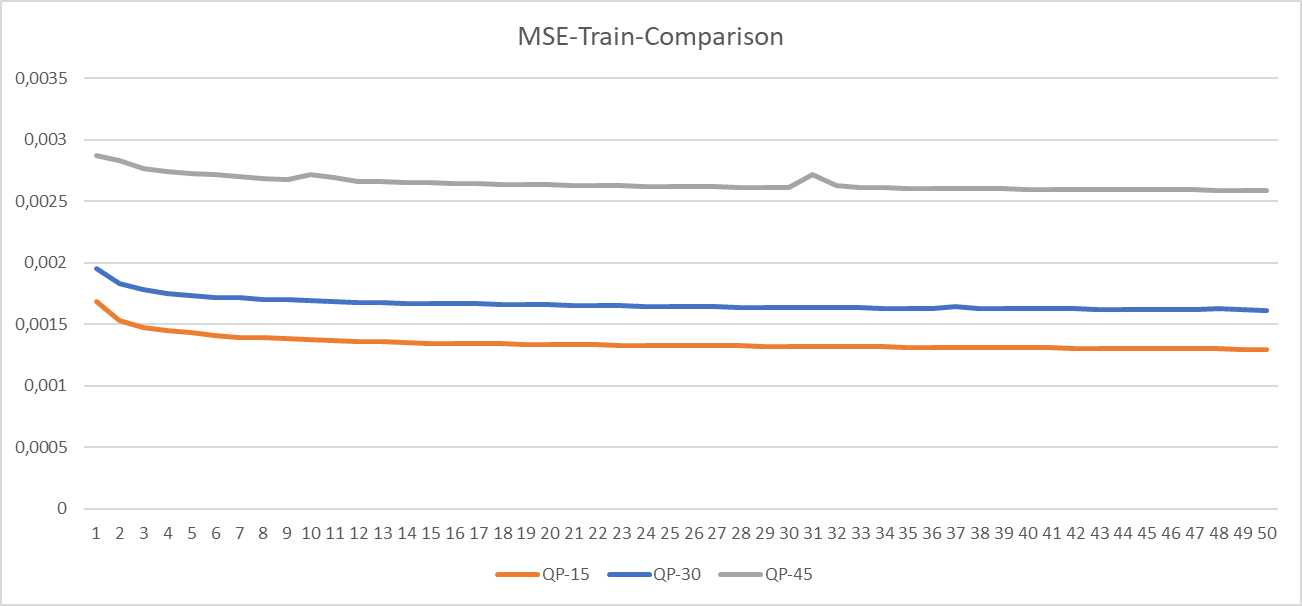
### Crowd run, Ducks, Take off, Park joy (400 frames per sequence)

* + - 3 resolutions:
      * 4k - 3840x2160
      * HD - 1920x1080
      * SD - 960x540
    - 3 QP values:
      * 15
      * 30
      * 45
    - Two options:
      * Deblocking enabled
      * Deblocking disabled
  + Many experiments were performed on QP 15, 30 and 45, training the network to upscale from SD to HD and to HD to 4K.
  + It was used an algorithm based on the importance sampling technique applied on the entropy values distribution of all crops in each frame.

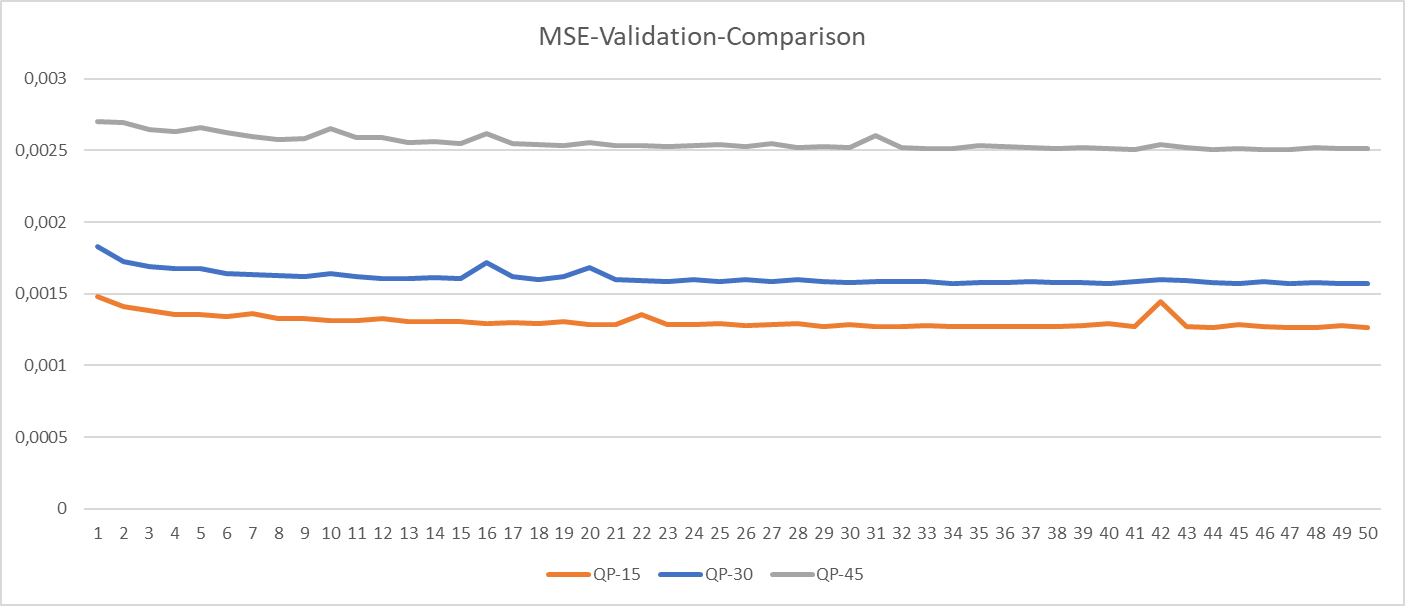
## Two experiments have been conducted using Deblocking filter On and Deblocking filler OFF. The following parameter have been used: crop size 128x128, batch size 2, learning rate 10e-5 and 50 epochs.

* Taking as input the SD resolution images encoded with Quantization Parameter 15, 30 and 45 decoded with deblocking filter, and their corresponding output HD resolution image.

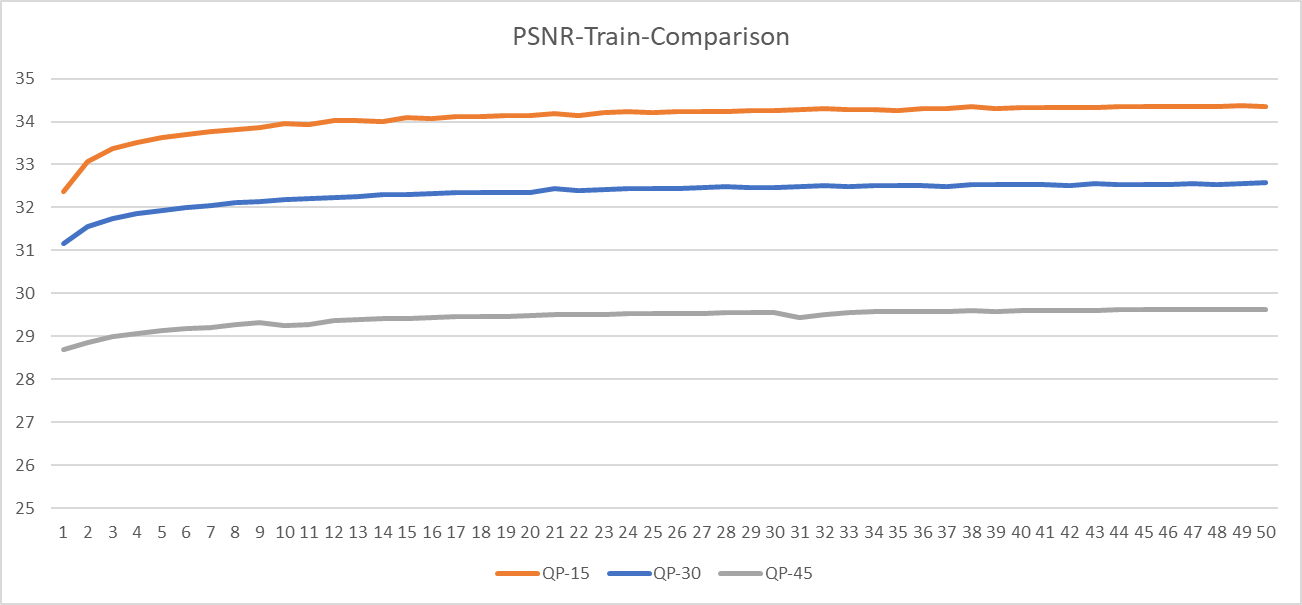
Below is shown also the loss function for training and validation, note that loss function (MSE metric) for training is smoothly decreasing during the training phase:

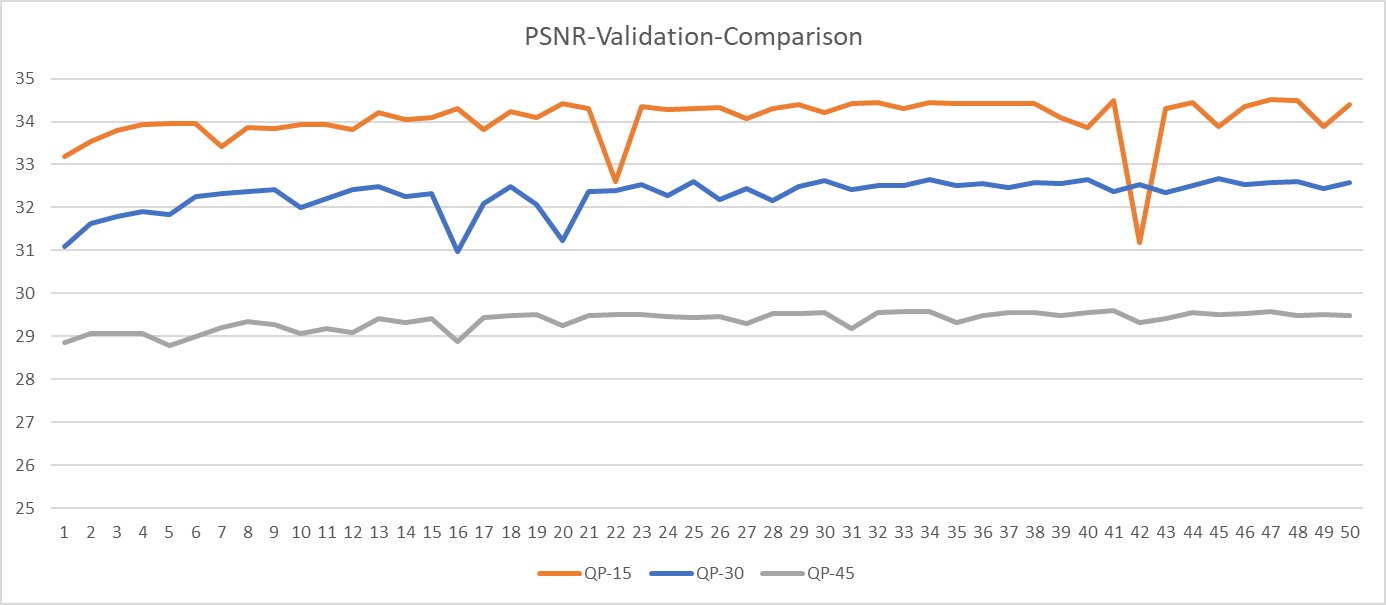


Similar behaviour is for the loss function for the validation dataset:

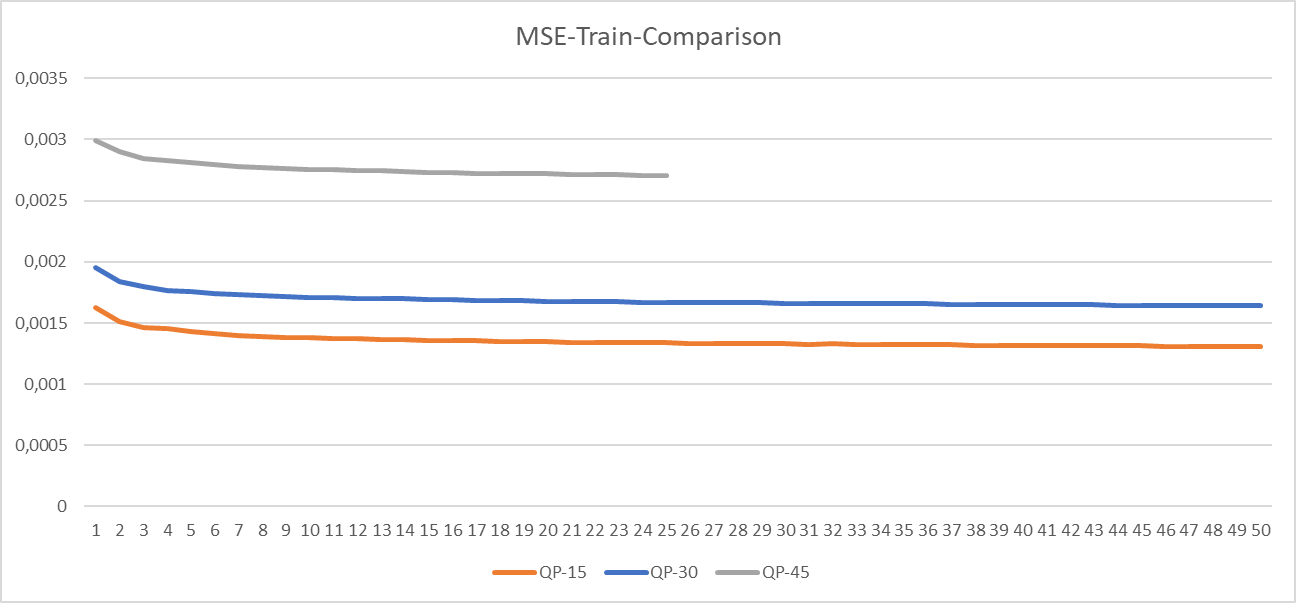


Regarding the PSNR metric:

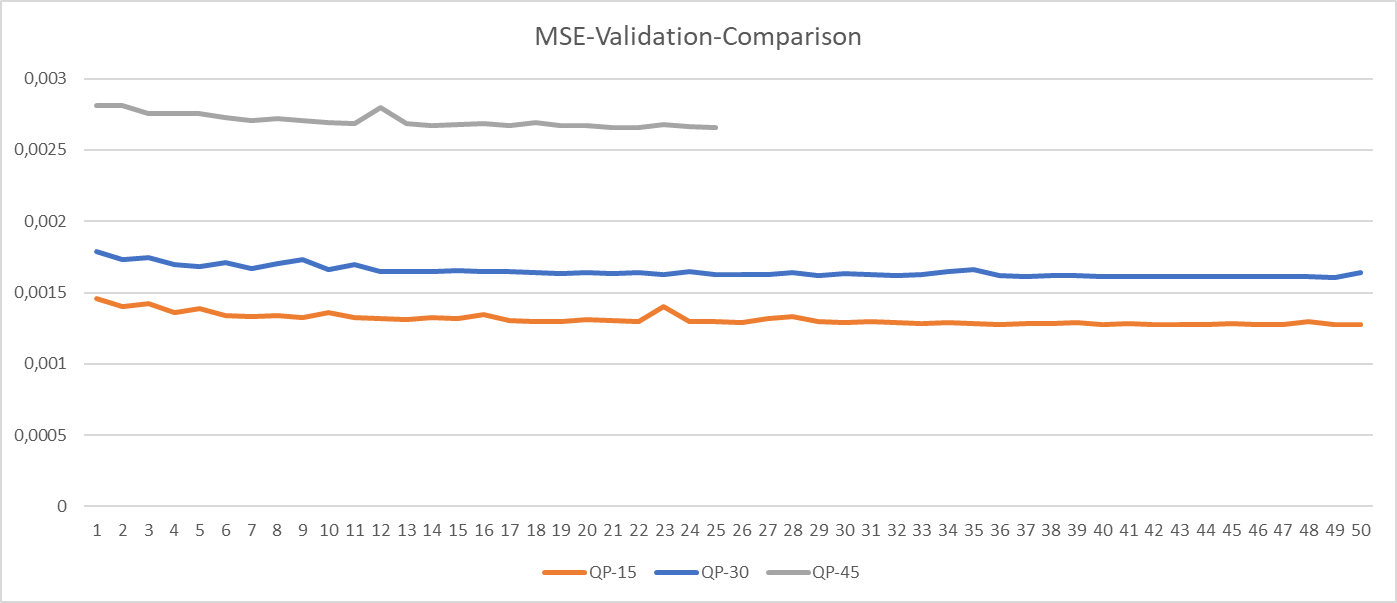


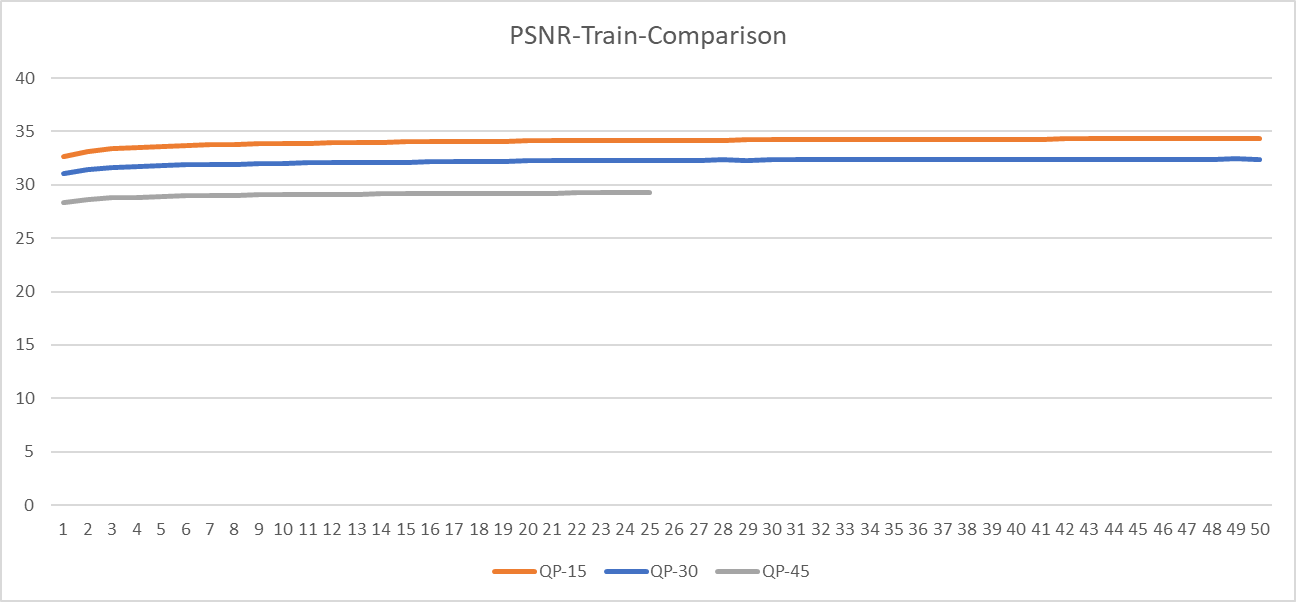


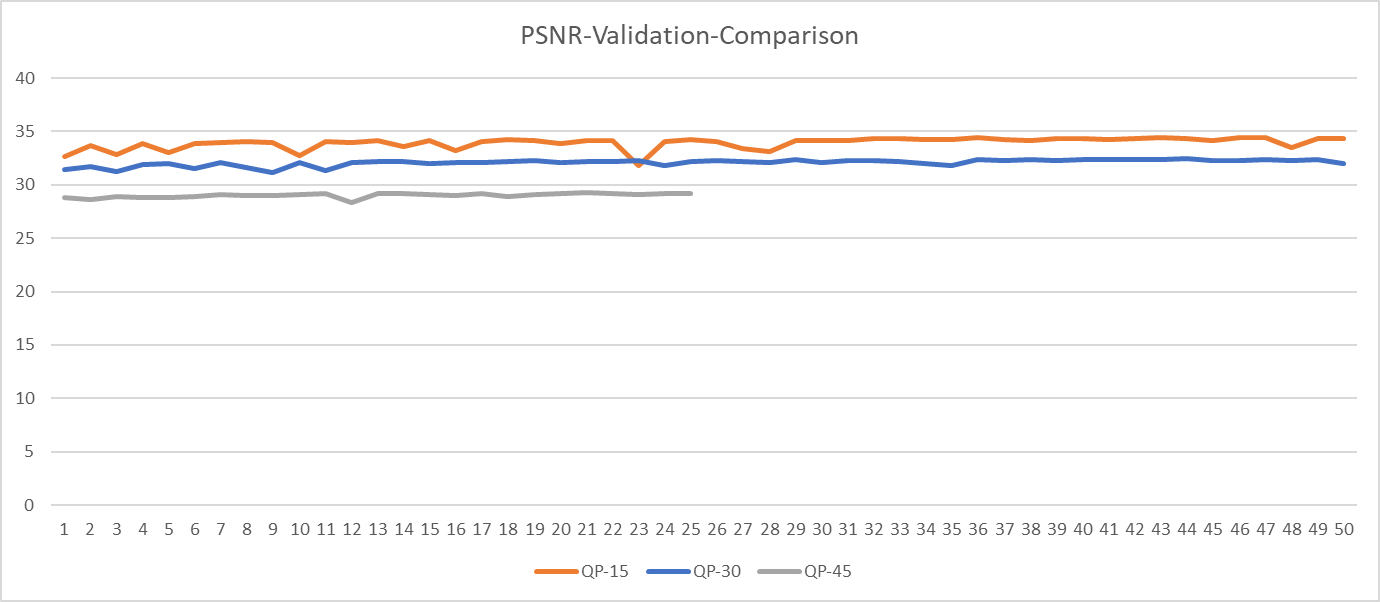
* The second experiment has produced the following results using Deblocking OFF. The loss function (MSE metric) for training is smoothly decreasing during the training phase:



Similar behaviour is for the loss function for the validation dataset:

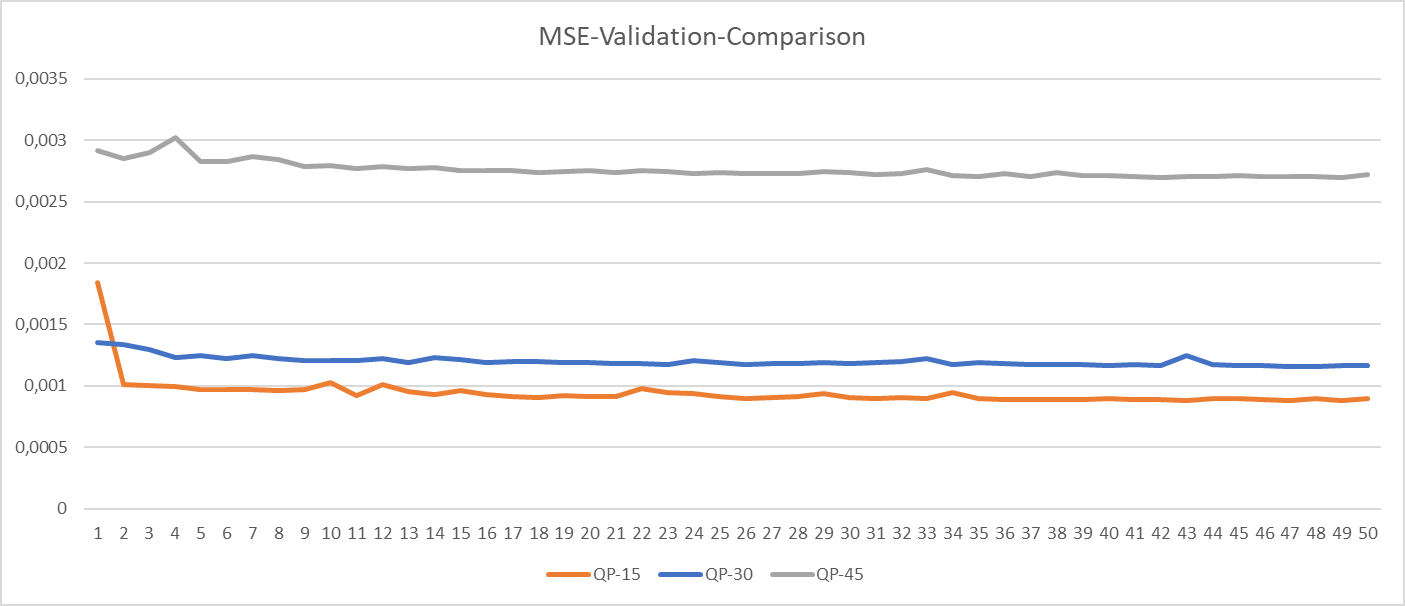


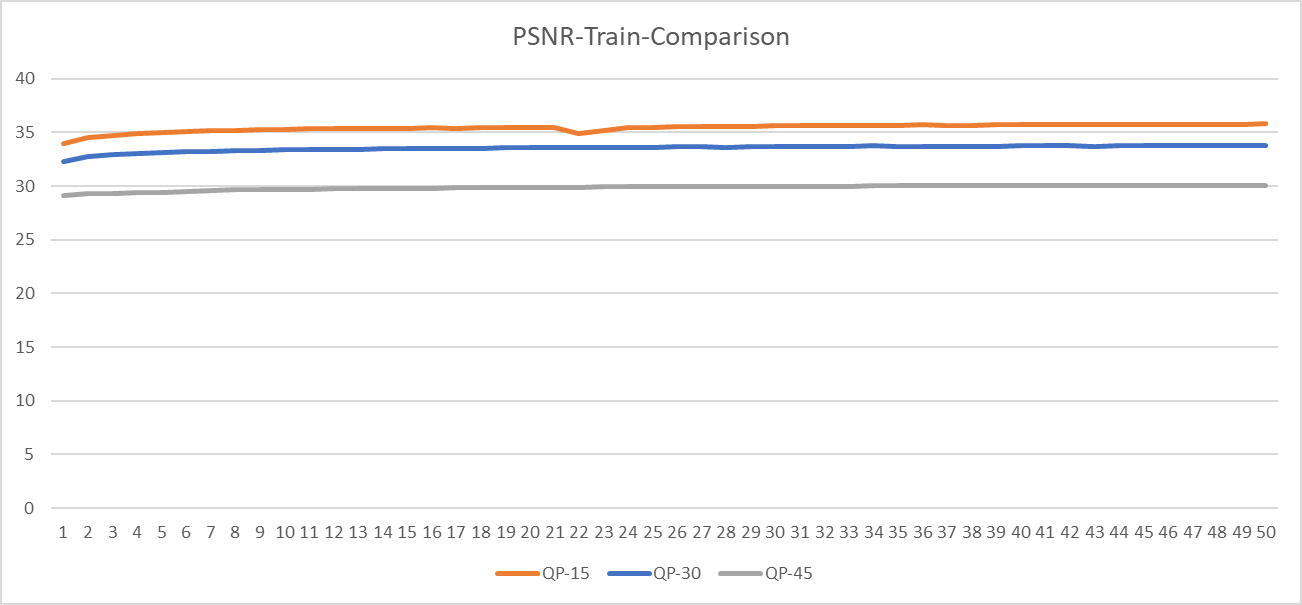


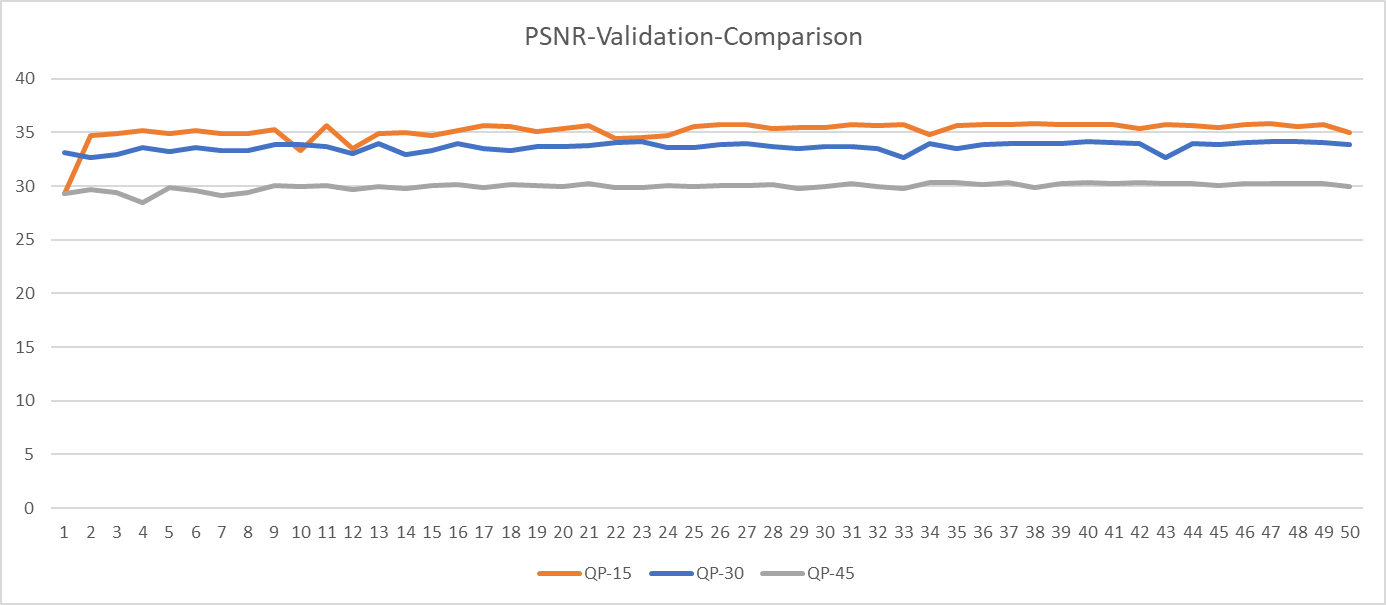


* Taking as input the HD resolution images encoded with Quantization Parameter 15, 30 and 45 decoded with deblocking filter ON, and their corresponding output 4K resolution images:









Code and dataset are available on MPAI-git and MPAI-ftp.

The next steps are:

|  |  |  |  |
| --- | --- | --- | --- |
| **Tool** | **Date** | **Topic** | **Who** |
| Intra prediction | 1 meeting cycle | Find a way to signal the best predictor between the NN and the EVC, to the decoder (maybe as a side information) | Attilio, Roberto |
| 1 meeting cycle | Build a database of 16x16 block size | Alessandra, Roberto |
| 2 meeting cycles | Deliver the results on further experiments on training | Alessandro, Francesco (Kebula), Giuseppe (Kebula), Francesco (CNR), Attilio |
| 2 meeting cycles | Measure the performances after training (BD-Rate) | Attilio |
| Super Resolution | 2 meeting cycles | Verify the generalization capability of the network on the test set | Alessandro, Francesco CRN, Francesco and Giuseppe Kebula |
| 2 meeting cycles | Produce a short tutorial to deploy and train the network | Francesco and Giuseppe |