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

|  |  |
| --- | --- |
| **N1694** | 2024/03/20 |
| **Source** | Requirements (EVC) |
| **Title** | Summary of results obtained by the MPAI-EVC project |
| **Target** | MPAI Community |

# Video material

A data set has been created by aggregating video sequences from different sources to provide sufficient content diversity. Several existing data sets have been considered and, as of today, the following collections have been selected:

1) BVI-DVC Part 1 (public)

2) The Ultravideo

3) The SVT Open Content

4) The Tencent Video

Thus, we have aggregated 350 4K sequences. 64 frames for each sequence have been used, totalling to 22400 frames.

# Super Resolution

In our super-resolution approach the picture before encoding is downscaled and then the super-resolution step is applied to the decoded picture to get the native resolution.



Figure 1 - Approach to test Super-resolution

Our work aims to:

1. Devise a novel approach for enhancing the training process of the state-of-the-art SR network.
2. Address the challenge of improving the performance of the trained SR network when dealing with images encoded with various codecs.

The methodology targets enhancing the SR network's performance when used to up-sample images of different resolutions. Moreover, it evaluates the effectiveness of utilizing the SR network compared to traditional upscale filters.

Our results related to SR are as follows:

For videos encoded with EVC Baseline Profile

* EVC HD24K -7,44 BD-rate gain (BD-PSNR\_Y)
* EVC SD2HD -21,62 BD-rate gain (BD-PSNR\_Y)

For videos encoded with VVC Main 10 profile

* VVC HD24K -6,36 BD-rate gain (BD-PSNR\_Y)
* VVC SD2HD -26,05 BD-rate gain (BD-PSNR\_Y)

# Super Resolution with Intra Prediction

We have investigated on whether it is possible to improve the performance of the Essential Video Coding (EVC) modified by enhancing/replacing existing video coding tools with AI-based tools keeping complexity increase to an acceptable level.



Figure 2 - Reference scheme of Encoder MPEG-5 EVC



Figure 3: Figure 2 - Reference scheme of Decoder MPEG-5 EVC

We investigated how replacing the two main video coding tools, such as directional mode and super resolution, with artificial intelligence-based tools can change the coding performance of the EVC coding standard. In this work, the performance of the single proposed artificial intelligence-based tool and the overall performance of the two proposed tools were analysed, showing a significant improvement in encoding performance compared to the baseline EVC video standard (Fig. 2).

The MPEG-5 EVC Baseline profile includes 5 intra prediction modes: DC, horizontal, vertical, and two diagonal modes for each CU. We proposed a learnable intra-prediction module realized as a Neural Network Auto-Encoder that can be easily used to add (or replace) an intra-prediction mode of the MPEG-5 EVC codec.

We have combined the intraframe coding capabilities of EVC enhanced by a neural network with SR (Fig. 3). To evaluate the effectiveness of this integration, we conducted an ablation study, the results of which provide valuable insights into the performance of these components and their combination.

As a point of reference, we included a baseline experiment that employed the MPEG5 codec with classical bicubic interpolation algorithm.

The results of these experiments were then compared and analysed to gauge the impact of our integrated approach, particularly in contrast to the baseline method. Results are shown in the table below.

Table 1 – Results of Intra-enhancement, Super-Resolution, and combined techniques.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **NN Intra Vs. MPEG5-EVC orig** | **MPEG5-EVC orig + SR Vs. MPEG5-EVC orig + bicubic** | **NN Intra + SR Vs. MPEG5-EVC orig + bicubic** |
| **SD2HD** | **-6.89%** | **-21,62%** | **-25,86%** |
| **HD24K** | **-6.89%** | **-7,44%** | **-11.99%** |