**Source:** Interdigital Finland Ory

**Title:** [FS\_AI4Media] New Neural Network hybrid coding use-case

**Agenda item:** 9.7

**Document for:** Discussion and Agreement

# 1 Introduction

One objective for the Study Item on “Artificial Intelligence (AI) and Machine Learning (ML) for Media” is to address current MPEG work on Deep Neural Network (DNN) for Video Coding. This work investigates two potential approaches: hybrid block-based coding with DNN (or hybrid coding) and End to End learning based coding (or End to End coding).

A recent contribution to the JVET group <https://jvet-experts.org/doc_end_user/documents/25_Teleconference/wg11/JVET-Y0239-v1.zip> describes the current technologies under investigation for neural-network-based video compression. In technologies using NN hybrid coding, existing video coding tools such as In-loop filtering, intra-prediction, or Inter-prediction are replaced or enhanced with NN based tools. Another approach introduces post-processing operations such as post-filtering, super resolution or HDR enhancement. These operations are placed outside the coding loop of the codec. Finally, a last approach consists in fully end-to-end solutions that do not rely on the codec itself.

A first contribution (S4aV220901) introducing neural network-based post-processing for a video use case was presented at the video ad-hoc session on 12 July. This new contribution proposes a revision of S4aV220901 including update on the exiting use-case to clarify the difference between both scenarios.

# 2 Proposed changes

--------------------------------------------- begin Change --------------------------------------------------------------------------

## 3.2 Video Quality Enhancement in Streaming

## 3.2.1 Sender-receiver approches

## 3.2.1.1 End-to-End neural network-based video coding

In this use case, the sender and receiver apply parts of an DNN model (e.g. an autoencoder model) to enhance the quality of a video stream. An example of an autoencoder DNN is depicted in figure 3.2.1:



**Figure 3.2.1**

The sender is typically represented by various media functions in the network, which processes the high-fidelity video using the down-scaling part of a pre-trained DNN model to generate intermediate data stream that is streamed together with a lower resolution encoding of the video. The receiver (UE) runs an inference algorithm (e.g. the up-scaling part of DNN model) on using the received intermediate data and video stream to produce a high-quality video for rendering.

The main scenario in this use case is about streaming intermediate data from the network for processing on the UE, involving AI/ML data distribution and operation splitting.

This use case covers all scenarios where intermediate data stream needs to be sent to the receiver, in addition to a low-resolution video

## 3.2.1.2 Neural network based post-processing for video coding

A neural network (NN) applies post-processing to a decoded video sequence to enhance the quality of the decoded frames. The post-processing is performed outside the coding loop and does not impact the decoding process of the video. Possible post-processing algorithms include:

* Post-filtering: where the output of the video decoder is provided as input to a NN to improve the quality of the decoded frames. Such improvements include removal of video coding artifacts, subjective quality enhancement, etc.
* Super resolution: where a NN is used to increase the resolution of the output video sequence when the resolution of the display is greater than the resolution of the decoded frames. The use of NN-based approaches in super resolution resampling process increases the quality of the resulting resampled frames.
* NN-based HDR enhancement: a NN is applied for example to enhance a SDR video into an HDR-looking video.

In contrast to 3.2.1, this approach does not use an intermediate data stream.



Figure 3.2.2 Neural network based post-processing for video coding use-case

Figure 3.2.2 depicts a neural-network-based post-processing use-case where pre-trained NN models are used at the receiver to post-process the decoded video to improve the quality. The video encoder processes the input video source to produce and send content-related metadata to the receiver, based on video/image or block, for example. The content-related metadata can be used to select a pre-trained NN model to be applied to a piece of content and to activate or not the selected NN model on it.

--------------------------------------------- End Change ----------------------------------------------------------------------------

# 3 Proposal

We propose to add the text of clause 3.2 of this contribution to the Permanent as a new clause 3.2.2 and to update the original in a new clause 3.2.1 to clarify the differences between the two use-cases.