|  |  |
| --- | --- |
|  | Moving Picture, Audio and Data Coding by Artificial Intelligencewww.mpai.community |
| **Public document** |
| **N2026** | 2024/10/23 |
| **Source** | Requirements (EVC) |
| **Title** | Use Cases and Functional Requirements: Up-sampling Filter for Video applications (MPAI-UFV) |
| **Target** | MPAI-49 |

Contents

[Foreword 1](#_Toc179366531)

[1 Introduction 1](#_Toc179366532)

[2 Scope of UFV Use Cases and Functional Requirements 1](#_Toc179366533)

[3 Use Cases 1](#_Toc179366534)

[4 Functional Requirements 1](#_Toc179366535)

[5 Definitions 1](#_Toc179366536)

[6 References 1](#_Toc179366537)

# Foreword

This document, issued by Moving Picture, Audio, and Data Coding by Artificial Intelligence (MPAI), collects Use Cases and Functional Requirements relevant to Technical Specification: Up-sampling Filter for Video applications (MPAI-UFV)

MPAI is an international non-profit organisation having the mission to develop standards for Artificial Intelligence (AI)-enabled data coding and technologies facilitating integration of data coding components into Information and Communication Technology (ICT) systems [1]. The MPAI Patent Policy [2] guides the accomplishment of the mission.

# Introduction

Established in September 2020, MPAI has developed eleven Technical Specifications relevant to its mission such as execution environment of multi-component AI applications, portable avatar format, object and scene description, neural network watermarking, context-based audio enhancements, multimodal human-machine conversation and communication, company performance prediction, metaverse, and governance of the MPAI ecosystem. Five Technical Specifications have been adopted by IEEE without modification and four more one more are in the pipeline. Several other standard projects – such as AI for Health, online gaming and XR Venues – are under way and are expected to deliver specifications in the next few months.

MPAI specifications are the result of a process whose main steps are:

1. Development of functional requirements in an open environment.
2. Adoption of “commercial requirements” (Framework Licence) by MPAI principal members setting main elements of the future licence to be issued by standard essential patents holders.
3. Publication of a Call for Technologies referring the two sets of requirements inviting the submission of contributions by parties who accept to licence their technologies according to the Framework Licence, if their technologies are accepted to be part of the target Technical Specification.

This document is the Use Cases and Functional Requirements related to the planned Technical Up-sampling Filter for Video applications (MPAI-UFV) – in the following called UFV – developed in the context of work of the MPAI AI-Enhanced Video Coding (EVC) group.

# Scope of UFV Use Cases and Functional Requirements

This document collects the Use Cases and Functional Requirements for a super-resolution approach, in which case the image prior to encoding is downscaled and then the super-resolution step is applied to the decoded image to achieve native resolution (Fig. 1).



Figure 1 - Super-resolution-based reference model

# Use Cases

**Use Case 1 – Enhancing visuals in real-time: Super-resolution on eSports streaming**

A user is watching in streaming an eSports tournament. The super-resolution tool analyses the incoming video frames and intelligently enhances them. This results in sharper images, reduced pixelation, and improved overall visual quality saving the user bandwidth. For instance, even when streaming at lower resolutions (e.g., 720p), super-resolution can make the visuals appear closer to native 1080p or even 4K quality.

**Use Case 2 – Expanding channel capacity with super-resolution**

A regional TV broadcaster wants to expand its channel lineup to attract a broader audience. However, the available bandwidth for transmission is limited, and adding more physical channels is not feasible due to regulatory constraints. The broadcaster decides to downscale the content before broadcasting it, effectively doubling the amount of content broadcast without using additional physical channels. Super-resolution at end user ensures that even lower-resolution channels maintain acceptable quality. For instance, a viewer with a 4K TV receives the UHD version of the channel even if it was transmitted in HD resolution, allowing broadcasters to optimise channel capacity, improve quality and adapt to different viewer needs.

**Use case 3 – UAV Skylens**

In the metropolis of Turin, security and surveillance are paramount. The network of unmanned aerial vehicles (UAVs) patrols the city's skies, ensuring the safety of citizens, monitoring traffic and safeguarding critical infrastructure. However, there’s a catch: the live video downlinks from these drones suffer from low resolution, hindering effective threat detection and response.

The city of Turin decides to use SR to improve the resolution of live UAV video downlinks while maintaining low latency. By leveraging advanced algorithms and neural networks, SR enhances the visual fidelity of surveillance footage, empowering security personnel with clearer, more detailed imagery.

# Functional Requirements

MPAI has carried out the following experiments:

* using EVC and VVC encoded sequences.
* SR network trained with EVC and VVC
* Tests of SR-based up-sampling of luminance only sequences for
	+ SD to HD
	+ HD to 4K
* In the following configurations
	+ Training using EVC-encoded and testing EVC-encoded sequences
	+ Training using VVC-encoded Testing VVC-encoded sequences
	+ Training using EVC-encoded ­­ Testing VVC-encoded sequences
	+ Training using VVC­-encoded Testing EVC-encoded sequences
* PSNR of decoded and SR-up-sampled vs original sequences improved by ~20% compared to bicubic-up-sampling.
* Performance using neural networks trained on the sequences encoded with the same codec improved performance by ~1%
* Results were confirmed using colour (objective) and subjective (experts viewing)

The proponent should define a codec that allows a downscaled resolution encoded/decoded stream to be upscaled by the decoder to bring the video to a higher resolution.

The key performance requirements for the video coding proposal are as follows:

1. Reducing the complexity of the MPAI-UFV neural network measured by TBD. Respondents should provide their reduced-complexity neural network or provide their complexity reduction algorithm. Proposals will be considered if the complexity of the proposed neural network is reduced by a factor of 10 and the performance loss is less than 5%.
2. The measures are performed on a set of test sequences provided in the call (available on GitLab to respondents).

The proposed codec shall support rectangular picture formats that will include all commonly used video resolutions, ranging at least from SD to 8Kx4K.

The proposed codec shall support:

1. YCbCr colour spaces with 4:2:0 sampling, 10 bits per component.
2. High dynamic range and wide colour gamut.
3. YCbCr/RGB 4:4:4 and YCbCr 4:2:2.
4. Bit depths up to 10 bits per component.
5. Progressive scan.
6. Test conditions: Intra, Random access, Low delay, P pictures only.
7. QP Values: 22, 27, 32, 37, 42.
8. GOP structure: 32.

Fixed and variable frame rates shall be supported, with upper limits specified by levels.

The proposed codec shall support encoding of the full variety of video content characteristics encountered in the intended applications. This includes (electronic and film) camera-captured scenes, text and graphics mixed into a camera-captured video source, rendered animation content, rendered computer graphics, etc.

Respondents with proposals accepted for consideration are required to join MPAI.

MPAI experts will develop a Test Model using the algorithms submitted and the Test Model will be collaboratively optimised.

The performance of the Test Model will be compared with the best neural network proposed without a disclosed associated algorithm and the better of the two selected for the MPA-UFV Technical Specification.

The standard will specify a method to enhance the performance of up-sampling filter based on trained neural networks. This will be in comparison to current state of the art techniques, with the goal of improving the up-sampling quality across different types of video sequences.

# Definitions

Table 1 gives Terms and Definitions for MPAI-UFV.

Table 1 - Terms and Definitions

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Activation Function | A function (for example, ReLU or generalized divisive normalization, GDN) that takes in the weighted sum of all of the inputs from the previous layer and then generates and passes an output value to the next layer. |
| Artificial Intelligence | A machine program that can solve sophisticated tasks. Machine learning is a sub-field of artificial intelligence. Deep Learning is a subset of machine learning. |
| Backpropagation | Algorithm for performing gradient descent on neural networks. First, the output values of each node are calculated in a forward pass. Then, the partial derivative of the error with respect to each parameter is calculated in a backward pass through the graph. |
| Complexity |  |
| Cross-validation | A mechanism for estimating how well a model will generalize to new data by testing the model against one or more non-overlapping data subsets withheld from the training set. |
| Dataset | A collection of sequences |
| Deep Neural Network  | A type of neural network containing multiple hidden layers. |
| Epoch | A full training pass over the entire dataset such that each example has been seen once. |
| Fine tuning | Perform a secondary optimization to adjust the parameters of an already trained model to fit a new problem.  |
| Generalization | Refers to your model's ability to make correct predictions on new, previously unseen data as opposed to the data used to train the model. |
| hyperparameter | The "knobs" that you tweak during successive runs of training a model. |
| Inference | Refers to the process of making predictions by applying the trained model to unlabeled examples. In this document it is used as a synonym for testing to maintain consistency with video encoding. |
| KMAC/pixel |  |
| Learning rate | A scalar used to train a model via gradient descent. During each iteration, the gradient descent algorithm multiplies the learning rate by the gradient. The resulting product is called the gradient step. |
| Mean Absolute Error (MAE) | MAE Mean Absolute Error |
| Mean Squared Error (MSE) | MSE Mean Squared Error |
| Multiscale Structural Similarity (MS-SSIM) | MS-SSIM Multiscale Structural Similarity |
| Peak signal-to-noise ratio (PSNR) | PSNR Peak signal-to-noise ratio |
| Shallow Neural Network | A neural network that usually have only one hidden layer. |
| Super resolution | The process of generating an image with a higher resolution than its source. |

# References

1. MPAI [Statutes](https://mpai.community/about/statutes/)
2. MPAI [Patent Policy](https://mpai.community/about/the-mpai-patent-policy/)
3. MPAI-EVC <https://mpai.community/standards/mpai-evc/>
4. MPAI; [Call for Technologies: Up-sampling Filter for Video Applications (MPAI-UFV)](https://mpai.community/standards/mpai-ufv/v1-0/call-for-technologies/); N2025
5. MPAI; [Framework Licence: Use Cases and Functional Requirements: Up-sampling Filter for Video Applications (MPAI-UFV)](https://mpai.community/standards/mpai-ufv/v1-0/framework-licence/); N2027
6. MPAI; [Template for Responses: Up-sampling Filter for Video Applications (MPAI-UFV)](https://mpai.community/standards/mpai-ufv/v1-0/template-for-responses/); N2028