|  |  |  |
| --- | --- | --- |
|  | | Moving Picture, Audio and Data Coding by Artificial Intelligence  www.mpai.community |
| **Public document** | | |
| **N1802** | 2024/05/15 | |
| **Source** | Requirements(EVC) | |
| **Title** | Draft Use Cases and Functional Requirements: Up-sampling Filter for Video applications (MPAI-UFV) | |
| **Target** | MPAI Community | |

Contents

[Foreword 1](#_Toc166672265)

[1 Introduction 1](#_Toc166672266)

[2 Scope of UFV Use Cases and Functional Requirements 2](#_Toc166672267)

[3 Use Cases 2](#_Toc166672268)

[4 Functional Requirements 3](#_Toc166672269)

[5 Definitions 4](#_Toc166672270)

[6 References 5](#_Toc166672271)

[Annex A: Information Form 6](#_Toc166672272)

[Annex B: Evaluation Sheet 7](#_Toc166672273)

[Annex C: Check list of data formats proposed by a respondent 10](#_Toc166672274)

[Annex D: Technologies that may require specific testing 11](#_Toc166672275)

[Annex E: Mandatory text in responses 12](#_Toc166672276)

# 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 up-sampling approach, in which case the image prior to encoding is downscaled and then the up-sampling step is applied to the decoded image to achieve native resolution (Fig. 1).

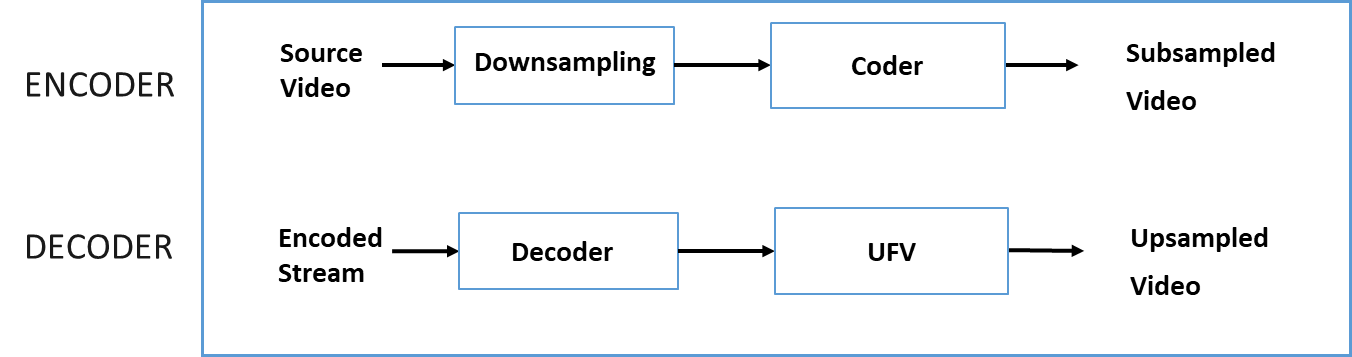


Figure 1 - Framework related to the UFV

# Use Cases

**Use Case 1 – Enhancing visuals in real-time: Up-sampling on eSports streaming**

A user is watching in streaming an eSports tournament. The up-sampling 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), up-sampling can make the visuals appear closer to native 1080p or even 4K quality.

**Use Case 2 – Expanding channel capacity with up-sampling**

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. Up-sampling 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

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. 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 eventual 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 | It is a term used to describe a neural network that usually have only one hidden layer. |
| Super resolution |  |

# 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-EVC project report and plans: <https://mpai.community/wp-content/uploads/2024/04/N1739-MPAI-EVC-Evidence-Project-report-and-plan.docx>

# Annex A: Information Form

This information form is to be filled in by a Respondent to this UFV Call for Technologies.

1. Title of the proposal.
2. Organisation: company name, position, e-mail of contact person.
3. What are the main functionalities of your proposal?
4. Does your proposal provide or describe a formal specification and APIs?
5. Will you provide a demonstration to show how your proposal meets the evaluation criteria?

# Annex B: Evaluation Sheet

NB: This evaluation sheet will be filled out by Evaluation Team members.

**Proposal title:**

**Main functionalities:**

**Response summary:** (a few lines)

**Comments on relevance to the CfT (Requirements):**

**Evaluation table:**

*Table 2 – Assessment of submission features*

|  |  |
| --- | --- |
| Note 1 | *Table 3* gives the semantics of submission features. |
| Note 2 | Evaluation Elements indicate the elements used by the evaluator in assessing the submission. |
| Note 3 | Final Assessment indicates the ultimate assessment based on the Evaluation Elements. |

|  |  |  |
| --- | --- | --- |
| **Submission features** | **Evaluation Elements** | **Final Assessment** |
| Completeness of description |  |  |
| Understandability |  |  |
| Extensibility |  |  |
| Use of standard technology |  |  |
| Efficiency |  |  |
| Test cases |  |  |
| Maturity of reference implementation |  |  |
| Relative complexity |  |  |

**Content of the criteria table cells:**

Evaluation facts should mention:

1. Not supported / partially supported / fully supported.
2. What supports these facts: submission/presentation/demo.
3. The summary of the facts themselves, e.g., very good in one way, but weak in another.

Final assessment should mention:

1. Possibilities to improve or add to the proposal, e.g., any missing or weak features.
2. How sure the evaluators are, i.e., evidence shown, very likely, very hard to tell, etc.
3. Global evaluation (Not Applicable/ --/ - / + / ++)

**New Use Cases/Requirements Identified:**

(Please describe)

**Evaluation summary:**

1. **Main strong points, qualitatively:**
2. **Main weak points, qualitatively:**
3. **Overall evaluation:** (0/1/2/3/4/5)

0: could not be evaluated

1: proposal is not relevant.

2: proposal is relevant but requires significant more work.

3: proposal is relevant, but with a few changes.

4: proposal has some very good points, so it is a good candidate for standard.

5: proposal is superior in its category, very strongly recommended for inclusion in standard.

**Additional remarks:** (points of importance not covered above.)

The submission features in *Table 2* are explained in the following *Table 3*.

*Table 3 – Explanation of submission features*

|  |  |
| --- | --- |
| **Submission**  **features** | **Criteria** |
| Completeness of description | Evaluators should:   1. Compare the list of requirements (Annex C) with the submission. 2. Check if respondents have described in sufficient detail how the requirements are supported by the proposal.   Note: Submissions will be judged for the merit of what is proposed. A submission on a single technology that is excellent may be considered instead of a submission that is complete but has a less performing technology. |
| Understandability | Evaluators should identify items that are demonstrably unclear (incon­sistencies, sentences with dubious meaning etc.) |
| Extensibility | Evaluators should check if respondent has proposed extensions to the Use Cases.  Note: Extensibility is the capability of the proposed solution to support use cases that are not supported by current requirements. |
| Use of standard Technology | Evaluators should check if new technologies are proposed where widely adopted technologies exist. If this is the case, the merit of the new tech­nology shall be proved. |
| Efficiency | Evaluators should assess power consumption, computational speed, computational complexity. |
| Test cases | Evaluators should report whether a proposal contains suggestions for testing the technologies proposed. |
| Maturity of reference implementation | Evaluators should assess the maturity of the proposal.  Note1: Maturity is measured by the completeness, i.e., having all the necessary information and appropriate parts of the HW/SW implementation of the submission disclosed.  Note2: If there are parts of the implementation that are not disclosed but demonstrated, they will be considered if and only if such components can be replicated. |
| Relative complexity | Evaluators should identify issues that would make it difficult to implement the proposal compared to the state of the art. |
| Support of non MPAI-UFV use cases | Evaluators should check whether the technologies proposed can demonstrably be used in other significantly different use cases. |

# Annex C: Check list of data formats proposed by a respondent

Note: The numbers in the first column refer to the Functional Requirementsof [4].

Table 5 - Table of response areas

|  |  |
| --- | --- |
| **UFV use cases** | **Response** |
| Use Case 1 – Enhancing visuals in real-time: Up-sampling on eSports streaming | Y/N |
| Use Case 2 – Expanding channel capacity with up-sampling | Y/N |
| Use Case 3 – UAV Skylens | Y/N |
| **UFV requirements areas** | **Response** |
| 1. The Functional Requirements apply to the …. For instance: | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |
|  | Y/N |

Respondent should in any case review the equivalent list in the table of contents of [4].

# Annex D: Technologies that may require specific testing

Table 6 will be compiled based on the responses received.

Table 6 – Functional Requirements that may require specific testing

|  |  |  |
| --- | --- | --- |
| **Section** | **Technology** | **Nature of Test** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Annex E: Mandatory text in responses

**A response to Call for Technologies: Up-sampling Filter for Video applications (MPAI-UFV) shall mandatorily include the following text**

*<Company/Member>* submits this technical document in response to UFV (Nxyz1).

*<Company/Member>* explicitly agrees to the steps of the MPAI standards development process defined in Annex 1 to the [MPAI Statutes](https://mpai.community/about/statutes/) (N421), in particular *<Company/Member>* declares that  *<Company/Member>* or its successors will make available the terms of the Licence related to its Essential Patents according to Framework Licence: Up-sampling Filter for Video applications (MPAI-UFV) (Nxyz3), alone or jointly with other IPR holders after the approval of the planned UFV by the General Assembly and in no event after commercial implementations of UFV become available on the market.

**In case the respondent is a non-MPAI member, the submission shall mandatorily include the following text:**

If (a part of) this submission is identified for inclusion in a specification, *<Company>* understands that *<Company>* will be requested to immediately join MPAI and that, if *<Company>* elects not to join MPAI, this submission will be discarded.

**Subsequent technical contribution shall mandatorily include this text**

*<Member>* submits this document to MPAI as a con­tribution to the development of the planned UFV.

*<Member>* explicitly agrees to the steps of the MPAI standards development process defined in Annex 1 to the [MPAI Statutes](https://mpai.community/about/statutes/) (N421), in particular  *<Company>* declares that *<Company>* or its successors will make available the terms of the Licence related to its Essential Patents according to the UFV Framework Licence (Nxyz3), alone or jointly with other IPR holders after the approval of the UFV Technical Specification by the General Assembly and in no event after UFV commercial implementations become available on the market.