 ISO/IEC JTC 1/SC 29/AG 5 N122

**ISO/IEC JTC 1/SC 29/AG 5**

**MPEG Visual Quality Assessment**

**Convenorship: DE**

**Document type:** Project

**Title:** **Call for learning-based video codecs for study of quality assessment, version 2**

**Status:** Approved

**Date of document:** 2024-05-27

**Source:** ISO/IEC JTC 1/SC 29/AG 5

**Expected action:** Comment/Reply

**Action due date:** 2024-07-08

**No. of pages:** 3 (without this cover page)

**Email of Convenor:** wien@lfb.rwth-aachen.de

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29ag5>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/AG 5**

**MPEG VISUAL QUALITY ASSESSMENT**

**ISO/IEC JTC 1/SC 29/AG 5 N122**

**Rennes, FR – April 2024**

|  |  |
| --- | --- |
| **Title** | Call for learning-based video codecs for study of quality assessment, version 2 |
| **Source** | AG 5 MPEG visual quality assessment |
| **Status** | Approved |
| **Serial Number** | 23974 |
| **Authors** | Yan Ye (Alibaba), Mathias Wien (RWTH Aachen University) |

# Introduction

ISO/IEC JTC 1/SC 29/AG 5 MPEG visual quality assessment has been studying the correlation between subjective quality and a variety of objective quality metrics for video content [1]. In order to facilitate this study, AG 5 has recently established the “compressed video for study of quality metrics” (CVQM) dataset [2]. The CVQM dataset contains 33 original (i.e. uncompressed) video sequences, as well as the compressed versions of these sequences that cover a wide range of bit rate vs. reconstructed quality tradeoff points. So far, the compressed video bitstreams in the CVQM dataset have been coded using the High Efficiency Video Coding (HEVC) and Versatile Video Coding (VVC) standards [3][4] developed by the Joint Video Experts Team (JVET) of ISO/IEC JTC 1/SC 29 and ITU-T SG16. The HEVC and VVC standards are based on the traditional block-based hybrid video coding framework. Common coding artefacts produced by such a system include blocking, blurring, ringing, etc., due to the use of block-based prediction, block transforms, and quantization in the hybrid coding framework (somewhat mitigated by in-loop filtering processes such as deblocking filtering).

Due to their recent rapid advancement, artificial intelligence and deep neural network technologies have found their use in many applications, including in machine learning-based video compression. Many of the learning-based video compression algorithms are trained in an end-to-end manner, and are based on system architectures fundamentally different from the hybrid video coding scheme [5][6][7]. Thus, it can be reasonably anticipated that the distortion characteristics of the reconstructed video generated by learning-based compression schemes may meaningfully diverge from those stemming from traditional coding schemes. In order to facilitate deeper understanding of such distortions and their potential impact on visual quality, AG 5 is issuing this Call for learning-based video codecs for study of quality assessment. This Call seeks responses from researchers who have developed learning-based video codecs and are willing to share compressed bitstreams with AG 5 for investigation of how distortions unique to learning-based video codecs may impact visual quality assessment. Details on the information requested from prospective responses are provided in Section 2. This Call was first released in November 2023 [8]. In this second version of the Call, further clarifications are provided to proponents regarding expectation on network models used in their proposals. Future updates to this Call are anticipated in order to reflect the continuous evolution of learning-based video compression algorithms.

AG 5 welcomes input contributions in response to this Call. Upon evaluation of the received input contributions, AG 5 will invite qualified response(s) to submit compressed bitstreams for inclusion into the CVQM dataset. Factors to be considered when determining the qualifications of a response include but are not limited to the following:

* Whether the compression algorithm in the response is likely to produce distortions unique to learning-based video codecs (e.g., autoencoders, transformers, and generative networks);
* Whether the compression performance is expected to be comparable or superior to that of widely-used codecs such as HEVC and VVC under test conditions commonly used in JVET;
* Whether the neural network models and the encoding/decoding software of the response can be made available to AG 5 for reproducibility of coding performance, coded bitstreams and reconstructed video.

It is noted that the purpose of the CVQM dataset is to provide a collection of processed video sequences with relevant distortion characteristics. There is no intent to deliver any ranking of compression schemes with respect to their compression performance. Upon inclusion of a proposed learning-based codec into the CVQM, AG 5 plans to conduct formal subjective tests on the reconstructed video for inclusion in the dataset, and to subsequently release and potentially formally publish the subjective test results in the context of studying quality metrics.

# Requirements of this Call

Responses to this Call are requested to be registered as an AG 5 input contribution to the 14th AG 5 meeting (see timeline in Section 3). The following information is requested of a response:

* **Summary statement:** please tell us why AG 5 should consider the proposed learning-based video codec, e.g., does your algorithm achieve superior or competitive coding performance? Have you observed unique distortion characteristics produced by your algorithm? To what extent does your algorithm represent a major category of research work on learning-based video compression?
* **Algorithm description:** please provide a detailed description of the system architecture and neural network models to facilitate in-depth understanding of the algorithm. Use and citation of scientific publications for this purpose are encouraged.
* **Compression performance:** please provide compression performance data, preferably in meaningful comparison to well-known MPEG codecs such as HEVC and VVC. It is recommended to use configurations similar to those outlined in JVET common test conditions for traditional video codecs [8][9] (e.g., not to disable or change certain parameter settings for the HEVC or VVC anchors just because your proposed learning-based video codec is not capable of providing the same capability, such as coding video in the 4:2:0 chroma format). Further, rather than requiring a different network model for each of the rate points in the CVQM dataset, it is preferable (though not required) for a network model in the proposal to be able to cover a range of bit rates;
* **Information regarding software copyright:** please tell us what kind of software copyright license can be provided for the neural network models, the encoding and decoding software, etc., that are used in your proposal. In the context of this Call, there is no intent to modify the software or the models. The purpose of requesting copyright licensing information is for AG 5 to assess the level of reproducibility and potential public release of the test results and the reconstructed video sequences.

# Timeline and contact

Those interested in responding to this Call are requested to contact AG 5 Convenor Mathias Wien at [wien@lfb.rwth-aachen.de](mailto:wien@lfb.rwth-aachen.de) prior to the next AG 5 meeting to express their intent to submit a response. The upcoming 16th AG 5 meeting is scheduled as a face-to-face meeting in Sapporo, Japan during July 12 to July 19, 2024. The deadline for registration and upload of an input document that corresponds to your response to this Call is 2024-07-08 (Monday) 23:59h UTC.

# References

1. ISO/IEC JTC 1/SC 29/AG 5, “Learning-based quality metrics and 2D video databases under study (v3),” doc. no. AG5N00074, December 2022.
2. ISO/IEC JTC 1/SC 29/AG 5, “Dataset of compressed video for study of quality metrics (draft 1),” doc. no. AG5N00095, September 2023.
3. ITU-T Rec. H.265 and ISO/IEC 23008-2, “High efficiency video coding,” September 2023.
4. ITU-T Rec. H.266 and ISO/IEC 23090-3, “Versatile video coding,” September 2023.
5. D. Liu, et. al., “Deep learning-based video coding: A review and a case study,” ACM Computing Surveys, Vol. 53, Issue 1, February 2020.
6. S. Ma, et. al., “Image and video compression with neural networks: A review,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 30, Issue 6, June 2020.
7. X. Pan, et. al., “Brief introduction of learning-based video compression,” MPEG doc. no. m65256, October 2023.
8. ISO/IEC JTC 1/SC 29/AG 5, “Call for learning-based video codecs for study of quality assessment, version 1,” doc. no. AG5N00104, November 2024.
9. F. Bossen, X. Li, V. Seregin, K. Sharman, K. Sühring, “VTM and HM common test conditions and software reference configurations for SDR 4:2:0 10 bit video,” JVET doc. no. JVET-AB2010.
10. M. Karczewicz, Y. Ye, “Common test conditions and evaluation procedures for enhanced compression tool testing,” JVET doc. no. JVET-AE2017.