|  |
| --- |
| **INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ORGANISATION INTERNATIONALE DE NORMALISATION ISO/IEC JTC 1/SC 29/WG 5 MPEG JOINT VIDEO EXPERTS TEAM WITH ITU-T SG 16** |
| **ISO/IEC JTC 1 / SC 29 / WG 5 N 266** |
| **Online – 17–26 January 2024** |
| |  |  | | --- | --- | | **Source:** | **Convenor (Jens-Rainer Ohm)** | | **Title:** | **Call for training materials for neural network-based video coding tool development** | | **Type:** | **General** | | **Subtype:** | **Other** | | **Status:** | **Approved** | | **Date:** | **2024-02-09** | | **Expected Action:** | **Info** | | **Action due date:** | **N/A** | | **Pages:** | **3** (not including this cover page) | | **Email of convenor:** | **ohm @ ient . rwth-aachen . de** | | **Committee URL:** | **https://sd.iso.org/documents/ui/#!/browse/iso/iso-iec-jtc-1/iso-iec-jtc-1-sc-29/iso-iec-jtc-1-sc-29-wg-5** | |

|  |  |
| --- | --- |
| **Joint Video Experts Team (JVET)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29**  33rd Meeting, by teleconference, 17–26 January 2024 | Document: JVET-AG2036 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Call for training materials for neural network-based video coding tool development** | | |
| *Status:* | Output document to JVET | | |
| *Author(s) or Contact(s):* | E. Alshina, F. Galpin, S. Liu, and M. Wien | Email: | [elena.alshina@huawei.com](about:blank)  [franck.galpin@interdigital.com](mailto:franck.galpin@interdigital.com)  [shanl@global.tencent.com](mailto:shanl@global.tencent.com)  [wien@lfb.rwth-aachen.de](mailto:wien@lfb.rwth-aachen.de) |
| *Source:* | Call coordinators | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This document provides a draft call for video training materials intended for usage in the development of neural network-based video coding technologies.

# Introduction

JVET started exploration on neural network-based video coding (NNVC) tools and the potential towards standardization in early 2018. The development of this new type of coding algorithms requires training on large dataset. Some progress of NNVC has been achieved both in developing compression algorithms and creating common training databases. Through this exploration study, JVET has learned that both quality and quantity of training data are critically important for development of efficient learning-based methods for video coding. JVET anticipates that further progress of NNVC would benefit from extended video training dataset.

In order to enable more rigorous activities of NNVC development, testing and training materials are sought. One of the contents of high interest for NNVC development is computer generated content featuring high degree of complex motion. Owners of such video materials are therefore kindly invited to consider providing their contents to JVET for the purpose of development, testing and training. Technical specifications of the materials that we are seeking are given in Section 2. If you are willing to provide such material, please refer to the logistics and ownership information given in Section 3.

# Training material sought

Suitable training material for the development of NNVC tools is sought. The following attributes are the main focus for the training material.

## Sequence formats and frame rates

Training materials in the following sequence formats (picture resolutions) are sought.

* UHD 8K
* UHD 4K
* 1080p
* 720p

For neural network methods training data augmentation is typically used. Patches are extracted from higher resolution material by cropping or/and downsampling. It is preferred that the licence terms allow downsampling, cropping, change of format, compression and other types of processing of the provided material during training.

A wide range of frame rates such as 24/50/60/72/100/120 frames per second and beyond are sought. Preservation of native frame rates and picture resolutions is desirable. Uncompressed material or material with a low amount of compression (visually lossless) are desirable. Heavy compression should be avoided. Changes from native resolution or/and frame rate, as well as light compression (if used) should be described.

## Colour spaces and colour sampling

## It is desirable that the test material be available in an RGB colour space in full-resolution colour sampling format (4:4:4). However, suitable training material in a YCC colour space and with a lower colour sampling format such as 4:2:0 will also be welcome. The colour space applicable to the submitted material and potentially applied conversion procedures from a source format are requested to be documented

## Bit depth

Training material is desired to have a bit depth of 10 bits per component or higher, such as 12, 14 or 16 bits per component. Training material with bit depth of 8 bits per component is also sought.

## Metadata

Metadata that are available should be provided. This could include:

* Information about camera type or generation method (for synthetic video).
* Extrinsic (position) and intrinsic camera parameters from calibration or external measurements.
* Information about illumination / lighting conditions.
* Information about lens focus and opening.
* Information about any processing applied (e.g. colour grading).
* Information about chroma sample position (in YCC).
* Information about colour space.

## Scanning methods

The test material must be progressively scanned (not interlace-scanned).

## Sequence length

The length of the training material should be at least 1 second long, and preferably longer.

## Content quality

The sequences should be captured with cameras that reflect the state of the art in the application domain that the sequences are intended for, e.g. in terms of low-noise properties. Synthetic video data are also welcome.

## Content diversity

The test material should cover a large variety of characteristics and content types, including content that is challenging for a typical video codec. The following are some possible examples of such characteristics:

* irregular camera motion, zoom, rotation, and change of camera viewpoint (pan, dolly etc.),
* non-rigid object motion, irregular motion trajectories,
* challenging texture structures,
* saturated colour,
* large contrast/dynamic range,
* slow illumination changes, as well as flash,
* different focus planes within the same picture,
* fades/cross fades,
* time lapse,
* unstable or stabilized camera motion,
* film grain and/or noise (from movie/TV/surveillance/watermarking typical content),
* wide field of view (fish-eye/360 degree video etc.), and
* non-camera captured content (animation, gaming, screen content, synthetic video etc.)

Typical content from important use cases is also desired to cover wider range of sequence characteristics.

Content examples include sports (such as soccer, basketball), nature (such as falling snow, rain, moving plants and animals), people and faces, music video, performances, scenes from movies, video with scrolling text, gaming, computer generated content, user-generated content, panoramic video, VR, surveillance.

# Logistics

Prospective contributors of test sequences should refer to the following contact:

**Mathias Wien**

Convenor of ISO/IEC JTC 1/SC 29/AG 5 MPEG Visual Quality Assessment

[wien@lfb.rwth-aachen.de](mailto:wien@lfb.rwth-aachen.de)

Institute of Imaging and Computer Vision

RWTH Aachen University

52056 Aachen

Germany

Tel. +49-241-80-27867

Various delivery formats can be negotiated.

Please consider clarifying conditions and copyrights under which the sequences can at least be used by members of MPEG and VCEG or by other standardization bodies that may cooperate with JVET during the development of standards. We would prefer allowance of usage in the context of our work that is as unlimited as possible (e.g. allowing cropping, resampling, format changes, re-hosting and redistribution, public demonstrations of technology, snapshots and test results in academic publications etc.) Allowance of usage beyond that purpose is desirable, but not mandatory. Please clarify any restrictions that may apply; these will be documented in JVET databases of training materials. If needed, the above contact is able to provide you with example copyright agreements that have been used in similar cases in the past.