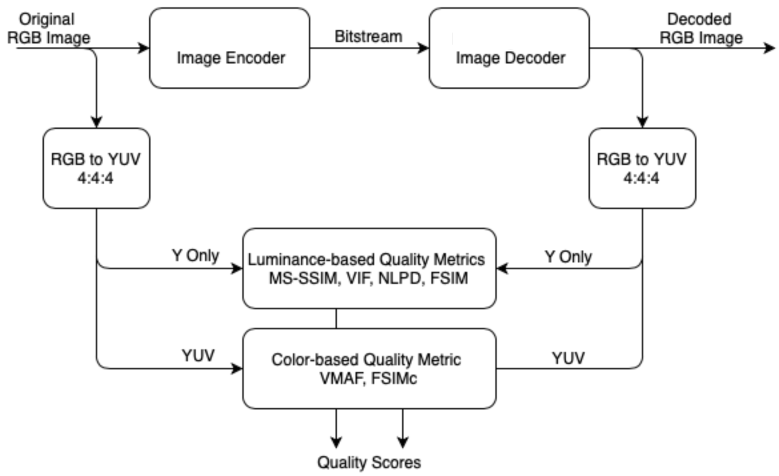


*Figure 1 – Encoding-decoding pipeline for image coding architeture.*

# 1 Target Rates

Target bitrates for the objective evaluations include $b_1$, $b_2$, $b_3$, … bpp. The maximum bitrate deviation from the target bitrate should not exceed x% (e.g. 15%). The proponents must declare for every test image which target bitrate their decoder and models can reach, and in case of deviation of the target bitrate, the proposed RD point may not be considered for evaluation. The target bitrates for the subjective evaluations will be a subset of the target bitrates for the objective evaluations and will depend on the complexity of the test images.

The bitrates specified should account for the total number of bits necessary for generating the encoded file (or files) out of which the decoder can reconstruct a lossy version of the entire image. The main rate metric is the number of bits per pixel (bpp) defined as:

$$BPP = \frac{N\_TOT\_BITS}{N\_TOT\_PIXELS}$$

where N_TOT_BITS is the number of bits for the compressed representation of the image and N_TOT_PIXELS is the number of pixels in the image.

# 2 Objective Quality Evaluation

Objective quality testing shall be done by computing several quality metrics, including MS-SSIM, VMAF, VIFP, NLPD, FSIM, between compressed and original image sequences, at the target bitrates mentioned in the precious Section. This section defines the objective image quality metrics that will be used for the assessment of image coding solutions.

## 2.1 MS-SSIM Definition and Computation

Multi-Scale Structural SIMilarity (MS-SSIM) [1] is one of the most well-known image quality evaluation algorithms and computes relative quality scores between the reference and distorted images by comparing details across resolutions, providing high performance for image codecs. The MS-SSIM [1] is more flexible than single-scale methods such as SSIM by including variations of image resolution and viewing conditions. Also, the MS-SSIM metric introduces an image synthesis-based approach to calibrate the parameters that weight the relative importance between different scales. A high score expresses better image quality.

The source code of this metric can be downloaded at this link:
https://ece.uwaterloo.ca/~z70wang/research/iwssim/.

## 2.2 VMAF Definition and Computation

The Video Multimethod Assessment Fusion (VMAF) metric [2] developed by Netflix is focused on artifacts created by compression and rescaling and estimates the quality score by computing scores from several quality assessment algorithms and fusing them with a support vector machine (SVM). Even if this metric is specific for videos, it can also be used to evaluate the quality of single images and has been proved that performs reasonably well for image codecs. Since the metric takes as input raw images in the YUV color space format, the PNG (RGB color space) images are converted to the YUV 4:4:4 format using FFMPEG (BT.709 primaries). A higher score of this metric indicates better image quality.

The source code of this metric can be downloaded at this link:
https://github.com/Netflix/vmaf

## 2.3 VIF Definition and Computation

The Visual Information Fidelity (VIF) [3] measures the loss of human perceived information in some degradation process, e.g. image compression. VIF exploits the natural scene statistics to evaluate information fidelity and is related to the Shannon mutual information between the degraded and original pristine image. The VIF metric operates in the wavelet domain and many experiments found that the metric values agree well with the human response, which also occurs for image codecs. A high score expresses better image quality.

The source code of this metric can be downloaded at this link:
https://live.ece.utexas.edu/research/Quality/VIF.htm

## 2.4 NLP Definition and Computation

The Normalized Laplacian Pyramid (NLPD) is an image quality metric [4] based on two different aspects associated with the human visual system: local luminance subtraction and local contrast gain control. NLP exploits a Laplacian pyramid decomposition and a local normalization factor. The metric value is computed in the normalized Laplacian domain, this means that the quality of the distorted image relative to its reference is the root mean squared error in some weight-normalized Laplacian domain. A lower score express better image quality.

The source code of this metric can be downloaded at this link:
http://www.cns.nyu.edu/~lcv/NLPyr/

### 2.5 FSIM Definition and Computation

The feature similarity (FSIM) metric [5] is based on the computation of two low level features that play complementary roles in the characterization of the image quality and reflects different aspects of the human visual system: 1) the phase congruency (PC), which is a dimensionless feature that accounts for the importance of the local structure and the image gradient magnitude (GM) feature to account for contrast information. Both color and luminance versions of the FSIM metric will be used. A high metric value express better image quality.

The source code of this metric can be downloaded at this link:

https://www4.comp.polyu.edu.hk/~cslzhang/IQA/FSIM/FSIM.htm

## 3 Subjective Quality Evaluation

To evaluate the selected coding solutions, a subjective quality assessment methodology should be used. This type of assessment is especially critical since the type of artifacts that the image compression solutions produce may be significantly different from those in standard image codecs. Subjective quality evaluation of the compressed images will be performed on the test dataset.

The Double Stimulus Continuous Quality Scale (DSCQS) methodology will be used, where subjects watch side by side the original image and the impaired decoded image and both are scored in a continuous scale. This scale is divided into five equal lengths which correspond to the normal ITU-R five-point quality scale, notably Excellent, Good, Fair, Poor and Bad. This method requires the assessment of both original and impaired versions of each test image. The observers are not told which one is the reference image and the position of the reference image is changed in pseudo-random order. The subjects assess the overall quality of the original and decoded images by inserting a mark on a vertical scale. The vertical scales are printed in pairs to accommodate the double presentation of each test picture.

The subjective test methodology will follow BT500.13 [6] and a randomized presentation order for the stimuli, as described in ITU-T P.910 [7] will be used; the same content is never displayed consecutively. There is no presentation or voting time limit. A training session should be organized before the experiment to familiarize participants with artefacts and distortions in the test images. At least, three training images will be used before actual scoring.

The images used for subjective evaluation are a subset of the test dataset images and its number will be selected depending on the number of proposals that will be subjectively evaluated. Moreover, four bitrate points covering a wide range of qualities will be used in the subjective evaluation and an expert viewing session may be organized to select bitrates, namely, to cover a significant range of qualities. The images to be used in the subjective evaluation will correspond to crops of the decoded images such that relevant coding artifacts are included.

To perform the tests, a semi-controlled crowdsourcing setup framework and/or a more controlled lab environment procedure can be used to show the images according to the DSCQS methodology. The semi-controlled crowdsourcing setup has been proven in the past its reliability, i.e. maintains a low variance of the scores [8]. The Amazon Mechanical Turk or other similar platform will be used for crowdsourcing. The QualityCrowd2 [9] software and Amazon Mechanical Turk (or other similar platform) will be used for crowdsourcing. Due to the COVID-19 pandemic, subjective evaluation may be only performed following a crowdsourcing approach.

# 4 Complexity evaluation

The following complexity metrics should be computed:

- Number of parameters (weights) of the proposed model and their precision.
- Running time with CPU only.
- Running time with GPU enabled.
- Specifications of the CPU and GPU.

While the first three metrics characterize the coding solution itself, the last one depend on the coding solution implementation and the running platform.

These complexity metrics should be accounted during testing (encoding and decoding processes). The complexity of the training process is less relevant for the purpose of evaluating the image coding solution and may optionally be reported.

# 5 Anchors Generation

This section describes the anchor generation process. As anchors, JPEG, JPEG 2000 and HEVC will be used. The list of anchors may be reduced if the number of proposals is too high.

- JPEG (ISO/IEC 10918-1 | ITU-T Rec. T.81)
- JPEG 2000 (ISO/IEC 15444-1 | ITU-T Rec. T.800)
- HEVC Intra (ISO/IEC 23008-2 | ITU-T Rec. H.265)

Information on available software and configurations to be used for these anchors is described next. The target bitrates for the *objective* evaluations are the same as Section 4.

## 5.1 JPEG Anchor

JPEG does not specify a rate allocation mechanism allowing to target a specific bitrate. Hence, an external rate control loop is required to achieve the targeted bitrate. The following conditions apply:

- Available software: JPEG XT reference software, v1.53
  - Available at http://jpeg.org/jpegxt/software.html.
  - License: GPLv3
- Command-line examples (to use within the rate-control loop):
  - jpeg -q [QUALITY_PARAMETER] -h -qt 3 -s 1x1,2x2,2x2 -oz [INPUTFILE] [OUTPUTFILE]
  - where the h is to optimize Huffman tables -qt 3 to select visually improved quantization tables, -s 1x1,2x2,2x2 to use 420 subsampling and -oz to use trellis quantization.

## 5.2 JPEG 2000 Anchor

The JPEG 2000 anchor generation should support two configurations: 1) PSNR optimized; and 2) Visually optimized. A target rate can be specified using the –rate [bpp] parameter. The following conditions apply:

- Available software: Kakadu, v7.10.2
  - Available at http://www.kakadusoftware.com.
  - License: demo binaries freely available for non-commercial use

- Command-line examples:
  - **MSE weighted:** kdu_compress -i [INPUTFILE] -o [OUTPUTFILE] -rate [BPP] Qstep=0.001 -tolerance 0 -full -precise
  - **Visually weighted:** kdu_compress -i [INPUTFILE] -o [OUTPUTFILE] -rate [BPP] Qstep=0.001 -tolerance 0 -full -precise -no_weights
  - **Decoding:** kdu_expand -i [INPUTFILE .mj2] -o [OUTPUTFILE .yuv] -precise

### 5.3 HEVC Intra Anchor

For HEVC Intra, an external rate control loop is required to achieve targeted bitrate. The HEVC RD performance for the target bitrates are obtained with the following conditions:

- Available software: HEVC Test Model (HM 16.16)
  - Available at https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.20+SCM-8.8/
  - License: BSD
- FFMPEG will be used to convert the PNG (RGB) to YUV files following the BT.709 primaries according to:
  - ffmpeg -hide_banner -i input.png -pix_fmt yuv444p10le -vf scale=out_color_matrix=bt709 -color_primaries bt709 -color_trc bt709 -colorspace bt709 -y output.yuv
- Configuration files to be used are available here:
  - [https://jpegai.github.io/public/encoder_intra_main_scc_10.cfg](https://jpegai.github.io/public/encoder_intra_main_scc_10.cfg)

## 6 Naming Convention for Decoded Images

The PNG decoded files should adhere to the following naming convention:

<TEAMID>_<IMGID>_TE_<RES>_8bit_sRGB_<BR>.png

with:

- TEAMID is the registration team ID attributed with 2 digits
- IMGID is an identification of the image with 5 digits
- TE is a fixed value which represents it is a test image
- RES is the spatial resolution (width x height)
- Bit depth (which must be 8 bits always)
- Color space (which must be sRGB)
- BR target bitrate for decoded images: YXX (e.g. 1.25 bpp would be '125' and 0.05 would be 005)

## References


[1] Z. Wang, E. P. Simoncelli, and A.C. Bovik, "Multiscale structural similarity for image quality assessment," Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, November 2003.

[2] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy and M. Manohara, "Toward A Practical Perceptual Video Quality Metric", [Online], Available here.

[3] H.R. Sheikh and A. C. Bovik, "Image Information and Visual Quality," IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, Canada, August 2004.

[4] L. Zhang, L. Zhang, X. Mou, D. Zhang, “FSIM: a Feature Similarity Index for Image Quality Assessment,” IEEE Transactions on Image Processing, vol. 20, no. 8, pp. 2378-2386, August 2011.
[5] V. Laparra, J. Ballé, A. Berardino, and E. P. Simoncelli, “Perceptual Image Quality Assessment using a Normalized Laplacian Pyramid”, S&T Symposium on Electronic Imaging: Conf. on Human Vision and Electronic Imaging, San Francisco, CA, USA, February 2016.
[6] ITU-R Recommendation BT.500-13, “Methodology for the subjective assessment of the quality of television pictures,” International Telecommunications Union, Geneva, Switzerland, 2012.
[7] ITU-T Recommendation P. 910, “Subjective video quality assessment methods for multimedia applications,” International Telecommunication Union, Geneva, 2008.
[8] D. Saupe, F. Hahn, V. Hosu, I. Zingman, M. Rana, and S. Li, “Crowdworkers proven useful: a comparative study of subjective video quality assessment,” International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, June 2016.
[9] C. Keimel, J. Habigt, C. Horch, and K. Diepold, "Qualitycrowd: a framework for crowd-based quality evaluation," Picture Coding Symposium, Krakow, Poland, May 2012.