ISO/IEC JTC 1/SC29/WG1 **N91015**
91st Meeting, Online, 19-23 April 2021

ISO/IEC JTC 1/SC 29/WG 1
(ITU-T SG16)

Coding of Still Pictures

JBIG
Joint Bi-level Image Experts Group

JPEG
Joint Photographic Experts Group

TITLE: JPEG AI First Draft of Call for Proposals

SOURCE: WG1

PROJECT: JPEG AI

STATUS: Draft

REQUESTED ACTION: For discussion

DISTRIBUTION: WG1

Contact:
ISO/IEC JTC 1/SC 29/WG 1 Convener – Prof. Touradj Ebrahimi
EPFL/STI/IEL/GR-EB, Station 11, CH-1015 Lausanne, Switzerland
Tel: +41 21 693 2606, Fax: +41 21 693 7600, E-mail: Touradj.Ebrahimi@epfl.ch

ISO/IEC JTC 1/SC29/WG1 N91015
91st Meeting, Online, 19-23 April 2021

First Draft JPEG AI Call for Proposals

Summary

The scope of JPEG AI is the creation of a learning-based image coding standard offering a **single-stream**, **compact** compressed domain representation, targeting both **human visualization**, with significant compression efficiency improvement over image coding standards in common use at equivalent subjective quality, and effective performance for **image processing and computer vision tasks**, with the goal of supporting a **royalty-free baseline**.

JPEG AI targets a wide range of applications such as cloud storage, visual surveillance, autonomous vehicles and devices, image collection storage and management, live monitoring of visual data and media distribution. The objective is to design a coding solution that requires significant compression efficiency improvement over coding standards in common use at equivalent subjective quality and an effective compressed domain processing for machine learning based image processing and computer vision tasks. Other key requirements include hardware/software implementation-friendly encoding and decoding, support for 8- and 10-bit depth, efficient coding of images with text and graphics and progressive decoding.

1. Introduction

Nowadays, image coding is a fundamental technology in our society, used billions of times per day, by a very large percentage of the world population. This includes not only personal pictures, many widely diffused in social networks, but also professional pictures used in many applications and services, such as in stock photo and video streaming sites (e.g. movie covers). Moreover, visual surveillance systems with multiple cameras often capture, analyze and store images, especially when relevant events occur. Also, the current explosion of imaging data brings the need for efficient mining and analysis tools for relevant tasks in the compressed domain.

Since the image resolution and target quality have been growing, their uncompressed size is also growing, thus critically asking for more efficient image coding solutions to facilitate transmission and storage. In this context, lossy image coding solutions capable to achieve higher compression, and thus larger rate savings are necessary. Simultaneously, an efficient compressed domain representation should be pursued not only for human visualization but also for machine image processing and computer vision.

Learning-based solutions are the state of the art for several computer vision tasks, such as those requiring high-level understanding of image semantics, e.g., image classification and object segmentation, but also image processing tasks, such as image denoising and super-resolution. Moreover, learning-based image coding solutions, namely those exploiting deep neural networks, can achieve better compression efficiency than available conventional image coding solutions, e.g., JPEG, JPEG 2000 and HEVC Intra [1]. The impact is that both compression and visual processing tasks can be efficiently performed with a compact image representation model able to represent the wide variety of visual content that is available today.

With its successful and widely adopted coding standards, WG1 is well positioned to undertake a standardization activity to develop a standard for a JPEG AI learning-based image coding system, especially to apply machine learning tools to achieve substantially better compression efficiency than existing image coding systems, along with features desirable for an efficient distribution and consumption of images.

2. Scope

Image coding standards provide interoperability between codecs built by different manufacturers and are nowadays the basis of many products in communication technology.

The scope of JPEG AI is the creation of a learning-based image coding standard offering a single-stream, compact compressed domain representation, targeting both human visualization, with significant compression efficiency improvement over image coding standards in common use at equivalent subjective quality, and effective performance for image processing and computer vision tasks, with the goal of supporting a royalty-free baseline.

The coding system to standardize will likely follow an end-to-end learning-based architecture, notably where analysis and synthesis transforms are learned using an appropriate loss function. It shall also allow to obtain

an efficient compressed domain representation useful not only for visualization, but also for machine image processing and computer vision tasks. Figure 1 shows the high-level JPEG AI framework, which is fully described in the JPEG AI Use Cases and Requirements document (WG1N91014); it includes three pipelines: standard image reconstruction, compressed domain computer vision processing and compressed domain image processing, all from the latent representation that is obtained after entropy decoding.

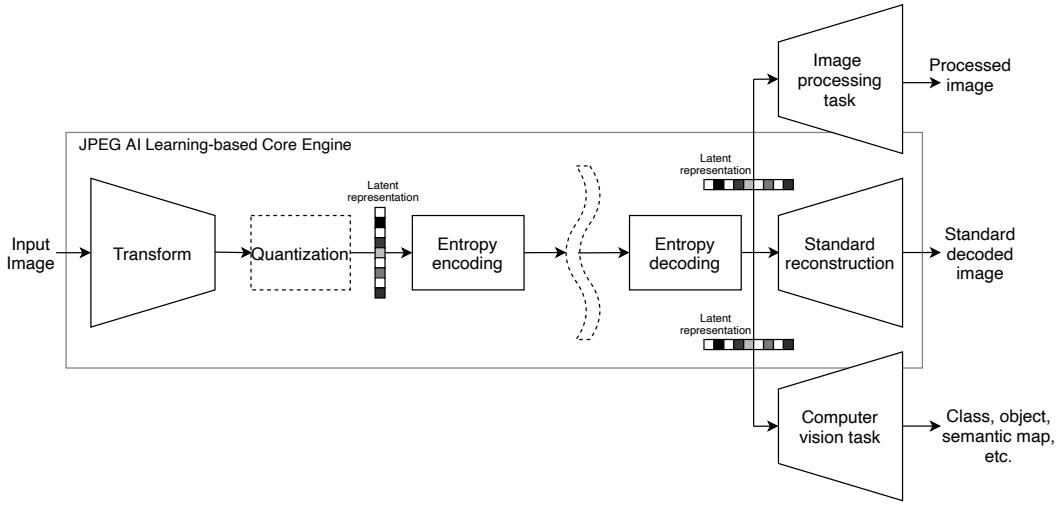

Fig. 1: JPEG AI learning-based image coding framework.

Considering this context, this Call for Proposals (CfP) on JPEG AI Learning-based Image Coding Technologies solicits technical contributions that demonstrate efficient compression of images as well as effective performance for image processing and computer vision tasks.

3. Use Cases and Requirements

This Call for Proposals addresses several use cases:

- Cloud storage
- Visual surveillance
- Autonomous vehicles and devices
- Image collection storage and management
- Live monitoring of visual data
- Media distribution
- Television broadcast distribution and editing

Detailed information on these use cases and derived requirements are contained in the JPEG AI Use Cases and Requirements document (WG1N91014).

4. Submission Requirements

Proponents are asked to submit detailed technical description on the entire image codec, as well as encoder and decoder implementations in software, and the decoded test images. Regarding training material, proponents must use the JPEG AI training dataset, which has a large set of images which can be used for standardization purposes, due to the permissive licensing (CC0); optionally the proponents may use additional data but, in this case, they have to include it in the submission.

Proponents are expected to code the provided test material, to be provided along with their codec, targeting the specified bitrates. In all cases, participants are required to submit material to validate the performance of their submission according to the procedure outlined below, notably:

- A detailed description of the coding algorithm, methodologies, as compression performance alone is not the only evaluation criterion. This item corresponds to a technical document to be submitted to JPEG.
- The codec training procedure must be described as faithful as possible. Expect that WG1 members will reproduce your results by performing the training procedure, obtaining a model and running the encoder and decoder.
- A decoder implementation in a form allowing stand-alone inference/testing on a standard computer (both CPU and GPU) in a reasonable amount of time, preferably in source code form. Decoded images should be in the sRGB color space in the PNG format with 8 or 10 bits per component.
- Compressed bitstreams and corresponding decoded images.
- Information about the running time for some CPU and GPU platform as well as the model size.

Contributors are also expected to provide to JPEG sufficient rights to allow usage of the provided software for the purpose of evaluation. The evaluation process may need to crop and/or clip the provided images to make them suitable for subjective evaluation. The submission requirements for the proposed solutions are detailed in Annex A.

5. Submission Assessment Criteria

The JPEG AI Common Test Conditions (CTC) document defines the training dataset (the test dataset will be provided after), benchmarking codecs, coding conditions (especially target bitrates) and a set of reliable objective quality metrics and subjective assessment procedures. The CTC will allow to exhaustively evaluate multiple aspects of the proposed learning-based image codecs to fully understand their strengths and weaknesses, notably regarding already available image coding technology. The JPEG AI CTC document will be available at the 92nd meeting as a separate document.

The JPEG committee plans to select technologies to be included in the JPEG AI standard based on satisfying the requirements as well as compliance to the JPEG AI CTC, such as target bitrates. The evaluation will be

made based on the results obtained through the evaluation procedure documented in JPEG AI CTC. Several criteria and results will be used for the evaluation of the submissions and the decision-making process:

- Subjective quality evaluation for standard reconstruction with the proposed codecs under the CTC. The decoded images to be evaluated will be obtained by running the provided decoder with submitted bitstreams.
- Objective quality evaluation for standard reconstruction with the quality metrics defined in the CTC.
- Complexity evaluation of both encoding and decoding process.
- Performance evaluation for computer vision and image processing tasks as defined in the CTC.

6. Timeline

The intended timeline for the evaluation of the proposals is the following:

April 2021 **91st JPEG Meeting**	First Draft JPEG AI Call for Proposals (this document) and JPEG AI Use Cases and Requirements document (WG1N91014).
July 2021 **92nd JPEG Meeting**	Second Draft JPEG AI Call for Proposals (this document), JPEG AI Use Cases and Requirements document (WG1N92XXX) and JPEG AI Common Test Conditions (WG1N92XXX.).
16-22 October 2021 **93rd JPEG Meeting**	Final JPEG AI Call for Proposals. Release of the training and validation parts of the datasets.
5th November 2021	Proposal registration.
10th December 2021	Submission of decoder implementation with some fixed model. No (re)training is allowed after this date.
15th December 2021	Release of the test dataset for proponents to code.
10th January 2021	CTC dry run of objective and subjective performance assessment with anchors.
17-21 January 2022 **94th JPEG Meeting**	Analysis of the results of the dry run, may issue final recommendations for proposal evaluation.
30th January 2021	Submission of bitstreams and decoded images for the test dataset. Objective and subjective evaluation of all the proposals starts.
23-29 April 2022 **95th JPEG Meeting**	JPEG AI proposals submission. Presentation and discussion of the proposals at JPEG meeting. Attendance is mandatory for proponents.

The intended timeline for the standardization process is as follows:

2022-07	WD
2023-01	CD
2023-07	DIS
2024-01	IS

7. Submission Composition and Requirements

7.1 Submission Elements

This CfP invites proponents to submit technology contributions that fulfill the scope and objectives according to the timeline presented above. Proponents are expected to present their proposals at the 95th WG1 meeting, April 2022.

Proponents are also reminded that they are expected to contribute to the standardization process, as described in Section 7.4, and attend meetings and present their findings.

7.2 Submission requirements

A submission shall consist of the elements specified in Annex A. All the elements to be submitted, excluding the decoded images, should be uploaded to the WG1 document registry. For the decoded images, instructions will be provided after the registration. Those proponents without access to the registry should contact the WG1 members listed in Section 9.

7.3 IPR conditions (ISO/IEC Directives)

Proponents are advised that this call is being made in the framework and subject to the common patent policy of ITU-T/ITU-R/ISO/IEC and other established policies of these standardization organizations. The contact persons named in Section 9 can assist potential submitters in identifying the relevant policy information.

7.4 Royalty-free goal

The royalty-free patent licensing commitments made by contributors to previous standards, e.g. JPEG 2000 Part 1, have arguably been instrumental to their success. JPEG expects that similar commitments would be helpful for the adoption of a future JPEG AI image coding standard.

7.5 Contribution to Standardization

Proponents are informed that based on the submitted proposals, a standard specification will be created. If they submit a proposal and (part of) the proposed technology is accepted for inclusion in the standard, they will hence have to attend subsequent WG1 meetings and contribute to the creation of the different standard

documents and reference software. Within this process, evolution and changes are possible as several technologies may be combined to obtain a better performing solution.

8. JPEG AI e-mail reflector information

E-mail reflector: jpeg-ai

To subscribe to the reflector, please visit http://listregistration.jpeg.orgor in case of problems contact lists@jpeg.org.

9. Contacts

Touradj Ebrahimi (JPEG Convener)
Email: Touradj.Ebrahimi@epfl.ch

Fernando Pereira (JPEG Requirements Sub-Group Chair)
Email: fp@lx.it.pt

João Ascenso and Elena Alshina (JPEG AI AhG Chairs)
Email: joao.ascenso@lx.it.pt and elena.alshina@huawei.com

References

[1] ISO/IEC JTC 1/SC29/WG1 N89022, "Report on the JPEG AI Call for Evidence Results", 89th Meeting, Online, October 2020.

ANNEX A – SUBMISSION REQUIREMENTS

The process to evaluate proposals will be done following the timeline defined in Section 6.

A.1. Proposal description

Each proposal must include a detailed technical description of the entire image codec, namely:

- Key features of the proposal, including the target quality range and covered bitrates.
- High-level description of the proposal, including the encoder/decoder architecture
- Detailed architectures for every module, training procedure including loss function used, strategies used to handle key parts, such as the non-differentiable nature of quantization and bitrate allocation.
- Model size if applicable, which corresponds to the number of weights (and precision of each weight) in the encoder and decoder.
- Running time (encoder and decoder) for some CPU and GPU platform. Include all the details of the platform (CPU model, clock rate and memory, GPU model and brand) but also the deep learning framework (e.g. Tensorflow or Pytorch). The recommended platform for GPU is NVIDIA 2080 Ti, but you may use other one.

This description shall be in Word document and PDF format. The presentation must clearly explain how the proposed algorithm meets the requirements defined for this call. Providing insufficient details might jeopardize the adoption of the proposed technology.

A.2. Codec implementation, codestreams and decoded material

The following additional elements must be submitted by all proposals:

- Standalone executable package: docker file with all the libraries and tools to run the encoder rand decoder with the submitted code-streams and preferably decoder in source code form. Proponents shall provide the command-line parameters intended to be used for encoding/decoding, and shall provide scripts to run their executable in the objective evaluation framework detailed below. All the information to run the decoder shall be provided, such as command line parameters, configuration files. If binaries are used, they should correspond to statically linked Linux executables with all required libraries and system dependencies.
- Code-streams corresponding to the encoded test images to be used for decoding.
- Decoded test images for objective and subjective evaluation. All test images will be made available to proponents after the decoder submission.
- Any additional dataset used in addition to the JPEG AI dataset.
- The decoder should correctly decode any codestream generated by the submitted encoder.

A.3. Verification model source code

Proponents agree to release source code to serve as (part of) a Verification Model (VM), written in a high-level language, such as Python or C++, if parts of their technology are selected in the evaluation process. The source code provided may include some project files needed to support compilation. Source code shall be documented and understandable. All libraries used by the source code shall be either public or provided in source code form with ISO/IEC and ITU-T compliant terms. The source code to be provided should run on a variety of operating systems (at least on Linux) and hardware GPU boards.