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: Final Call for Proposals for JPEG AI

SOURCE: WG1

PROJECT: JPEG AI (ISO/IEC 6048)

STATUS: Approved

REQUESTED ACTION: For distribution

DISTRIBUTION: Public

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
Table of Contents

Summary 3
1. Introduction 4
2. Scope 5
3. Use Cases and Requirements 6
4. Evaluation Conditions and Procedures 6
5. Timeline 7
6. Proposal Composition and Requirements 8
   6.1. Proposal Components 8
   6.2. Proposal Registration and Delivery 10
   6.3. IPR Conditions (ISO/IEC Directives) 10
   6.4. Contribution to Standardization 10
7. Royalty-free Goal 11
8. Participation 11
9. Contacts 11

References 11

ANNEX A. Submission requirements 12
   A.1. Codec implementation, bitstreams and decoded material 12
   A.2. Verification model source code 12

ANNEX B. Registration Form 13
Final Call for Proposals for JPEG AI

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 offers 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.

The JPEG AI project is a joint standardization effort between ISO/IEC JTC1/SC29/WG1 and ITU-T SG16.
Final Call for Proposals for JPEG AI

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 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, e.g., 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, HEVC Intra and VVC 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 learning-based image coding system, nicknamed JPEG AI, especially to apply machine learning tools to achieve substantially better compression efficiency than existing image coding solutions, along with features desirable for an efficient distribution and consumption of images.
The detailed timeline for the JPEG AI project is defined in Section 5 of this document. The intended overall timeline for the standardization process is as follows:

<table>
<thead>
<tr>
<th>Year</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>2022-10</td>
<td>WD</td>
</tr>
<tr>
<td>2023-04</td>
<td>CD</td>
</tr>
<tr>
<td>2023-10</td>
<td>DIS</td>
</tr>
<tr>
<td>2024-04</td>
<td>IS</td>
</tr>
</tbody>
</table>

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 [2]; it includes three pipelines: standard image reconstruction, compressed-domain image processing and compressed-domain computer vision 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 final Call for Proposals (CfP) for JPEG AI solicits technical contributions that demonstrate efficient compression of images as well as effective performance for image processing and computer vision tasks using a learning-based coding approach.

3. **Use Cases and Requirements**

This Call for Proposals addresses several use cases, notably:

- 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 [2].

4. **Evaluation Conditions and Procedures**

The JPEG AI Common Training and Test Conditions (CTTC) document defines the training dataset, benchmarking codecs, coding conditions (especially target bitrates) and a set of reliable objective quality metrics and subjective assessment procedures [3]. The test dataset will be selected and provided by the JPEG experts João Ascenso, Thomas Richter and Fernando Pereira not involved in any proposal. The CTTC 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 CTTC document is released at the 94th meeting (January 2022) as an output document.

JPEG plans to select technologies to be included in the JPEG AI standard based on satisfying the JPEG AI requirements [2] as well as compliance to the JPEG AI CTTC [3], such as target bitrates. The evaluation will be made based on the results obtained through the evaluation conditions and procedures documented in the JPEG AI CTTC [3]. Several criteria and metrics will be used for the evaluation of submissions and the decision-making process:

- Subjective quality evaluation of the proposed submission for standard reconstruction according to the CTTC. The subjective evaluation results will be the primary attribute for the decision-making process.
- Objective quality evaluation for standard reconstruction with the quality metrics defined in the CTTC.
• Complexity evaluation of both encoding and decoding process according to the metrics defined in the CTTC.
• Performance evaluation for the image processing and computer vision tasks as defined in the CTTC.

The device interoperability requirement states that performance difference between submission operating in different platforms should not be greater than 0.5% BD-rate. While it is accepted to not meet this requirement for the CfP submission, this requirement is mandatory to be met for inclusion in the JPEG AI WD/CD and reference software.

5. Timeline

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

<table>
<thead>
<tr>
<th>Deadline</th>
<th>Action</th>
</tr>
</thead>
</table>
| **January 2022** 94th JPEG Meeting | Release of the final JPEG AI Call for Proposals (WG1N100095).  
Release of the final JPEG AI Use Cases and Requirements (WG1N100094).  
Release of the final JPEG AI Common Training and Test Conditions (WG1N100106).  
Release of the training and validation parts of the datasets, anchor generation software and quality assessment framework. |
| **25th February 2022** | JPEG AI proposals’ registration deadline.  
Registrations should be sent to the JPEG Convener, JPEG Requirements Subgroup Chair, JPEG Image, Coding & Quality Chair (see Section 9) with the form available in Annex B. |
| **10th March 2022** | **CfP 1st phase deadline:** Submission of proposed decoder implementation with some fixed coding model.  
No (re)training is allowed after this date. |
| **15th March 2022** | Release of the test datasets for standard reconstruction tasks (and proponents to code). |
| **30th March 2022** | Release of the test datasets for image processing and computer vision tasks (and proponents to code). |
| **20th April 2022** | Anchors generation and cross-check by JPEG.  
Objective quality assessment and dry run of subjective quality assessment with anchors for standard reconstruction, according to the CTTC. |
| **25-29 April 2022** 95th JPEG Meeting | Analysis of the results of the dry run; may issue final recommendations for JPEG AI proposals’ evaluation. |
Assign cross-checkers for each CfP submission.

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
</table>
| 2nd May 2022          | **CfP 2nd phase deadline**: Submission of proposals’ bitstreams and decoded images and/or labels for the test datasets.  
                         | Start of objective and subjective evaluation of JPEG AI proposals. |
| 23rd May 2022         | Results for decoder cross-check are reported to JPEG.                |
| 18th July 2022        | **CfP 3rd phase deadline**: Submission of source code for the encoder, decoder, training procedure and the proposal description. |
| 25-29 July 2022       | Presentation and discussion of the JPEG AI proposals at JPEG meeting. Attendance is mandatory for proponents. |

6. **Proposal Composition and Requirements**

6.1. **Proposal Components**

Proponents are asked to submit a detailed technical description of the entire image codec, as well as encoder and decoder implementations in software, and the decoded test images. The description shall be in Word document and PDF format. Providing insufficient details might jeopardize the selection of the proposed technology for the collaborative phase of the standardization process.

Regarding training material, proponents must use the JPEG AI training dataset (if a sub-set of the training dataset is used for training, then it should be clearly described), which has a large set of images that can be used for standardization purposes, due to their permissive licensing (CC0); optionally, the proponents may use additional data but, in this case, they have to describe the used data in detail (including the licensing conditions) and include the data and its detailed description in the proposal submission. Moreover, proponents should give to JPEG enough rights to use the submitted dataset in the standardization process. JPEG strongly encourages that, in addition to standard reconstruction (sub-section A), at least one extra image processing or/and computer vision (sub-section B) task is supported by the proposal submission.

A. **Standard Reconstruction Task**

For standard reconstruction, proponents are expected to code the provided test material, to be provided to JPEG along with the decoder implementation, targeting the bitrates specified as mandatory in the CTTC [3]. In all cases, participants are required to submit materials to validate the performance of their submission for the standard reconstruction task 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 with the following information:
Key features of the proposal, including how the core and desirable requirements [2] for this CfP are addressed by the proposed technology.

- High-level description of the proposal, including the encoder(s) and the decoder(s) architectures, the target quality range, the supported bitrates, and the bitrate allocation method used.

- Training methodology adopted, described as faithfully as possible, including how the non-differentiable nature of quantization is addressed. Proponents should expect that JPEG members will reproduce the proposals’ results by performing the training procedure, obtaining a model, and running the encoder and decoder using the provided software.

- RD performance evaluation for standard reconstruction according to the CTTC conditions using all objective quality metrics specified in the CTTC.

- Complexity evaluation according to the CTTC.

- 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 with clear instructions how to build, run software and configure it for CPU and GPU. Decoded images from this implementation should be in the sRGB color space and PNG format, with 8 or 10 bits per component (same as original).

- Compressed bitstreams and corresponding decoded images. The decoding of submitted bitstreams will be made by each proponent in a cross-check fashion, this means that proponent A will decode the bitstreams of proponent B and measure the bitstream size and objective quality according to the CTTC and vice-versa. Decoded images from proponents will be used for subjective quality evaluation unless problems are detected during cross-check. More precisely, in the cross-check process, the difference between the JPEG obtained decoded images (by running the provided decoder with submitted bitstreams) and the decoded images submitted by the participants will be measured. Cross-check is considered to be successful if the BD-rate difference between proponent and cross-checker results is within 0.5%.

- Information about the complexity of the submitted technology according to the CTTC.

B. Image Processing and Computer Vision Tasks

To meet the JPEG AI scope and offer a single-stream compact-domain representation that is useful for multiple purposes, the same encoder as for the standard reconstruction task should be used. Moreover, to obtain the latent representation that can be used by a compressed-domain processor (i.e., the networks responsible for the standard reconstruction, image processing and computer vision tasks), proponents are expected to perform entropy decoding as for standard reconstruction (see Fig. 1). In all cases, participants are required to submit material to validate the performance of their submission for one or more tasks according to the procedure outlined below, notably:
• A detailed description of each compressed-domain processor, namely:
  o Key features, namely which image processing or computer vision task is supported.
  o Training process and all aspects related to the network architecture.
  o Decoder anchor and compressed-domain processor’s performance; for image processing
tasks, this corresponds to the quality and rate for every JPEG AI test image and, for the
computer vision tasks, this corresponds to the rate and accuracy for the test dataset according
to the metrics defined in the CTTC.
  o Complexity evaluation of the decoder anchor and compressed-domain processor according to
the CTTC.
• Compressed-domain processor software implementations for one or more tasks 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.
• Compressed bitstreams and corresponding processed images (for image processing task) or decoded
labels (for computer vision tasks).

For all tasks, proponents are also expected to provide JPEG with sufficient rights to allow usage of the
provided software for the purpose of evaluation. Instructions or/and software header for the submitted
software package should include information about usage rights.

The submission requirements for the proposed solutions are detailed in Annex A.

6.2. Proposal Registration and Delivery
Proponents are requested to register proposals by the deadline specified in Section 5 of this document (25th
February 2022). Registration shall be made by filling the form available in Annex B and sending it by e-mail
to the JPEG Convener, JPEG Requirements Subgroup Chair and the JPEG Image, Coding & Quality Chair,
see contacts in Section 9.

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

6.4. 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 JPEG 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.

7. **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.

8. **Participation**

The Ad Hoc Group on JPEG AI was established at the 82nd JPEG meeting. All interested parties are encouraged to subscribe to the AhG email reflector as follows:

E-mail reflector: jpeg-ai

To subscribe to the reflector, please visit [http://listregistration.jpeg.org](http://listregistration.jpeg.org) or in case of problems contact lists@jpeg.org.

9. **Contacts**

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

Fernando Pereira (JPEG Requirements Subgroup Chair)
Email: fp@lx.it.pt

Thomas Richter (JPEG Image, Coding & Quality Subgroup Chair)
Email: thorfdbgiis@gmail.com

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

**References**


ANNEX A. Submission Requirements

The process to evaluate proposals will be done following the timeline defined in Section 5.

A.1. Codec Implementation, Bitstreams and Decoded Material

The following additional elements must be submitted by all proposals:

- Standalone software package: sufficient description how to build and run the encoder (optionally) and decoder with the submitted bitstreams and preferably decoder in source code form; all the libraries and tools needed should also be included. Proponents shall provide the command-line parameters intended to be used for encoding/decoding for both CPU and GPU, and all necessary scripts and configuration files. If binaries are used, they should correspond to statically linked Linux executables with all required libraries and system dependencies.
- Bitstreams 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 training dataset.

The decoder should correctly decode any bitstream generated by the submitted encoder.

A.2. 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 by 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, namely CPU processors and GPU boards.
ANNEX B. Registration Form

Proposal name:
__________________________________________________________

Proponents’ names and contact information:

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Choose all that apply with a cross:

**Submission tracks** | Selection
---|---
A. Standard Reconstruction (mandatory) | X
B. Compressed Domain Super-resolution
C. Compressed Domain Denoising
D. Compressed Domain Image Classification

Training dataset submission (Yes/No): _____________

Machine learning platform (e.g. Tensorflow, PyTorch or other), including version:
_________________________________________________________________

Any notes for the cross-checker: