White Paper on

JPEG AI Scope and Framework

Version 1.0
January 2021

Editors:

João Ascenso – Instituto Superior Técnico – Instituto de Telecomunicações, Portugal
Evgeniy Upenik – École Polytechnique Fédérale de Lausanne, Switzerland

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: convenor@jpeg.org

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's population. This includes not only personal pictures, many widely distributed in social networks, but also in professional 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 frequently, especially when events of interest occur. Also, the growing explosion of imaging data justifies the need for efficient mining and analysis technologies conducted in the compressed domain.

Since image resolution and target quality have been increasing, the uncompressed size of images is also increasing, thus critically asking for efficient image coding solutions to further facilitate transmission and storage. In this context, lossy image coding solutions able to achieve high compression ratios, and thus large rate savings are necessary. Simultaneously, an efficient compressed domain representation should be pursued not only for visualization, but also for effective image processing and computer vision tasks.

Scope

The scope of the 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 Framework

Learning-based image coding solutions have already shown that they can achieve substantially better compression efficiency than existing conventional solutions, namely by exploiting advanced machine learning tools, such as deep neural networks [1]. In particular, it has been shown that, when compared to JPEG, JPEG 2000 and HEVC Intra, learning-based coding solutions can provide better perceptual quality, for some target bitrates, both in terms of appropriate perceptual objective quality metrics and subjective assessment scores [2]. Besides their high compression efficiency, learning-based image coding solutions may be adapted with little extra effort to image processing and computer vision tasks without the need for full decoding, i.e., without performing image reconstruction. This contrasts with classical image codecs, which when used in image processing and computer vision pipelines, need to perform full decoding of the compressed bitstream to obtain a pixel-based representation.

Figure 1 shows the high-level JPEG AI framework, highlighting the three pipelines. The input to the learning-based image coding framework is a digital image and the output bitstream may be processed for human

visualization by performing entropy decoding and standard reconstruction, thus producing a standard decoded image. As shown in Figure 1, the standard reconstruction may be skipped since the latent representation produced by the encoder contains the necessary information not only for decoding but also to perform image processing and computer vision tasks at the decoder side (after entropy decoding). These tasks are carried out on the latent representation, directly extracted from the original image and not from the (lossy) decoded image. This intrinsically feature-rich latent representation can be used in two main ways: 1) to perform an image processing task, such as targeting the enhancement or modification of the image, where a processed image is produced, for example with increased resolution, contrast, etc.; and 2) to perform a computer vision task where high-level semantic information is extracted, e.g., to generate classes, labels, regions, etc.

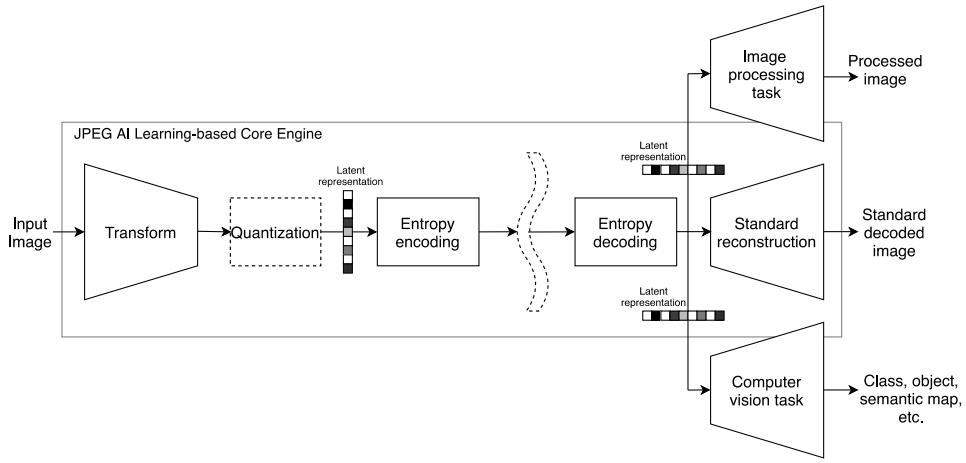

Fig. 1: JPEG AI learning-based image coding framework.

Key Tasks

Following the JPEG AI scope, the coded bitstream will have a triple-purpose, thus offering compelling advantages for applications where an image processing task aims to enhance or modify the image or where semantic (or higher-level) information needs to be extracted from large amounts of visual data. This may have a significant impact on image processing and computer vision tasks, which may be performed with lower complexity by using as input the compressed domain representation instead the original or decoded images. Some examples of relevant image processing tasks are:

- Super-resolution
- Denoising
- Low-light enhancement
- Color correction
- Exposure compensation
- Inpainting

Relevant examples of computer vision tasks are:

- Image retrieval and classification
- Object detection, recognition and identification

- Semantic segmentation
- Event detection and action recognition
- Face detection and recognition

Example Use Cases

While efficient image coding is relevant for multiple applications, the new learning-based image coding approach may be especially relevant for the following use cases:

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

The list of identified use cases will allow the derivation of the requirements to be met by the JPEG AI standard; the current version of the JPEG AI Use Cases and Requirements document is available at the JPEG web site at jpeg.org [3].

To stay informed on JPEG AI activities, please regularly consult our website at jpeg.org and/or subscribe to the JPEG AI e-mail reflector.

References

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

[2] ISO/IEC JTC 1/SC29/WG1 N85013, "Performance Evaluation of Learning based Image Coding Solutions and Quality Metrics", 85th JPEG Meeting, San Jose, USA, November 2019.

[3] ISO/IEC JTC 1/SC29/WG1 N90021, " JPEG AI Use Cases and Requirements ", 90th JPEG Meeting, Online, January 2021.