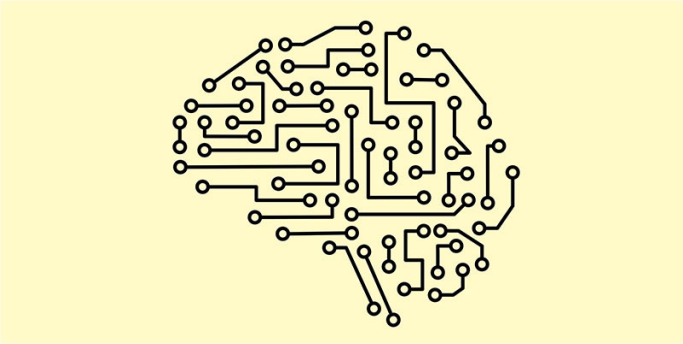


JPEG AI Learning based Image Coding: its Current Status and its Roadmap

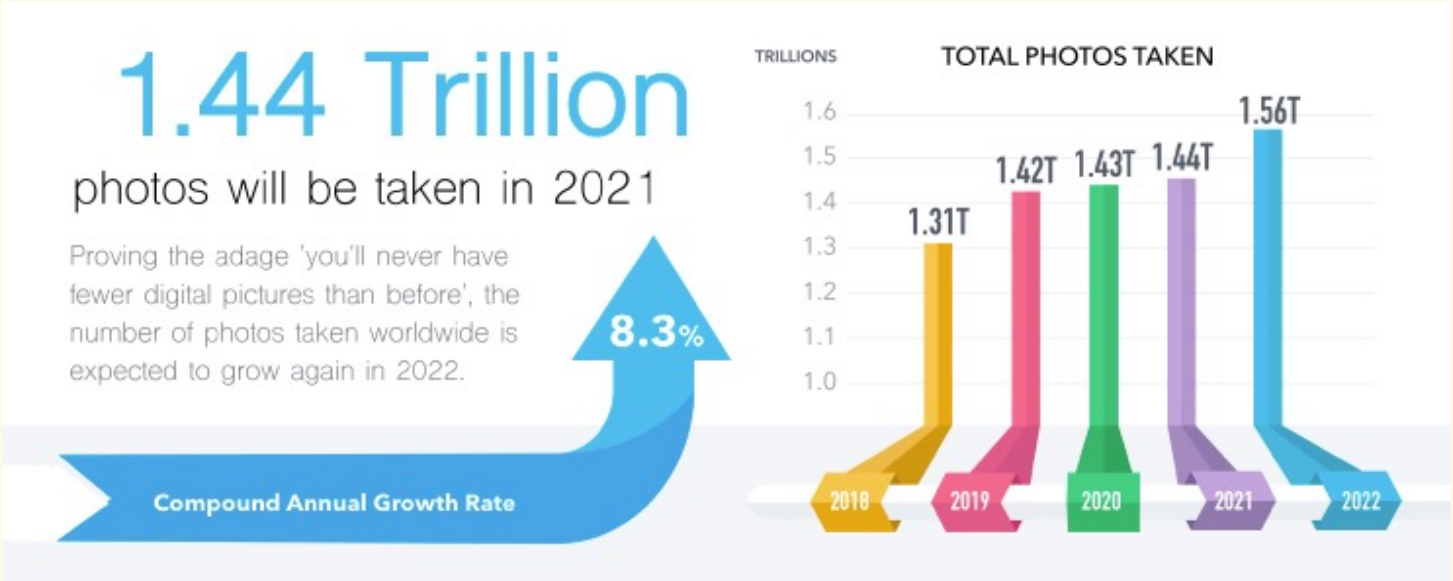
João Ascenso, Instituto Superior Técnico, Lisbon

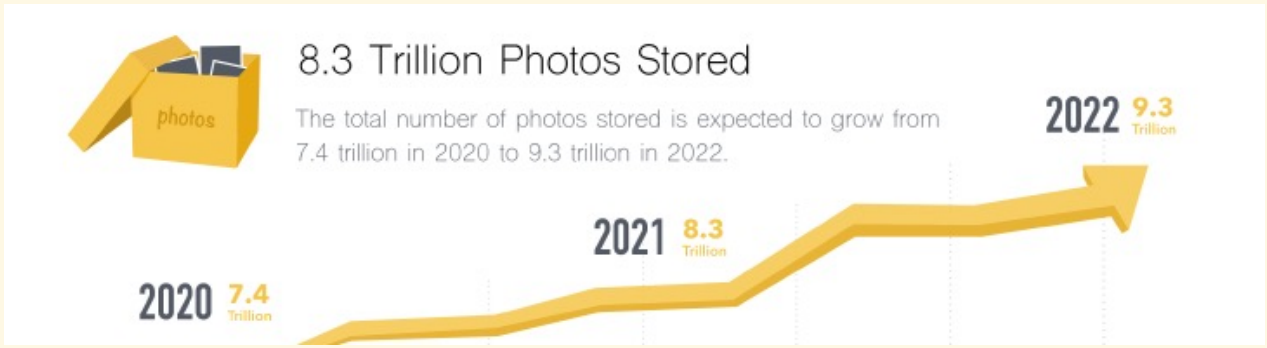
Summary

- Context and Motivation
- JPEG AI Uses Cases and Requirements
- JPEG AI Call for Proposals
- Conclusions and Future Work

1 Context and Motivation

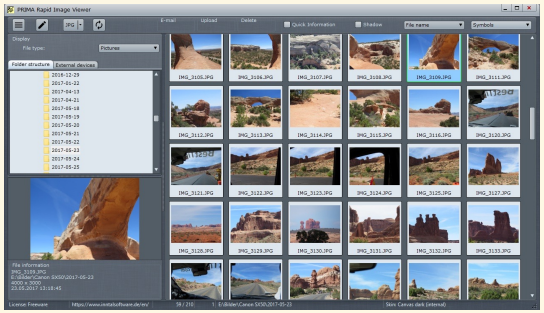
1,440,000,000,000 Images in 2021



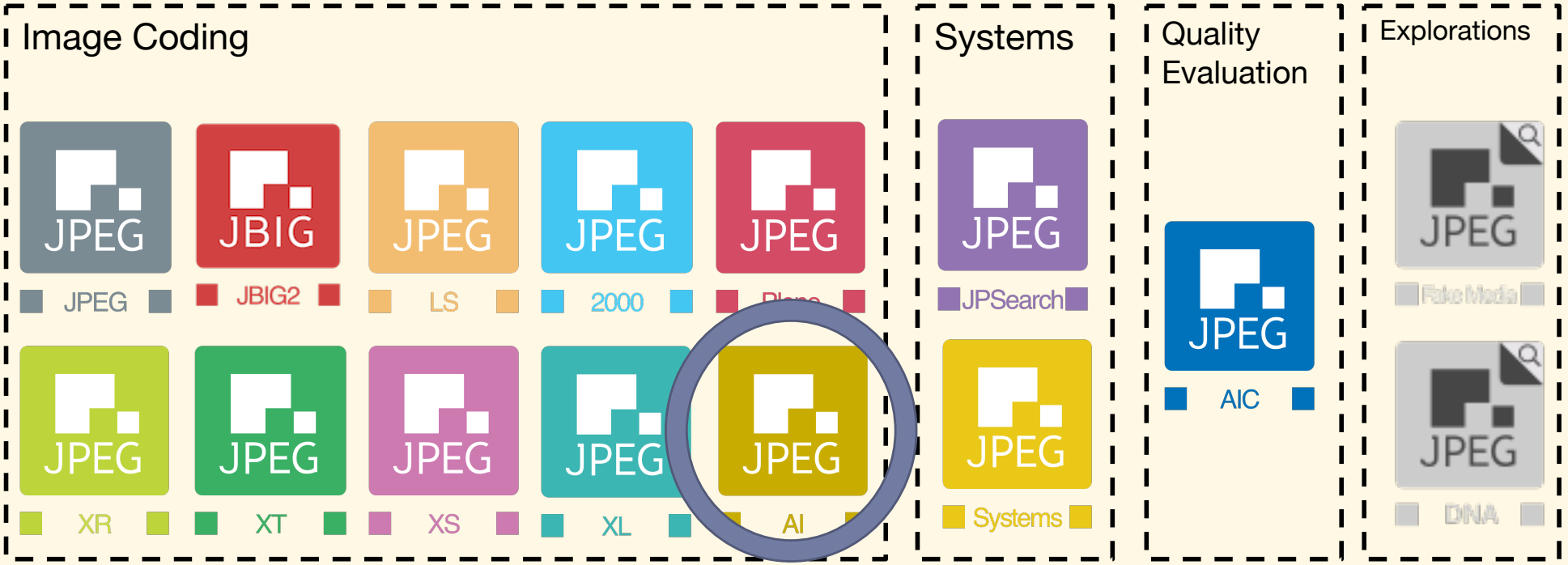


https://blog.mylio.com/how-many-photos-will-be-taken-in-2021-stats/

Rich Ecosystem for Image Technologies



JPEG Family of Standards



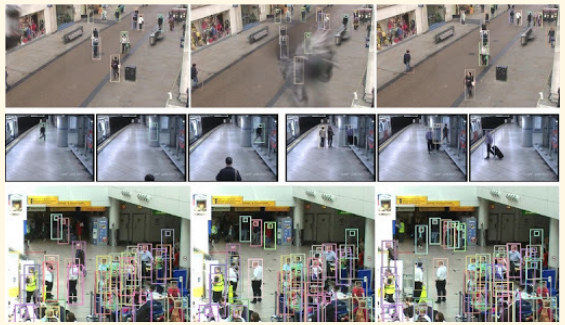
JPEG AI Achievements

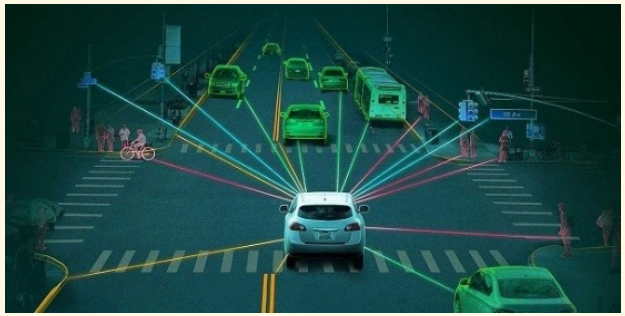
- JPEG AI Project (ISO/IEC 6048) within the JPEG standardization group aims to develop and standardize learning-based image compression
  - Being considered for joint standardization between SC29/WG1 and ITU-T SG16
  - Active since 2019
- Call for Evidence combined with MMSP Workshop Grand Challenge
  - 6 codecs submitted (out of 8 registered)
- Call for Proposals will be issued during January 2022
- Some relevant public documents:
  - White Paper on JPEG AI Scope and Framework
  - JPEG AI Call for Proposals
  - JPEG AI Uses Cases and Requirements
  - JPEG AI Training and Test Conditions
  - More information at: https://jpeg.org/jpegai/

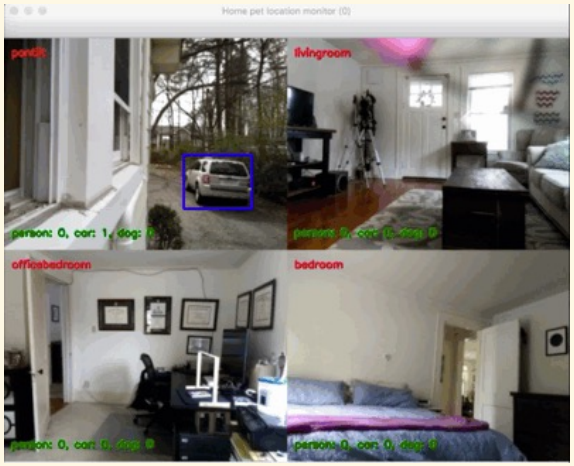
2 JPEG AI Uses Cases and Requirements

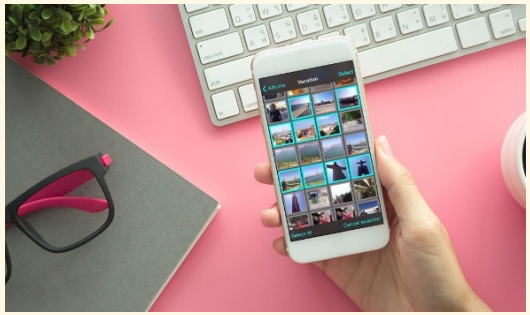
JPEG AI Use Cases

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









Application-driven Requirements

- High coding efficiency is important for many applications such as cloud storage or media distribution

- Content understanding is vital for many applications such as visual surveillance, autonomous vehicles, image collection management, etc
  - Objects may need to be recognized
  - Images may need to be classified for organization purposes
  - Actions or events may need to be recognized

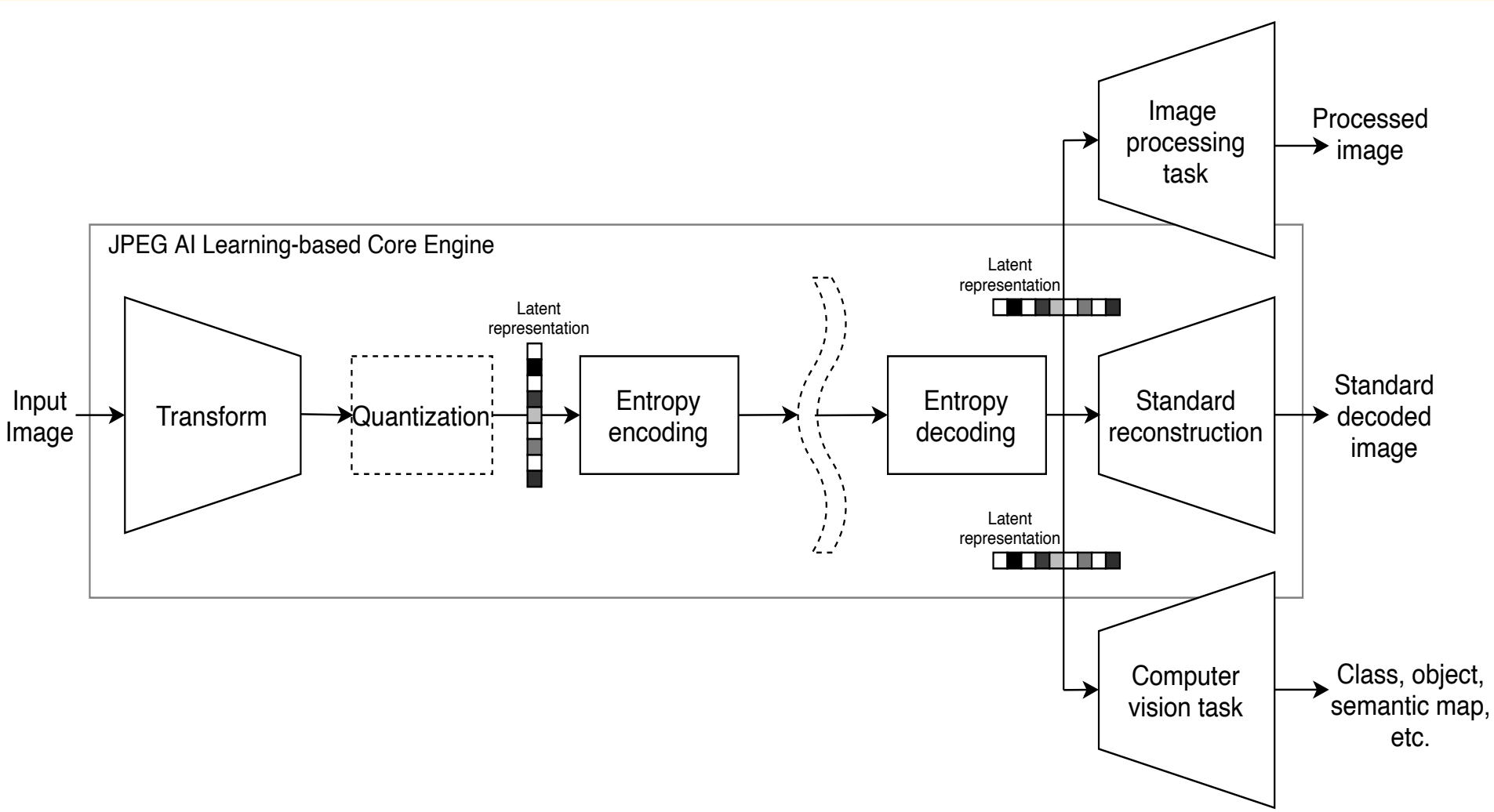
- Content is not consumed by humans in the same way. Not always the aim is to reproduce the original image as close as possible:
  - Noise can be reduced
  - Resolution can be enhanced
  - Colors can be corrected

JPEG AI Scope

The JPEG AI scope 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, as well as effective performance for image processing and computer vision tasks, with the goal of supporting a royalty-free baseline

- Advantages:
  - Same compressed stream is useful for decoding as well as image processing and computer vision tasks
  - Reduces the resources needed to perform image processing and computer vision tasks
  - Allows performing processing and computer vision tasks using features extracted from the original instead of the lossy decoded images

JPEG AI Framework



JPEG AI: Three Pipelines

- Standard JPEG AI decoding
- Image processing tasks:
  - Super-resolution
  - Denoising
  - Low-light enhancement
  - Color correction
  - Exposure compensation
  - Inpainting
- Computer vision tasks:
  - Image retrieval and classification
  - Object detection, recognition and identification
  - Semantic segmentation
  - Event detection and action recognition
  - Face detection and recognition

JPEG AI Core Requirements

- Effective compressed domain image processing and computer vision tasks
- Significant compression efficiency improvement over coding standards in common use at equivalent subjective quality
- Reconstructed images with both high subjective quality and high fidelity as measured by full reference objective quality metrics and double stimulus subjective assessment protocols
- Reconstruction reproducibility, from the same bitstream, if decoders in different platforms (CPU and GPU) provide different decoded images, it should not be greater than around 0.5% of BD-rate
- Hardware platform agnostic, encoder and decoder should be implementable in a wide range of hardware platforms.
- Hardware/software implementation-friendly encoding and decoding (in terms of parallelization, memory, complexity, and power consumption)
- Support for 8- and 10-bit depth
- Support for efficient coding of images with text and graphics
- Support for progressive decoding

JPEG AI Desirable Requirements

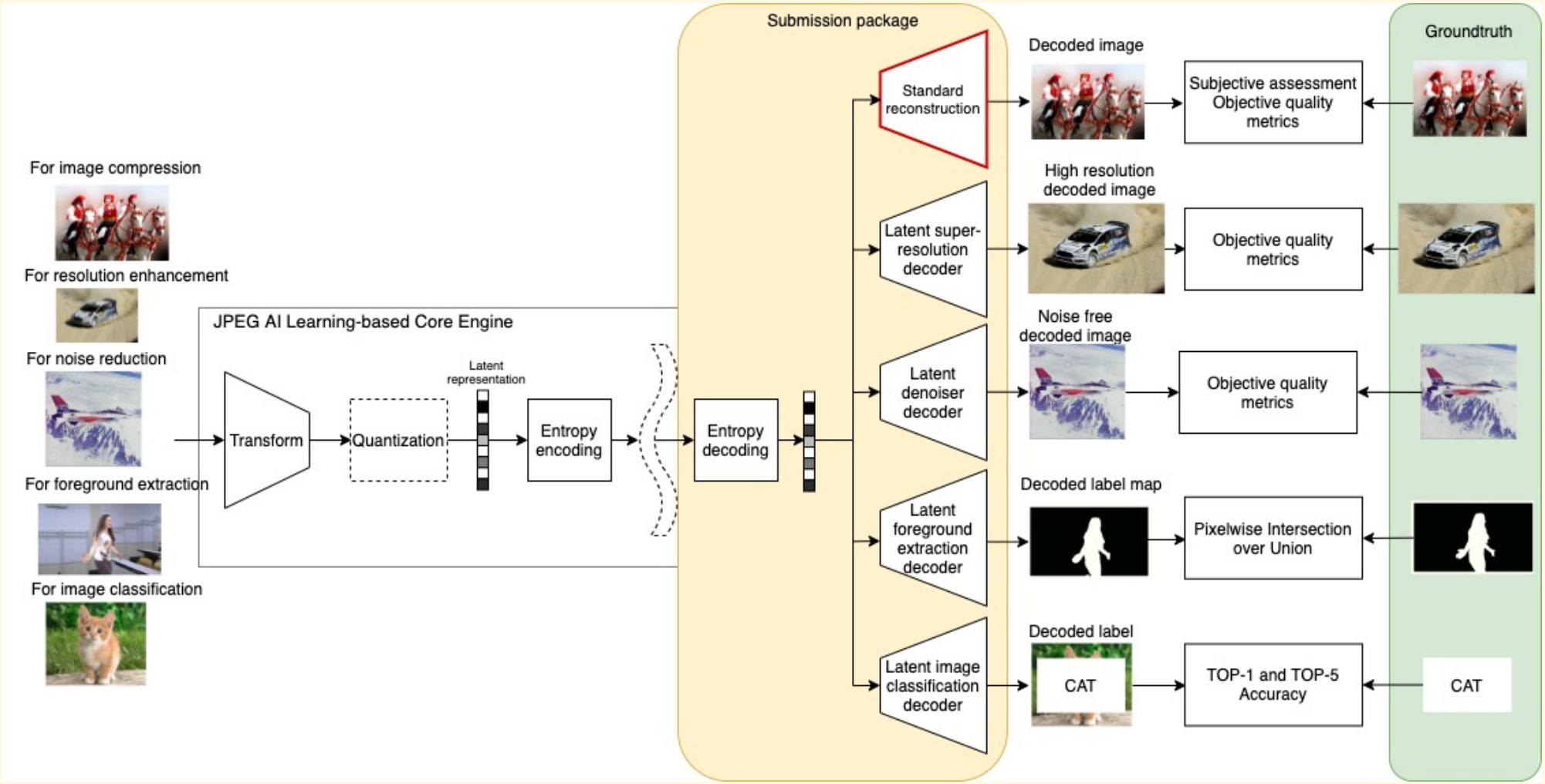
- Support for higher bit depth (e.g., 12 to 16-bit integer and floating-point HDR) images
- Support for region of interest-based coding
- Support for progressive decoding up to lossless
- Support for lossless alpha channel/transparency coding
- Support for animated image sequences
- Support for wide color gamut coding
- Support for different color representations
- Support for very low file size image coding (e.g. 64 × 64 pixel images)
- Support for a low-complexity profile - low encode/decode time even on resource-constrained hardware (e.g., mobile devices)
- Minimal generation loss when lossy compression is applied multiple times

3 JPEG AI Call for Proposals

JPEG AI Timeline

April 2021 91st JPEG Meeting	First Draft JPEG AI Call for Proposals (WG1N91015) and JPEG AI Use Cases and Requirements document (WG1N91014).
July 2021 92nd JPEG Meeting	Second Draft JPEG AI Call for Proposals (WG1N92014), JPEG AI Use Cases and Requirements document (WG1N92022) and JPEG AI Common Training and Test Conditions (WG1N92048.).
October 2021 93rd JPEG Meeting	Third Draft JPEG AI Call for Proposals, JPEG AI Use Cases and Requirements document and JPEG AI Common Training and Test Conditions.
January 2022 94th JPEG Meeting	Final JPEG AI Call for Proposals. Release of the training and validation parts of the datasets.
5th February 2022	Proposal registration.
10th March 2022	Submission of decoder implementation with some fixed model. No (re)training is allowed after this date.
15th March 2022	Release of the test datasets for proponents to code.
10th April 2022	CTTC dry run of objective and subjective performance assessment with anchors.
April 2022 95th JPEG Meeting	Analysis of the results of the dry run, may issue final recommendations for proposal evaluation.
30th April 2022	Submission of bitstreams and decoded images for the test datasets. Objective and subjective evaluation of all the proposals starts.
23-29 July 2022 96th JPEG Meeting	JPEG AI proposals submission. Presentation and discussion of the proposals at JPEG meeting. Attendance is mandatory for proponents.

CfP JPEG AI Pipelines



Standard Reconstruction I

Evaluation Conditions

- Subjective quality evaluation for standard reconstruction
  - Double Stimulus Continuous Quality Scale (DSCQS) methodology
- Objective quality evaluation for standard reconstruction using JPEG AI defined quality assessment metrics
  - MS-SSIM, IW-SSIM, VMAF, VIFP, PSNR-HVS-M, NLPD and FSIM
- Complexity evaluation of both encoding and decoding process
  - Number of parameters, model precision, running time, MAC operations, etc.
- Device interoperability requirement
  - 0.5% BD rate mismatch between CPU and GPU

Standard Reconstruction II

Proposal Composition and Requirements

- Must use JPEG AI training dataset, optionally proponents may use other data that must be made available to JPEG
- Detailed description of the coding algorithm, methodologies and training procedure
- Encoder (optionally) or decoder implementation in a form allowing stand-alone inference/testing on a standard computer
- Compressed bitstreams and corresponding decoded images
- Complexity assessment of the submission

Compress Domain Image Classification

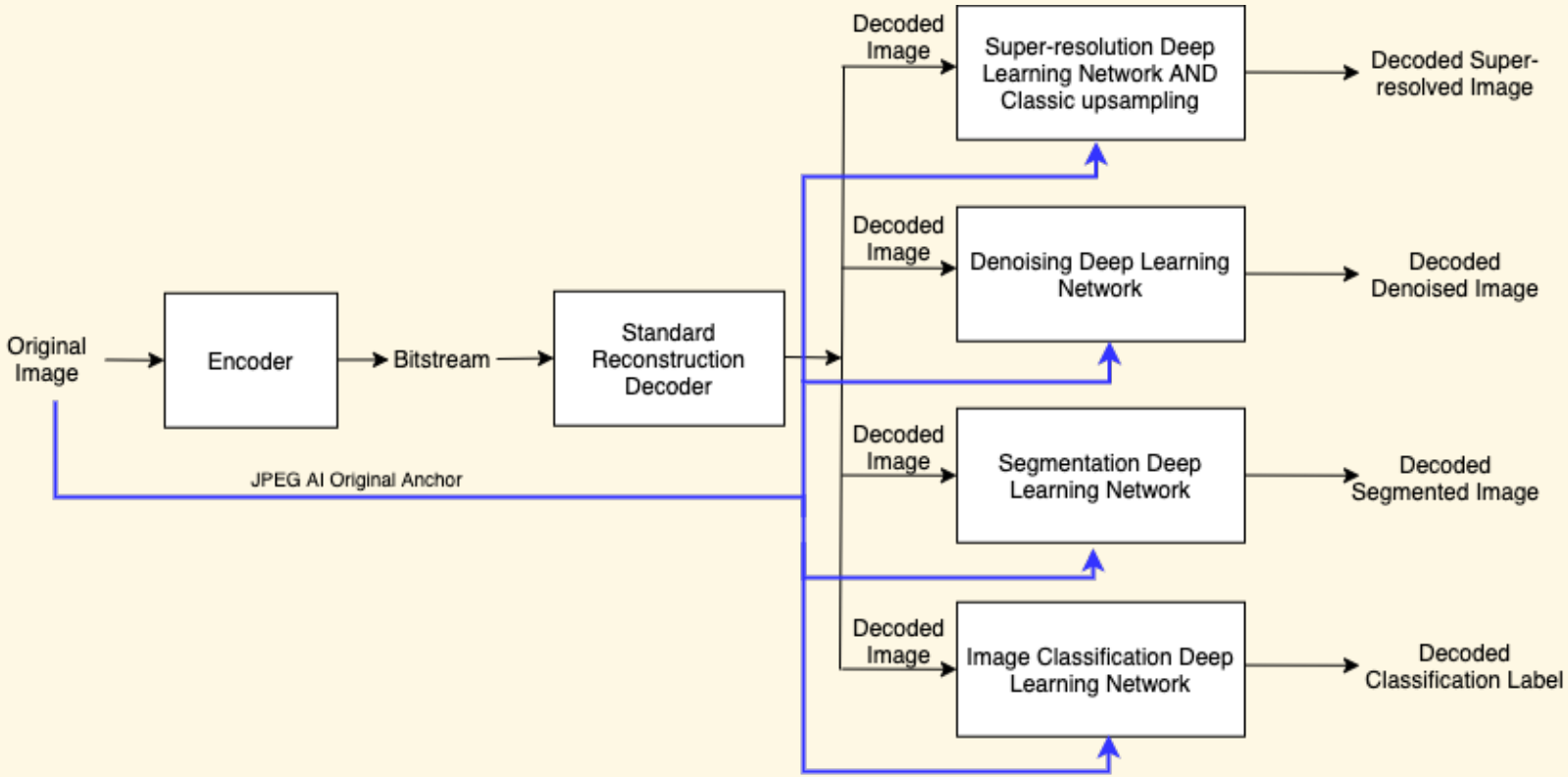
- Objective: Image classification performed directly on the latent representations produced by learning-based image codecs
  - Compute a label of an image from a pre-defined set of 1000 classes
- Learning-based image codec: proponent submission for standard reconstruction
- Training dataset: ImageNet 2012 dataset
- Four target bitrates 0.12, 0.25, 0.50, 0.75 bpp (bit per pixel) should be reported
- Performance metrics:
  - Top-1 accuracy: probability of the label of the top-1 image (with highest confidence) being the true label
  - Top-5 accuracy: probability of the label of the top-5 images (with highest confidence) being the true label
- Complexity assessment similar to the standard reconstruction

Compress Domain Super Resolution

- Objective: Super resolution performed directly in the latent representations produced by learning-based image codecs
- Learning-based image codec: proponent submission for standard reconstruction
- Training dataset: JPEG AI dataset
- Anchor neural networks:
  - DNN-based WDSR network with the pretrained WDSRx4 model
  - Classical up-sampling with a Lanczos interpolation filter with window size of 3 and 8
- Down-sampling of original high-resolution JPEG AI test images will be down-sampled by a factor of 4 using Bilinear/Bicubic/Spline/Lanczos3 interpolation
- Bitrates, performance metrics and complexity assessment: same as standard reconstruction

Computer Vision and Image Processing Tasks I

- Original anchor: Processing task is applied to the original images, before any compression, to assess the performance without any compression artifacts
- Decoded anchor: Processing task is applied to fully decoded RGB images, i.e., from the decoded pixel-wise representation



4 Conclusions and Future Work

Conclusions

- Existing deep learning-based coding solutions have now clearly outperformed the benchmarks
  - Better than HEVC, WebP, JPEG2000 and JPEG
  - Similar performance compared to VVC
- Learning based image compression solutions hold the promise of better compression efficiency but that's only one aspect!
- Main challenge is to have a multi-purpose bitstream (THE visual language) that is good for a multitude of visual tasks!
  - Not only image compression but for content understanding and image enhancement!
  - Providing several benefits for a wide range of applications and use cases


The End