JPEG AI: Methodologies for Evaluation of Learning-based Image Codecs

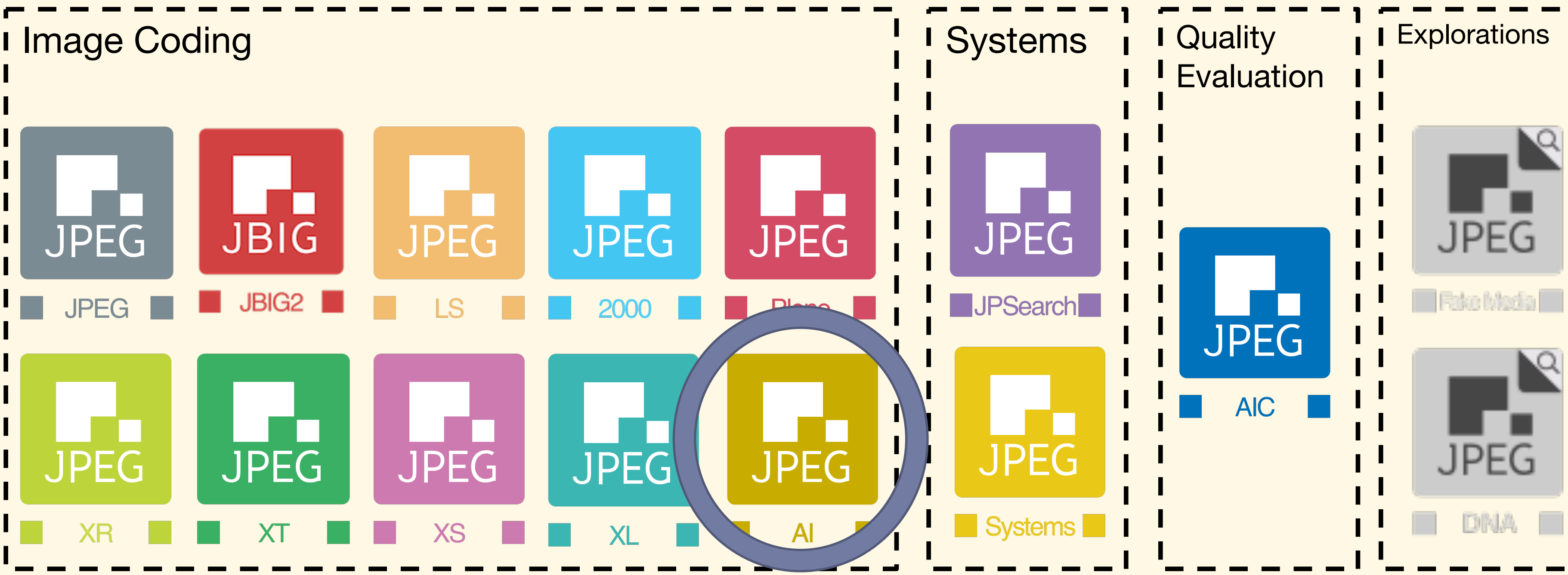
João Ascenso, IST-IT

Summary

- The JPEG-AI Project
- Common Training and Testing Conditions for Human Visualization
- Common Training and Testing Conditions for Image Processing and Computer Vision Tasks
- Conclusions and Future Work

1 The JPEG AI Project

JPEG Family of Standards



JPEG AI

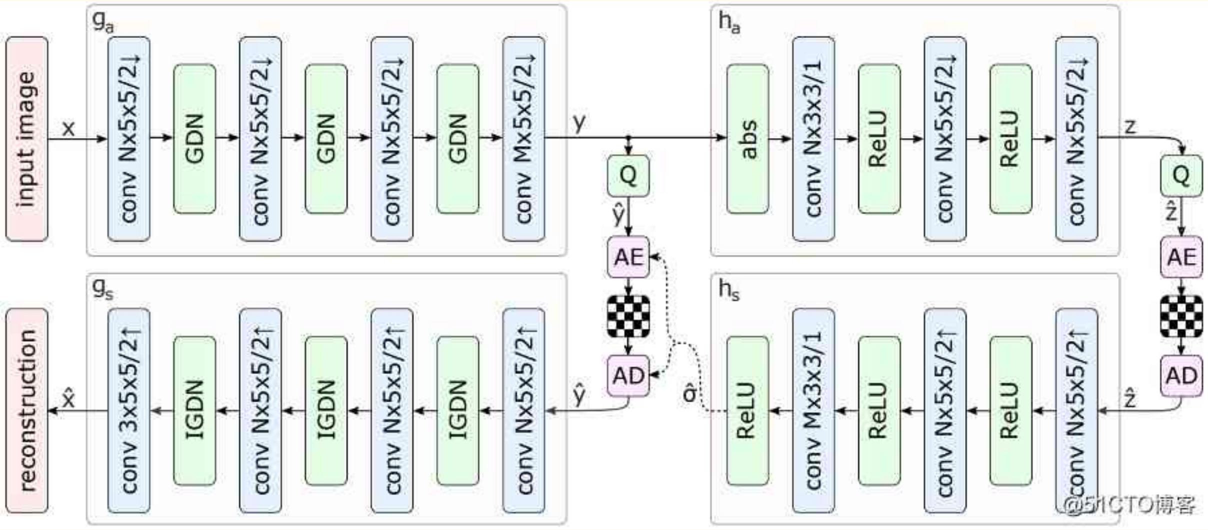
- JPEG AI Project (ISO/IEC 6048) within the JPEG standardization group aims to develop and standardize learning-based image compression
  - Being considered for joint standardization between SC29/WG1 and ITU-T SG16
  - Active since 2019
- Call for Evidence combined with MMSP Workshop Grand Challenge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

Deep Image Compression

- Exploiting the power of neural network models
  - Which are excellent in extrapolating visual data

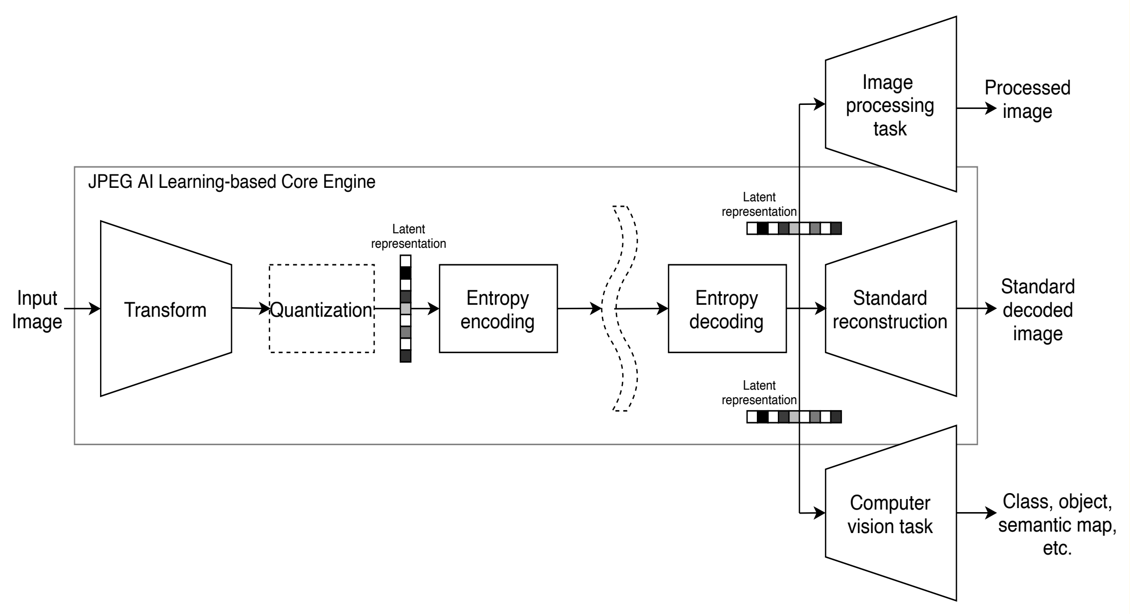
- End-to-end learned image compression
  - Learning non-linear transformations
  - Learning quantization levels and entropy coding models

Auto-encoder driven image compression



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



JPEG AI Timeline

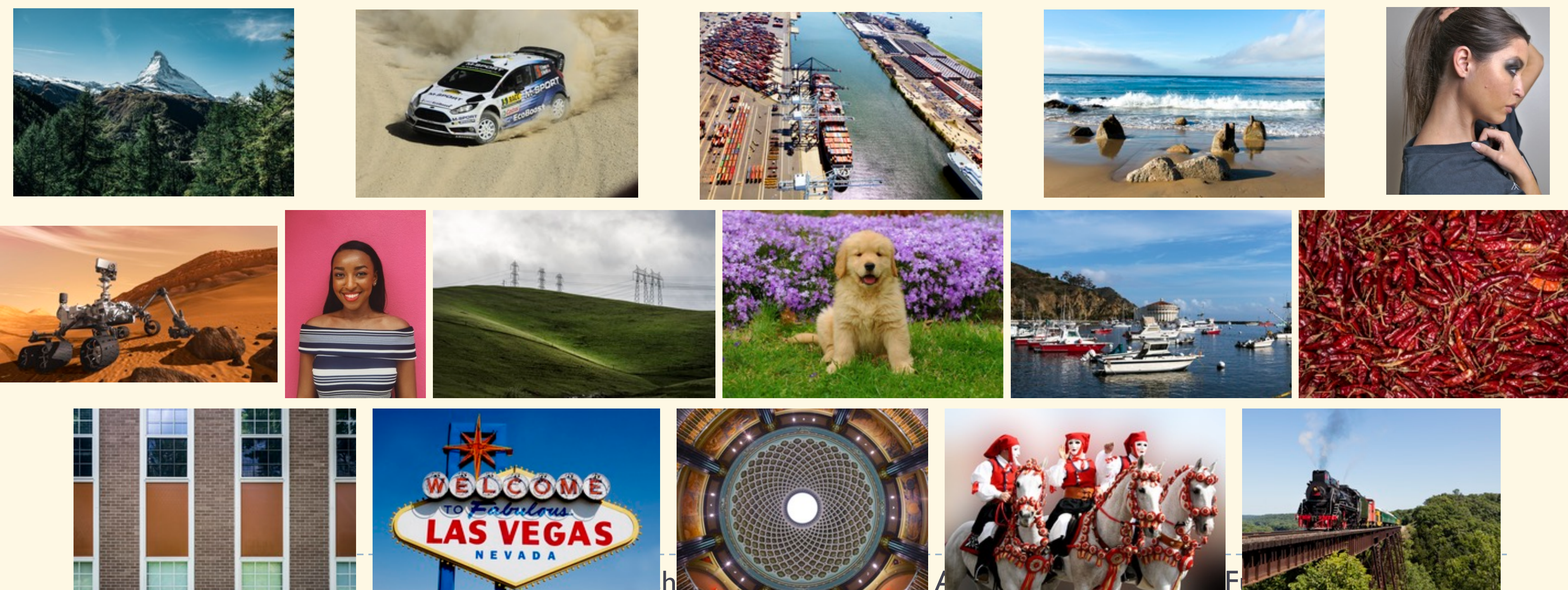
April 2021 91st JPEG Meeting	First Draft JPEG AI Call for Proposals (WG1N91015) and JPEG AI Use Cases and Requirements document (WG1N91014).
July 2021 92nd JPEG Meeting	Second Draft JPEG AI Call for Proposals (WG1N92014), JPEG AI Use Cases and Requirements document (WG1N92022) and JPEG AI Common Training 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
January 2022 94th JPEG Meeti	sets.
5th February 202	
10th March 2022	fixed model. No
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.

One meeting cycle left to have:
- Final Call for Proposals
- Final Common Training and Test Conditions

2 Common Training and Testing Conditions for Human Visualization

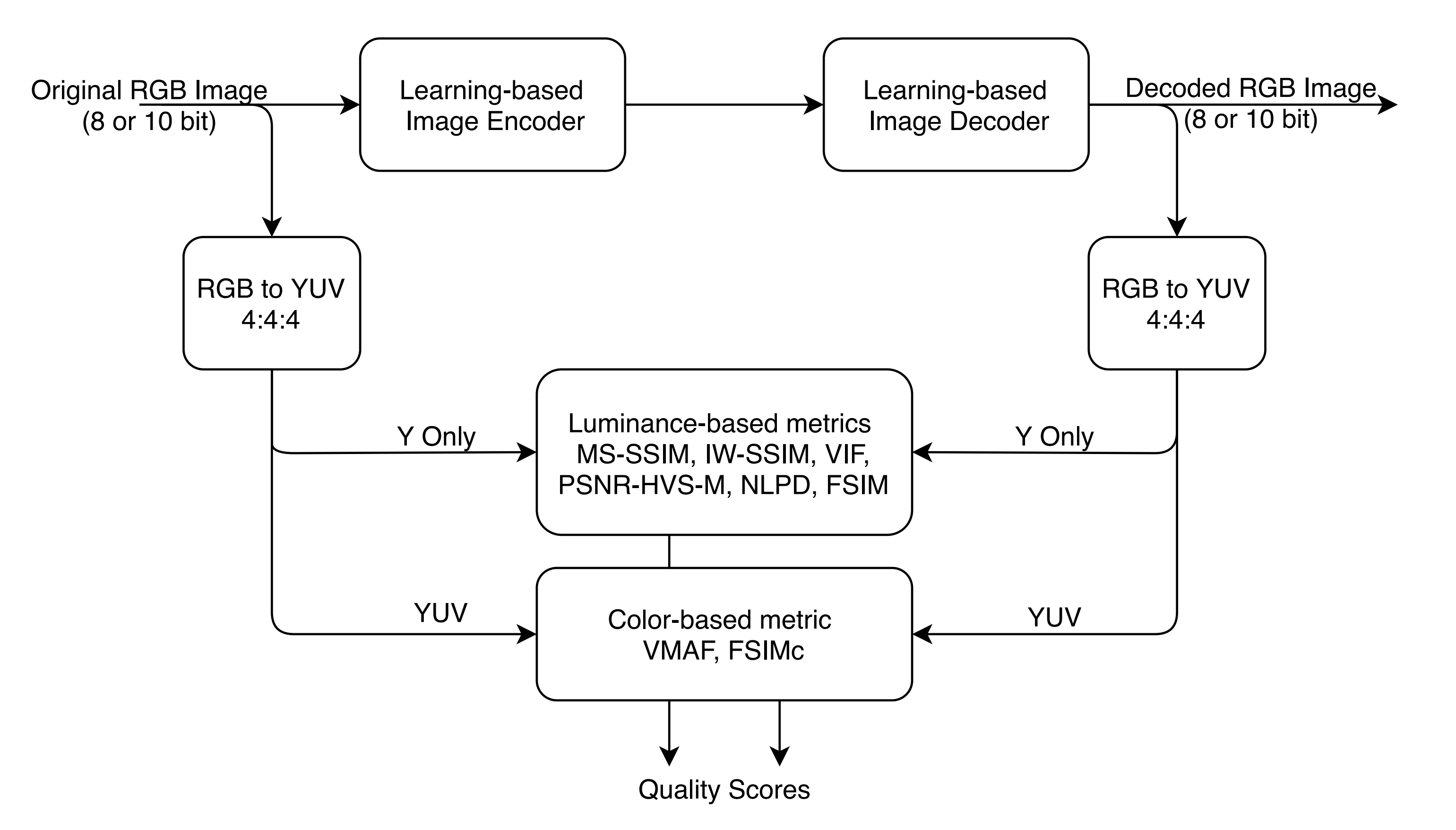
JPEG AI Dataset

- Training/validation dataset: 5264/350 images
- Proponents may also use another training dataset (optionally) which must be made available for standardization
- Test images are kept hidden until some appropriate stage, to avoid being used for training or validation



Evaluation Procedure

- Objective quality evaluation using JPEG AI defined quality assessment metrics
  - MS-SSIM, IW-SSIM, VMAF, VIFP, PSNR-HVS-M, NLPD and FSIM



JPEG AI Objective Quality Assessment Framework

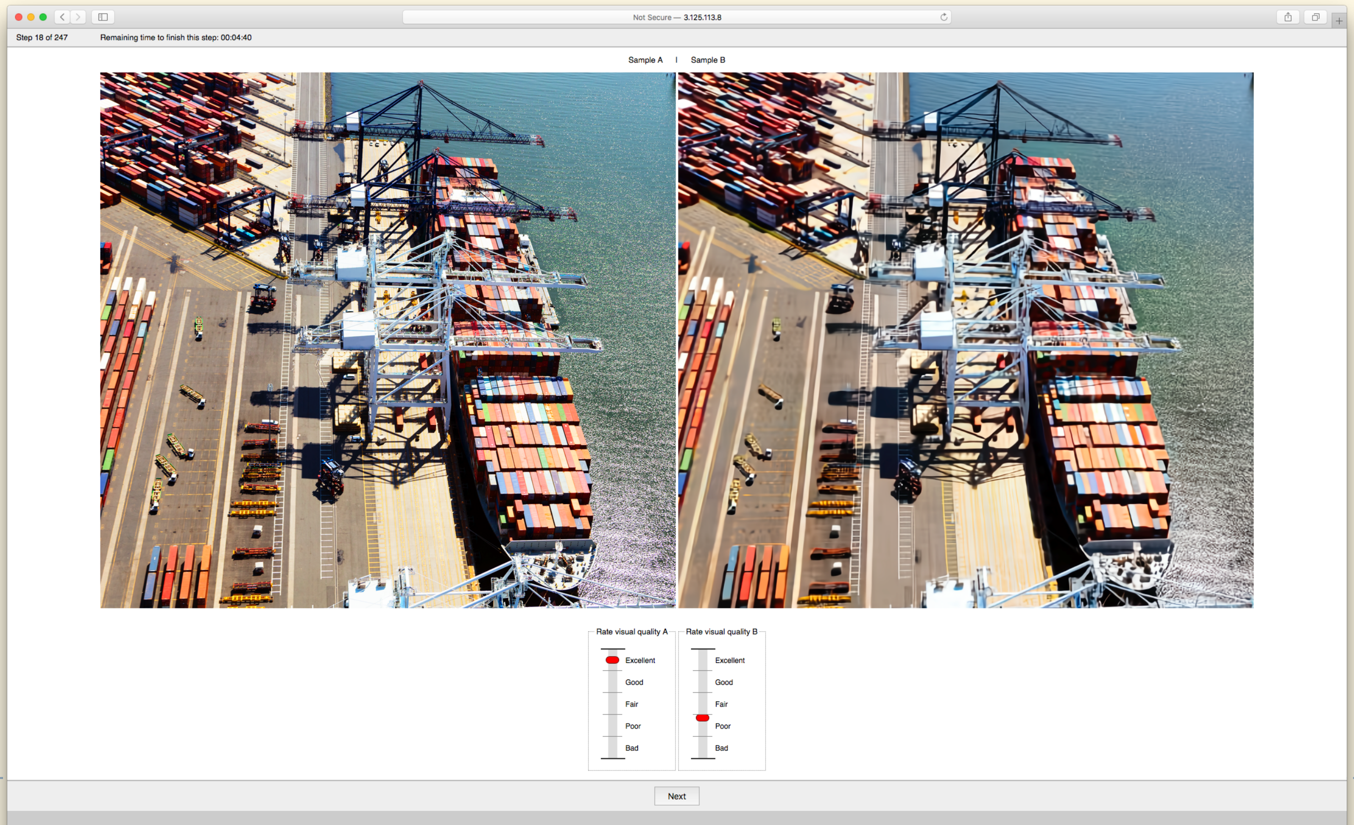
- Objective Quality Assessment Framework with reference implementation of all JPEG AI quality metrics defined
  - Cross-checked, including correlation performance assessment
  - Outputs CSV/TXT statistics
  - Supports several types of color conversions
  - Available in JPEG Gitlab https://gitlab.com/wg1/jpeg-ai/jpeg-ai-qaf

Target Rates

- Target bitrates for the objective evaluations include 0.03, **0.06, 0.12, 0.25, 0.50, 0.75**, 1.00, 1.50, and 2.00 bpp
- The maximum bitrate deviation above the target bitrate should not exceed 10%
- BD rate computation uses the mandatory rates
- Target bitrates for the subjective evaluations will be a subset of the target bitrates for the objective evaluations

Subjective Evaluation

- Double Stimulus Continuous Quality Scale (ITU Rec. BT-500) methodology:
  - Reference and the impaired stimuli are shown side by side in randomized order
  - Both reference and impaired stimuli quality are assessed by subjects
  - Difference between these two scores is then used to quantify changes in quality



Complexity Evaluation and Device Interoperabiliy

- Complexity evaluation of both encoding and decoding process
  - Number of parameters (weights) for the size of the largest model. Total number of parameters for all models, including models for all mandatory rate points.
  - Model precision, that can assume floating-point, fixed-point or integer with N bits. The N value used must be included.
  - Running time with CPU only (mandatory) and with GPU enabled (recommended), for both encoder and decoder.
  - MAC operations, number of Multiply Accumulate operations per sample (kilo), for encoder (submitted bitstreams) and decoder (worst case) operations.
  - Minimum GPU Memory Size (per Model) for encoding and decoding.

- Device interoperability requirement:
  - Encode with CPU and decode with GPU
  - Account for the BD rate mismatch which should be less than 0.5%

JPEG AI Anchors

- JPEG (ISO/IEC 10918-1 | ITU-T Rec. T.81)
  - JPEG XT reference software
- JPEG 2000 (ISO/IEC 15444-1 | ITU-T Rec. T.800)
  - Kakadu software
  - PSNR optimized and visually optimized
- HEVC Intra (ISO/IEC 23008-2 | ITU-T Rec. H.265)
  - HEVC Test Model (HM 16.20)
- VVC Intra (ISO/IEC 23090-3 | ITU-T Rec. H.266)
  - VVC Test Model (VTM 11.1)
- FFMPEG for the PNG (RGB) to YUV files conversion

3 Common Training and Testing Conditions for Image Processing and Computer Vision Tasks

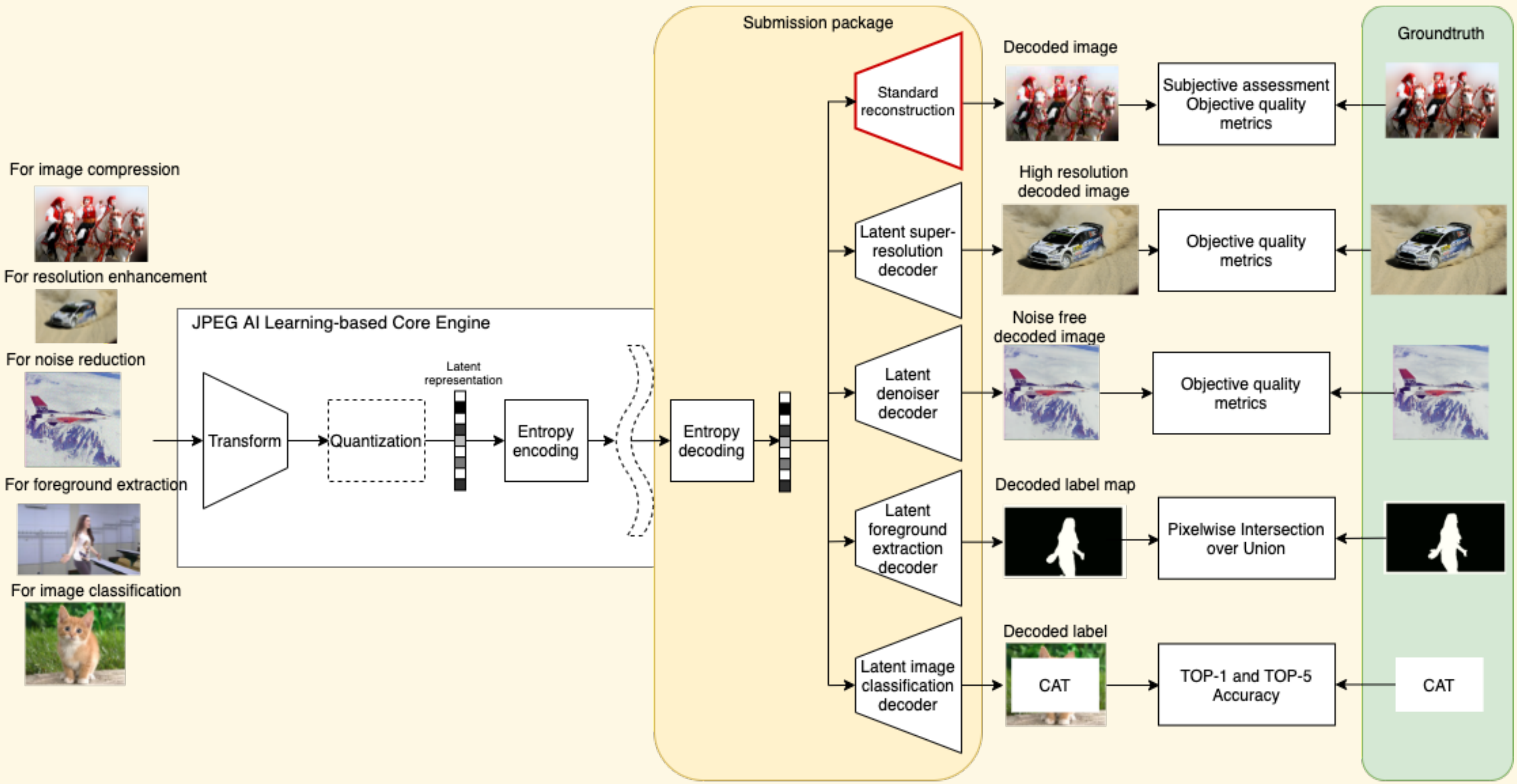
JPEG AI Tasks

- Tasks under consideration:
  - Compressed Domain Image Classification
  - Compressed Domain Real-time Foreground Extraction
  - Compressed Domain Super-Resolution
  - Compressed Domain Denoising

- Following the vision of a multi-purpose bitstream !

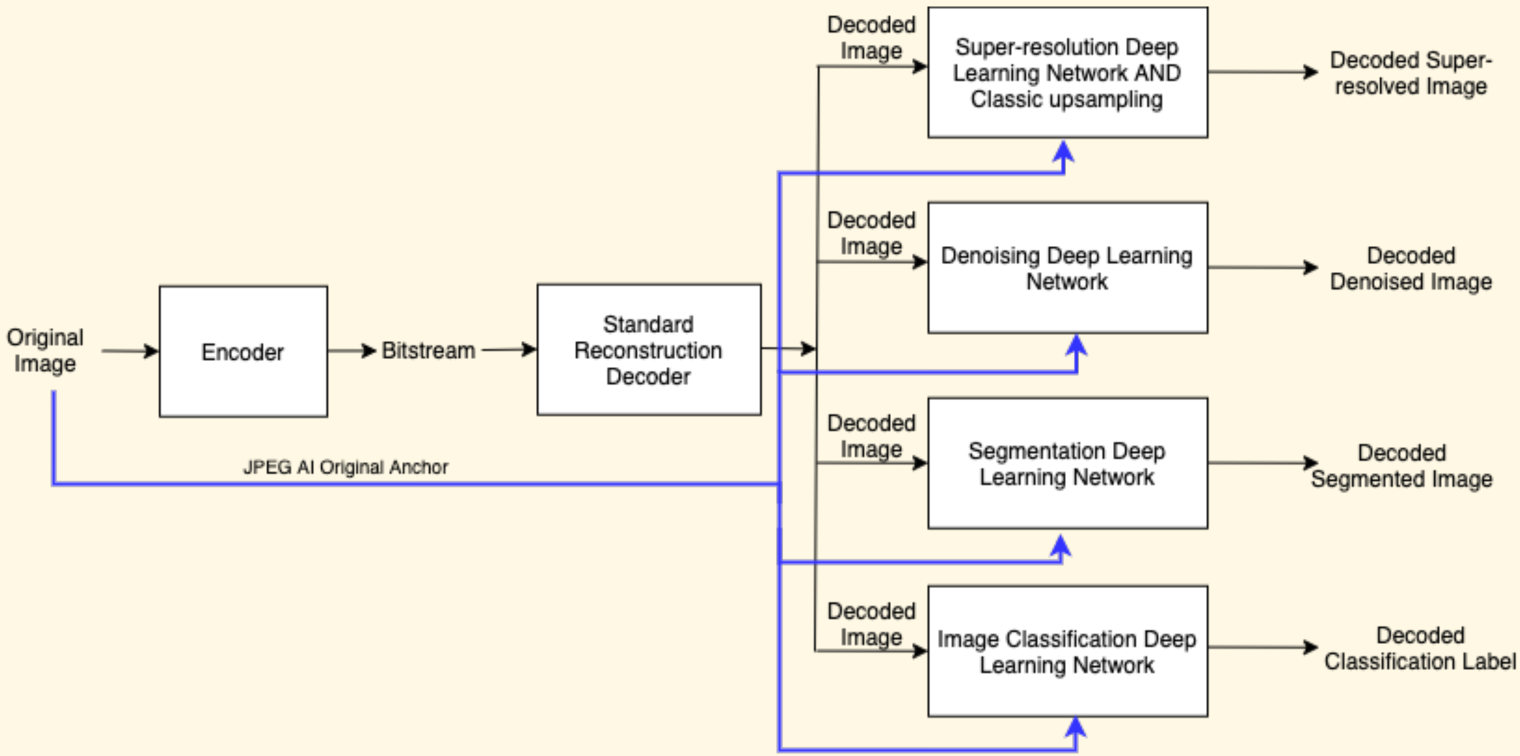
- All tasks performed directly on the latent representations produced by learning-based image codecs
  - Requires to perform entropy decoding only
  - Reduces the computational complexity needed to perform these tasks
  - Features extracted from the original are used instead from the lossy decoded images

JPEG AI Pipelines



JPEG AI Anchors

- 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

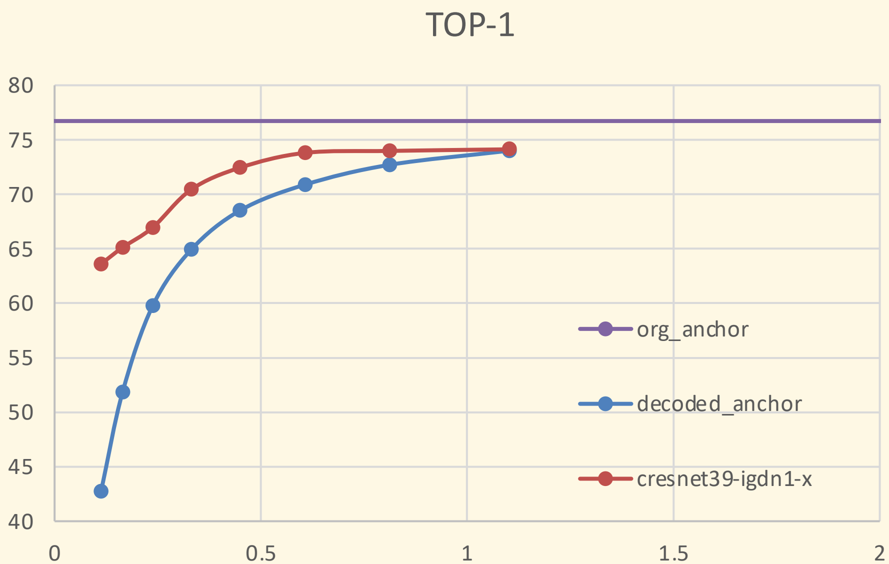


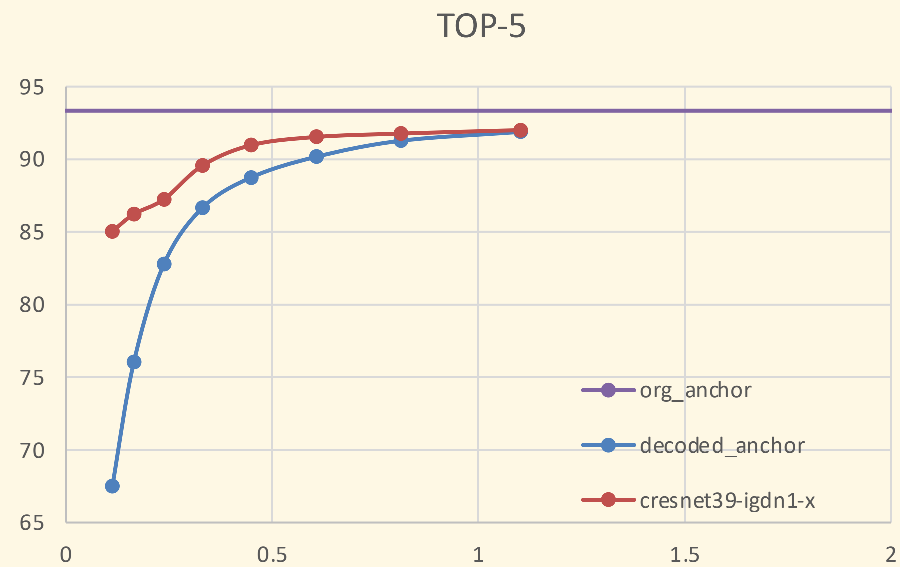
Compress Domain Image Classification

- 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
- Anchor method: pre-trained Resnet-50
- Training dataset: ImageNet 2012 dataset
- Bitrates: 0.15 to 1.8 bpp
- 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

Rate Accuracy Performance

- JPEG AI Exploration Studies have showed great potential for this task:
  - High performance especially at low bitrates
  - Less complexity to perform image classification from the latent representation
- Performance results are for the MS-SSIM loss function using the Ballé et al. Hyperprior image codec



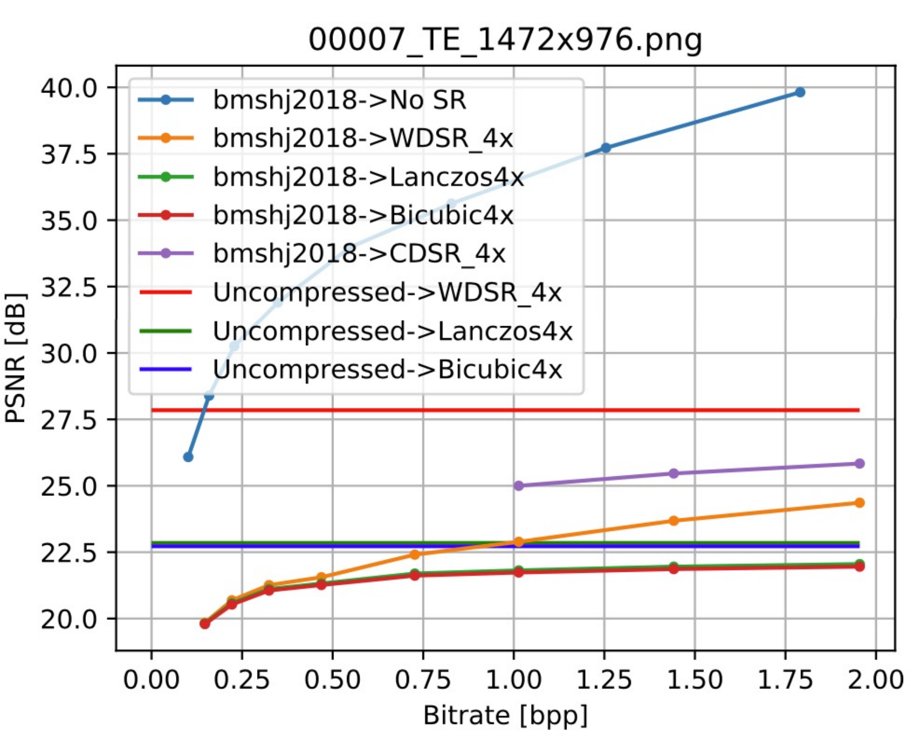


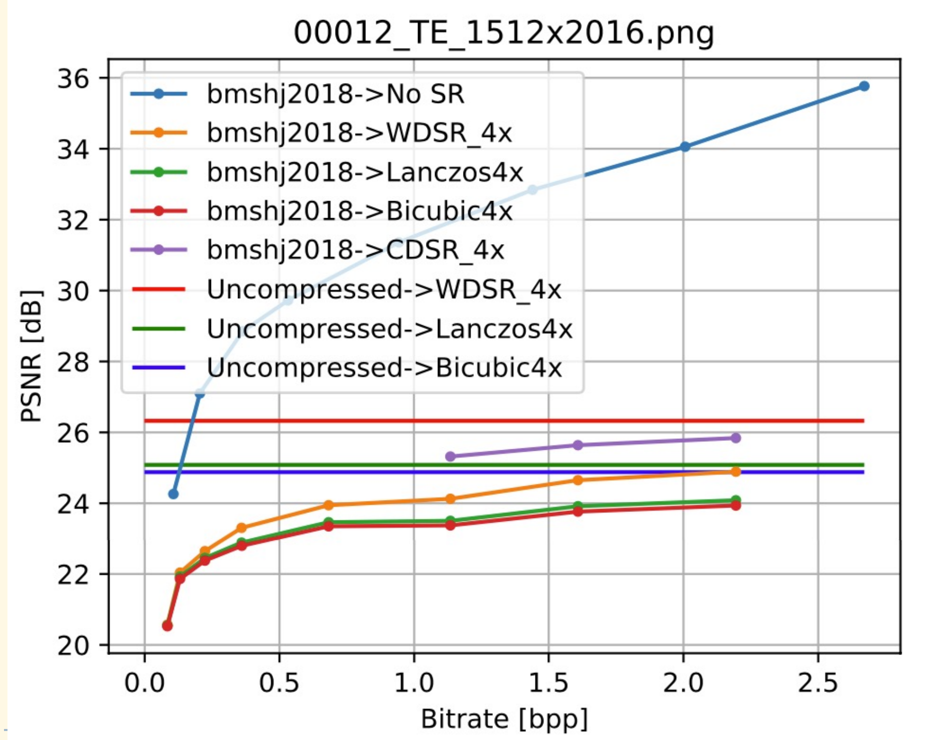
Compress Domain Super Resolution

- 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 methods:
  - 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 and performance metrics: same as standard reconstruction

Performance Evaluation

- Test high-resolution images are down-scaled by a factor of four using bicubic interpolation
- JPEG AI Exploration Studies have showed than compress domain super resolution provides higher performance than decoding anchor





4 Conclusions and Future Work

Conclusions and Future Work

- Evaluation procedure for standard reconstruction must include several aspects:
  - Coding conditions and anchors
  - Subjective and objective assessment
  - Complexity assessment
  - Device interoperability

- Evaluation procedure for image processing tasks must include other aspects:
  - Anchor pipelines
  - More training datasets
  - Other performance metrics

Future Work

- Finish Common Training and Test Conditions
- Call for Proposal submissions evaluation
- Feel free to participate!

Many thanks to all those who have contributed to these slides,
Touradj Ebrahimi, Fernando Pereira, Michela Testolina, Evgeniy Upenik, Alexander Karabutov and Elena Alshina.