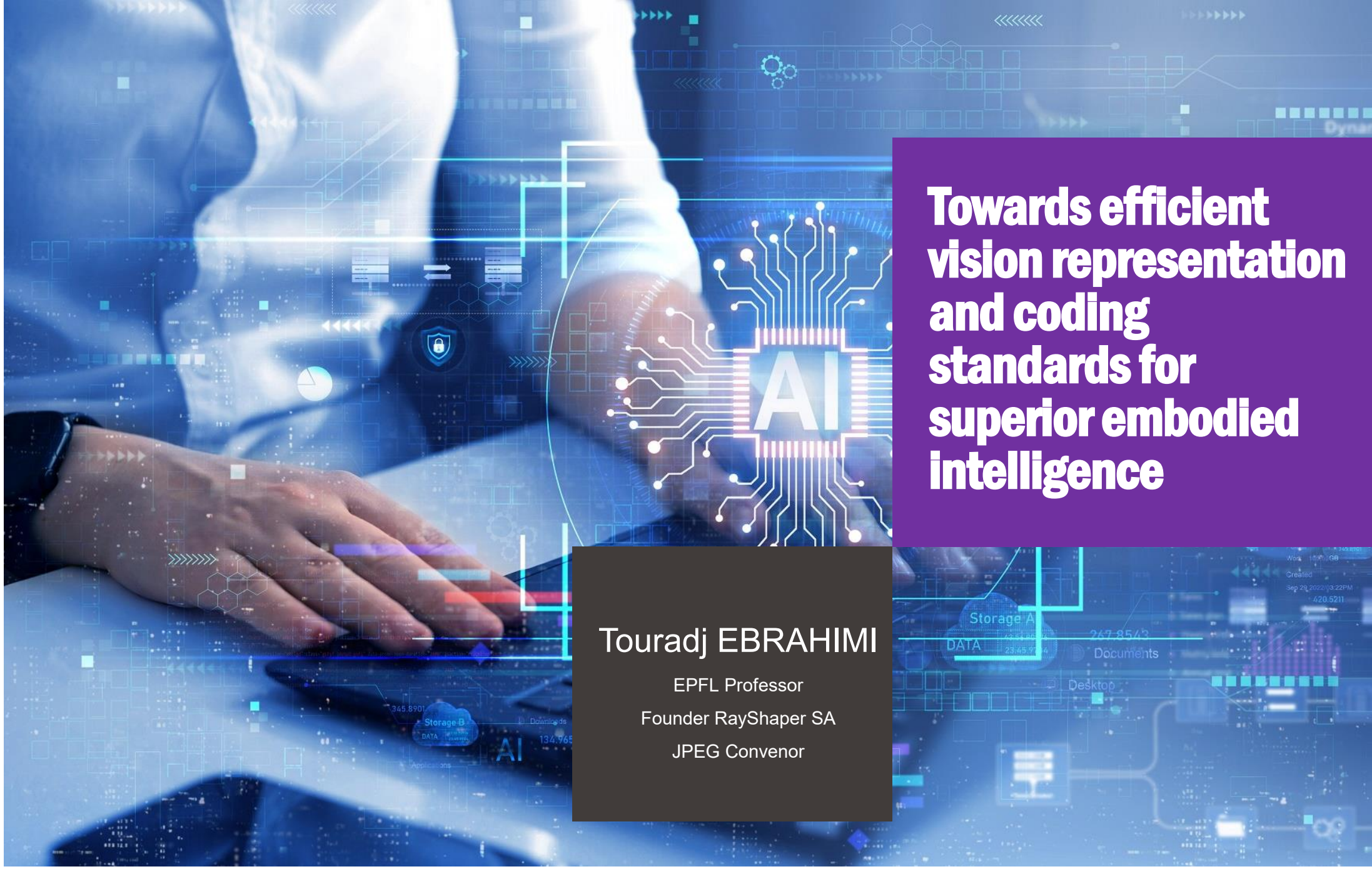


Towards efficient vision representation and coding standards for superior embodied intelligence

Touradj EBRAHIMI

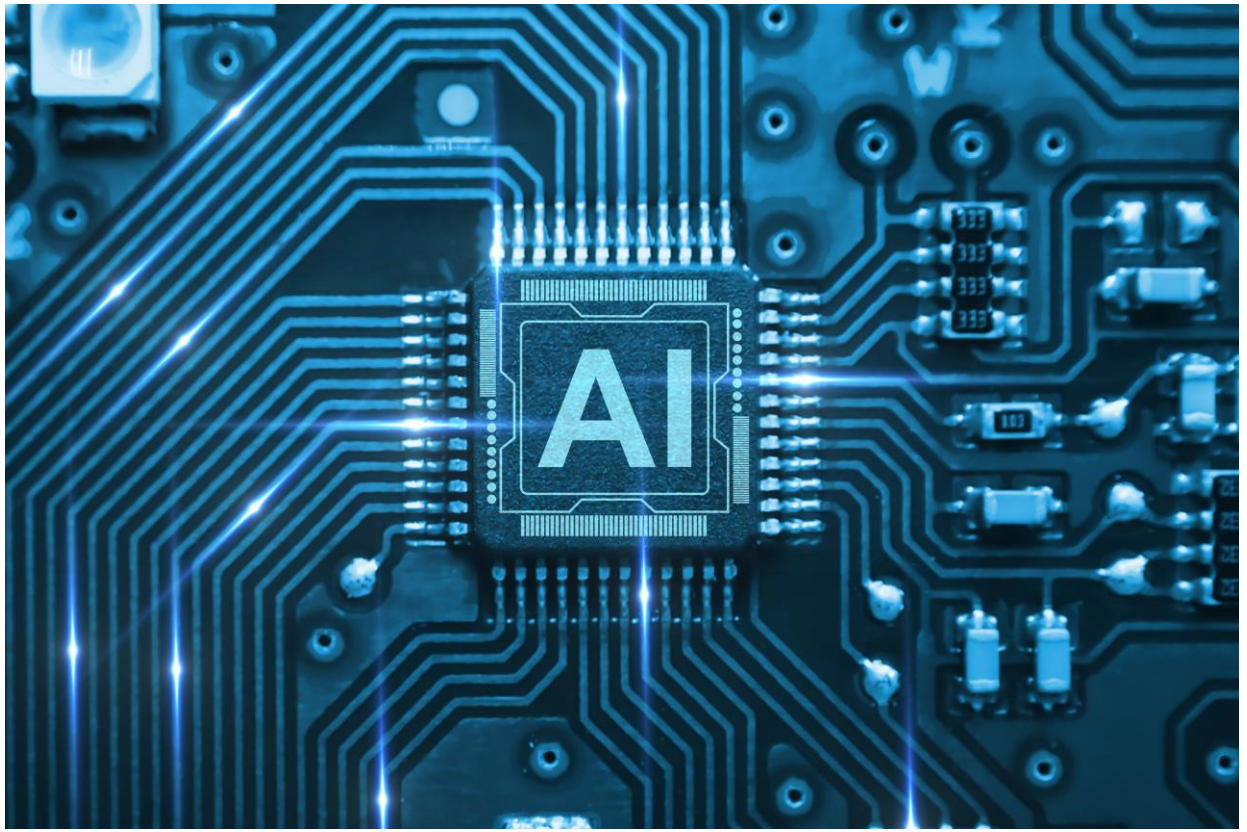
EPFL Professor

Founder RayShaper SA

JPEG Convenor

Challenges in embodied AI applications

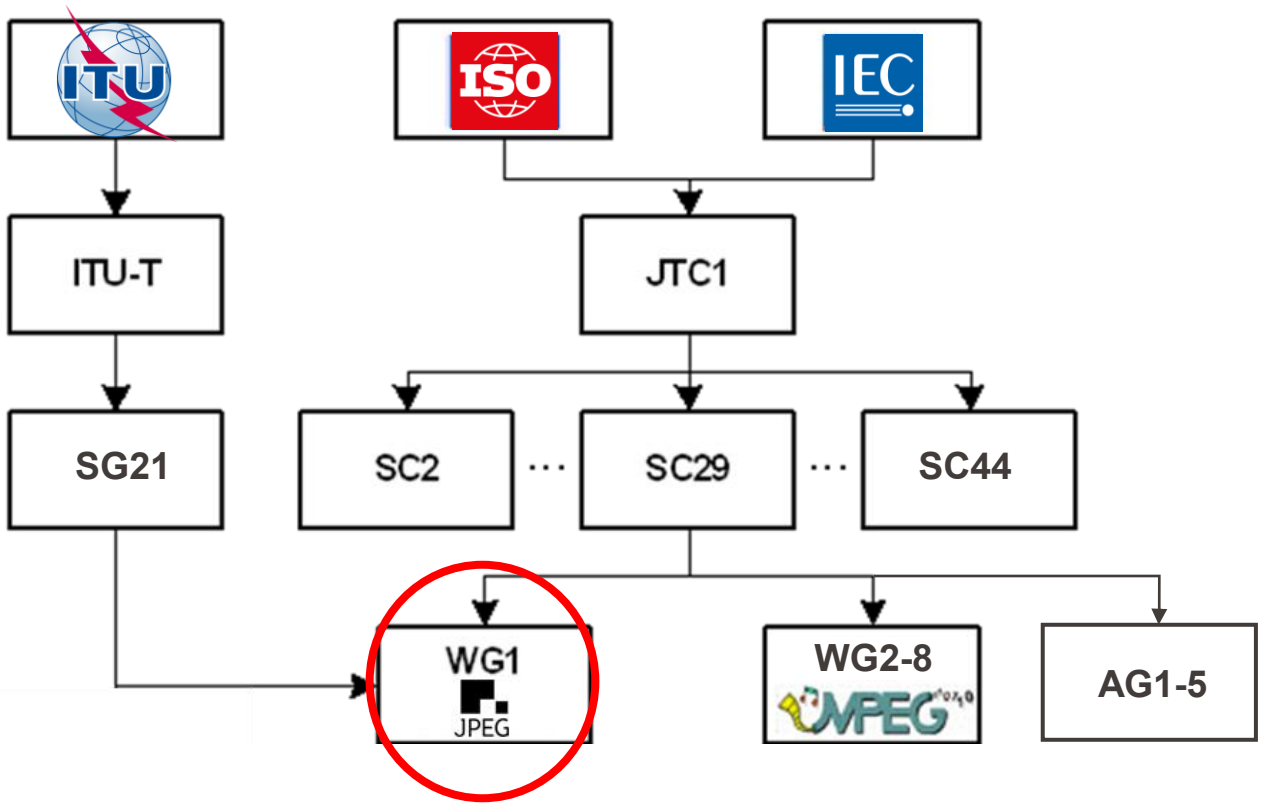
- Efficient data representation
- Efficient resource management:
  - Bandwidth
  - Complexity
  - Power consumption
- Latency
- Privacy and Security
- Efficient formats
- Performance metrics
- …



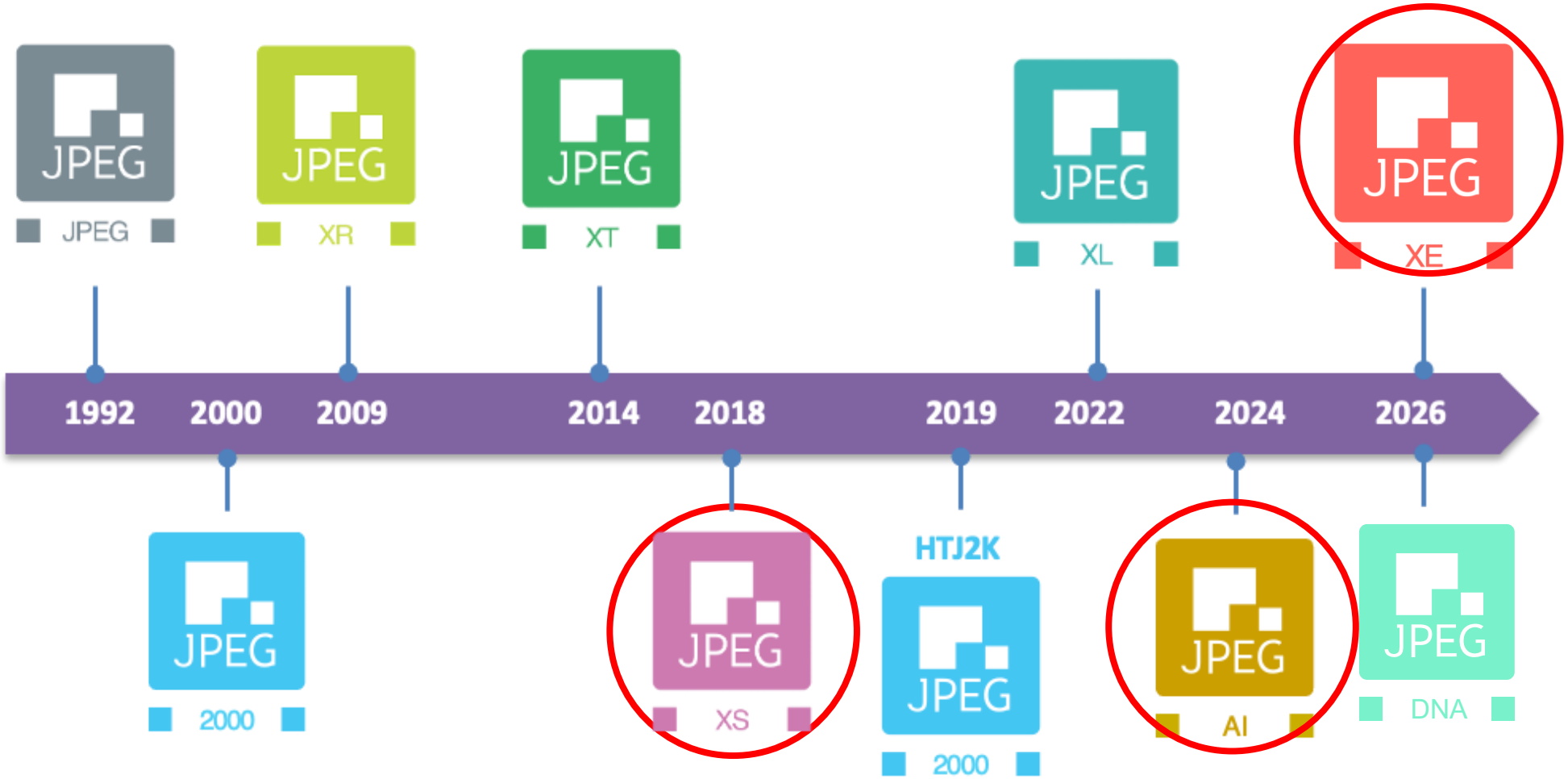
The most obvious challenge in embodied AI

- Interoperability will play a key role in the success of Embodied Intelligence
  - International Standards, as opposed to proprietary solutions dominated by a few
  - Leveraging existing International Standards, as opposed to re-inventing the wheel
  - Joint standardization efforts, as opposed to multiple similar standards for the same purposes

JPEG's perspective on Embodied AI challenges



A continuously evolving family of standards



Scope of JPEG XS

The scope of JPEG XS is to provide
a low-latency lightweight image coding system,
replacing uncompressed image and video everywhere.

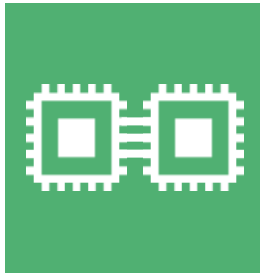
JPEG XS is the first coding standard that offers
high-quality contents with REDUCED complexity and ultra-low latency

Applications

IP Production
Remote production
Remote VAR

Cloud processing
Storage,
Connectivity

Automotive
connectivity



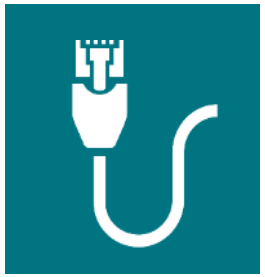
Chip-to-chip
interface

Medical

Displays
& Mobile

AR/VR HMD
Gaming

Wireless
(60GhZ, WiFi-6E,
5G)



AV-over-IP
KVMs

Digital
photography
(raw)

EVERYWHERE where uncompressed is the norm

Key benefits

Open standard

Series of open standards (ISO 21122-x)

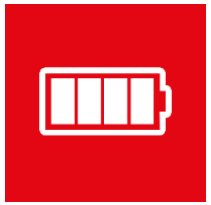
Visually lossless

Fully transparent at 3 bpp (for lossless & near-lossless coding)

Visually lossless at 1.5 bpp

Additional TDC profiles for screen and desktop content

Constant quality



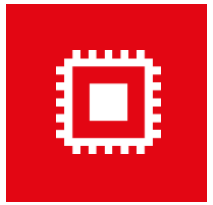
Low power

No external memory, only a few internal SRAMs

Better quality pixels

More pixels, higher bit depth, high frame rates, all at the cost of baseband HD

HDR capability



Ultra-low complexity

Low logic & low memory in ASIC or FPGA

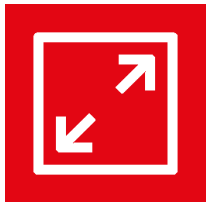
Highly parallelizable for CPU & GPU

Microsecond latency

Line-based processing (or even less than a line)

Any platform

IP-cores & Libraries for CPU, GPU, FPGA, and ASIC



Multi-res handling

Multiple layers of resolution in the compressed domain

Error concealment

To packet losses for unpredictable networks (WiFi, 60Ghz, 5G)

Industry adoption

Supported by multiple industry standards

JPEG XS and supporting standards

ITEM	DESCRIPTION
JPEG XS (ISO/IEC 21122)	-1 : Core coding system
	-2 : Profiles and buffer models
	-3 : Transport and container formats (color, video,...)
	-4 : Conformance testing
	-5 : Reference software
IETF RTP	IETF RFC9134 JPEG XS RTP payload
ST 2110-22	Encapsulation of compressed video in ST 2110 (for Media & ProAV) **VSF TR08** provides recommendations for Intra Facilities, Inter-Facilities & IPMX,
TS format	MPEG-2 Transport Stream (TS) wrapper for JPEG XS **VSF TR07** provides recommendation for WAN transmission
MXF format	MXF video file encapsulation (SMPTE ST 2124)
HEIF format	High Efficiency Image File Format, *for storing of mixed image and video content*
JXS format	JPEG XS File format, *for storing of single images*
MP4 format	ISO Base Media File format (ISOBMFF), *for storing video file*
NMOS protocol	AMWA BCP-006-01 v1.0.0 – NMOS with JPEG XS
GenICam **GenDC**	Adds support in GenICam, USB3 Vision, GigE Vision, and many more specifications that rely on GenDC

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**

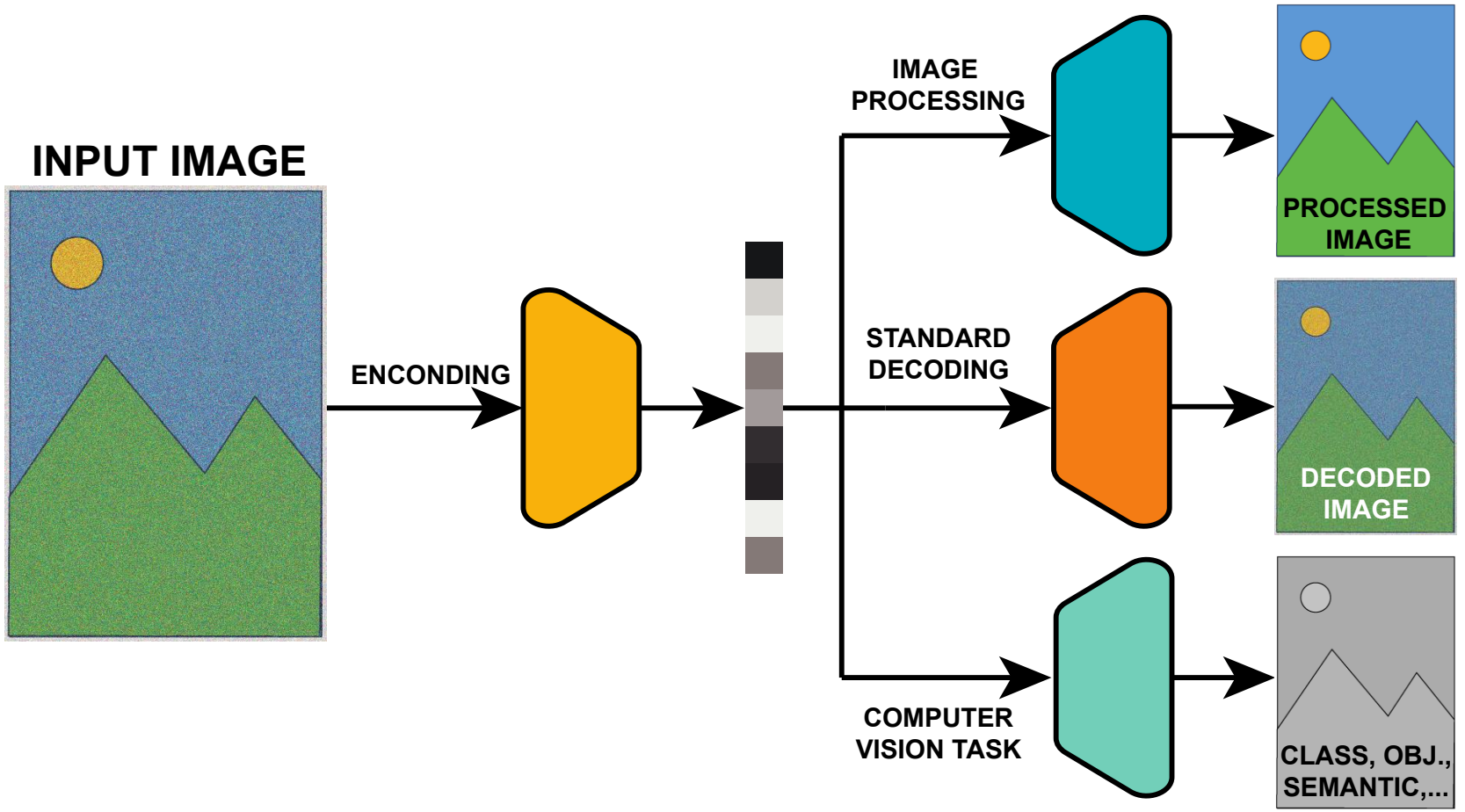
The JPEG AI is a joint standardization effort between ISO/IEC JTC1/SC29/WG1 and ITU-T SG21

JPEG AI is the first coding standard that leverages artificial intelligence for superior coding efficiency while taking into account the requirements of AI-powered content processing and analytics

Requirements

- **High coding efficiency** is important for many applications such as cloud storage or media distribution
  - AI-powered architectures have demonstrated to be particularly efficient in coding
- **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 enhancement** is desirable in many applications such as in media distribution
  - Noise can be reduced
  - Resolution can be increased
  - Colors can be corrected
- Separate **content coding for machines** and **content coding for humans** hinders interoperability and increases complexity in applications that require both
  - Why two codecs when one can do both?
  - In many applications, the same content is consumed by both humans and machines

JPEG AI triple-purpose framework



JPEG AI strategy

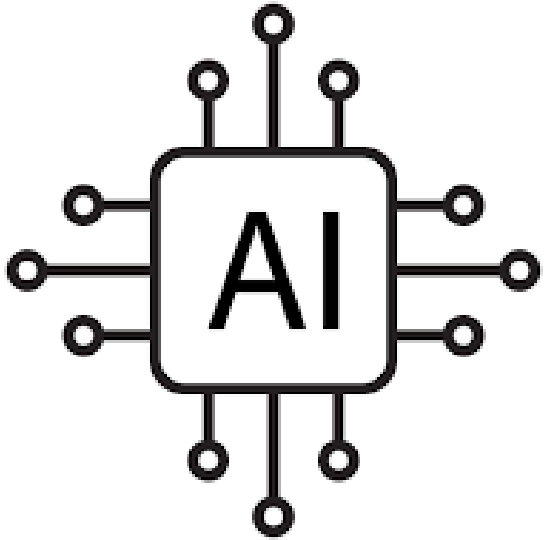
VERSION 1 (already published)

- Version 1 addresses several (but not all) JPEG AI ‘core’ and ‘desirable’ requirements with emphasis on compression efficiency.

VERSION 2 (in progress)

- Version 2 will address/include:
  1. JPEG AI requirements not yet addressed in Version 1, e.g. processing and computer vision tasks
  2. Significantly improved tools for JPEG AI requirements already addressed in Version 1, e.g. compression efficiency.

An observation about today's AI-powered content coding paradigm

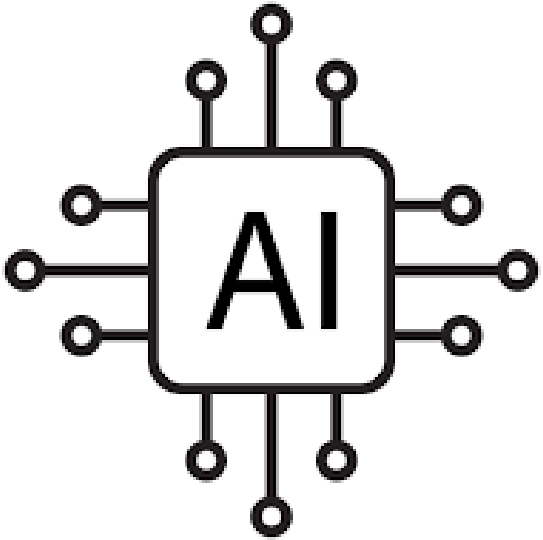


Not inspired by biological sensing

Inspired by biological computation (autoencoder and DNN)

An observation about today's AI-powered image coding paradigm

Image

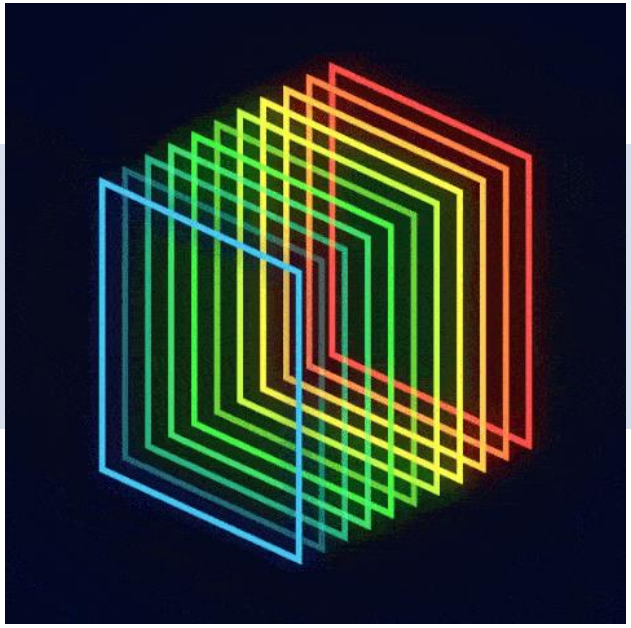


~~Not inspired~~ by biological sensing

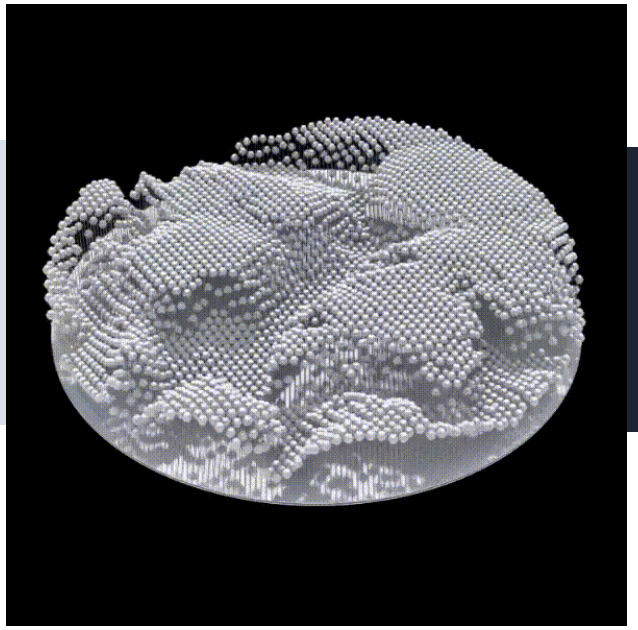
Inspired

Inspired by biological computation (autoencoder and DNN)

Event sensing



FRAME-BASED



EVENT-BASED

1. Sequential static pictures (frames)
2. Clock-driven (pre-defined frame rate)
3. Requires minimum exposure time
4. Fixed amounts of data in raw format
5. Appealing to human consumption

1. Continuous events (asynchronous intelligent pixels)
2. Scene-driven (1µs time resolution - 10’000 fps equivalent)
3. Little exposure time (120dB HDR / 40mlux low light sensitivity)
4. Amount of data varies with scene dynamics (10x to 1000x less)
5. Efficient data for machine vision

JPEG XE Scope

The scope of JPEG XE is the creation and development of a standard to represent Events in an efficient way allowing interoperability between sensing, storage, and processing, targeting machine vision applications.

The JPEG XE is planned to become a joint standardization effort between ISO/IEC JTC1/SC29/WG1 and ITU-T SG21."

JPEG XE will be the first international standard for coding of events

Requirements

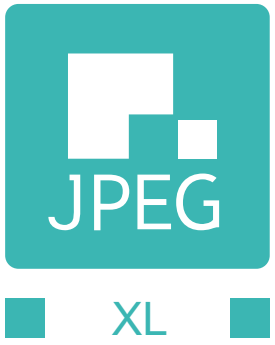
- **Efficient coding, storage, compression** and **transmission** of event data
  - Better compression
  - Low memory requirements
  - Low power requirements
  - Low end-to-end latency
  - Low complexity
  - Recovery of absolute time-base
  - Interoperability with neural network processing architectures
  - Random access
- **Interoperability** with major **interface** specifications and protocols
  - Ease of **adoption** by **industry protocols** (DCMI, MIPI, …)

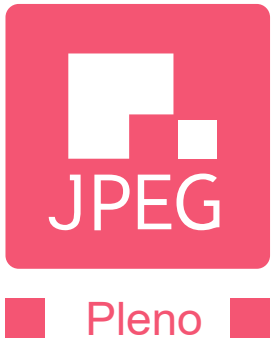
JPEG XE strategy

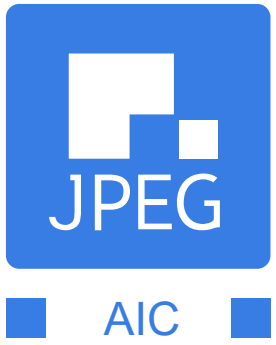
- Lossless coding
  - Urgently needed in many of today's use cases
  - Two scenarios
    - Constrained scenario: Extreme low computational complexity, willing to sacrifice compression efficiency (phase 1)
    - Unconstrained scenario: More computational resources, aim for the highest possible compression efficiency (phase 2)
- Lossy coding (phase 3)
  - Desired by many use cases and tasks in the near future
    - Tasks allow assessing the impact of lossy compression.
    - Difficult to assess the impact of loss. Requires dedicated new metrics

Last words…

- JPEG has several standards and specifications that are relevant to embodied intelligence applications
- In addition to those mentioned in this talk, other relevant standards and specifications include:
  - JPEG XL
  - JPEG Systems Part 4: Privacy and Security
  - JPEG Systems Part 5: JPEG Universal Metadata Box Format (JUMBF)
  - JPEG Pleno Learning Based Point Cloud Coding
  - JPEG AIC: Assessment of Image Coding







More information

Prof. Touradj Ebrahimi

Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL)

Touradj.Ebrahimi@epfl.ch

RayShaper

touradj@rayshapersa.ch