**3GPPSA4-e (AH) Video SWG post 131S4aV250025**

**Online, March 18, 2025 In revision of S4aV250014**

**Source: Huawei, Hisilicon**

**Title: [FS\_AI4Media] pCR on related work in TR 26.927**

**Agenda item: 9.6**

**Document for: Agreement**

**1. Introduction**

At the last meeting SA4#131 an LS from JPEG was received about JPEG AI becoming an international standard. In this paper we introduce descriptive text to add to the related work section of this report based on this LS.

**2. Proposal**

It is proposed to agree the following changes to the 3GPP draft TR 26.927 V0.11.0

\* \* \* First Change \* \* \*

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TR 22.874: "Study on traffic characteristics and performance requirements for AI/ML model transfer".

[3] Video-Based Sign Language Digit Recognition for the Thai Language: A New Dataset and Method Comparisons. https://www.scitepress.org/Papers/2023/116437/116437.pdf

[4] Cao, Junli, et al. "Real-time neural light field on mobile devices." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[5] 3GPP TS 22.261: "Service requirements for the 5G system (5GS)".

[6] 3GPP TR 23.700-80: "Study on 5G system support for AI/ML-based services".

[7] 3GPP TS 23.501: "System architecture for the 5G System (5GS)".

[8] 3GPP TS 23.502: "Procedures for the 5G System (5GS)".

[9] 3GPP TS 23.503: "Policy and charging control framework for the 5G System (5GS)".

[10] 3GPP TS 23.288: "Architecture enhancements for 5G System (5GS) to support network data analytics services".

[11] 3GPP TS 26.501: "5G Media Streaming (5GMS); General description and architecture".

[12] 3GPP TS 22.501: "System architecture for the 5G System (5GS)".

[13] 3GPP TS 23.228: "IP Multimedia Subsystem (IMS); Stage 2".

[14] Cunningham, P., Cord, M., Delany, S.J. (2008). Supervised Learning. In: Cord, M., Cunningham, P. (eds) Machine Learning Techniques for Multimedia. Cognitive Technologies. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-75171-7\_2.

[15] Open Neural Network Exchange (ONNX), https://onnx.ai.

[16] The Khronos NNEF Working Group, "Neural Network Exchange Format", https://www.khronos.org/registry/NNEF/specs/1.0/nnef-1.0.5.html.

[17] "ISO/IEC 15938-17:2024 Information technology - Multimedia content description interface - Part 17: Compression of neural networks for multimedia content description and analysis”, Edition 2, ISO/IEC, January 2024.

[18] Y. Matsubara, R. Yang, M. Levorato and S. Mandt, "Supervised Compression for Resource-Constrained Edge Computing Systems," 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2022, pp. 923-933, doi: 10.1109/WACV51458.2022.00100.

[19] Y. Matsubara, D. Callegaro, S. Singh, M. Levorato and F. Restuccia, "BottleFit: Learning Compressed Representations in Deep Neural Networks for Effective and Efficient Split Computing," 2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Belfast, United Kingdom, 2022, pp. 337-346, doi: 10.1109/WoWMoM54355.2022.00032.

[20] 3GPP TR 26.847: "Evaluation of Artificial Intelligence and Machine learning in 5G media services".

[21] TensorFlow. (2024). TensorFlow: An open-source machine learning platform. Retrieved from https://www.tensorflow.org/

[22] PyTorch. (2024). PyTorch: An open-source machine learning library. Retrieved from https://pytorch.org/

[23] 3GPP TS 26.511: "5G Media Streaming (5GMS); Profiles, codecs and formats".

[24] 3GPP TS 26.113: "Real-Time Media Communication; Protocols and APIs.

[25] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia Telephony; Media handling and interaction".

[26] AI Model Efficiency Toolkit (AIMET), https://github.com/quic/aimet

[27] Fraunhofer NNCodec: https://github.com/fraunhoferhhi/nncodec.

[28] "Application and Verification of NNC in Different Use Cases", MPEG document N0450, MPEG Video Coding ISO/IEC JTC 1/SC 29/WG 04, January 2024. https://www.mpeg.org/wp-content/uploads/mpeg\_meetings/145\_OnLine/w23554.zip

[29] "White Paper on Neural Network Compression", MPEG document N139, ISO/IEC JTC 1/SC 29/AG 03, January 2024. https://www.mpeg.org/wp-content/uploads/mpeg\_meetings/145\_OnLine/w23564.zip

[30] ISO/IEC JTC 1/SC 29/WG 04 N0598, "Common Test and Training Conditions for FCM", 2024-12-13.

[31] 3GPP TS 26.506: "5G Real-time Media Communication Architecture (Stage 2)".

[XX] ITU-T T.840.1| ISO/IEC 6048-1 Information technology — JPEG AI learning-based image coding system

[XX1] ISO/IEC 6048-3 Information technology — JPEG AI learning-based image coding system reference software

NOTE: software is currently also available on <https://gitlab.com/wg1/jpeg-ai/jpeg-ai-reference-software>

[XX2] JPEG AI Is Coming: What You Need to Know - Streaming Learning Center Available: https://streaminglearningcenter.com/codecs/jpeg-ai-is-coming-what-you-need-to-know.html

\* \* \* Second Change (all new text) \* \* \*

### 4.3.X JPEG AI: learning-based image coding system

The scope of JPEG AI [XX] is the creation of a learning-based image coding standard offering a single-stream, compressed domain representation with significant compression efficiency improvement over image coding standards in common use at equivalent subjective quality. The JPEG AI image compression standard is based on neural networks. It is Multi-task standard, in addition to compression it also targets computer vision tasks such as enhancement or detection of objects in compressed domain. The standard was completed in January 2025 under ISO/IEC 6048-1/ITU-T T.840.1 [XX].Figure 4.3.X-1 shows the high-level JPEG AI encoder, highlighting key pipelines. In a JPEG AI encoder, the input image chroma and luma are separated and processed with a transform, which aims to de-correlate the input image information typically using convolutional layers of a neural network, each one followed by nonlinear activation layers; each convolutional layer consists of learnable filters where some of them also perform spatial down-sampling. At this stage, the so-called latent representation (or latent code) is obtained, which can be understood as a compact representation of the input image. The statistical redundancy present in the latent representation is exploited by the entropy coding engine to produce the final bitstream to be transmitted or stored.



Figure 4.3.X-1 High level encoder diagram for JPEG AI

*Orig. Image*

Encoder *ID*=0,1

*y*

*stream-z*

*stream-r*

$$\hat{r}$$

$$\hat{z}$$

*Rec. Image*

$$\hat{y}$$

Entropy Dec.

AD

AE

Entropy Enc.

Latent domain Prediction

$$\hat{r}$$

$$\hat{z}$$

Latent domain Prediction

CDF-r

CDF-z

Decoder *ID*=0,1,2

Filters

$$σ$$

*Sender Receiver*

*x*0[*H,W*]

*x*1[*H*/*sver*,*W*/*shor* ]

*x*2[*H*/*sver*,*W*/*shor* ]

$\hat{x}$0[*H,W*]

$\hat{x}$1[*H*/*sver*,*W*/*shor* ]

$\hat{x}$2[*H*/*sver*,*W*/*shor* ]

Convert

Convert-1

*x*Y[1,*H,W*]

*x*UV[2, *H*/*cver*,*W*/*chor* ]

$\hat{x}$Y[1,*H,W*]

$\hat{x}$UV[2, *H*/*cver*,*W*/*chor* ]

Figure 4.3.X-2 Encoder and Decoder diagram of JPEG AI

Figure 4.3.X-2 shows the combined encoder and decoder diagram of JPEG AI [XX] (similar to other coding specifications JPEG AI only defines the decoder part normatively). The encoder receives a source image and converts it to the coding format. This includes colour separation to primary and secondary component, which are compressed using the same sequence of steps: analysis transform, encoder, latent prediction, entropy and residual coding.

The analysis transform (marked as ‘*EncoderID* on Figure 4.3.X-2) produces latent representation of the image *y*. This latent representation is further compressed to tensor *z*, which carries information about latent domain prediction and entropy parameters. The tensor *z* is quantized to $\hat{z}$ and compressed by arithmetic coder AE with entropy parameters which are part of trained model (known to both encoder and decoder). The latent domain prediction is subtracted from *y*, and the residual *r* obtained is quantized to $\hat{r}$ and encoded by arithmetic coder with entropy parameters produced by the decoder from the tensor.

The decoder performs same steps for primary and secondary components, including: a) parsing tensor$ \hat{z}$ (entropy parameters are part of the trained model), b) processing by the decoder to produce entropy parameters for residual decoding, c) processing, by the decoder to produce explicitly signalled components of prediction and latent domain prediction, d) forming reconstructed latent representation of the image $\hat{y}$. The synthesis transform for the primary component operates independently of the secondary component synthesis transform.

The synthesis transform (marked as ‘*DecoderID*’ on Figure 4.3.X-2)) processes the latent image representation to create the reconstructed image, which is later processed by series of enhancement filters and converted from coded picture format to output picture format.The latent domain prediction and entropy decoder model are identical for all analysis transform (*EncoderID*=0,1)andsynthesis transforms(*DecoderID*=0,1,2)*.*

In addition to the specification, the JPEG consortium has made reference and conformance software available. At the time of this report the software is available on <https://gitlab.com/wg1/jpeg-ai>/jpeg-ai-reference-software, the software is also planned as ISO publication ISO/IEC 6048-3 (at time of writing as draft international standard DIS) [XX1].In addition, early implementations of JPEG AI are already emerging on mobile devices, with demos shown running on smartphones like the Huawei Mate50 Pro with a Qualcomm Snapdragon 8+ Gen1 chipset, showcasing high-resolution 4K image decoding, tiling, and arbitrary resolution decoding [XX2].

JPEG AI shows improved image coding results compared to state of the art anchors. In addition, it supports the split (in the network) inference task, evaluation was performed on using the compressed JPEG AI images for computer vision tasks. The setup is shown in Figure 4.3.X-3. In this experiment the resnet-50 was used for the image classification task on either the original image (after resize and crop), the decoded images (after resize and crop) and the latent space (using another classifier for the compressed domain inference). At 0.7 bit per pixel both top-1 and top-5 results under 2 and under 1 percent less accurate compared to inference on the uncompressed image respectively (these results are based on independent evaluation following a JPEG CfP on this topic of computer vision task performance).



Figure 4.3.X-3 Evaluation of JPEG AI for split inference/in the network computer vision task.

The usage of “latent domain” representation for computer vision task is another potential benefit of the JPEG AI standardThis still needs more exploration and additional results can be added when they become available*.* Latent space based classification could become more relevant once more JPEG AI images are available.

\* \* \* End of Changes \* \* \*