

ISO/IEC 6048-1:2025-09 (E)

Information technology - JPEG AI learning-based image coding system - Part 1: Core coding system

Contents

Page

1	Scope	1
2	Normative references	1
2.1	Identical Recommendations International Standards	1
2.2	Paired Recommendations International Standards equivalent in technical content	1
2.3	Additional references	1
3	Definitions.....	1
4	Abbreviations	3
5	Conventions.....	4
5.1	General.....	4
5.2	Arithmetic operators	4
5.3	Logical operators.....	5
5.4	Relational operators	5
5.5	Bit-wise operators	5
5.6	Assignment operators.....	5
5.7	Range notation	6
5.8	Mathematical functions.....	6
5.9	Order of operation precedence	6
5.10	Variables, syntax elements and tables.....	7
5.11	Text description of logical operations	8
5.12	Processes	9
6	Image representation, tensor operations and neural network	9
6.1	Image representation.....	9
6.2	Tensor operations and neural network design elements.....	9
6.2.1	Average pooling 8.....	9
6.2.2	Batch normalization	9
6.2.3	Bicubic upsampling.....	10
6.2.4	Concatenation of two tensors	10
6.2.5	Concatenation of three tensors	10
6.2.6	Convolution layer.....	10
6.2.7	Convolution-based attention block.....	10
6.2.8	Cropping layer.....	11
6.2.9	Deinteger function.....	11
6.2.10	Downsampling layer	11
6.2.11	Downshuffle operation.....	11
6.2.12	Grouped convolution layer.....	12
6.2.13	Layer normalization	12
6.2.14	Latent combine block.....	12
6.2.15	Lightweight residual block.....	13
6.2.16	Matrix transpose.....	13
6.2.17	Matrix multiplication	13
6.2.18	Multiplication of tensors by a constant	13
6.2.19	Nearest neighbour downsampling.....	13
6.2.20	Padding layer	14
6.2.21	Per element absolute value.....	14
6.2.22	Per element addition of tensors	14
6.2.23	Per element multiplication of tensors.....	14
6.2.24	Quantized convolution	14
6.2.25	Quantized transposed convolution	14
6.2.26	Rectified exponential unit	15

6.2.27	Rectified linear unit.....	15
6.2.28	Rectified linear unit 6.....	15
6.2.29	Reshape.....	15
6.2.30	Residual activation unit.....	15
6.2.31	Residual block.....	16
6.2.32	Shuffle layer.....	16
6.2.33	Sigmoid.....	16
6.2.34	Soft-max.....	16
6.2.35	Tensor chunk.....	17
6.2.36	Tensor normalization.....	17
6.2.37	Transformer-based attention block.....	17
6.2.38	Transformer-based attention module.....	18
6.2.39	Transposed convolution.....	19
6.2.40	Unshuffle layer.....	19
6.2.41	Upsampling layer.....	19
6.2.42	Upshuffle operation.....	19
6.2.43	Zero padding.....	20
7	Overview.....	20
7.1	Learnable neural network models.....	20
7.2	An overview on the encoding and decoding processes.....	20
7.3	Functional overview on the encoding processes.....	22
7.4	Functional overview on the decoding process.....	23
7.5	Encoder requirements.....	24
7.6	Decoder requirements.....	24
7.7	Models and weights.....	24
8	Syntax and semantics.....	25
8.1	Method of specifying syntax in tabular form.....	25
8.2	Specification of the parsing process.....	26
8.2.1	Specification of syntax functions and descriptors.....	26
8.2.2	Parsing process for k-th order Exp-Golomb codes.....	27
8.2.3	Mapping process for signed Exp-Golomb codes.....	28
9	Colour component separation.....	28
9.1	General.....	28
9.2	Conversion to coding format.....	28
9.3	Example of conversion to coding colour space for RGB input.....	29
9.4	Example of conversion to coding colour space for YCbCr input.....	29
9.5	Downsampling for secondary component.....	29
9.6	Upsampling of secondary component.....	29
9.7	Conversion to output picture format.....	30
9.8	Conversion from coding colour space to output image colour space.....	30
10	Latent domain tiles.....	31
10.1	Neural network modules processing with tiles.....	31
10.1.1	Tiling processing.....	31
10.1.2	Tiles size and location computation.....	32
10.1.3	Tiling of input tensor.....	33
10.1.4	Merging tiles into output tensor.....	34
10.1.5	Overlap amount determination for tile.....	34
11	Synthesis transform.....	35
11.1	Analysis transform network (informative).....	35
11.2	Synthesis transform with latent domain tiles.....	37
11.3	Synthesis transform network.....	37
12	Codestream structure and entropy coder.....	41
12.1	Codestream layout.....	41
12.2	Substream syntax.....	43
12.2.1	Substream syntax structure.....	43
12.2.2	Semantics of substreams.....	44

12.3	Picture header.....	44
12.3.1	Picture header syntax	44
12.3.2	Picture header semantics.....	46
12.4	Residual and hyper tensor coding	50
12.4.1	Hyper tensor decoding	50
12.4.2	Quality map information decoder	53
12.4.3	Residual data decoding	55
12.5	Entropy coder me-tANS.....	58
12.5.1	General.....	58
12.5.2	Initialization of the codestream segment.....	59
12.5.3	Data decoding	60
12.5.4	Pre-calculation steps for me-tANS.....	61
12.6	Tools header.....	62
12.6.1	General.....	62
12.6.2	Tools header syntax table.....	62
12.6.3	Tools information semantics.....	65
12.7	User defined information	67
12.7.1	General.....	67
12.7.2	User defined information syntax table	67
12.7.3	User defined information semantics.....	67
12.8	Rendering information	67
12.8.1	General.....	67
12.8.2	Rendering information syntax.....	67
12.8.3	Rendering information semantics.....	68
13	Entropy parameters decoding.....	69
13.1	General.....	69
13.2	Hyper encoder.....	69
13.2.1	Hyper encoder process	69
13.2.2	Single component hyper encoder.....	69
13.3	Hyper scale decoder.....	70
13.3.1	Hyper scale decoder process.....	70
13.3.2	Single-component hyper scale decoder.....	70
13.4	Sigma scale	71
13.5	Adaptive sigma scale	71
13.6	Sigma quantization.....	71
14	Latent domain prediction and residual	72
14.1	General.....	72
14.2	Hyper decoder.....	72
14.2.1	Hyper decoder process	72
14.2.2	Hyper decoder with tiling.....	72
14.2.3	Single-component hyper decoder.....	73
14.3	Latent tensor reconstruction.....	73
14.3.1	Latent tensor reconstruction with tiling	73
14.3.2	Latent reconstruction.....	74
14.3.3	Multistage context modelling.....	74
15	Variable rate support	78
15.1	General.....	78
15.2	Control parameters and gain tensor derivation.....	78
15.3	Gain unit.....	79
15.4	Inverse gain unit.....	79
16	Mask and scale tools	79
16.1	General.....	79
16.2	Residual and variance scaling (RVS).....	80
16.2.1	RVS process.....	80
16.2.2	RVS scaling tensor generation	80
16.2.3	Precalculation of RVS tables	80
16.2.4	Residual scale.....	81
16.2.5	Inverse residual scale	81

16.3	SKIP mode.....	82
16.3.1	SKIP mode process	82
16.3.2	Cube flag generation	82
16.3.3	SKIP mask generation.....	82
16.4	Latent scale before synthesis (LSBS).....	82
16.4.1	LSBS process	82
16.4.2	Pre-calculation of LSBS tables	82
16.4.3	Decoder LSBS operation.....	83
17	Enhancement filter technologies	83
17.1	General.....	83
17.2	Secondary component re-sampling in absence of adaptive linear filter	84
17.3	Adaptive linear filter	84
17.3.1	Adaptive linear filter process	84
17.3.2	Adaptive linear filter parameters updating process	86
17.3.3	Adaptive linear filter tiling process.....	86
17.4	Inter channel correlation information filter.....	87
17.4.1	Inter channel correlation information filter process	87
17.4.2	Inter Channel Correlation Information filter for one tile.....	88
17.5	Non-linear chroma enhancement filter.....	90
17.5.1	Non-linear chroma enhancement filter process.....	90
17.5.2	Non-linear filter parameters tiling process.....	91
17.6	Luma edge filtering (LEF) filter.....	91
18	Constant tables	92
18.1	General.....	92
18.1.1	Probability distribution tables for hyper-tensor coding.....	92
18.1.2	Probability distribution tables for residual coding	93
18.2	Trained model parameters.....	94
18.3	Entropy coder constants	94
	Bibliography.....	95

List of Tables

	<i>Page</i>
Table 1 – Operation precedence from highest (at top of table) to lowest (at bottom of table)	7
Table 2 – Definition of tensor sizes.....	21
Table 3 – Supported colour sampling modes and scaling factors.....	21
Table 4 – Structure of directories in the electronic attachment	25
Table 5 – Syntax in tabular form	26
Table 6 – Bit strings with "prefix" and "suffix" bits and assignment to codeNum ranges (informative)	27
Table 7 – Exp-Golomb bit strings and codeNum in explicit form and used as ue(v) (informative).....	28
Table 8 – Assignment of syntax element to codeNum for signed Exp-Golomb coded syntax elements se(v) (informative).....	28
Table 9 – Codestream markers code assignment, symbol, description and use	43
Table 10 – Mapping table between parsed index and sigmaIdxQ.....	53
Table 11 – Additional beta-displacement for local quality control	55
Table 12 – icciModelShortlist table of parameters	65
Table 13 – icciModelLonglist of parameters.....	66
Table 14 – Parameters <i>ThrRVSLog</i> for RVS scaling tensor generation.....	80
Table 15 – Tabulated values <i>sclRVSlin</i> for RVS scaling tensor generation	80
Table 16 – Tabulated values <i>sclRVSlin</i> for RVS scaling tensor generation.....	81
Table 17 – Tabulated values <i>sclRVSLog</i> for RVS scaling tensor generation	81

Table 18 – Variable threshold_lsbs for LSBS scaling tensor generation	83
Table 19 – Variables scale0_lsbs and scale1_lsbs for LSBS scaling tensor generation	83
Table 20 – Lists of filter candidates in adaptive linear filter (cand[X][Y][4])	87
Table 21 – Forward DWT kernels	88
Table 22 – Inverse DWT kernels	89
Table 23 – Luma edge filter intensity	92
Table 24 – Luma edge filter thresholds	92

List of Figures

	<i>Page</i>
Figure 6-1 – Convolution-based attention block	11
Figure 6-2 – Downshuffle operation (illustration for the case of C being equal to 1)	12
Figure 6-3 – Latent Combine Block	13
Figure 6-4 – Lightweight residual block (LRB)	13
Figure 6-5 – Residual activation unit	16
Figure 6-6 – Residual block (RB) unit	16
Figure 6-7 – Transformer-based attention block	18
Figure 6-8 – Transformer-based attention module	19
Figure 6-9 – Upshuffle operation (illustration for the case of C being equal to 1)	20
Figure 7-1 – General JPEG AI encoder and decoder diagram	20
Figure 7-2 – General JPEG AI encoder structure	22
Figure 7-3 – General JPEG AI decoder structure	23
Figure 10-1 – Latent tiles example	32
Figure 11-1 – Secondary component input pre-processing	35
Figure 11-2 – Analysis transform network for <i>EncoderID</i> = 0	36
Figure 11-3 – Analysis transform network for <i>EncoderID</i> = 1	36
Figure 11-4 – Synthesis transform network for <i>decoderID</i> = 0	38
Figure 11-5 – Synthesis transform network for <i>decoderID</i> = 1	39
Figure 11-6 – Synthesis transform network for <i>decoderID</i> = 2	40
Figure 12-1 – Codestream layout	42
Figure 12-2 – Codestream segment decoding with me-tANS	59
Figure 13-1 – Hyper encoder structure	70
Figure 13-2 – Hyper scale decoder structure	70
Figure 14-1 – Hyper decoder structure	73
Figure 14-2 – Diagram of the multistage context model structure	74
Figure 14-3 – Diagram of the MCM0 structure	75
Figure 14-4 – Diagram of the MCM1 structure	76
Figure 14-5 – Diagram of the MCM2 structure	77

Figure 14-6 – Diagram of the MCM3 structure.....	78
Figure 17-1 – Architecture of the enhancement filter technologies	84
Figure 17-2 – Diagram of the primary component guided adaptive linear filter, wherein SS is equal to scalever * scalehor.....	85
Figure 17-3 – Architecture of ICCI sub-network	88
Figure 17-4 – ICCI processing module	89
Figure 17-5 – 1D ResBlock processing block	89
Figure 17-6 – Example implementation of EFE non-linear filter.....	90
Figure 17-7 – Example of an implementation of the LEF filter	91