

ISO/IEC FDIS 21794-6:2024(en)

ISO/IEC JTC 1/SC 29

Secretariat: JISC

Date: 2025-01-1704-02

Information technology — Plenoptic image coding system (JPEG
Pleno) —

Part 6:
Learning-based point cloud coding

FDIS stage

Edited DIS - MUST BE USED FOR FINAL DRAFT

Formatted: Left

Style Definition

Formatted: zzCover large

Formatted: Right: 1.5 cm, Bottom: 1 cm, Gutter: 0 cm,
Section start: New page, Header distance from edge:
1.27 cm, Footer distance from edge: 1.27 cm

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Regular, Font: Bold

Formatted: Cover Title_A2

Formatted: Regular, Font: Bold

Formatted: Left

Formatted: HeaderCentered

Formatted: Font: Bold

© ISO/IEC 2024/2025

Formatted: Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Default Paragraph Font

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

Formatted: French (Switzerland)

Formatted: French (Switzerland)

Formatted: French (Switzerland)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

ISO/IEC FDIS 21794-6

<https://standards.iteh.ai/catalog/standards/iso/2ecc3dc1-81db-47be-a2c0-1606ba8819ed/iso-iec-fdis-21794-6>

Formatted: FooterPageRomanNumber

Contents

Foreword..... viii

Introduction..... ix

1 Scope..... 1

2 Normative references 1

3 Terms and definitions..... 2

4 Symbols and abbreviated terms 5

5 Conventions..... 10

6 General..... 12

7 Organization of the document..... 19

Annex A (normative) File format 20

Annex B (normative) Geometry codestream syntax 27

Annex C (normative) Deep learning-based decoding 35

Annex D (normative) Colour decoding 39

Annex E (normative) Synthesis transform 43

Annex F (normative) Hyper decoders..... 51

Annex G (normative) Entropy decoding..... 57

Annex H (normative) Binarization 60

Annex I (normative) Block merging 62

Annex J (normative) Up-sampling..... 65

Annex K (normative) Deep learning-based super-resolution 66

Bibliography 73

Foreword..... vii

Introduction..... viii

1 Scope..... 1

2 Normative references 1

3 Terms, definitions and abbreviations..... 2

3.1 Terms and definitions..... 2

4 Symbols and abbreviated terms 4

4.1 Symbols..... 4

4.2 Abbreviated terms 7

5 Conventions..... 7

5.1 Naming conventions for numerical values..... 7

5.2 Operators..... 8

5.2.1 Arithmetic operators..... 8

5.2.2 Logical operators..... 8

5.2.3 Relational operators..... 8

5.2.4 Set operators..... 9

5.2.5 Precedence order of operators..... 9

Formatted: Font: Bold, English (United Kingdom)

Formatted: HeaderCentered, Left

Formatted: Font: Bold, English (United Kingdom)

Formatted: Font: Bold

Formatted: Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Font: 10 pt

Formatted: FooterCentered, Left, Space Before: 0 pt, Tab stops: Not at 17.2 cm

Formatted: Font: 10 pt

Formatted: Font: 11 pt

Formatted: FooterPageRomanNumber, Left, Space After: 0 pt, Tab stops: Not at 17.2 cm

5.2.6 Mathematical functions.....	9
6 General.....	10
6.1 Point cloud geometry representations.....	10
6.2 Multiple learnable neural network models.....	10
6.3 Functional description of the encoding process.....	10
Figure 1 Generic JPEG Pleno point cloud encoder architecture.....	11
6.4 Functional description of the decoding process.....	11
Figure 2 Generic JPEG Pleno Point Cloud decoder architecture.....	12
6.5 Encoder requirements.....	12
6.6 Decoder requirements.....	12
6.7 Trained models and parameters.....	13
Table 1 Directory structure in electronic attachment.....	13
7 Organization of the document.....	14
Annex A (normative) File format.....	15
A.1 General.....	15
A.2 Organization of the JPEG Pleno Point Cloud superbox.....	15
Figure A.1 Hierarchical organization of a JPEG Pleno Point Cloud superbox.....	16
A.3 Defined boxes.....	17
A.3.1 General description.....	17
Table A.1 Defined boxes.....	17
A.3.2 JPEG Pleno Point Cloud Header box.....	17
A.3.2.1 General.....	17
Figure A.2 Organization of the contents of a JPEG Pleno Point Cloud Header box.....	18
A.3.3 Point Cloud Header box.....	18
A.3.3.1 General.....	18
Figure A.3 Organization of the contents of a Point Cloud Header box.....	19
Table A.2 Format of the contents of the Point Cloud Header box.....	19
Table A.3 C values.....	20
A.3.4 JPEG Pleno Point Cloud Geometry Data box.....	20
A.3.4.1 General.....	20
A.3.5 JPEG Pleno Point Cloud Attribute Data box.....	20
A.3.5.1 General.....	20
Table A.4 Format of the contents of the JPEG Pleno Point Cloud Attribute Data box.....	20
Annex B (normative) Geometry codestream syntax.....	21
B.1 General.....	21
B.2 Markers, marker segments, and headers.....	21
Table B.1 Size parameters for the SLpcc, SLcob, SLpnt, and SLsob.....	22
B.3 Defined marker segments.....	22
Table B.2 List of defined marker segments.....	22

B.4 — Construction of the codestream	22
Figure B.1 — Codestream structure	23
B.5 — Delimiting markers and marker segments	23
B.5.1 — General	23
B.5.2 — Start of Codestream (SOC)	23
Table B.3 — Start of codestream parameter values	24
B.5.3 — Point Cloud Configuration (PCC)	24
Table B.4 — Format of the contents of configuration parameter set	24
B.5.4 — Coordinate Bitstream (COB)	25
Table B.5 — Format of the contents of the start of block	25
B.5.5 — Codestream Pointer Set (PNT)	25
Table B.6 — Format of the contents of the codestream pointer set	26
Table B.7 — Size parameters for Spnt	26
B.5.6 — Start of Block (SOB)	26
Table B.8 — Format of the contents of the start of block	26
B.5.7 — End of Codestream (EOC)	27
Table B.9 — Format of the contents of end of codestream	27
Annex C (normative) Deep learning-based decoding	28
C.1 — General	28
C.2 — Deep learning-based encoding	28
Figure C.1 — Deep learning coding model encoder architecture	29
C.3 — Deep learning-based decoding	29
Figure C.2 — Deep learning coding model decoder architecture	30
Annex D (normative) Colour decoding	31
D.1 — General	31
D.2 — Recolouring	31
D.3 — 3D to 2D projection	31
D.4 — 2D image encoding	32
D.5 — 2D image decoding	32
D.6 — 2D to 3D inverse projection	32
D.7 — Colour super-resolution	32
Annex E (normative) Synthesis transform	34
E.1 — General	34
E.2 — Analysis transform network	34
Figure E.1 — Analysis transform network architecture	35
E.3 — Synthesis transform network	35
Figure E.2 — Synthesis transform network architecture	36

Formatted: Font: Bold, English (United Kingdom)

Formatted: HeaderCentered, Left

Formatted: Font: Bold

Formatted: Font: Bold, English (United Kingdom)

Formatted: Font: 10 pt

Formatted: FooterCentered, Left, Space Before: 0 pt,
Tab stops: Not at 17.2 cm

Formatted: Font: 10 pt

Formatted: Font: 11 pt

Formatted: FooterPageRomanNumber, Left, Space
After: 0 pt, Tab stops: Not at 17.2 cm

E.4 — Sparse convolution layer	36
E.5 — Generative transposed sparse convolution layer	37
E.6 — Inception-residual block	37
Figure E.3 — Inception-residual block (IRB)	38
Annex F (normative) Hyper-decoders	39
F.1 — General	39
F.2 — Hyper-encoder network	39
Figure F.1 — Hyper-encoder network architecture	39
F.3 — Hyper-mean decoder network	40
Figure F.2 — Hyper-mean decoder network architecture	40
F.4 — Hyper-scale decoder network	40
Figure F.3 — Hyper-scale decoder network architecture	41
F.5 — Quantized sparse convolution layer	41
F.6 — Quantized generative transposed sparse convolution layer	42
Annex G (normative) Entropy decoding	44
G.1 — General	44
G.2 — Scale bound table computation	44
G.3 — Scale indexes computation	44
G.4 — Residual data decoding	44
G.5 — Hyper-latent representation decoding	45
G.6 — Entropy decoding: range variant of asymmetric numeral systems (rANS)	45
Annex H (normative) Binarization	47
H.1 — General	47
H.2 — Binarization optimization	47
Table H.1 — Binarization optimization procedure	47
H.3 — Binarization	48
Annex I (normative) Block merging	49
I.1 — General	49
I.2 — Block division	49
Figure I.1 — Example of the block division and corresponding <i>BlockPosition</i> coordinates	50
I.3 — Block merging	50
Annex J (normative) Up-sampling	51
J.1 — General	51
J.2 — Down-sampling	51
J.3 — Up-sampling	51
Annex K (normative) Deep learning-based super-resolution	52
K.1 — General	52

K.2 — Super-resolution network 52

Figure K.1 — Deep learning super-resolution model architecture 53

K.3 — Transposed sparse convolution layer 54

K.4 — Lightweight inception-residual block 54

Figure K.2 — Lightweight inception-residual block (LIRB) 55

Formatted: Font: Bold, English (United Kingdom)

Formatted: HeaderCentered, Left

Formatted: Font: Bold

Formatted: Font: Bold, English (United Kingdom)

iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

<https://standards.iteh.ai/catalog/standards/iso/2ecc3dc1-81db-47be-a2c0-1606ba8819ed/iso-iec-fdis-21794-6>

Formatted: Font: 10 pt

Formatted: FooterCentered, Left, Space Before: 0 pt, Tab stops: Not at 17.2 cm

Formatted: Font: 10 pt

Formatted: Font: 11 pt

Formatted: FooterPageRomanNumber, Left, Space After: 0 pt, Tab stops: Not at 17.2 cm

Formatted: HeaderCentered

Formatted: Font: Bold

Foreword

Formatted: Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

ISO (the International Organization for Standardization) ~~is a and IEC (the International Electrotechnical Commission) form the specialized system for~~ worldwide ~~federation of national standards~~ standardization. National bodies ~~(that are members of ISO member bodies). The work or IEC participate in the development of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International~~ ~~by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.~~

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ~~ISO documents~~ document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 ~~(see www.iso.org/directives 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).~~

~~Attention is drawn~~ ISO and IEC draw attention to the possibility that ~~some of the elements~~ implementation of this document may ~~be involve~~ the subject of (a) patent ~~rights. ISO(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and <https://patents.iec.ch>. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).~~

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see ~~www.iso.org/iso/foreword.html~~ www.iso.org/iso/foreword.html. In the IEC, see ~~www.iec.ch/understanding-standards~~.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

A list of all parts in the ISO/IEC 21794 series can be found on the ISO ~~website~~ and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at ~~www.iso.org/members.html~~ www.iso.org/members.html and www.iec.ch/national-committees.

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: FooterPageRomanNumber

Formatted: Font: Bold, English (United Kingdom)

Formatted: HeaderCentered, Left

Formatted: Font: Bold

Formatted: Font: Bold, English (United Kingdom)

Introduction

This ~~standard document~~ is part of a series of standards for a system known as JPEG Pleno. This set of standards facilitates the capture, representation, exchange and visualization of plenoptic imaging modalities. A plenoptic image modality can be a light field, point cloud or hologram, which are sampled representations of the plenoptic function in the form of, respectively, a vector function that represents the radiance of a discretized set of light rays, a collection of points with position and attribute information, or a complex wavefront. The plenoptic function describes the radiance in time and in space obtained by positioning a pinhole camera at every viewpoint in 3D spatial coordinates, every viewing angle and every wavelength, resulting in a 7D function.

JPEG Pleno specifies tools for coding these modalities while providing advanced functionality at system level, such as support for data and metadata manipulation, editing, random access and interaction, protection of privacy and ownership rights.

The scope of ~~Part 6 of the standard~~~~this document~~ is the specification of a learning-based coding standard for point clouds and associated attributes, offering a single-stream, compact compressed domain representation, supporting advanced flexible data access functionalities. In this context, learning-based refers to the use of machine learning technologies to learn an optimal compressed domain representation from supplied training data.

iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

ISO/IEC FDIS 21794-6

<https://standards.iteh.ai/catalog/standards/iso/2ecc3dc1-81db-47be-a2c0-1606ba8819ed/iso-iec-fdis-21794-6>

Formatted: Font: 10 pt

Formatted: FooterCentered, Left, Space Before: 0 pt, Tab stops: Not at 17.2 cm

Formatted: Font: 10 pt

Formatted: Font: 11 pt

Formatted: FooterPageRomanNumber, Left, Space After: 0 pt, Tab stops: Not at 17.2 cm

Information technology— Plenoptic image coding system (JPEG Pleno)— **Part 6: Learning-based point cloud coding**

Part 6:
Learning-based point cloud coding

1 Scope

This ~~standard document~~ defines the JPEG Pleno framework for learning-based point cloud coding.

This ~~standard document~~ is applicable to interactive human visualization, with competitive compression efficiency compared to state of the art point cloud coding solutions in common use, and effective performance for 3D processing and machine-related computer vision tasks, and has the goal of supporting a royalty-free baseline.

This ~~standard document~~ specifies a coded codestream format for storage of point clouds. ~~This standard also~~ provides information on the encoding tools. ~~The standard~~ It also defines extensions to the JPEG Pleno File Format and associated metadata descriptors that are specific to point cloud modalities.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 6048-1⁴

~~ISO/IEC 21794-1~~¹⁾, Information technology — ~~Plenoptic~~~~JPEG AI learning-based~~ image coding system (~~JPEG Pleno~~), — Part 1: ~~Framework~~~~Core coding system~~

~~REC. ITU-T T.800~~²⁾, ISO/IEC 15444-1²⁾, Information technology — JPEG 2000 image coding system — Part 1: Core coding system

~~REC. ITU-T T.801~~³⁾, ISO/IEC 15444-2³⁾, Information technology — JPEG 2000 image coding system — Part 2: Extensions

~~ISO/IEC 21794-1~~, *Information technology — Plenoptic image coding system (JPEG Pleno) — Part 1: Framework*

~~ISO/IEC 21794-2~~, Information technology — Plenoptic image coding system (JPEG Pleno) — Part 2: Light field coding

~~ISO/IEC 21794-3~~, Information technology — Plenoptic image coding system (JPEG Pleno) — Part 3: Conformance testing

⁴ Under preparation. Stage at the time of publication: ISO/IEC DIS 6048-1:2024.

¹⁾ Under preparation. Stage at the time of publication: ISO/IEC PRF 6048-1:2025.

²⁾ Similar to REC. ITU-T T.800 | ISO/IEC 15444-1

³⁾ Similar to REC. ITU-T T.801 | ISO/IEC 15444-2

Formatted: Left

Formatted: Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Adjust space between Latin and Asian text, Adjust space between Asian text and numbers, Tab stops: Not at 0.7 cm + 1.4 cm + 2.1 cm + 2.8 cm + 3.5 cm + 4.2 cm + 4.9 cm + 5.6 cm + 6.3 cm + 7 cm

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Adjust space between Latin and Asian text, Adjust space between Asian text and numbers, Tab stops: Not at 0.7 cm + 1.4 cm + 2.1 cm + 2.8 cm + 3.5 cm + 4.2 cm + 4.9 cm + 5.6 cm + 6.3 cm + 7 cm

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Default Paragraph Font

Formatted: Footer, Left, Space After: 0 pt, Tab stops: Not at 17.2 cm

Formatted: HeaderCentered

Formatted: Font: Bold

ISO/IEC 21794-4, Information technology — Plenoptic image coding system (JPEG Pleno) — Part 4: Reference software

Formatted

ISO/IEC 21794-5, Information technology — Plenoptic image coding system (JPEG Pleno) — Part 5: Holography

Formatted

ISO/IEC 23090-5:2023, Information technology — Coded representation of immersive media — Part 5: Visual volumetric video-based coding (V3C) and video-based point cloud compression (V-PCC)

Formatted

ISO/IEC 23090-9:2023, Information technology — Coded representation of immersive media — Part 9: Geometry-based point cloud compression

Formatted

ISO/IEC 60559, Information technology — Microprocessor Systems — Floating-Point arithmetic

Formatted

3 Terms, and definitions and abbreviations

Formatted: Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

For the purposes of this document, the terms and definitions given in ISO/IEC 21794-1, ISO/IEC 21794-2, ISO/IEC 21794-3, ISO/IEC 21794-4, ISO/IEC 21794-5, ISO/IEC 23090-5:2023, ISO/IEC 23090-9:2023, ISO/IEC 6048-1, REC. ITU-T T.800, ISO/IEC 15444-1, REC. ITU-T T.801, ISO/IEC 15444-2, ISO/IEC 60559 and the following apply.

Formatted

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at <https://www.iso.org/obp>

Formatted: Adjust space between Latin and Asian text, Adjust space between Asian text and numbers, Tab stops: Not at 0.7 cm + 1.4 cm + 2.1 cm + 2.8 cm + 3.5 cm + 4.2 cm + 4.9 cm + 5.6 cm + 6.3 cm + 7 cm

— IEC Electropedia: available at <https://www.electropedia.org/>

3.1 Terms and definitions

3.1 3.1.1

point

fundamental element of a point cloud comprising a position specified as 3D spatial coordinates and colour attributes

Formatted: TermNum2, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

3.2 3.1.2

point cloud

unordered list of points

3.3 3.1.3

neural network layer

tensor operation that contains trainable parameters which receives and outputs a tensor

3.4 3.1.4

neural network module

set of neural network layers

3.5 3.1.5

neural network model

specified sequence of neural network modules (also called architecture) and corresponding trained parameters

3.6 3.1.6

trainable parameters

parameters of a neural network layer whose values require a training process based on input ground truth data to be set

Formatted

Formatted: FooterPageNumber, Space After: 0 pt, Line spacing: single

Formatted: Font: 11 pt

3.7
3.1.7**trained parameters**

parameters of a neural network model whose values have been set by a training process based on input ground truth data

3.8
3.1.8**dense tensor**

representation of a 3D block as a regular array with four dimensions: horizontal, vertical, depth, and channel dimension

3.9
3.1.9**sparse tensor**

as opposed to a dense tensor, a sparse tensor is a representation of a 3D block, where only non-zero elements are represented as a set of indices (or coordinates) \mathcal{C} and associated values (or features) \mathcal{F}

Note 1 to entry: The set of coordinates \mathcal{C} is represented as a matrix $\mathcal{C} \in \mathbb{Z}^{P \times 3}$, $\mathcal{C} \in \mathbb{Z}^{P \times 3}$ and the associated features \mathcal{F} are represented as a matrix $\mathcal{F} \in \mathbb{R}^{P \times N}$, $\mathcal{F} \in \mathbb{R}^{P \times N}$, where P is the number of non-zero elements and N is the number of channels. The remaining elements of a sparse tensor are zeros.

3.1.10 **Note 2 to entry:** Opposed to a dense tensor.

3.10

latent tensor

intermediate representation of point cloud data during encoding or decoding processes, as a sparse tensor

3.11
3.1.11**standard deviations tensor**

tensor of unsigned 16 bits integers, used for entropy coding, denoted as σ

3.12
3.1.12**concatenation of sparse tensors**

process where the features of two sparse tensors are concatenated

3.13
3.1.13**element-wise addition of sparse tensors**

process where the features of two sparse tensors are added element-wise

3.14
3.1.14**sparse convolution layer**

three-dimensional sparse convolution denoted as $SpConv(K_{ver} \times K_{hor} \times K_{dep}, N_{in}, N_{out}, s \downarrow)$

Formatted: Font: 11 pt, Bold

Formatted: HeaderCentered, Left

Formatted: Font: 11 pt, Bold

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Term(s), Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Font: Bold

Formatted: Term(s), Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Term(s), Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Note, Adjust space between Latin and Asian text, Adjust space between Asian text and

Formatted: TermNum2, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Term(s), Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Term(s), Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Term(s), Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Term(s), Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Font: 10 pt

Formatted: FooterCentered, Left, Space Before: 0 pt, Tab stops: Not at 17.2 cm

Formatted: Font: 10 pt

Formatted: Font: 11 pt

Formatted: FooterPageNumber, Left, Space After: 0 pt, Tab stops: Not at 17.2 cm

Formatted: Font: Bold

Formatted: HeaderCentered

3.15

3.1.15

transposed sparse convolution layer

three-dimensional transposed sparse convolution denoted as $TSpConv(K_{ver} \times K_{hor} \times K_{dep}, N_{in}, N_{out}, s \uparrow)$
 $TSpConv(K_{ver} \times K_{hor} \times K_{dep}, N_{in}, N_{out}, s \uparrow)$

Formatted: Term(s), Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

3.16 3.1.16

generative transposed sparse convolution layer

three-dimensional generative transposed sparse convolution denoted as
 $GTSpConv(K_{ver} \times K_{hor} \times K_{dep}, N_{in}, N_{out}, s \uparrow)$
 $GTSpConv(K_{ver} \times K_{hor} \times K_{dep}, N_{in}, N_{out}, s \uparrow)$

Formatted: TermNum2, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

3.17

3.1.17

quantized sparse convolution layer

three-dimensional quantized sparse convolution is denoted as
 $qSpConv(K_{ver} \times K_{hor} \times K_{dep}, N_{in}, N_{out}, s \downarrow, d, p)$
 $qSpConv(K_{ver} \times K_{hor} \times K_{dep}, N_{in}, N_{out}, s \downarrow, d, p)$

Formatted: Term(s), Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

3.18 3.1.18

quantized generative transposed sparse convolution layer

three-dimensional quantized generative transposed sparse convolution is denoted as
 $qGTSpConv(K_{ver} \times K_{hor} \times K_{dep}, N_{in}, N_{out}, s \uparrow, d, p)$
 $qGTSpConv(K_{ver} \times K_{hor} \times K_{dep}, N_{in}, N_{out}, s \uparrow, d, p)$

Formatted: TermNum2, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

3.19

3.1.19

matrix multiplication

matrix multiplication (denoted as \times) receives two two-dimensional arrays $input_1[h_{out}, L]$ and $input_2[L, w_{out}]$.

Formatted: Term(s), Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Note 1 to entry: This module produces a two-dimensional array $output$ of size $[h_{out}, w_{out}]$.

Formatted: Note, Adjust space between Latin and Asian text, Adjust space between Asian text and

For $i=0, \dots, h_{out}-1$ and $j=0, \dots, w_{out}-1$:

$$output[i, j] = \sum_{l=0}^{L-1} input_1[i, l] \cdot input_2[l, j]$$

3.1.20

$$output[i, j] = \sum_{l=0}^{L-1} input_1[i, l] \cdot input_2[l, j]$$

3.20

rectified linear unit

rectified linear unit is denoted as $ReLU$.

Note 1 to entry: This element-wise function is defined as:

$$ReLU(x) = \begin{cases} x, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Formatted: Term(s), Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Note, Adjust space between Latin and Asian text, Adjust space between Asian text and

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: FooterPageNumber, Space After: 0 pt, Line spacing: single

Formatted: Font: 11 pt