



(51) International Patent Classification:

H04N 19/132 (2014.01) H04N 19/184 (2014.01)
H04N 19/85 (2014.01) G06N 3/08 (2006.01)
H04N 19/50 (2014.01)

(21) International Application Number:

PCT/KR2019/010645

(22) International Filing Date:

21 August 2019 (21.08.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

10-2018-0125406 19 October 2018 (19.10.2018) KR
PCT/KR2019/004171
08 April 2019 (08.04.2019) KR
10-2019-0053248 07 May 2019 (07.05.2019) KR
10-2019-0062583 28 May 2019 (28.05.2019) KR

(71) Applicant: SAMSUNG ELECTRONICS CO., LTD.

[KR/KR]; 129, Samsung-ro, Yeongtong-Gu, Suwon-Si, Gyeonggi-do 16677 (KR).

(72) Inventors: KIM, Jaehwan; 129, Samsung-ro, Yeongtong-Gu, Suwon-Si, Gyeonggi-do 16677 (KR).

LEE, Jongseok; 129, Samsung-ro, Yeongtong-Gu, Suwon-Si, Gyeonggi-do 16677 (KR). JEON, Sunyoung; 129, Samsung-ro, Yeongtong-Gu, Suwon-Si, Gyeonggi-do 16677 (KR).

(KR). **CHOI, Kwangpyo**; 129, Samsung-ro, Yeongtong-Gu, Suwon-Si, Gyeonggi-do 16677 (KR). **CHOI, Minseok**; 129, Samsung-ro, Yeongtong-Gu, Suwon-Si, Gyeonggi-do 16677 (KR). **DINH, Quockhanh**; 129, Samsung-ro, Yeongtong-Gu, Suwon-Si, Gyeonggi-do 16677 (KR). **PARK, Youngo**; 129, Samsung-ro, Yeongtong-Gu, Suwon-Si, Gyeonggi-do 16677 (KR).

(74) Agent: **Y.P.LEE, MOCK & PARTNERS**; 12F Daelim Acrotel, 13 Eonju-ro 30-gil, Gangnam-Gu, Seoul 06292 (KR).

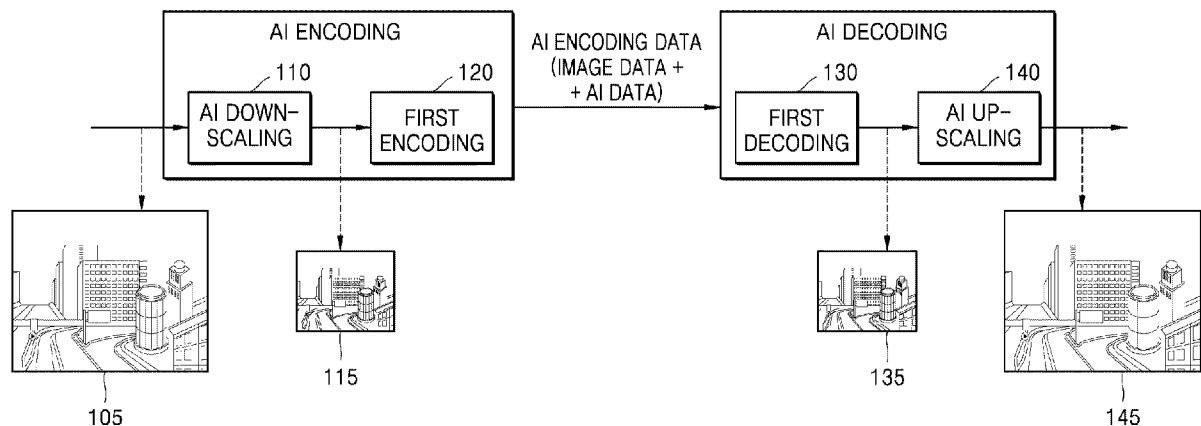
(81) Designated States (unless otherwise indicated, for every kind of national protection available):

AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available):

ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

(54) Title: METHODS AND APPARATUSES FOR PERFORMING ARTIFICIAL INTELLIGENCE ENCODING AND ARTIFICIAL INTELLIGENCE DECODING ON IMAGE



(57) Abstract: Provided is an artificial intelligence decoding apparatus includes: a memory; and a processor configured to execute the one or more instructions stored in the memory, the processor is configured to: obtain AI data related to AI down-scaling an original image to a first image; obtain image data corresponding to an encoding result on the first image; obtain a second image corresponding to the first image by performing a decoding on the image data; obtain deep neural network (DNN) setting information among a plurality of DNN setting information from the AI data; and obtain, by an up-scaling DNN, a third image by performing the AI up-scaling on the second image, the up-scaling DNN being configured with the obtained DNN setting information, wherein the plurality of DNN setting information comprises a parameter used in the up-scaling DNN, the parameter being obtained through joint training of the up-scaling DNN and a down-scaling DNN.



MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

Description

Title of Invention: METHODS AND APPARATUSES FOR PERFORMING ARTIFICIAL INTELLIGENCE ENCODING AND ARTIFICIAL INTELLIGENCE DECODING ON IMAGE

Technical Field

- [1] The disclosure relates to an image processing field. More particularly, the disclosure relates to methods and apparatuses for encoding and decoding an image based on artificial intelligence (AI).

Background Art

- [2] An image is stored in a recording medium or transmitted via a communication channel in a form of a bitstream after being encoded via a codec following a certain data compression standard, such as the Moving Picture Expert Group (MPEG) standard.
- [3] With the development and supply of hardware capable of reproducing and storing a high resolution and high quality image, the need for a codec capable of effectively encoding and decoding the high resolution and high quality image has increased.

Disclosure of Invention

Technical Problem

- [4] When an image of large resolution is encoded and decoded, large bitrate is needed.

Solution to Problem

- [5] According to an aspect of the present disclosure, an artificial intelligence (AI) decoding apparatus includes: a memory storing one or more instructions; and a processor configured to execute the one or more instructions stored in the memory, wherein the processor is configured to: obtain AI data related to AI down-scaling an original image to a first image, the AI data comprising at least one of information related to the first image and information about a difference between the original image and the first image; obtain image data corresponding to an encoding result on the first image; obtain a second image corresponding to the first image by performing a decoding on the image data; obtain deep neural network (DNN) setting information among a plurality of DNN setting information from the AI data, the DNN setting information being for performing AI up-scaling on the second image; and obtain, by an up-scaling DNN, a third image by performing the AI up-scaling on the second image, the up-scaling DNN being configured with the obtained DNN setting information, wherein the plurality of DNN setting information comprises a parameter used in the up-scaling DNN, the parameter being obtained through joint training of the up-scaling

DNN and a down-scaling DNN, and wherein the down-scaling DNN is used to obtain the first image from the original image.

Advantageous Effects of Invention

[6] According to artificial intelligence (AI) encoding and AI decoding methods and apparatuses according to embodiments, an image may be processed with a low bitrate via AI-based image encoding and decoding.

[7] However, effects achievable by the AI encoding and AI decoding methods and apparatuses according to embodiments are not limited thereto, and other effects not mentioned here will become apparent to one of ordinary skill in the art from following description.

Brief Description of Drawings

[8] A brief description of each drawing is provided to more fully understand the drawing recited in the present specification.

[9] FIG. 1 is a diagram for describing an artificial intelligence (AI) encoding process and an AI decoding process, according to an embodiment.

[10] FIG. 2 is a block diagram of a configuration of an AI decoding apparatus according to an embodiment.

[11] FIG. 3 is a diagram showing a second deep neural network (DNN) for performing AI up-scaling on a second image.

[12] FIG. 4 is a diagram for describing a convolution operation by a convolution layer.

[13] FIG. 5 is a table showing a mapping relationship between several pieces of image-related information and several pieces of DNN setting information.

[14] FIG. 6 is a diagram showing a second image including a plurality of frames.

[15] FIG. 7 is a block diagram of a configuration of an AI encoding apparatus according to an embodiment.

[16] FIG. 8 is a diagram showing a first DNN for performing AI down-scaling on an original image.

[17] FIG. 9 is a diagram for describing a method of training a first DNN and a second DNN.

[18] FIG. 10 is a diagram for describing a training process of a first DNN and a second DNN by a training apparatus.

[19] FIG. 11 is a diagram of an apparatus for performing AI down-scaling on an original image and an apparatus for performing AI up-scaling on a second image.

[20] FIG. 12 is a flowchart of an AI decoding method according to an embodiment.

[21] FIG. 13 is a flowchart of an AI encoding method according to an embodiment.

Best Mode for Carrying out the Invention

[22] According to an aspect of the present disclosure, an artificial intelligence (AI)

decoding apparatus includes: a memory storing one or more instructions; and a processor configured to execute the one or more instructions stored in the memory, wherein the processor is configured to: obtain AI data related to AI down-scaling an original image to a first image, the AI data comprising at least one of information related to the first image and information about a difference between the original image and the first image; obtain image data corresponding to an encoding result on the first image; obtain a second image corresponding to the first image by performing a decoding on the image data; obtain deep neural network (DNN) setting information among a plurality of DNN setting information from the AI data, the DNN setting information being for performing AI up-scaling on the second image; and obtain, by an up-scaling DNN, a third image by performing the AI up-scaling on the second image, the up-scaling DNN being configured with the obtained DNN setting information, wherein the plurality of DNN setting information comprises a parameter used in the up-scaling DNN, the parameter being obtained through joint training of the up-scaling DNN and a down-scaling DNN, and wherein the down-scaling DNN is used to obtain the first image from the original image.

- [23] The AI data may include the information about the difference between the original image and the first image, and the processor may be further configured to obtain the DNN setting information for performing AI up-scaling on the second image to match the third image with the difference between the original image and the first image.
- [24] The AI data may include the information related to the first image, and the processor may be further configured to obtain the DNN setting information mapped to the information related to the first image, based on a mapping relationship between several image-related information and the plurality of DNN setting information, wherein the information related to the first image comprises at least one of a resolution, a bitrate or a codec type.
- [25] The image data may include quantization parameter information used in the decoding, and the processor may be further configured to obtain, based on the quantization parameter information and the information related to the first image, the DNN setting information.
- [26] The obtained DNN setting information may include parameters of a filter kernel, the filter kernel may be associated with at least one convolution layer, and the up-scaling DNN may include the at least one convolution layer.
- [27] The processor may be further configured to set the up-scaling DNN with the obtained DNN setting information instead of DNN setting information set in the up-scaling DNN, when the DNN setting information set in the up-scaling DNN is different from the obtained DNN setting information.
- [28] The up-scaling DNN may be trained based on quality loss information, wherein the

quality loss information may be correspond to a comparison of a training image output from the up-scaling DNN and an original training image before AI down-scaling is performed.

[29] The quality loss information may be used in training of the down-scaling DNN.

[30] When parameters of a DNN of any one of the up-scaling DNN and the down-scaling DNN are updated during a training process, parameters of a DNN of the other one may be updated.

[31] According to another aspect of the present disclosure, a system includes: an AI encoding apparatus including a down-scaling DNN, the down-scaling DNN being configured to be trained based on: structural loss information corresponding to a comparison of a first training image output from the down-scaling DNN and a reduced training image, complexity loss information corresponding to a spatial complexity of the first training image, and quality loss information corresponding to a comparison of the original training image and a third training image output from the up-scaling DNN; and an AI decoding apparatus including the up-scaling DNN, the up-scaling DNN being configured to be trained based on the quality loss information.

[32] According to another aspect of the present disclosure, an artificial intelligence (AI) encoding apparatus includes: a memory storing one or more instructions; and a processor configured to execute the one or more instructions stored in the memory to: obtain, by a down-scaling deep neural network (DNN) performing AI down-scaling on an original image, a first image, the down-scaling DNN being configured with DNN setting information; encode the first image to obtain image data; and transmit the image data and AI data for selecting DNN setting information of an up-scaling DNN, the AI data comprising at least one of information related to the first image and information about a difference between the original image and the first image, wherein the up-scaling DNN is configured to perform AI up-scaling on a second image, wherein the second image is obtained by decoding the image data, and wherein the DNN setting information of the down-scaling DNN and the DNN setting information of the up-scaling DNN comprises a parameter used in the down-scaling DNN and the up-scaling DNN, the parameter being obtained through joint training of the up-scaling DNN and a down-scaling DNN.

[33] The down-scaling DNN and the up-scaling DNN may be trained based on quality loss information, wherein the quality loss information may be correspond to a comparison of a training image output from the up-scaling DNN and an original training image before AI down-scaling is performed.

[34] According to another aspect of the present disclosure, an artificial intelligence (AI) decoding method of an image, the AI decoding method includes: obtaining AI data related to AI down-scaling an original image to a first image, the AI data comprising at

least one of information related to the first image and information about a difference between the original image and the first image; obtaining image data corresponding to an encoding result on the first image; obtaining a second image corresponding to the first image by performing a decoding on the image data; obtaining deep neural network (DNN) setting information among a plurality of DNN setting information from the AI data, the DNN setting information being for performing AI up-scaling on the second image; and obtaining, by an up-scaling DNN, a third image by performing the AI up-scaling on the second image, the up-scaling DNN being configured with the obtained DNN setting information, wherein the plurality of DNN setting information comprises a parameter used in the up-scaling DNN, the parameter being obtained through joint training of the up-scaling DNN and a down-scaling DNN, and wherein the down-scaling DNN is used to obtain the first image from the original image.

- [35] According to another aspect of the present disclosure, an artificial intelligence (AI) encoding method of an image, the AI encoding method includes: obtaining, by a down-scaling deep neural network (DNN) performing AI down-scaling on an original image, a first image, the down-scaling DNN being configured with DNN setting information; encoding the first image to obtain image data; and transmitting the image data and AI data for selecting DNN setting information of an up-scaling DNN, the AI data comprising at least one of information related to the first image and information about a difference between the original image and the first image, wherein the up-scaling DNN is configured to perform AI up-scaling on a second image, wherein the second image is obtained by decoding the image data, and wherein the DNN setting information of the down-scaling DNN and the DNN setting information of the up-scaling DNN comprises a parameter used in the down-scaling DNN and the up-scaling DNN, the parameter being obtained through joint training of the up-scaling DNN and a down-scaling DNN.

Mode for the Invention

- [36] As the disclosure allows for various changes and numerous examples, particular embodiments will be illustrated in the drawings and described in detail in the written description. However, this is not intended to limit the disclosure to particular modes of practice, and it will be understood that all changes, equivalents, and substitutes that do not depart from the spirit and technical scope of the disclosure are encompassed in the disclosure.
- [37] In the description of embodiments, certain detailed explanations of related art are omitted when it is deemed that they may unnecessarily obscure the essence of the disclosure. Also, numbers (for example, a first, a second, and the like) used in the description of the specification are merely identifier codes for distinguishing one element

from another.

- [38] Also, in the present specification, it will be understood that when elements are "connected" or "coupled" to each other, the elements may be directly connected or coupled to each other, but may alternatively be connected or coupled to each other with an intervening element therebetween, unless specified otherwise.
- [39] In the present specification, regarding an element represented as a "unit" or a "module", two or more elements may be combined into one element or one element may be divided into two or more elements according to subdivided functions. In addition, each element described hereinafter may additionally perform some or all of functions performed by another element, in addition to main functions of itself, and some of the main functions of each element may be performed entirely by another component.
- [40] Also, in the present specification, an 'image' or a 'picture' may denote a still image, a moving image including a plurality of consecutive still images (or frames), or a video.
- [41] Also, in the present specification, a deep neural network (DNN) is a representative example of an artificial neural network model simulating brain nerves, and is not limited to an artificial neural network model using a specific algorithm.
- [42] Also, in the present specification, a 'parameter' is a value used in an operation process of each layer forming a neural network, and for example, may include a weight used when an input value is applied to a certain operation expression. Here, the parameter may be expressed in a matrix form. The parameter is a value set as a result of training, and may be updated through separate training data when necessary.
- [43] Also, in the present specification, a 'first DNN' indicates a DNN used for artificial intelligence (AI) down-scaling an image, and a 'second DNN' indicates a DNN used for AI up-scaling an image.
- [44] Also, in the present specification, 'DNN setting information' includes information related to an element constituting a DNN. 'DNN setting information' includes the parameter described above as information related to the element constituting the DNN. The first DNN or the second DNN may be set by using the DNN setting information.
- [45] Also, in the present specification, an 'original image' denotes an image to be an object of AI encoding, and a 'first image' denotes an image obtained as a result of performing AI down-scaling on the original image during an AI encoding process. Also, a 'second image' denotes an image obtained via first decoding during an AI decoding process, and a 'third image' denotes an image obtained by AI up-scaling the second image during the AI decoding process.
- [46] Also, in the present specification, 'AI down-scale' denotes a process of decreasing resolution of an image based on AI, and 'first encoding' denotes an encoding process according to an image compression method based on frequency transformation. Also,

'first decoding' denotes a decoding process according to an image reconstruction method based on frequency transformation, and 'AI up-scale' denotes a process of increasing resolution of an image based on AI.

[47] FIG. 1 is a diagram for describing an AI encoding process and an AI decoding process, according to an embodiment.

[48] As described above, when resolution of an image remarkably increases, the throughput of information for encoding and decoding the image is increased, and accordingly, a method for improving efficiency of encoding and decoding of an image is required.

[49] As shown in FIG. 1, according to an embodiment of the disclosure, a first image 115 is obtained by performing AI down-scaling 110 on an original image 105 having high resolution. Then, first encoding 120 and first decoding 130 are performed on the first image 115 having relatively low resolution, and thus a bitrate may be largely reduced compared to when the first encoding and the first decoding are performed on the original image 105.

[50] In particular, in FIG. 1, the first image 115 is obtained by performing the AI down-scaling 110 on the original image 105 and the first encoding 120 is performed on the first image 115 during the AI encoding process, according to an embodiment. During the AI decoding process, AI encoding data including AI data and image data, which are obtained as a result of AI encoding is received, a second image 135 is obtained via the first decoding 130, and a third image 145 is obtained by performing AI up-scaling 140 on the second image 135.

[51] Referring to the AI encoding process in detail, when the original image 105 is received, the AI down-scaling 110 is performed on the original image 105 to obtain the first image 115 of certain resolution or certain quality. Here, the AI down-scaling 110 is performed based on AI, and AI for the AI down-scaling 110 needs to be trained jointly with AI for the AI up-scaling 140 of the second image 135. This is because, when the AI for the AI down-scaling 110 and the AI for the AI up-scaling 140 are separately trained, a difference between the original image 105 which is an object of AI encoding and the third image 145 reconstructed through AI decoding is increased.

[52] In an embodiment of the disclosure, the AI data may be used to maintain such a joint relationship during the AI encoding process and the AI decoding process. Accordingly, the AI data obtained through the AI encoding process may include information indicating an up-scaling target, and during the AI decoding process, the AI up-scaling 140 is performed on the second image 135 according to the up-scaling target verified based on the AI data.

[53] The AI for the AI down-scaling 110 and the AI for the AI up-scaling 140 may be embodied as a DNN. As will be described later with reference to FIG. 9, because a first

DNN and a second DNN are jointly trained by sharing loss information under a certain target, an AI encoding apparatus may provide target information used during joint training of the first DNN and the second DNN to an AI decoding apparatus, and the AI decoding apparatus may perform the AI up-scaling 140 on the second image 135 to target resolution based on the provided target information.

[54] Regarding the first encoding 120 and the first decoding 130 of FIG. 1, information amount of the first image 115 obtained by performing AI down-scaling 110 on the original image 105 may be reduced through the first encoding 120. The first encoding 120 may include a process of generating prediction data by predicting the first image 115, a process of generating residual data corresponding to a difference between the first image 115 and the prediction data, a process of transforming the residual data of a spatial domain component to a frequency domain component, a process of quantizing the residual data transformed to the frequency domain component, and a process of entropy-encoding the quantized residual data. Such first encoding 120 may be performed via one of image compression methods using frequency transformation, such as MPEG-2, H.264 Advanced Video Coding (AVC), MPEG-4, High Efficiency Video Coding (HEVC), VC-1, VP8, VP9, and AOMedia Video 1 (AV1).

[55] The second image 135 corresponding to the first image 115 may be reconstructed by performing the first decoding 130 on the image data. The first decoding 130 may include a process of generating the quantized residual data by entropy-decoding the image data, a process of inverse-quantizing the quantized residual data, a process of transforming the residual data of the frequency domain component to the spatial domain component, a process of generating the prediction data, and a process of reconstructing the second image 135 by using the prediction data and the residual data. Such first decoding 130 may be performed via an image reconstruction method corresponding to one of image compression methods using frequency transformation, such as MPEG-2, H.264 AVC, MPEG-4, HEVC, VC-1, VP8, VP9, and AV1, which is used in the first encoding 120.

[56] The AI encoding data obtained through the AI encoding process may include the image data obtained as a result of performing the first encoding 120 on the first image 115, and the AI data related to the AI down-scaling 110 of the original image 105. The image data may be used during the first decoding 130 and the AI data may be used during the AI up-scaling 140.

[57] The image data may be transmitted in a form of a bitstream. The image data may include data obtained based on pixel values in the first image 115, for example, residual data that is a difference between the first image 115 and prediction data of the first image 115. Also, the image data includes information used during the first encoding 120 performed on the first image 115. For example, the image data may

include prediction mode information, motion information, and information related to quantization parameter used during the first encoding 120. The image data may be generated according to a rule, for example, according to a syntax, of an image compression method used during the first encoding 120, among MPEG-2, H.264 AVC, MPEG-4, HEVC, VC-1, VP8, VP9, and AV1.

[58] The AI data is used in the AI up-scaling 140 based on the second DNN. As described above, because the first DNN and the second DNN are jointly trained, the AI data includes information enabling the AI up-scaling 140 to be performed accurately on the second image 135 through the second DNN. During the AI decoding process, the AI up-scaling 140 may be performed on the second image 135 to have targeted resolution and/or quality, based on the AI data.

[59] The AI data may be transmitted together with the image data in a form of a bitstream. Alternatively, according to an embodiment, the AI data may be transmitted separately from the image data, in a form of a frame or a packet. The AI data and the image data obtained as a result of the AI encoding may be transmitted through the same network or through different networks.

[60] FIG. 2 is a block diagram of a configuration of an AI decoding apparatus 100 according to an embodiment.

[61] Referring to FIG. 2, the AI decoding apparatus 200 according to an embodiment may include a receiver 210 and an AI decoder 230. The receiver 210 may include a communicator 212, a parser 214, and an outputter 216. The AI decoder 230 may include a first decoder 232 and an AI up-scaler 234.

[62] The receiver 210 receives and parses AI encoding data obtained as a result of AI encoding, and distinguishably outputs image data and AI data to the AI decoder 230.

[63] In particular, the communicator 212 receives the AI encoding data obtained as the result of AI encoding through a network. The AI encoding data obtained as the result of performing AI encoding includes the image data and the AI data. The image data and the AI data may be received through a same type of network or different types of networks.

[64] The parser 214 receives the AI encoding data received through the communicator 212 and parses the AI encoding data to distinguish the image data and the AI data. For example, the parser 214 may distinguish the image data and the AI data by reading a header of data obtained from the communicator 212. According to an embodiment, the parser 214 distinguishably transmits the image data and the AI data to the outputter 216 via the header of the data received through the communicator 212, and the outputter 216 transmits the distinguished image data and AI data respectively to the first decoder 232 and the AI up-scaler 234. At this time, it may be verified that the image data included in the AI encoding data is image data generated via a certain

codec (for example, MPEG-2, H.264 AVC, MPEG-4, HEVC, VC-1, VP8, VP9, or AV1). In this case, corresponding information may be transmitted to the first decoder 232 through the outputter 216 such that the image data is processed via the verified codec.

[65] According to an embodiment, the AI encoding data parsed by the parser 214 may be obtained from a data storage medium including a magnetic medium such as a hard disk, a floppy disk, or a magnetic tape, an optical recording medium such as CD-ROM or DVD, or a magneto-optical medium such as a floptical disk.

[66] The first decoder 232 reconstructs the second image 135 corresponding to the first image 115, based on the image data. The second image 135 obtained by the first decoder 232 is provided to the AI up-scaler 234. According to an embodiment, first decoding related information, such as prediction mode information, motion information, quantization parameter information, or the like included in the image data may be further provided to the AI up-scaler 234.

[67] Upon receiving the AI data, the AI up-scaler 234 performs AI up-scaling on the second image 135, based on the AI data. According to an embodiment, the AI up-scaling may be performed by further using the first decoding related information, such as the prediction mode information, the quantization parameter information, or the like included in the image data.

[68] The receiver 210 and the AI decoder 230 according to an embodiment are described as individual devices, but may be implemented through one processor. In this case, the receiver 210 and the AI decoder 230 may be implemented through an dedicated processor or through a combination of software and general-purpose processor such as application processor (AP), central processing unit (CPU) or graphic processing unit (GPU). The dedicated processor may be implemented by including a memory for implementing an embodiment of the disclosure or by including a memory processor for using an external memory.

[69] Also, the receiver 210 and the AI decoder 230 may be configured by a plurality of processors. In this case, the receiver 210 and the AI decoder 230 may be implemented through a combination of dedicated processors or through a combination of software and general-purpose processors such as AP, CPU or GPU. Similarly, the AI up-scaler 234 and the first decoder 232 may be implemented by different processors.

[70] The AI data provided to the AI up-scaler 234 includes information enabling the second image 135 to be processed via AI up-scaling. Here, an up-scaling target should correspond to down-scaling of a first DNN. Accordingly, the AI data includes information for verifying a down-scaling target of the first DNN.

[71] Examples of the information included in the AI data include difference information between resolution of the original image 105 and resolution of the first image 115, and

information related to the first image 115.

- [72] The difference information may be expressed as information about a resolution conversion degree of the first image 115 compared to the original image 105 (for example, resolution conversion rate information). Also, because the resolution of the first image 115 is verified through the resolution of the reconstructed second image 135 and the resolution conversion degree is verified accordingly, the difference information may be expressed only as resolution information of the original image 105. Here, the resolution information may be expressed as vertical/horizontal sizes or as a ratio (16:9, 4:3, or the like) and a size of one axis. Also, when there is pre-set resolution information, the resolution information may be expressed in a form of an index or flag.
- [73] The information related to the first image 115 may include information about at least one of a bitrate of the image data obtained as the result of performing first encoding on the first image 115 or a codec type used during the first encoding of the first image 115.
- [74] The AI up-scaler 234 may determine the up-scaling target of the second image 135, based on at least one of the difference information or the information related to the first image 115, which are included in the AI data. The up-scaling target may indicate, for example, to what degree resolution is to be up-scaled for the second image 135. When the up-scaling target is determined, the AI up-scaler 234 performs AI up-scaling on the second image 135 through a second DNN to obtain the third image 145 corresponding to the up-scaling target.
- [75] Before describing a method, performed by the AI up-scaler 234, of performing AI up-scaling on the second image 135 according to the up-scaling target, an AI up-scaling process through the second DNN will be described with reference to FIGS. 3 and 4.
- [76] FIG. 3 is a diagram showing a second DNN 300 for performing AI up-scaling on the second image 135, and FIG. 4 is a diagram for describing a convolution operation in a first convolution layer 310 of FIG. 3.
- [77] As shown in FIG. 3, the second image 135 is input to the first convolution layer 310. 3X3X4 indicated in the first convolution layer 310 shown in FIG. 3 indicates that a convolution process is performed on one input image by using four filter kernels having a size of 3x3. Four feature maps are generated by the four filter kernels as a result of the convolution process. Each feature map indicates inherent characteristics of the second image 135. For example, each feature map may represent a vertical direction characteristic, a horizontal direction characteristic, or an edge characteristic, etc of the second image 135.
- [78] A convolution operation in the first convolution layer 310 will be described in detail

with reference to FIG. 4.

[79] One feature map 450 may be generated through multiplication and addition between parameters of a filter kernel 430 having a size of 3x3 used in the first convolution layer 310 and corresponding pixel values in the second image 135. Because four filter kernels are used in the first convolution layer 310, four feature maps may be generated through the convolution operation using the four filter kernels.

[80] I1 through I49 indicated in the second image 135 in FIG. 4 indicate pixels in the second image 135, and F1 through F9 indicated in the filter kernel 430 indicate parameters of the filter kernel 430. Also, M1 through M9 indicated in the feature map 450 indicate samples of the feature map 450.

[81] In FIG. 4, the second image 135 includes 49 pixels, but the number of pixels is only an example and when the second image 135 has a resolution of 4 K, the second image 135 may include, for example, 3840x2160 pixels.

[82] During a convolution operation process, pixel values of I1, I2, I3, I8, I9, I10, I15, I16, and I17 of the second image 135 and F1 through F9 of the filter kernels 430 are respectively multiplied, and a value of combination (for example, addition) of result values of the multiplication may be assigned as a value of M1 of the feature map 450. When a stride of the convolution operation is 2, pixel values of I3, I4, I5, I10, I11, I12, I17, I18, and I19 of the second image 135 and F1 through F9 of the filter kernels 430 are respectively multiplied, and the value of the combination of the result values of the multiplication may be assigned as a value of M2 of the feature map 450.

[83] While the filter kernel 430 moves along the stride to the last pixel of the second image 135, the convolution operation is performed between the pixel values in the second image 135 and the parameters of the filter kernel 430, and thus the feature map 450 having a certain size may be generated.

[84] According to the present disclosure, values of parameters of a second DNN, for example, values of parameters of a filter kernel used in convolution layers of the second DNN (for example, F1 through F9 of the filter kernel 430), may be optimized through joint training of a first DNN and the second DNN. As described above, the AI up-scaler 234 may determine an up-scaling target corresponding to a down-scaling target of the first DNN based on AI data, and determine parameters corresponding to the determined up-scaling target as the parameters of the filter kernel used in the convolution layers of the second DNN.

[85] Convolution layers included in the first DNN and the second DNN may perform processes according to the convolution operation process described with reference to FIG. 4, but the convolution operation process described with reference to FIG. 4 is only an example and is not limited thereto.

[86] Referring back to FIG. 3, the feature maps output from the first convolution layer

310 may be input to a first activation layer 320.

- [87] The first activation layer 320 may assign a non-linear feature to each feature map. The first activation layer 320 may include a sigmoid function, a Tanh function, a rectified linear unit (ReLU) function, or the like, but is not limited thereto.
- [88] The first activation layer 320 assigning the non-linear feature indicates that at least one sample value of the feature map, which is an output of the first convolution layer 310, is changed. Here, the change is performed by applying the non-linear feature.
- [89] The first activation layer 320 determines whether to transmit sample values of the feature maps output from the first convolution layer 310 to the second convolution layer 330. For example, some of the sample values of the feature maps are activated by the first activation layer 320 and transmitted to the second convolution layer 330, and some of the sample values are deactivated by the first activation layer 320 and not transmitted to the second convolution layer 330. The intrinsic characteristics of the second image 135 represented by the feature maps are emphasized by the first activation layer 320.
- [90] Feature maps 325 output from the first activation layer 320 are input to the second convolution layer 330. One of the feature maps 325 shown in FIG. 3 is a result of processing the feature map 450 described with reference to FIG. 4 in the first activation layer 320.
- [91] 3X3X4 indicated in the second convolution layer 330 indicates that a convolution process is performed on the feature maps 325 by using four filter kernels having a size of 3x3. An output of the second convolution layer 330 is input to a second activation layer 340. The second activation layer 340 may assign a non-linear feature to input data.
- [92] Feature maps 345 output from the second activation layer 340 are input to a third convolution layer 350. 3X3X1 indicated in the third convolution layer 350 shown in FIG. 3 indicates that a convolution process is performed to generate one output image by using one filter kernel having a size of 3x3. The third convolution layer 350 is a layer for outputting a final image and generates one output by using one filter kernel. According to an embodiment of the disclosure, the third convolution layer 350 may output the third image 145 as a result of a convolution operation.
- [93] There may be a plurality of pieces of DNN setting information indicating the numbers of filter kernels of the first, second, and third convolution layers 310, 330, and 350 of the second DNN 300, a parameter of filter kernels of the first, second, and third convolution layers 310, 330, and 350 of the second DNN 300, and the like, as will be described later, and the plurality of pieces of DNN setting information should be connected to a plurality of pieces of DNN setting information of a first DNN. The connection between the plurality of pieces of DNN setting information of the second

DNN and the plurality of pieces of DNN setting information of the first DNN may be realized via joint training of the first DNN and the second DNN.

[94] In FIG. 3, the second DNN 300 includes three convolution layers (the first, second, and third convolution layers 310, 330, and 350) and two activation layers (the first and second activation layers 320 and 340), but this is only an example, and the numbers of convolution layers and activation layers may vary according to an embodiment. Also, according to an embodiment, the second DNN 300 may be implemented as a recurrent neural network (RNN). In this case, a convolutional neural network (CNN) structure of the second DNN 300 according to an embodiment of the disclosure is changed to an RNN structure.

[95] According to an embodiment, the AI up-scaler 234 may include at least one arithmetic logic unit (ALU) for the convolution operation and the operation of the activation layer described above. The ALU may be implemented as a processor. For the convolution operation, the ALU may include a multiplier that performs multiplication between sample values of the second image 135 or the feature map output from previous layer and sample values of the filter kernel, and an adder that adds result values of the multiplication. Also, for the operation of the activation layer, the ALU may include a multiplier that multiplies an input sample value by a weight used in a pre-determined sigmoid function, a Tanh function, or an ReLU function, and a comparator that compares a multiplication result and a certain value to determine whether to transmit the input sample value to a next layer.

[96] Hereinafter, a method, performed by the AI up-scaler 234, of performing the AI up-scaling on the second image 135 according to the up-scaling target will be described.

[97] According to an embodiment, the AI up-scaler 234 may store a plurality of pieces of DNN setting information settable in a second DNN.

[98] Here, the DNN setting information may include information about at least one of the number of convolution layers included in the second DNN, the number of filter kernels for each convolution layer, or a parameter of each filter kernel. The plurality of pieces of DNN setting information may respectively correspond to various up-scaling targets, and the second DNN may operate based on DNN setting information corresponding to a certain up-scaling target. The second DNN may have different structures based on the DNN setting information. For example, the second DNN may include three convolution layers based on any piece of DNN setting information, and may include four convolution layers based on another piece of DNN setting information.

[99] According to an embodiment, the DNN setting information may only include a parameter of a filter kernel used in the second DNN. In this case, the structure of the second DNN does not change, but only the parameter of the internal filter kernel may change based on the DNN setting information.

- [100] The AI up-scaler 234 may obtain the DNN setting information for performing AI up-scaling on the second image 135, among the plurality of pieces of DNN setting information. Each of the plurality of pieces of DNN setting information used at this time is information for obtaining the third image 145 of pre-determined resolution and/or pre-determined quality, and is trained jointly with a first DNN.
- [101] For example, one piece of DNN setting information among the plurality of pieces of DNN setting information may include information for obtaining the third image 145 of resolution twice higher than resolution of the second image 135, for example, the third image 145 of 4 K (4096x2160) twice higher than 2 K (2048x1080) of the second image 135, and another piece of DNN setting information may include information for obtaining the third image 145 of resolution four times higher than the resolution of the second image 135, for example, the third image 145 of 8 K (8192x4320) four times higher than 2 K (2048x1080) of the second image 135.
- [102] Each of the plurality of pieces of DNN setting information is obtained jointly with DNN setting information of the first DNN of an AI encoding apparatus 600 of FIG. 6, and the AI up-scaler 234 obtains one piece of DNN setting information among the plurality of pieces of DNN setting information according to an enlargement ratio corresponding to a reduction ratio of the DNN setting information of the first DNN. In this regard, the AI up-scaler 234 may verify information of the first DNN. In order for the AI up-scaler 234 to verify the information of the first DNN, the AI decoding apparatus 200 according to an embodiment receives AI data including the information of the first DNN from the AI encoding apparatus 600.
- [103] In other words, the AI up-scaler 234 may verify information targeted by DNN setting information of the first DNN used to obtain the first image 115 and obtain the DNN setting information of the second DNN trained jointly with the DNN setting information of the first DNN, by using information received from the AI encoding apparatus 600.
- [104] When DNN setting information for performing the AI up-scaling on the second image 135 is obtained from among the plurality of pieces of DNN setting information, input data may be processed based on the second DNN operating according to the obtained DNN setting information.
- [105] For example, when any one piece of DNN setting information is obtained, the number of filter kernels included in each of the first, second, and third convolution layers 310, 330, and 350 of the second DNN 300 of FIG. 3, and the parameters of the filter kernels are set to values included in the obtained DNN setting information.
- [106] In particular, parameters of a filter kernel of 3x3 used in any one convolution layer of the second DNN of FIG. 4 are set to {1, 1, 1, 1, 1, 1, 1, 1, 1}, and when DNN setting information is changed afterwards, the parameters are replaced by {2, 2, 2, 2, 2, 2, 2, 2, 2,

2} that are parameters included in the changed DNN setting information.

- [107] The AI up-scaler 234 may obtain the DNN setting information for AI up-scaling from among the plurality of pieces of DNN setting information, based on information included in the AI data, and the AI data used to obtain the DNN setting information will now be described.
- [108] According to an embodiment, the AI up-scaler 234 may obtain the DNN setting information for AI up-scaling from among the plurality of pieces of DNN setting information, based on difference information included in the AI data. For example, when it is verified that the resolution (for example, 4 K (4096x2160)) of the original image 105 is twice higher than the resolution (for example, 2 K (2048x1080)) of the first image 115, based on the difference information, the AI up-scaler 234 may obtain the DNN setting information for increasing the resolution of the second image 135 two times.
- [109] According to another embodiment, the AI up-scaler 234 may obtain the DNN setting information for AI up-scaling the second image 135 from among the plurality of pieces of DNN setting information, based on information related to the first image 115 included in the AI data. The AI up-scaler 234 may pre-determine a mapping relationship between image-related information and DNN setting information, and obtain the DNN setting information mapped to the information related to the first image 115.
- [110] FIG. 5 is a table showing a mapping relationship between several pieces of image-related information and several pieces of DNN setting information.
- [111] Through an embodiment according to FIG. 5, it will be determined that AI encoding and AI decoding processes according to an embodiment of the disclosure do not only consider a change of resolution. As shown in FIG. 5, DNN setting information may be selected considering resolution, such as standard definition (SD), high definition (HD), or full HD, a bitrate, such as 10 Mbps, 15 Mbps, or 20 Mbps, and codec information, such as AV1, H.264, or HEVC, individually or collectively. For such consideration of the resolution, the bitrate and the codec information, training in consideration of each element should be jointly performed with encoding and decoding processes during an AI training process (see FIG. 9).
- [112] Accordingly, when a plurality of pieces of DNN setting information are provided based on image-related information including a codec type, resolution of an image, and the like, as shown in FIG. 5 according to training, the DNN setting information for AI up-scaling the second image 135 may be obtained based on the information related to the first image 115 received during the AI decoding process.
- [113] In other words, the AI up-scaler 234 is capable of using DNN setting information according to image-related information by matching the image-related information at the left of a table of FIG. 5 and the DNN setting information at the right of the table.

- [114] As shown in FIG. 5, when it is verified, from the information related to the first image 115, that the resolution of the first image 115 is SD, a bitrate of image data obtained as a result of performing first encoding on the first image 115 is 10 Mbps, and the first encoding is performed on the first image 115 via AV1 codec, the AI up-scaler 234 may use A DNN setting information among the plurality of pieces of DNN setting information.
- [115] Also, when it is verified, from the information related to the first image 115, that the resolution of the first image 115 is HD, the bitrate of the image data obtained as the result of performing the first encoding is 15 Mbps, and the first encoding is performed via H.264 codec, the AI up-scaler 234 may use B DNN setting information among the plurality of pieces of DNN setting information.
- [116] Also, when it is verified, from the information related to the first image 115, that the resolution of the first image 115 is full HD, the bitrate of the image data obtained as the result of performing the first encoding is 20 Mbps, and the first encoding is performed via HEVC codec, the AI up-scaler 234 may use C DNN setting information among the plurality of pieces of DNN setting information, and when it is verified that the resolution of the first image 115 is full HD, the bitrate of the image data obtained as the result of performing the first encoding is 15 Mbps, and the first encoding is performed via HEVC codec, the AI up-scaler 234 may use D DNN setting information among the plurality of pieces of DNN setting information. One of the C DNN setting information and the D DNN setting information is selected based on whether the bitrate of the image data obtained as the result of performing the first encoding on the first image 115 is 20 Mbps or 15 Mbps. The different bitrates of the image data, obtained when the first encoding is performed on the first image 115 of the same resolution via the same codec, indicates different qualities of reconstructed images. Accordingly, a first DNN and a second DNN may be jointly trained based on certain image quality, and accordingly, the AI up-scaler 234 may obtain DNN setting information according to a bitrate of image data indicating the quality of the second image 135.
- [117] According to another embodiment, the AI up-scaler 234 may obtain the DNN setting information for performing AI up-scaling on the second image 135 from among the plurality of pieces of DNN setting information considering both information (prediction mode information, motion information, quantization parameter information, and the like) provided from the first decoder 232 and the information related to the first image 115 included in the AI data. For example, the AI up-scaler 234 may receive quantization parameter information used during a first encoding process of the first image 115 from the first decoder 232, verify a bitrate of image data obtained as an encoding result of the first image 115 from AI data, and obtain DNN

setting information corresponding to the quantization parameter information and the bitrate. Even when the bitrates are the same, the quality of reconstructed images may vary according to the complexity of an image. A bitrate is a value representing the entire first image 115 on which first encoding is performed, and the quality of each frame may vary even within the first image 115. Accordingly, DNN setting information more suitable for the second image 135 may be obtained when prediction mode information, motion information, and/or a quantization parameter obtainable for each frame from the first decoder 232 are/is considered together, compared to when only the AI data is used.

[118] Also, according to an embodiment, the AI data may include an identifier of mutually agreed DNN setting information. An identifier of DNN setting information is information for distinguishing a pair of pieces of DNN setting information jointly trained between the first DNN and the second DNN, such that AI up-scaling is performed on the second image 135 to the up-scaling target corresponding to the down-scaling target of the first DNN. The AI up-scaler 234 may perform AI up-scaling on the second image 135 by using the DNN setting information corresponding to the identifier of the DNN setting information, after obtaining the identifier of the DNN setting information included in the AI data. For example, identifiers indicating each of the plurality of DNN setting information settable in the first DNN and identifiers indicating each of the plurality of DNN setting information settable in the second DNN may be previously designated. In this case, the same identifier may be designated for a pair of DNN setting information settable in each of the first DNN and the second DNN. The AI data may include an identifier of DNN setting information set in the first DNN for AI down-scaling of the original image 105. The AI up-scaler 234 that receives the AI data may perform AI up-scaling on the second image 135 by using the DNN setting information indicated by the identifier included in the AI data among the plurality of DNN setting information.

[119] Also, according to an embodiment, the AI data may include the DNN setting information. The AI up-scaler 234 may perform AI up-scaling on the second image 135 by using the DNN setting information after obtaining the DNN setting information included in the AI data.

[120] According to an embodiment, when pieces of information (for example, the number of convolution layers, the number of filter kernels for each convolution layer, a parameter of each filter kernel, and the like) constituting the DNN setting information are stored in a form of a lookup table, the AI up-scaler 234 may obtain the DNN setting information by combining some values selected from values in the lookup table, based on information included in the AI data, and perform AI up-scaling on the second image 135 by using the obtained DNN setting information.

- [121] According to an embodiment, when a structure of DNN corresponding to the up-scaling target is determined, the AI up-scaler 234 may obtain the DNN setting information, for example, parameters of a filter kernel, corresponding to the determined structure of DNN.
- [122] The AI up-scaler 234 obtains the DNN setting information of the second DNN through the AI data including information related to the first DNN, and performs AI up-scaling on the second image 135 through the second DNN set based on the obtained DNN setting information, and in this case, memory usage and throughput may be reduced compared to when features of the second image 135 are directly analyzed for up-scaling.
- [123] According to an embodiment, when the second image 135 includes a plurality of frames, the AI up-scaler 234 may independently obtain DNN setting information for a certain number of frames, or may obtain common DNN setting information for entire frames.
- [124] FIG. 6 is a diagram showing the second image 135 including a plurality of frames.
- [125] As shown in FIG. 6, the second image 135 may include frames t_0 through t_n .
- [126] According to an embodiment, the AI up-scaler 234 may obtain DNN setting information of a second DNN through AI data, and perform AI up-scaling on the frames t_0 through t_n based on the obtained DNN setting information. In other words, the frames t_0 through t_n may be processed via AI up-scaling based on common DNN setting information.
- [127] According to another embodiment, the AI up-scaler 234 may perform AI up-scaling on some of the frames t_0 through t_n , for example, the frames t_0 through t_a , by using 'A' DNN setting information obtained from AI data, and perform AI up-scaling on the frames t_{a+1} through t_b by using 'B' DNN setting information obtained from the AI data. Also, the AI up-scaler 234 may perform AI up-scaling on the frames t_{b+1} through t_n by using 'C' DNN setting information obtained from the AI data. In other words, the AI up-scaler 234 may independently obtain DNN setting information for each group including a certain number of frames among the plurality of frames, and perform AI up-scaling on frames included in each group by using the independently obtained DNN setting information.
- [128] According to another embodiment, the AI up-scaler 234 may independently obtain DNN setting information for each frame forming the second image 135. In other words, when the second image 135 includes three frames, the AI up-scaler 234 may perform AI up-scaling on a first frame by using DNN setting information obtained in relation to the first frame, perform AI up-scaling on a second frame by using DNN setting information obtained in relation to the second frame, and perform AI up-scaling on a third frame by using DNN setting information obtained in relation to the third

frame. DNN setting information may be independently obtained for each frame included in the second image 135, according to a method of obtaining DNN setting information based on information (prediction mode information, motion information, quantization parameter information, or the like) provided from the first decoder 232 and information related to the first image 115 included in the AI data described above. This is because the mode information, the quantization parameter information, or the like may be determined independently for each frame included in the second image 135.

- [129] According to another embodiment, the AI data may include information about to which frame DNN setting information obtained based on the AI data is valid. For example, when the AI data includes information indicating that DNN setting information is valid up to the frame t_a , the AI up-scaler 234 performs AI up-scaling on the frames t_0 through t_a by using DNN setting information obtained based on the AI data. Also, when another piece of AI data includes information indicating that DNN setting information is valid up to the frame t_n , the AI up-scaler 234 performs AI up-scaling on the frames t_{a+1} through t_n by using DNN setting information obtained based on the other piece of AI data.
- [130] Hereinafter, the AI encoding apparatus 600 for performing AI encoding on the original image 105 will be described with reference to FIG. 7.
- [131] FIG. 7 is a block diagram of a configuration of the AI encoding apparatus 600 according to an embodiment.
- [132] Referring to FIG. 7, the AI encoding apparatus 600 may include an AI encoder 610 and a transmitter 630. The AI encoder 610 may include an AI down-scaler 612 and a first encoder 614. The transmitter 630 may include a data processor 632 and a communicator 634.
- [133] In FIG. 7, the AI encoder 610 and the transmitter 630 are illustrated as separate devices, but the AI encoder 610 and the transmitter 630 may be implemented through one processor. In this case, the AI encoder 610 and the transmitter 630 may be implemented through an dedicated processor or through a combination of software and general-purpose processor such as AP, CPU or graphics processing unit GPU. The dedicated processor may be implemented by including a memory for implementing an embodiment of the disclosure or by including a memory processor for using an external memory.
- [134] Also, the AI encoder 610 and the transmitter 630 may be configured by a plurality of processors. In this case, the AI encoder 610 and the transmitter 630 may be implemented through a combination of dedicated processors or through a combination of software and a plurality of general-purpose processors such as AP, CPU or GPU. The AI down-scaler 612 and the first encoder 614 may be implemented through different

processors.

- [135] The AI encoder 610 performs AI down-scaling on the original image 105 and first encoding on the first image 115, and transmits AI data and image data to the transmitter 630. The transmitter 630 transmits the AI data and the image data to the AI decoding apparatus 200.
- [136] The image data includes data obtained as a result of performing the first encoding on the first image 115. The image data may include data obtained based on pixel values in the first image 115, for example, residual data that is a difference between the first image 115 and prediction data of the first image 115. Also, the image data includes information used during a first encoding process of the first image 115. For example, the image data may include prediction mode information, motion information, quantization parameter information used to perform the first encoding on the first image 115, and the like.
- [137] The AI data includes information enabling AI up-scaling to be performed on the second image 135 to an up-scaling target corresponding to a down-scaling target of a first DNN. According to an embodiment, the AI data may include difference information between the original image 105 and the first image 115. Also, the AI data may include information related to the first image 115. The information related to the first image 115 may include information about at least one of resolution of the first image 115, a bitrate of the image data obtained as the result of performing the first encoding on the first image 115, or a codec type used during the first encoding of the first image 115.
- [138] According to an embodiment, the AI data may include an identifier of mutually agreed DNN setting information such that the AI up-scaling is performed on the second image 135 to the up-scaling target corresponding to the down-scaling target of the first DNN.
- [139] Also, according to an embodiment, the AI data may include DNN setting information settable in a second DNN.
- [140] The AI down-scaler 612 may obtain the first image 115 obtained by performing the AI down-scaling on the original image 105 through the first DNN. The AI down-scaler 612 may determine the down-scaling target of the original image 105, based on a predetermined standard.
- [141] In order to obtain the first image 115 matching the down-scaling target, the AI down-scaler 612 may store a plurality of pieces of DNN setting information settable in the first DNN. The AI down-scaler 612 obtains DNN setting information corresponding to the down-scaling target from among the plurality of pieces of DNN setting information, and performs the AI down-scaling on the original image 105 through the first DNN set in the obtained DNN setting information.

- [142] Each of the plurality of pieces of DNN setting information may be trained to obtain the first image 115 of pre-determined resolution and/or pre-determined quality. For example, any one piece of DNN setting information among the plurality of pieces of DNN setting information may include information for obtaining the first image 115 of resolution half resolution of the original image 105, for example, the first image 115 of 2 K (2048x1080) half 4 K (4096x2160) of the original image 105, and another piece of DNN setting information may include information for obtaining the first image 115 of resolution quarter resolution of the original image 105, for example, the first image 115 of 2 K (2048x1080) quarter 8 K (8192x4320) of the original image 105.
- [143] According to an embodiment, when pieces of information (for example, the number of convolution layers, the number of filter kernels for each convolution layer, a parameter of each filter kernel, and the like) constituting the DNN setting information are stored in a form of a lookup table, the AI down-scaler 612 may obtain the DNN setting information by combining some values selected from values in the lookup table, based on the down-scaling target, and perform AI down-scaling on the original image 105 by using the obtained DNN setting information.
- [144] According to an embodiment, the AI down-scaler 612 may determine a structure of DNN corresponding to the down-scaling target, and obtain DNN setting information corresponding to the determined structure of DNN, for example, obtain parameters of a filter kernel.
- [145] The plurality of pieces of DNN setting information for performing the AI down-scaling on the original image 105 may have an optimized value as the first DNN and the second DNN are jointly trained. Here, each piece of DNN setting information includes at least one of the number of convolution layers included in the first DNN, the number of filter kernels for each convolution layer, or a parameter of each filter kernel.
- [146] The AI down-scaler 612 may set the first DNN with the DNN setting information obtained for performing the AI down-scaling on the original image 105 to obtain the first image 115 of certain resolution and/or certain quality through the first DNN. When the DNN setting information for performing the AI down-scaling on the original image 105 is obtained from the plurality of pieces of DNN setting information, each layer in the first DNN may process input data based on information included in the DNN setting information.
- [147] Hereinafter, a method, performed by the AI down-scaler 612, of determining the down-scaling target will be described. The down-scaling target may indicate, for example, by how much is resolution decreased from the original image 105 to obtain the first image 115.
- [148] According to an embodiment, the AI down-scaler 612 may determine the down-scaling target based on at least one of a compression ratio (for example, a resolution

difference between the original image 105 and the first image 115, target bitrate, or the like), compression quality (for example, type of bitrate), compression history information, or a type of the original image 105.

[149] For example, the AI down-scaler 612 may determine the down-scaling target based on the compression ratio, the compression quality, or the like, which is pre-set or input from a user.

[150] As another example, the AI down-scaler 612 may determine the down-scaling target by using the compression history information stored in the AI encoding apparatus 600. For example, according to the compression history information usable by the AI encoding apparatus 600, encoding quality, a compression ratio, or the like preferred by the user may be determined, and the down-scaling target may be determined according to the encoding quality determined based on the compression history information. For example, the resolution, quality, or the like of the first image 115 may be determined according to the encoding quality that has been used most often according to the compression history information.

[151] As another example, the AI down-scaler 612 may determine the down-scaling target based on the encoding quality that has been used more frequently than a certain threshold value (for example, average quality of the encoding quality that has been used more frequently than the certain threshold value), according to the compression history information.

[152] As another example, the AI down-scaler 612 may determine the down-scaling target, based on the resolution, type (for example, a file format), or the like of the original image 105.

[153] According to an embodiment, when the original image 105 includes a plurality of frames, the AI down-scaler 612 may independently determine down-scaling target for a certain number of frames, or may determine down-scaling target for entire frames.

[154] According to an embodiment, the AI down-scaler 612 may divide the frames included in the original image 105 into a certain number of groups, and independently determine the down-scaling target for each group. The same or different down-scaling targets may be determined for each group. The number of frames included in the groups may be the same or different according to the each group.

[155] According to another embodiment, the AI down-scaler 612 may independently determine a down-scaling target for each frame included in the original image 105. The same or different down-scaling targets may be determined for each frame.

[156] Hereinafter, an example of a structure of a first DNN 700 on which AI down-scaling is based will be described.

[157] FIG. 8 is a diagram showing the first DNN 700 for performing AI down-scaling on the original image 105.

- [158] As shown in FIG. 8, the original image 105 is input to a first convolution layer 710. The first convolution layer 710 performs a convolution process on the original image 105 by using 32 filter kernels having a size of 5x5. 32 feature maps generated as a result of the convolution process are input to a first activation layer 720. The first activation layer 720 may assign a non-linear feature to the 32 feature maps.
- [159] The first activation layer 720 determines whether to transmit sample values of the feature maps output from the first convolution layer 710 to the second convolution layer 730. For example, some of the sample values of the feature maps are activated by the first activation layer 720 and transmitted to the second convolution layer 730, and some of the sample values are deactivated by the first activation layer 720 and not transmitted to the second convolution layer 730. Information represented by the feature maps output from the first convolution layer 710 is emphasized by the first activation layer 720.
- [160] An output 725 of the first activation layer 720 is input to a second convolution layer 730. The second convolution layer 730 performs a convolution process on input data by using 32 filter kernels having a size of 5x5. 32 feature maps output as a result of the convolution process are input to a second activation layer 740, and the second activation layer 740 may assign a non-linear feature to the 32 feature maps.
- [161] An output 745 of the second activation layer 740 is input to a third convolution layer 750. The third convolution layer 750 performs a convolution process on input data by using one filter kernel having a size of 5x5. As a result of the convolution process, one image may be output from the third convolution layer 750. The third convolution layer 750 generates one output by using the one filter kernel as a layer for outputting a final image. According to an embodiment of the disclosure, the third convolution layer 750 may output the first image 115 as a result of a convolution operation.
- [162] There may be a plurality of pieces of DNN setting information indicating the numbers of filter kernels of the first, second, and third convolution layers 710, 730, and 750 of the first DNN 700, a parameter of each filter kernel of the first, second, and third convolution layers 710, 730, and 750 of the first DNN 700, and the like, and the plurality of pieces of DNN setting information may be connected to a plurality of pieces of DNN setting information of a second DNN. The connection between the plurality of pieces of DNN setting information of the first DNN and the plurality of pieces of DNN setting information of the second DNN may be realized via joint training of the first DNN and the second DNN.
- [163] In FIG. 8, the first DNN 700 includes three convolution layers (the first, second, and third convolution layers 710, 730, and 750) and two activation layers (the first and second activation layers 720 and 740), but this is only an example, and the numbers of convolution layers and activation layers may vary according to an embodiment. Also,

according to an embodiment, the first DNN 700 may be implemented as an RNN. In this case, a CNN structure of the first DNN 700 according to an embodiment of the disclosure is changed to an RNN structure.

- [164] According to an embodiment, the AI down-scaler 612 may include at least one ALU for the convolution operation and the operation of the activation layer described above. The ALU may be implemented as a processor. For the convolution operation, the ALU may include a multiplier that performs multiplication between sample values of the original image 105 or the feature map output from previous layer and sample values of the filter kernel, and an adder that adds result values of the multiplication. Also, for the operation of the activation layer, the ALU may include a multiplier that multiplies an input sample value by a weight used in a pre-determined sigmoid function, a Tanh function, or an ReLU function, and a comparator that compares a multiplication result and a certain value to determine whether to transmit the input sample value to a next layer.
- [165] Referring back to FIG. 7, upon receiving the first image 115 from the AI down-scaler 612, the first encoder 614 may reduce an information amount of the first image 115 by performing first encoding on the first image 115. The image data corresponding to the first image 115 may be obtained as a result of performing the first encoding by the first encoder 614.
- [166] The data processor 632 processes at least one of the AI data or the image data to be transmitted in a certain form. For example, when the AI data and the image data are to be transmitted in a form of a bitstream, the data processor 632 may process the AI data to be expressed in a form of a bitstream, and transmit the image data and the AI data in a form of one bitstream through the communicator 634. As another example, the data processor 632 may process the AI data to be expressed in a form of bitstream, and transmit each of a bitstream corresponding to the AI data and a bitstream corresponding to the image data through the communicator 634. As another example, the data processor 632 may process the AI data to be expressed in a form of a frame or packet, and transmit the image data in a form of a bitstream and the AI data in a form of a frame or packet through the communicator 634.
- [167] The communicator 634 transmits AI encoding data obtained as a result of performing AI encoding, through a network. The AI encoding data obtained as the result of performing AI encoding includes the image data and the AI data. The image data and the AI data may be transmitted through a same type of network or different types of networks.
- [168] According to an embodiment, the AI encoding data obtained as a result of processes of the data processor 632 may be stored in a data storage medium including a magnetic medium such as a hard disk, a floppy disk, or a magnetic tape, an optical recording

medium such as CD-ROM or DVD, or a magneto-optical medium such as a floptical disk.

[169] Hereinafter, a method of jointly training the first DNN 700 and the second DNN 300 will be described with reference to FIG. 9.

[170] FIG. 9 is a diagram for describing a method of training the first DNN 700 and the second DNN 300.

[171] In an embodiment, the original image 105 on which AI encoding is performed through an AI encoding process is reconstructed to the third image 145 via an AI decoding process, and in order to maintain similarity between the original image 105 and the third image 145 obtained as a result of AI decoding, connectivity is between the AI encoding process and the AI decoding process is required. In other words, information lost in the AI encoding process needs to be reconstructed during the AI decoding process, and in this regard, the first DNN 700 and the second DNN 300 need to be jointly trained.

[172] For accurate AI decoding, ultimately, quality loss information 830 corresponding to a result of comparing a third training image 804 and an original training image 801 shown in FIG. 9 needs to be reduced. Accordingly, the quality loss information 830 is used to train both of the first DNN 700 and the second DNN 300.

[173] First, a training process shown in FIG. 9 will be described.

[174] In FIG. 9, the original training image 801 is an image on which AI down-scaling is to be performed and a first training image 802 is an image obtained by performing AI down-scaling on the original training image 801. Also, the third training image 804 is an image obtained by performing AI up-scaling on the first training image 802.

[175] The original training image 801 includes a still image or a moving image including a plurality of frames. According to an embodiment, the original training image 801 may include a luminance image extracted from the still image or the moving image including the plurality of frames. Also, according to an embodiment, the original training image 801 may include a patch image extracted from the still image or the moving image including the plurality of frames. When the original training image 801 includes the plurality of frames, the first training image 802, the second training image, and the third training image 804 also each include a plurality of frames. When the plurality of frames of the original training image 801 are sequentially input to the first DNN 700, the plurality of frames of the first training image 802, the second training image and the third training image 804 may be sequentially obtained through the first DNN 700 and the second DNN 300.

[176] For joint training of the first DNN 700 and the second DNN 300, the original training image 801 is input to the first DNN 700. The original training image 801 input to the first DNN 700 is output as the first training image 802 via the AI down-scaling, and

the first training image 802 is input to the second DNN 300. The third training image 804 is output as a result of performing the AI up-scaling on the first training image 802.

- [177] Referring to FIG. 9, the first training image 802 is input to the second DNN 300, and according to an embodiment, a second training image obtained as first encoding and first decoding are performed on the first training image 802 may be input to the second DNN 300. In order to input the second training image to the second DNN 300, any one codec among MPEG-2, H.264, MPEG-4, HEVC, VC-1, VP8, VP9, and AV1 may be used. In particular, any one codec among MPEG-2, H.264, MPEG-4, HEVC, VC-1, VP8, VP9, and AV1 may be used to perform first encoding on the first training image 802 and first decoding on image data corresponding to the first training image 802.
- [178] Referring to FIG. 9, separate from the first training image 802 being output through the first DNN 700, a reduced training image 803 obtained by performing legacy down-scaling on the original training image 801 is obtained. Here, the legacy down-scaling may include at least one of bilinear scaling, bicubic scaling, lanczos scaling, or stair step scaling.
- [179] In order to prevent a structural feature of the first image 115 from deviating greatly from a structural feature of the original image 105, the reduced training image 803 is obtained to preserve the structural feature of the original training image 801.
- [180] Before training is performed, the first DNN 700 and the second DNN 300 may be set to pre-determined DNN setting information. When the training is performed, structural loss information 810, complexity loss information 820, and the quality loss information 830 may be determined.
- [181] The structural loss information 810 may be determined based on a result of comparing the reduced training image 803 and the first training image 802. For example, the structural loss information 810 may correspond to a difference between structural information of the reduced training image 803 and structural information of the first training image 802. Structural information may include various features extractable from an image, such as luminance, contrast, histogram, or the like of the image. The structural loss information 810 indicates how much structural information of the original training image 801 is maintained in the first training image 802. When the structural loss information 810 is small, the structural information of the first training image 802 is similar to the structural information of the original training image 801.
- [182] The complexity loss information 820 may be determined based on spatial complexity of the first training image 802. For example, a total variance value of the first training image 802 may be used as the spatial complexity. The complexity loss information 820 is related to a bitrate of image data obtained by performing first encoding on the first

training image 802. It is defined that the bitrate of the image data is low when the complexity loss information 820 is small.

[183] The quality loss information 830 may be determined based on a result of comparing the original training image 801 and the third training image 804. The quality loss information 830 may include at least one of an L1-norm value, an L2-norm value, an Structural Similarity (SSIM) value, a Peak Signal-To-Noise Ratio-Human Vision System (PSNR-HVS) value, an Multiscale SSIM(MS-SSIM) value, a Variance Inflation Factor (VIF) value, or a Video Multimethod Assessment Fusion (VMAF) value regarding the difference between the original training image 801 and the third training image 804. The quality loss information 830 indicates how similar the third training image 804 is to the original training image 801. The third training image 804 is more similar to the original training image 801 when the quality loss information 830 is small.

[184] Referring to FIG. 9, the structural loss information 810, the complexity loss information 820 and the quality loss information 830 are used to train the first DNN 700, and the quality loss information 830 is used to train the second DNN 300. In other words, the quality loss information 830 is used to train both the first and second DNNs 700 and 300.

[185] The first DNN 700 may update a parameter such that final loss information determined based on the first through quality loss information 810 through 830 is reduced or minimized. Also, the second DNN 300 may update a parameter such that the quality loss information 830 is reduced or minimized.

[186] The final loss information for training the first DNN 700 and the second DNN 300 may be determined as Equation 1 below.

[187] [Equation 1]

[188] $\text{LossDS} = a \times \text{Structural loss information} + b \times \text{Complexity loss information} + c \times \text{Quality loss information}$

[189] $\text{LossUS} = d \times \text{Quality loss information}$

[190] In Equation 1, LossDS indicates final loss information to be reduced or minimized to train the first DNN 700, and LossUS indicates final loss information to be reduced or minimized to train the second DNN 300. Also, a, b, c and d may be pre-determined certain weights.

[191] In other words, the first DNN 700 updates parameters in a direction LossDS of Equation 1 is reduced, and the second DNN 300 updates parameters in a direction LossUS is reduced. When the parameters of the first DNN 700 are updated according to LossDS derived during the training, the first training image 802 obtained based on the updated parameters becomes different from a previous first training image 802 obtained based on not updated parameters, and accordingly, the third training image

804 also becomes different from a previous third training image 804. When the third training image 804 becomes different from the previous third training image 804, the quality loss information 830 is also newly determined, and the second DNN 300 updates the parameters accordingly. When the quality loss information 830 is newly determined, LossDS is also newly determined, and the first DNN 700 updates the parameters according to newly determined LossDS. In other words, updating of the parameters of the first DNN 700 leads to updating of the parameters of the second DNN 300, and updating of the parameters of the second DNN 300 leads to updating of the parameters of the first DNN 700. In other words, because the first DNN 700 and the second DNN 300 are jointly trained by sharing the quality loss information 830, the parameters of the first DNN 700 and the parameters of the second DNN 300 may be jointly optimized.

[192] Referring to Equation 1, it is verified that LossUS is determined according to the quality loss information 830, but this is only an example and LossUS may be determined based on at least one of the structural loss information 810 and the complexity loss information 820, and the quality loss information 830.

[193] Hereinabove, it has been described that the AI up-scaler 234 of the AI decoding apparatus 200 and the AI down-scaler 612 of the AI encoding apparatus 600 store the plurality of pieces of DNN setting information, and methods of training each of the plurality of pieces of DNN setting information stored in the AI up-scaler 234 and the AI down-scaler 612 will now be described.

[194] As described with reference to Equation 1, the first DNN 700 updates the parameters considering the similarity (the structural loss information 810) between the structural information of the first training image 802 and the structural information of the original training image 801, the bitrate (the complexity loss information 820) of the image data obtained as a result of performing first encoding on the first training image 802, and the difference (the quality loss information 830) between the third training image 804 and the original training image 801.

[195] In particular, the parameters of the first DNN 700 may be updated such that the first training image 802 having similar structural information as the original training image 801 is obtained and the image data having a small bitrate is obtained when first encoding is performed on the first training image 802, and at the same time, the second DNN 300 performing AI up-scaling on the first training image 802 obtains the third training image 804 similar to the original training image 801.

[196] A direction in which the parameters of the first DNN 700 are optimized may vary by adjusting the weights a, b, and c of Equation 1. For example, when the weight b is determined to be high, the parameters of the first DNN 700 may be updated by prioritizing a low bitrate over high quality of the third training image 804. Also, when the

weight c is determined to be high, the parameters of the first DNN 700 may be updated by prioritizing high quality of the third training image 804 over a high bitrate or maintaining of the structural information of the original training image 801.

[197] Also, the direction in which the parameters of the first DNN 700 are optimized may vary according to a type of codec used to perform first encoding on the first training image 802. This is because the second training image to be input to the second DNN 300 may vary according to the type of codec.

[198] In other words, the parameters of the first DNN 700 and the parameters of the second DNN 300 may be jointly updated based on the weights a , b , and c , and the type of codec for performing first encoding on the first training image 802. Accordingly, when the first DNN 700 and the second DNN 300 are trained after determining the weights a , b , and c each to a certain value and determining the type of codec to a certain type, the parameters of the first DNN 700 and the parameters of the second DNN 300 connected and optimized to each other may be determined.

[199] Also, when the first DNN 700 and the second DNN 300 are trained after changing the weights a , b , and c , and the type of codec, the parameters of the first DNN 700 and the parameters of the second DNN 300 connected and optimized to each other may be determined. In other words, the plurality of pieces of DNN setting information jointly trained with each other may be determined in the first DNN 700 and the second DNN 300 when the first DNN 700 and the second DNN 300 are trained while changing values of the weights a , b , and c , and the type of codec.

[200] As described above with reference to FIG. 5, the plurality of pieces of DNN setting information of the first DNN 700 and the second DNN 300 may be mapped to the information related to the first image. To set such a mapping relationship, first encoding may be performed on the first training image 802 output from the first DNN 700 via a certain codec according to a certain bitrate and the second training image obtained by performing first decoding on a bitstream obtained as a result of performing the first encoding may be input to the second DNN 300. In other words, by training the first DNN 700 and the second DNN 300 after setting an environment such that the first encoding is performed on the first training image 802 of a certain resolution via the certain codec according to the certain bitrate, a DNN setting information pair mapped to the resolution of the first training image 802, a type of the codec used to perform the first encoding on the first training image 802, and the bitrate of the bitstream obtained as a result of performing the first encoding on the first training image 802 may be determined. By variously changing the resolution of the first training image 802, the type of codec used to perform the first encoding on the first training image 802 and the bitrate of the bitstream obtained according to the first encoding of the first training image 802, the mapping relationships between the plurality of DNN setting in-

formation of the first DNN 700 and the second DNN 300 and the pieces of information related to the first image may be determined.

- [201] FIG. 10 is a diagram for describing training processes of the first DNN 700 and the second DNN by a training apparatus 1000.
- [202] The training of the first DNN 700 and the second DNN 300 described with reference FIG. 9 may be performed by the training apparatus 1000. The training apparatus 1000 includes the first DNN 700 and the second DNN 300. The training apparatus 1000 may be, for example, the AI encoding apparatus 600 or a separate server. The DNN setting information of the second DNN 300 obtained as the training result is stored in the AI decoding apparatus 200.
- [203] Referring to FIG. 10, the training apparatus 1000 initially sets the DNN setting information of the first DNN 700 and the second DNN 300, in operations S840 and S845. Accordingly, the first DNN 700 and the second DNN 300 may operate according to pre-determined DNN setting information. The DNN setting information may include information about at least one of the number of convolution layers included in the first DNN 700 and the second DNN 300, the number of filter kernels for each convolution layer, the size of a filter kernel for each convolution layer, or a parameter of each filter kernel.
- [204] The training apparatus 1000 inputs the original training image 801 into the first DNN 700, in operation S850. The original training image 801 may include a still image or at least one frame included in a moving image.
- [205] The first DNN 700 processes the original training image 801 according to the initially set DNN setting information and outputs the first training image 802 obtained by performing AI down-scaling on the original training image 801, in operation S855. In FIG. 10, the first training image 802 output from the first DNN 700 is directly input to the second DNN 300, but the first training image 802 output from the first DNN 700 may be input to the second DNN 300 by the training apparatus 1000. Also, the training apparatus 1000 may perform first encoding and first decoding on the first training image 802 via a certain codec, and then input the second training image to the second DNN 300.
- [206] The second DNN 300 processes the first training image 802 or the second training image according to the initially set DNN setting information and outputs the third training image 804 obtained by performing AI up-scaling on the first training image 802 or the second training image, in operation S860.
- [207] The training apparatus 1000 calculates the complexity loss information 820, based on the first training image 802, in operation S865.
- [208] The training apparatus 1000 calculates the structural loss information 810 by comparing the reduced training image 803 and the first training image 802, in

operation S870.

[209] The training apparatus 1000 calculates the quality loss information 830 by comparing the original training image 801 and the third training image 804, in operation S875.

[210] The initially set DNN setting information is updated in operation S880 via a back propagation process based on the final loss information. The training apparatus 1000 may calculate the final loss information for training the first DNN 700, based on the complexity loss information 820, the structural loss information 810, and the quality loss information 830.

[211] The second DNN 300 updates the initially set DNN setting information in operation S885 via a back propagation process based on the quality loss information 830 or the final loss information. The training apparatus 1000 may calculate the final loss information for training the second DNN 300, based on the quality loss information 830.

[212] Then, the training apparatus 1000, the first DNN 700, and the second DNN 300 may repeat operations S850 through S885 until the final loss information is minimized to update the DNN setting information. At this time, during each repetition, the first DNN 700 and the second DNN 300 operate according to the DNN setting information updated in the previous operation.

[213] Table 1 below shows effects when AI encoding and AI decoding are performed on the original image 105 according to an embodiment of the disclosure and when encoding and decoding are performed on the original image 105 via HEVC.

[214] [Table 1]

Content	Resolution	Frame Number	Information Amount (Bitrate) (Mbps)		Subjective Image Quality Score (VMAF)	
			HEVC	AI Encoding/ AI Decoding	HEVC	AI Encoding/ AI Decoding
Content_01	8K (7680×4320)	300 frames	46.3	21.4	94.80	93.54
Content_02			46.3	21.6	98.05	98.98
Content_03			46.3	22.7	96.08	96.00
Content_04			46.1	22.1	86.26	92.00
Content_05			45.4	22.7	93.42	92.98
Content_06			46.3	23.0	95.99	95.61
		Average	46.11	22.25	94.10	94.85

[215] As shown in Table 1, despite subjective image quality when AI encoding and AI decoding are performed on content including 300 frames of 8 K resolution, according to an embodiment of the disclosure, is higher than subjective image quality when encoding and decoding are performed via HEVC, a bitrate is reduced by at least 50%.

[216] FIG. 11 is a diagram of an apparatus 20 for performing AI down-scaling on the original image 105 and an apparatus 40 for performing AI up-scaling on the second

image 135.

- [217] The apparatus 20 receives the original image 105 and provides image data 25 and AI data 30 to the apparatus 40 by using an AI down-scaler 1124 and a transformation-based encoder 1126. According to an embodiment, the image data 25 corresponds to the image data of FIG. 1 and the AI data 30 corresponds to the AI data of FIG. 1. Also, according to an embodiment, the transformation-based encoder 1126 corresponds to the first encoder 614 of FIG. 7 and the AI down-scaler 1124 corresponds to the AI down-scaler 612 of FIG. 7.
- [218] The apparatus 40 receives the AI data 30 and the image data 25 and obtains the third image 145 by using a transformation-based decoder 1146 and an AI up-scaler 1144. According to an embodiment, the transformation-based decoder 1146 corresponds to the first decoder 232 of FIG. 2 and the AI up-scaler 1144 corresponds to the AI up-scaler 234 of FIG. 2.
- [219] According to an embodiment, the apparatus 20 includes a CPU, a memory, and a computer program including instructions. The computer program is stored in the memory. According to an embodiment, the apparatus 20 performs functions to be described with reference to FIG. 11 according to execution of the computer program by the CPU. According to an embodiment, the functions to be described with reference to FIG. 11 are performed by a dedicated hardware chip and/or the CPU.
- [220] According to an embodiment, the apparatus 40 includes a CPU, a memory, and a computer program including instructions. The computer program is stored in the memory. According to an embodiment, the apparatus 40 performs functions to be described with reference to FIG. 11 according to execution of the computer program by the CPU. According to an embodiment, the functions to be described with reference to FIG. 11 are performed by a dedicated hardware chip and/or the CPU.
- [221] In FIG. 11, a configuration controller 1122 receives at least one input value 10. According to an embodiment, the at least one input value 10 may include at least one of a target resolution difference for the AI down-scaler 1124 and the AI up-scaler 1144, a bitrate of the image data 25, a bitrate type of the image data 25 (for example, a variable bitrate type, a constant bitrate type, or an average bitrate type), or a codec type for the transformation-based encoder 1126. The at least one input value 10 may include a value pre-stored in the apparatus 20 or a value input from a user.
- [222] The configuration controller 1122 controls operations of the AI down-scaler 1124 and the transformation-based encoder 1126, based on the received input value 10. According to an embodiment, the configuration controller 1122 obtains DNN setting information for the AI down-scaler 1124 according to the received input value 10, and sets the AI down-scaler 1124 with the obtained DNN setting information. According to an embodiment, the configuration controller 1122 may transmit the received input

value 10 to the AI down-scaler 1124 and the AI down-scaler 1124 may obtain the DNN setting information for performing AI down-scaling on the original image 105, based on the received input value 10. According to an embodiment, the configuration controller 1122 may provide, to the AI down-scaler 1124, additional information, for example, color format (luminance component, chrominance component, red component, green component, or blue component) information to which AI down-scaling is applied and tone mapping information of a high dynamic range (HDR), together with the input value 10, and the AI down-scaler 1124 may obtain the DNN setting information considering the input value 10 and the additional information. According to an embodiment, the configuration controller 1122 transmits at least a part of the received input value 10 to the transformation-based encoder 1126 and the transformation-based encoder 1126 performs first encoding on the first image 115 via a bitrate of a certain value, a bitrate of a certain type, and a certain codec.

[223] The AI down-scaler 1124 receives the original image 105 and performs an operation described with reference to at least one of FIG. 1, 7, 8, 9, or 10 to obtain the first image 115.

[224] According to an embodiment, the AI data 30 is provided to the apparatus 40. The AI data 30 may include at least one of resolution difference information between the original image 105 and the first image 115, or information related to the first image 115. The resolution difference information may be determined based on the target resolution difference of the input value 10, and the information related to the first image 115 may be determined based on at least one of a target bitrate, the bitrate type, or the codec type. According to an embodiment, the AI data 30 may include parameters used during the AI up-scaling. The AI data 30 may be provided from the AI down-scaler 1124 to the apparatus 40.

[225] The image data 25 is obtained as the original image 105 is processed by the transformation-based encoder 1126, and is transmitted to the apparatus 40. The transformation-based encoder 1126 may process the first image 115 according to MPEG-2, H.264 AVC, MPEG-4, HEVC, VC-1, VP8, VP9, or VA1.

[226] A configuration controller 1142 controls an operation of the AI up-scaler 1144, based on the AI data 30. According to an embodiment, the configuration controller 1142 obtains the DNN setting information for the AI up-scaler 1144 according to the received AI data 30, and sets the AI up-scaler 1144 with the obtained DNN setting information. According to an embodiment, the configuration controller 1142 may transmit the received AI data 30 to the AI up-scaler 1144 and the AI up-scaler 1144 may obtain the DNN setting information for performing AI up-scaling on the second image 135, based on the AI data 30. According to an embodiment, the configuration controller 1142 may provide, to the AI up-scaler 1144, additional information, for

example, the color format (luminance component, chrominance component, red component, green component, or blue component) information to which AI up-scaling is applied, and the tone mapping information of HDR, together with the AI data 30, and the AI up-scaler 1144 may obtain the DNN setting information considering the AI data 30 and the additional information. According to an embodiment, the AI up-scaler 1144 may receive the AI data 30 from the configuration controller 1142, receive at least one of prediction mode information, motion information, or quantization parameter information from the transformation-based decoder 1146, and obtain the DNN setting information based on the AI data 30 and at least one of the prediction mode information, the motion information, and the quantization parameter information.

[227] The transformation-based decoder 1146 may process the image data 25 to reconstruct the second image 135. The transformation-based decoder 1146 may process the image data 25 according to MPEG-2, H.264 AVC, MPEG-4, HEVC, VC-1, VP8, VP9, or AV1.

[228] The AI up-scaler 1144 may obtain the third image 145 by performing AI up-scaling on the second image 135 provided from the transformation-based decoder 1146, based on the set DNN setting information.

[229] The AI down-scaler 1124 may include a first DNN and the AI up-scaler 1144 may include a second DNN, and according to an embodiment, DNN setting information for the first DNN and second DNN are trained according to the training method described with reference to FIGS. 9 and 10.

[230] FIG. 12 is a flowchart of an AI decoding method according to an embodiment.

[231] In operation S910, the AI decoding apparatus 200 receives AI encoding data including image data and AI data. The AI decoding apparatus 200 may receive the AI encoding data from the AI encoding apparatus 600 through a network. The AI decoding apparatus 200 may obtain the AI encoding data stored in a data storage medium.

[232] In operation S920, the AI decoding apparatus 200 obtains the second image 135 based on the image data. In particular, the AI decoding apparatus 200 reconstructs the second image 135 corresponding to the first image 115 by decoding the image data based on an image reconstruction method using frequency transform.

[233] In operation S930, the AI decoding apparatus 200 obtains DNN setting information for performing AI up-scaling on the second image 135, from among a pre-stored plurality of pieces of DNN setting information. Because each of the plurality of pieces of DNN setting information is jointly optimized with each of a plurality of pieces of DNN setting information used to perform AI down-scaling on the original image 105, the DNN setting information enabling AI up-scaling to be performed on the second

image 135 according to an up-scaling target matching a down-scaling target of the original image 105 needs to be selected.

[234] In operation S940, the AI decoding apparatus 200 obtains the third image 145 obtained by performing AI up-scaling on the second image 135, through a second DNN operating with the DNN setting information obtained in operation S930. The third image 145 may be output from the AI decoding apparatus 200 and displayed through a display device or may be displayed after being post-processed.

[235] When the DNN setting information is pre-set in the second DNN and the DNN setting information selected in operation S930 is different from the pre-set DNN setting information, the AI decoding apparatus 200 sets the second DNN to the selected DNN setting information.

[236] FIG. 13 is a flowchart of an AI encoding method according to an embodiment.

[237] In operation S1010, the AI encoding apparatus 600 obtains the first image 115 obtained by performing AI down-scaling the original image 105, through a first DNN.

[238] The AI encoding apparatus 600 may determine a down-scaling target based on a certain standard, and obtain DNN setting information corresponding to the down-scaling target from among a pre-stored plurality of pieces of DNN setting information. Also, the AI encoding apparatus 600 may perform AI down-scaling on the original image 105 through the first DNN operating according to the obtained DNN setting information.

[239] In operation S1020, the AI encoding apparatus 600 obtains image data by performing first encoding on the first image 115. In particular, the AI encoding apparatus 600 obtains the image data corresponding to the first image 115 by encoding the first image 115 based on an image compression method using frequency transform.

[240] In operation S1030, the AI encoding apparatus 600 transmits AI encoding data including the image data and AI data including information related to AI down-scaling. The AI data includes information for selecting DNN setting information of a second DNN for AI up-scaling of the second image 135. According to an embodiment, the AI encoding data may be stored in a data storage medium.

[241] As described above, because the first DNN and the second DNN are jointly trained, when the AI encoding apparatus 600 performs AI down-scaling on the original image 105 to a particular down-scaling target, the AI decoding apparatus 200 performs AI up-scaling on the second image 135 to an up-scaling target corresponding to the down-scaling target.

[242] Accordingly, the AI data includes information enabling the AI decoding apparatus 200 to perform AI up-scaling on the second image 135 to the up-scaling target corresponding to the down-scaling target of the original image 105. In particular, the AI data includes information used to obtain DNN setting information corresponding to the

up-scaling target.

[243] Upon receiving the AI data, the AI decoding apparatus 200 is able to infer or verify which DNN setting information is used by the AI encoding apparatus 600 to perform AI down-scaling on the original image 105, and accordingly, may obtain DNN setting information corresponding to the DNN setting information used to perform AI down-scaling, and perform AI up-scaling by using the obtained DNN setting information.

[244] Meanwhile, the embodiments of the disclosure described above may be written as computer-executable programs or instructions that may be stored in a medium.

[245] The medium may continuously store the computer-executable programs or instructions, or temporarily store the computer-executable programs or instructions for execution or downloading. Also, the medium may be any one of various recording media or storage media in which a single piece or plurality of pieces of hardware are combined, and the medium is not limited to a medium directly connected to a computer system, but may be distributed on a network. Examples of the medium include magnetic media, such as a hard disk, a floppy disk, and a magnetic tape, optical recording media, such as CD-ROM and DVD, magneto-optical media such as a optical disk, and ROM, RAM, and a flash memory, which are configured to store program instructions. Other examples of the medium include recording media and storage media managed by application stores distributing applications or by websites, servers, and the like supplying or distributing other various types of software.

[246] Meanwhile, a model related to the DNN described above may be implemented via a software module. When the DNN model is implemented via a software module (for example, a program module including instructions), the DNN model may be stored in a computer-readable recording medium.

[247] Also, the DNN model may be a part of the AI decoding apparatus 200 or AI encoding apparatus 600 described above by being integrated in a form of a hardware chip. For example, the DNN model may be manufactured in a form of an dedicated hardware chip for AI, or may be manufactured as a part of an existing general-purpose processor (for example, CPU or application processor) or a graphic-dedicated processor (for example GPU).

[248] Also, the DNN model may be provided in a form of downloadable software. A computer program product may include a product (for example, a downloadable application) in a form of a software program electronically distributed through a manufacturer or an electronic market. For electronic distribution, at least a part of the software program may be stored in a storage medium or may be temporarily generated. In this case, the storage medium may be a server of the manufacturer or electronic market, or a storage medium of a relay server.

[249] While one or more embodiments of the disclosure have been described with

reference to the figures, it will be understood by those of ordinary skill in the art that various changes in form and details may be made therein without departing from the spirit and scope as defined by the following claims.

Claims

- [Claim 1] An artificial intelligence (AI) decoding apparatus comprising:
a memory storing one or more instructions; and
a processor configured to execute the one or more instructions stored in the memory,
wherein the processor is configured to:
obtain AI data related to AI down-scaling an original image to a first image, the AI data comprising at least one of information related to the first image and information about a difference between the original image and the first image;
obtain image data corresponding to an encoding result on the first image;
obtain a second image corresponding to the first image by performing a decoding on the image data;
obtain deep neural network (DNN) setting information among a plurality of DNN setting information from the AI data, the DNN setting information being for performing AI up-scaling on the second image;
and
obtain, by an up-scaling DNN, a third image by performing the AI up-scaling on the second image, the up-scaling DNN being configured with the obtained DNN setting information,
wherein the plurality of DNN setting information comprises a parameter used in the up-scaling DNN, the parameter being obtained through joint training of the up-scaling DNN and a down-scaling DNN,
and
wherein the down-scaling DNN is used to obtain the first image from the original image.
- [Claim 2] The AI decoding apparatus of claim 1, wherein the AI data comprises the information about the difference between the original image and the first image, and the processor is further configured to obtain the DNN setting information for performing AI up-scaling on the second image to match the third image with the difference between the original image and the first image.
- [Claim 3] The AI decoding apparatus of claim 1, wherein the AI data comprises the information related to the first image, and the processor is further configured to obtain the DNN setting information mapped to the information related to the first image, based on a mapping relationship

- between several image-related information and the plurality of DNN setting information,
wherein the information related to the first image comprises at least one of a resolution, a bitrate or a codec type.
- [Claim 4] The AI decoding apparatus of claim 3, wherein the image data comprises quantization parameter information used in the decoding, and
wherein the processor is further configured to obtain, based on the quantization parameter information and the information related to the first image, the DNN setting information.
- [Claim 5] The AI decoding apparatus of claim 1, wherein the obtained DNN setting information comprises parameters of a filter kernel,
wherein the filter kernel is associated with at least one convolution layer, and
wherein the up-scaling DNN comprises the at least one convolution layer.
- [Claim 6] The AI decoding apparatus of claim 1, wherein the processor is further configured to set the up-scaling DNN with the obtained DNN setting information instead of DNN setting information set in the up-scaling DNN, when the DNN setting information set in the up-scaling DNN is different from the obtained DNN setting information.
- [Claim 7] The AI decoding apparatus of claim 1, wherein the up-scaling DNN is trained based on quality loss information,
wherein the quality loss information corresponds to a comparison of a training image output from the up-scaling DNN and an original training image before AI down-scaling is performed.
- [Claim 8] The AI decoding apparatus of claim 7, wherein the quality loss information is used in training of the down-scaling DNN.
- [Claim 9] The AI decoding apparatus of claim 8, wherein when parameters of a DNN of any one of the up-scaling DNN and the down-scaling DNN are updated during a training process, parameters of a DNN of the other one are updated.
- [Claim 10] A system comprising:
an AI encoding apparatus comprising a down-scaling DNN, the down-scaling DNN being configured to be trained based on:
structural loss information corresponding to a comparison of a first training image output from the down-scaling DNN and a reduced training image,

complexity loss information corresponding to a spatial complexity of the first training image, and
quality loss information corresponding to a comparison of the original training image and a third training image output from the up-scaling DNN; and
an AI decoding apparatus comprising the up-scaling DNN, the up-scaling DNN being configured to be trained based on the quality loss information.

[Claim 11]

An artificial intelligence (AI) encoding apparatus comprising:
a memory storing one or more instructions; and
a processor configured to execute the one or more instructions stored in the memory to:
obtain, by a down-scaling deep neural network (DNN) performing AI down-scaling on an original image, a first image, the down-scaling DNN being configured with DNN setting information;
encode the first image to obtain image data; and
transmit the image data and AI data for selecting DNN setting information of an up-scaling DNN, the AI data comprising at least one of information related to the first image and information about a difference between the original image and the first image,
wherein the up-scaling DNN is configured to perform AI up-scaling on a second image,
wherein the second image is obtained by decoding the image data, and
wherein the DNN setting information of the down-scaling DNN and the DNN setting information of the up-scaling DNN comprises a parameter used in the down-scaling DNN and the up-scaling DNN, the parameter being obtained through joint training of the up-scaling DNN and a down-scaling DNN.

[Claim 12]

The AI encoding apparatus of claim 11, wherein the down-scaling DNN and the up-scaling DNN are trained based on quality loss information,
wherein the quality loss information corresponds to a comparison of a training image output from the up-scaling DNN and an original training image before AI down-scaling is performed.

[Claim 13]

An artificial intelligence (AI) decoding method of an image, the AI decoding method comprising:
obtaining AI data related to AI down-scaling an original image to a first image, the AI data comprising at least one of information related to the

first image and information about a difference between the original image and the first image;
obtaining image data corresponding to an encoding result on the first image;
obtaining a second image corresponding to the first image by performing a decoding on the image data;
obtaining deep neural network (DNN) setting information among a plurality of DNN setting information from the AI data, the DNN setting information being for performing AI up-scaling on the second image;
and
obtaining, by an up-scaling DNN, a third image by performing the AI up-scaling on the second image, the up-scaling DNN being configured with the obtained DNN setting information,
wherein the plurality of DNN setting information comprises a parameter used in the up-scaling DNN, the parameter being obtained through joint training of the up-scaling DNN and a down-scaling DNN,
and
wherein the down-scaling DNN is used to obtain the first image from the original image.

[Claim 14]

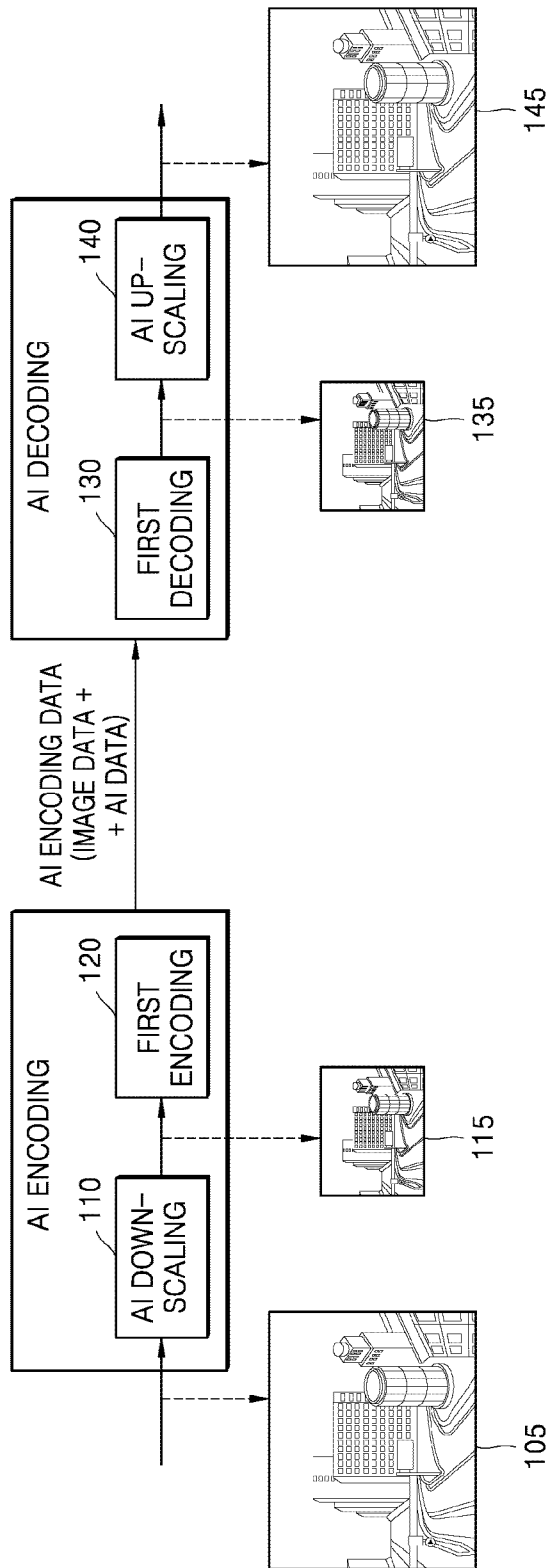
An artificial intelligence (AI) encoding method of an image, the AI encoding method comprising:
obtaining, by a down-scaling deep neural network (DNN) performing AI down-scaling on an original image, a first image, the down-scaling DNN being configured with DNN setting information;
encoding the first image to obtain image data; and
transmitting the image data and AI data for selecting DNN setting information of an up-scaling DNN, the AI data comprising at least one of information related to the first image and information about a difference between the original image and the first image,
wherein the up-scaling DNN is configured to perform AI up-scaling on a second image,
wherein the second image is obtained by decoding the image data, and
wherein the DNN setting information of the down-scaling DNN and the DNN setting information of the up-scaling DNN comprises a parameter used in the down-scaling DNN and the up-scaling DNN, the parameter being obtained through joint training of the up-scaling DNN and a down-scaling DNN.

[Claim 15]

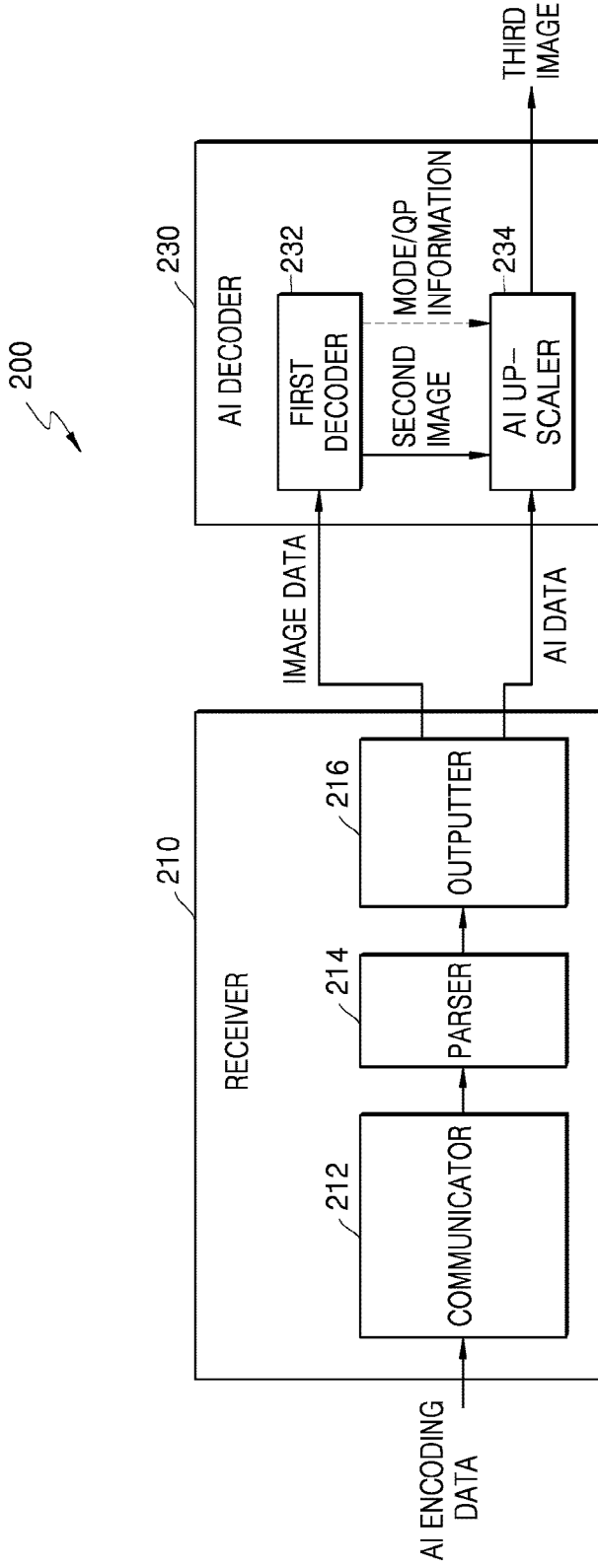
A non-transitory computer-readable recording medium having recorded

thereon a program which, when executed by a computer, performs the artificial intelligence (AI) decoding method of claim 13.

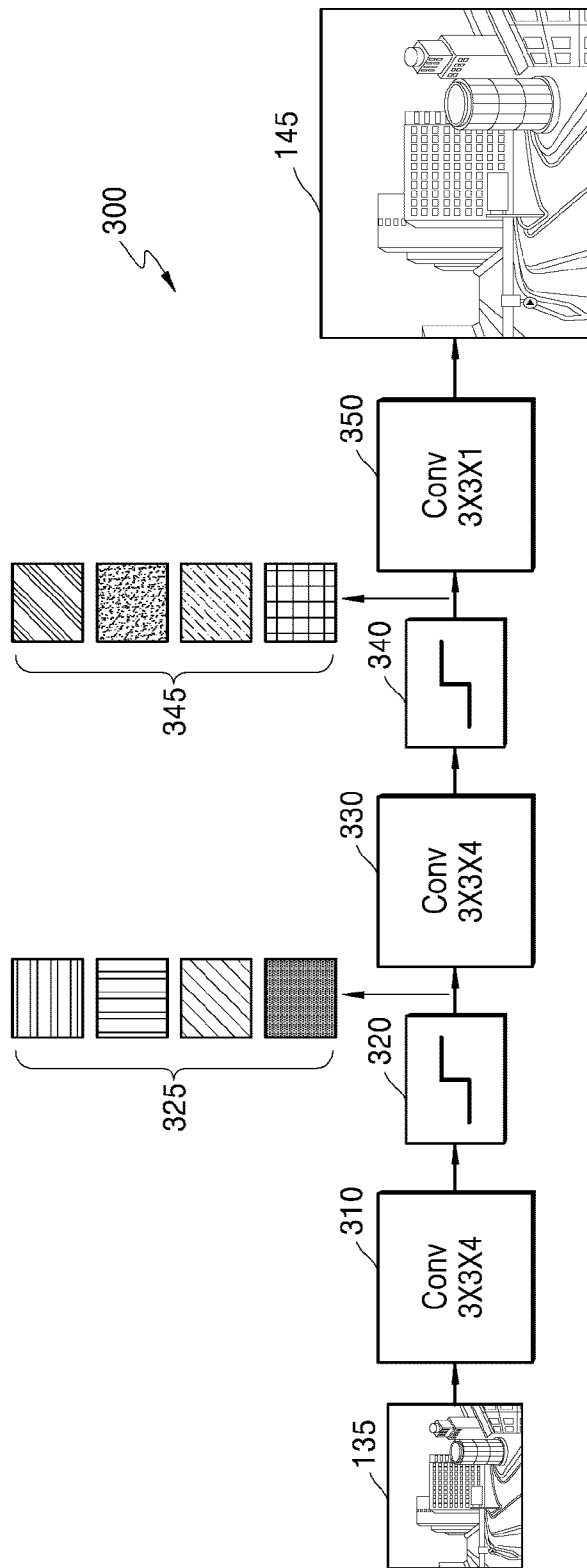
[Fig. 1]



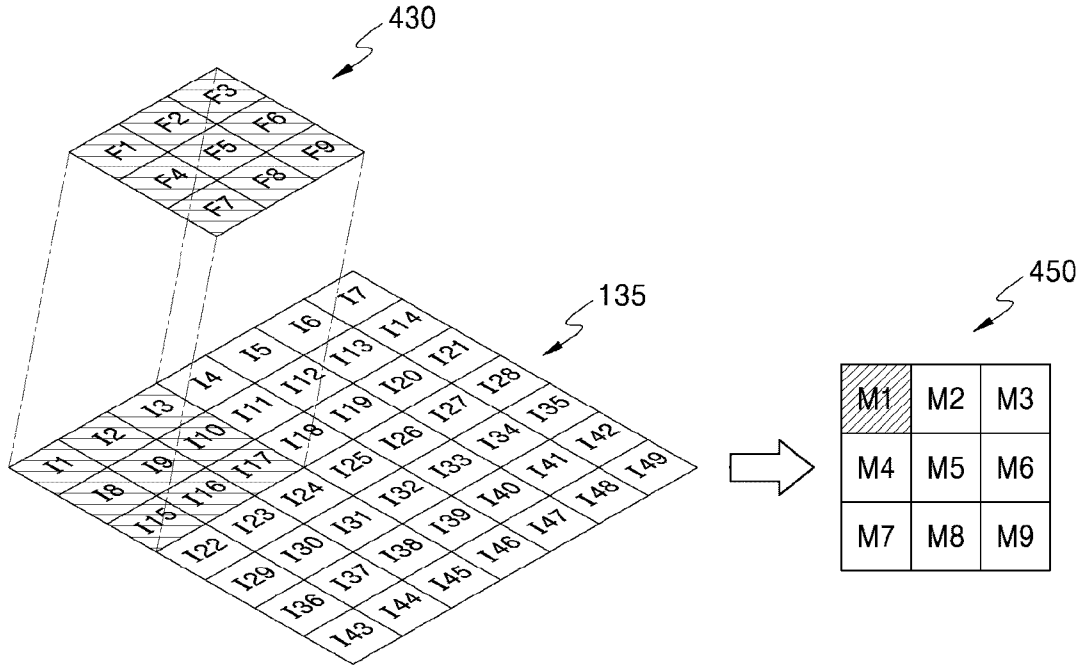
[Fig. 2]



[Fig. 3]



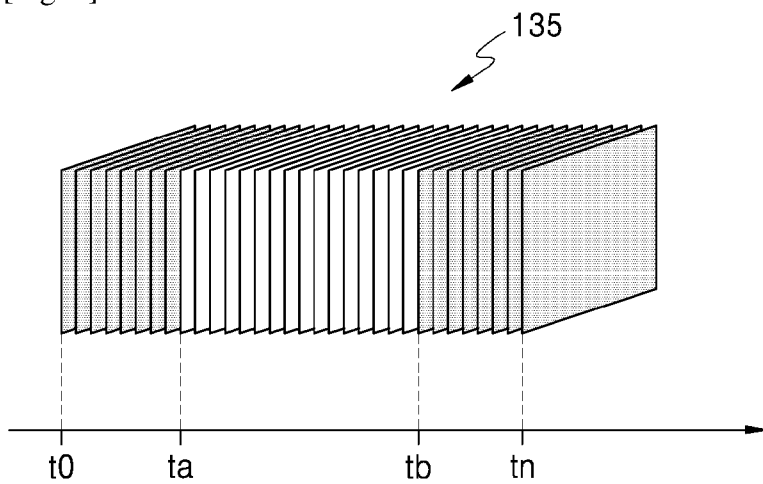
[Fig. 4]



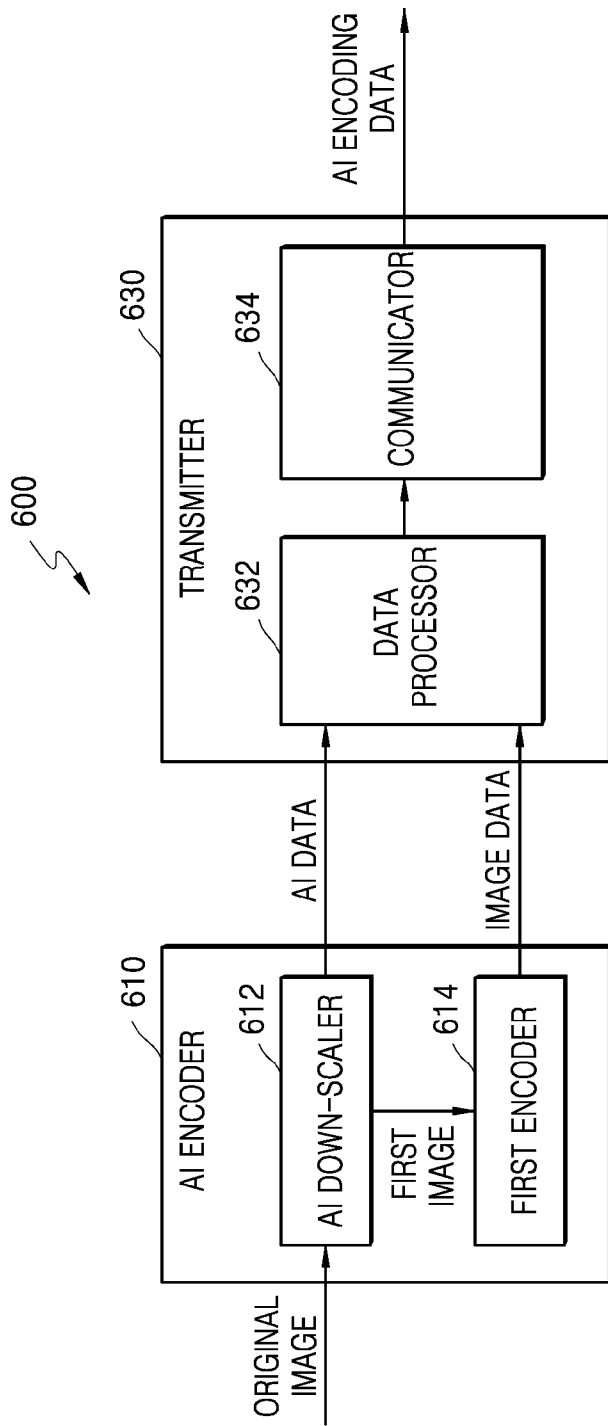
[Fig. 5]

IMAGE-RELATED INFORMATION	DNN SETTING INFORMATION
SD, 10Mbps, AV1	A DNN SETTING INFORMATION
HD, 15Mbps, H.264	B DNN SETTING INFORMATION
Full HD, 20Mbps, HEVC	C DNN SETTING INFORMATION
Full HD, 15Mbps, HEVC	D DNN SETTING INFORMATION

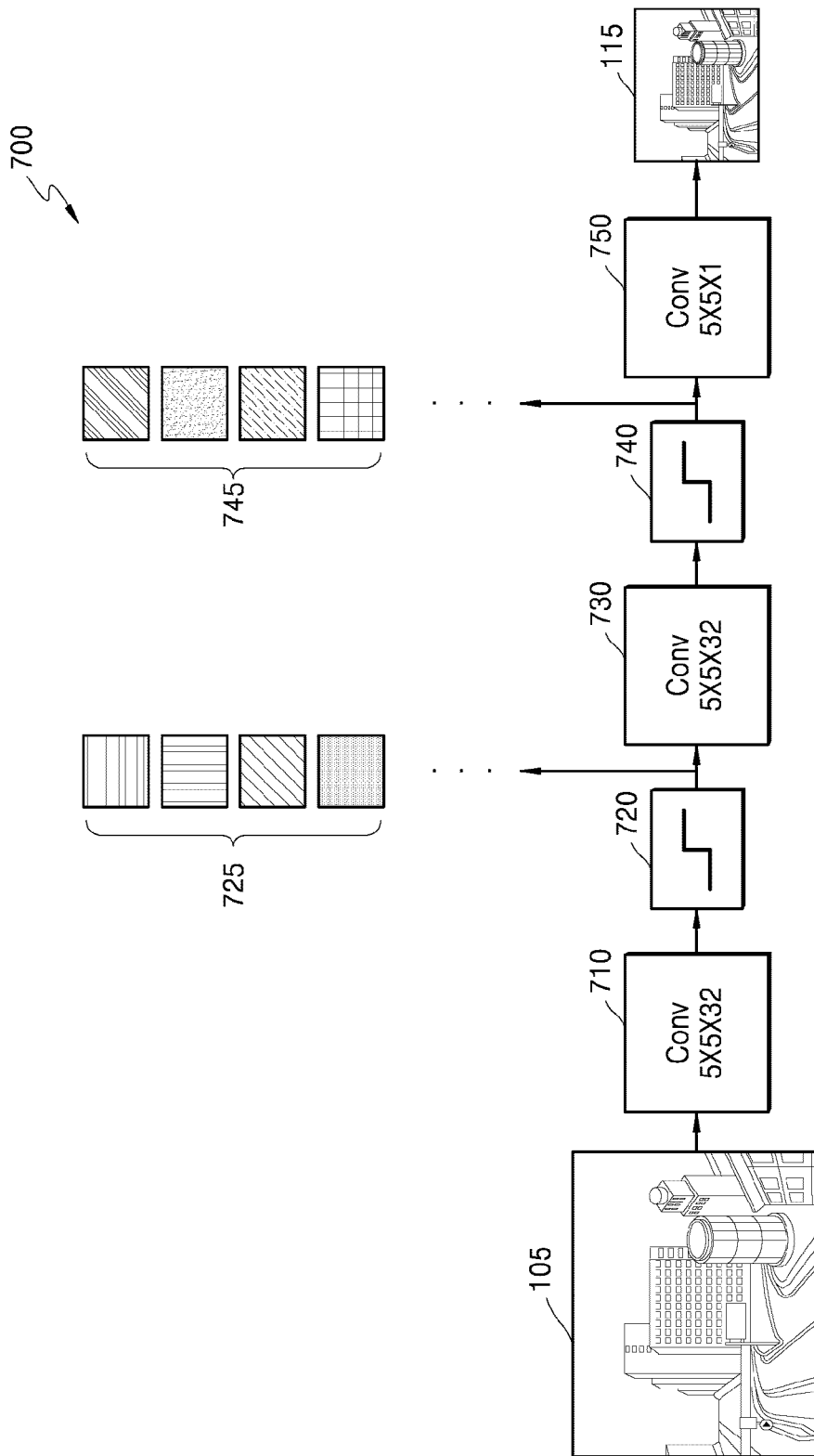
[Fig. 6]



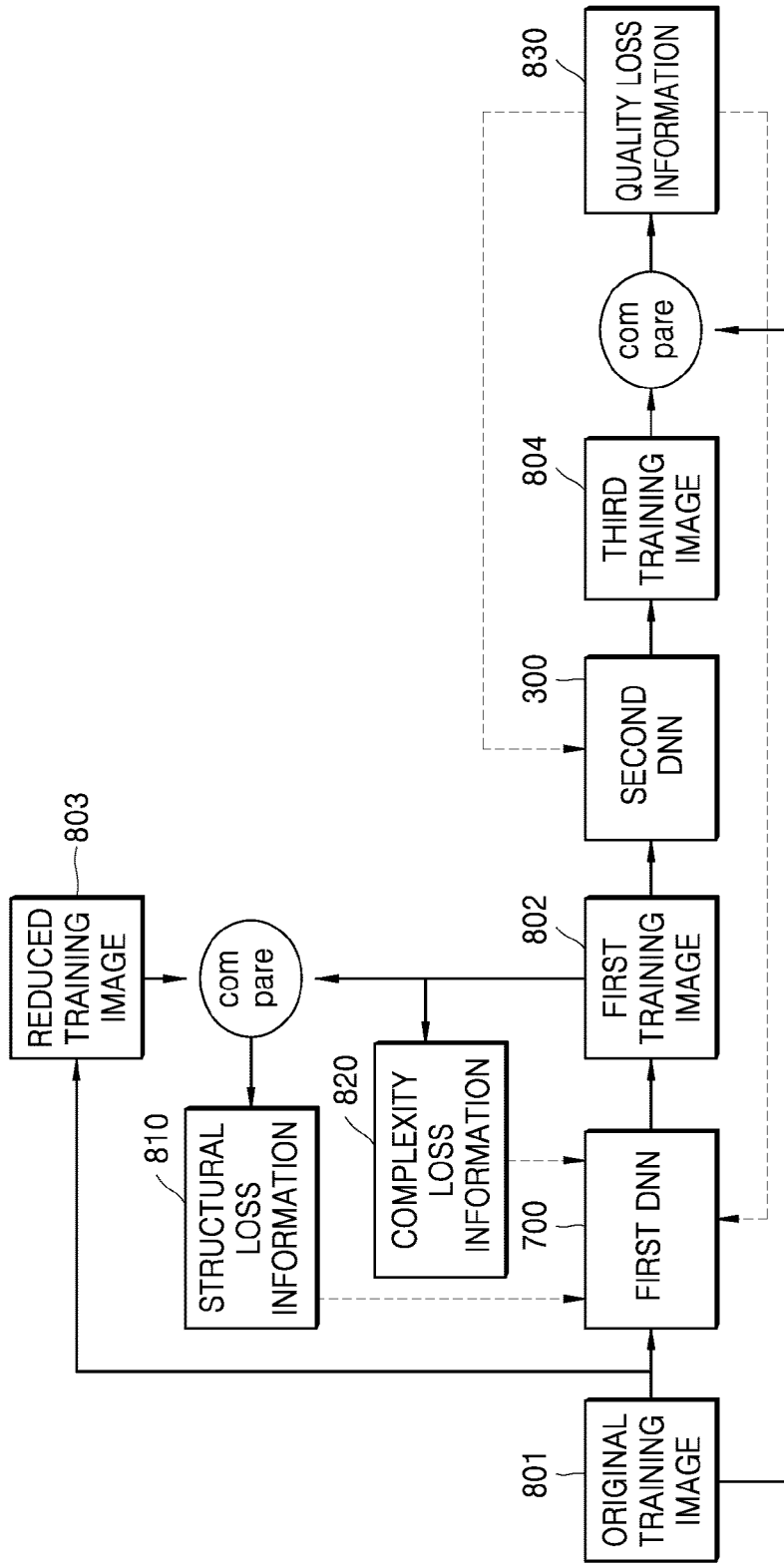
[Fig. 7]



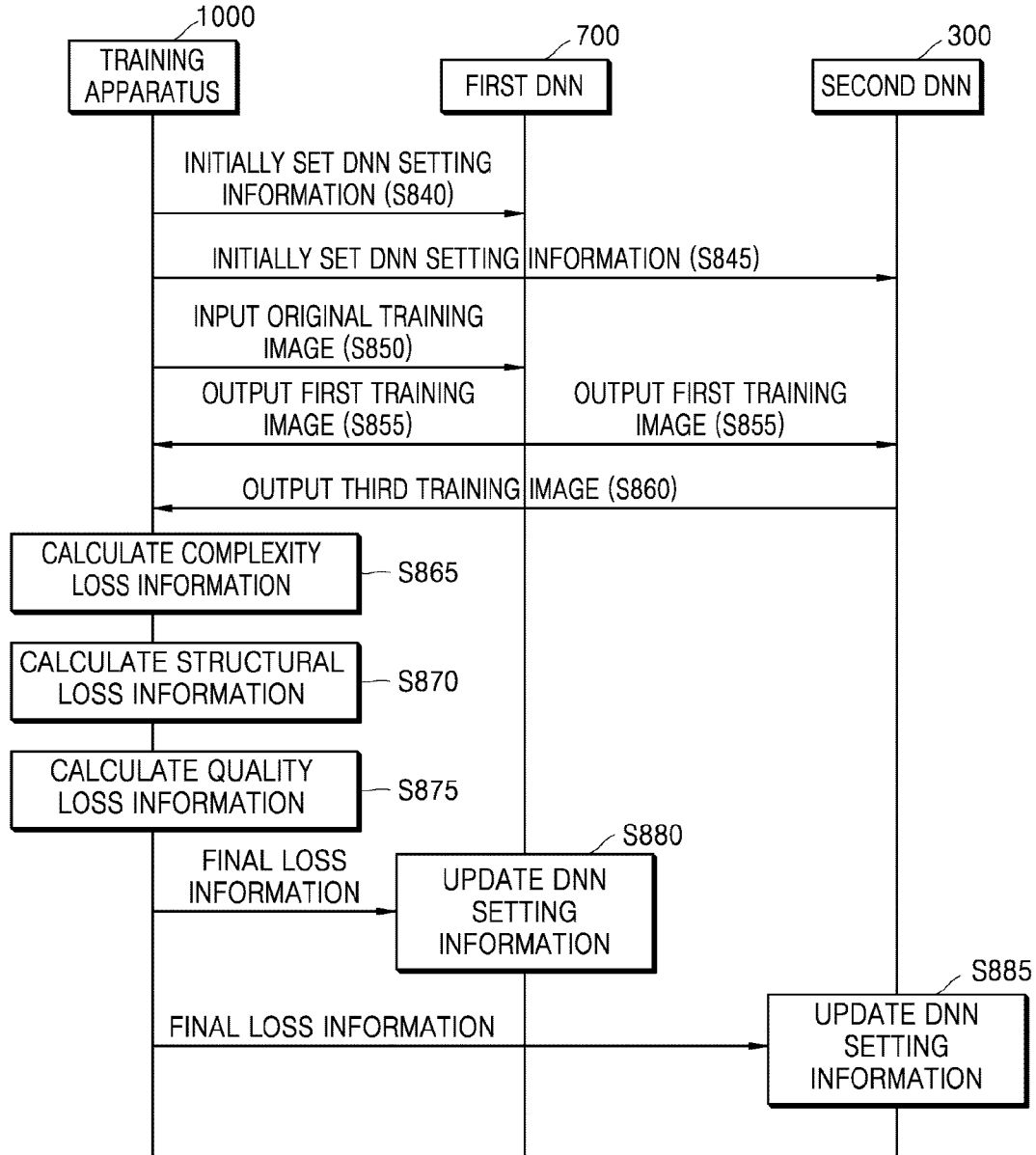
[Fig. 8]



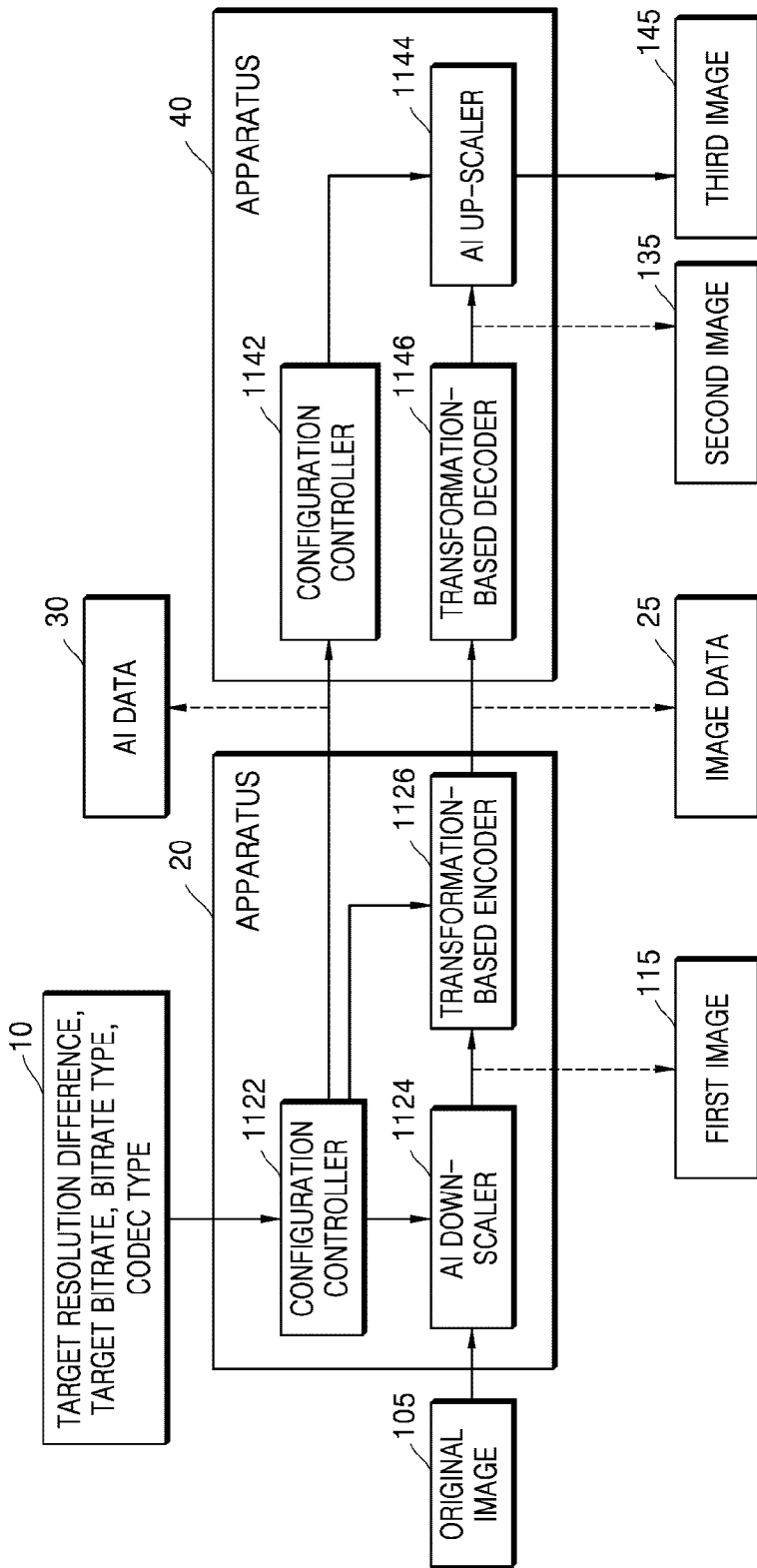
[Fig. 9]



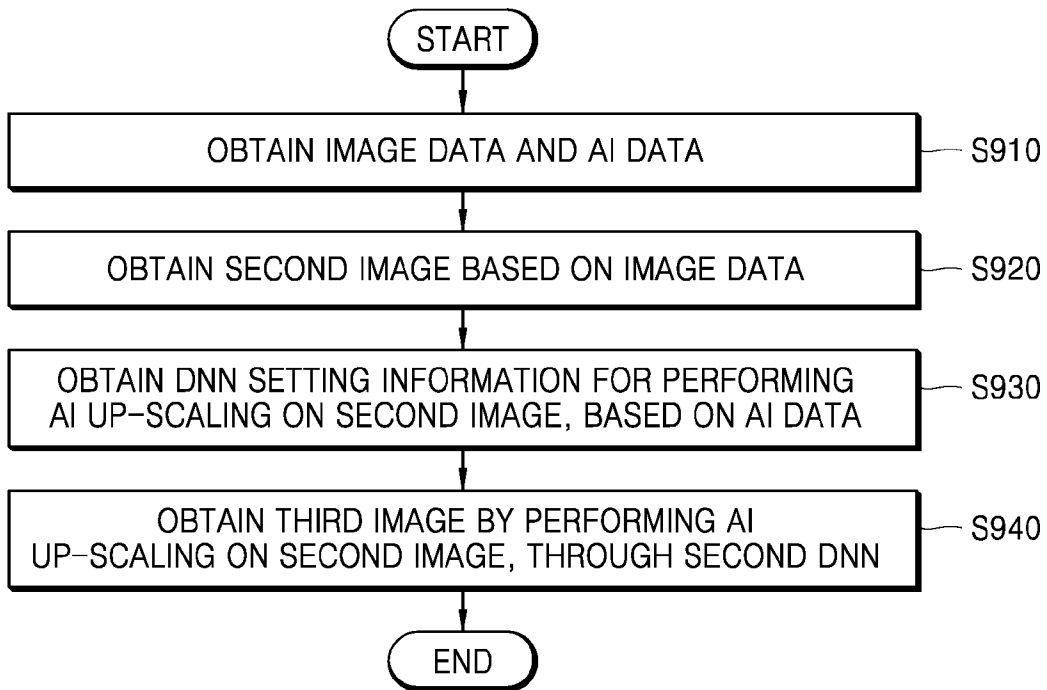
[Fig. 10]



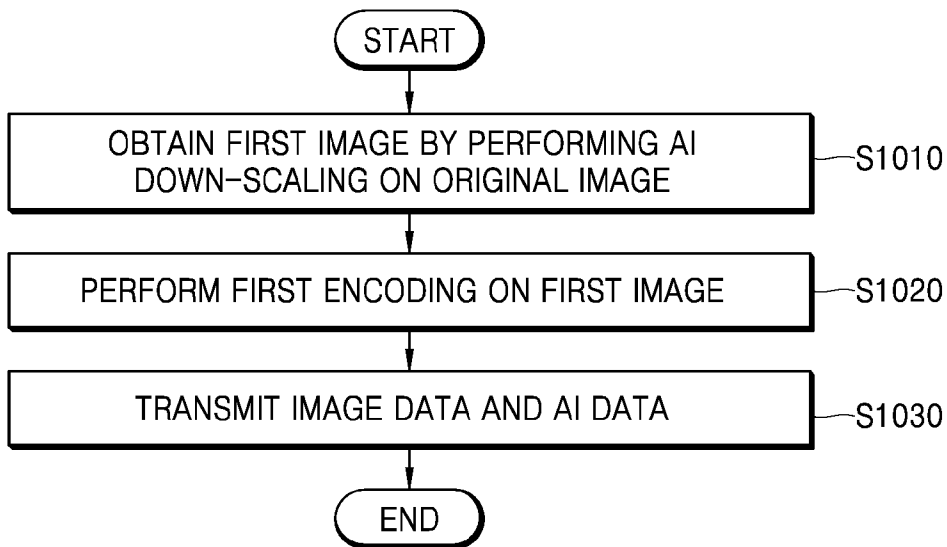
[Fig. 11]



[Fig. 12]



[Fig. 13]



A. CLASSIFICATION OF SUBJECT MATTER**H04N 19/132(2014.01)i, H04N 19/85(2014.01)i, H04N 19/50(2014.01)i, H04N 19/184(2014.01)i, G06N 3/08(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N 19/132; H04N 19/117; H04N 19/124; H04N 19/147; H04N 19/176; H04N 19/50; H04N 19/85; H04N 21/2343; H04N 19/184; G06N 3/08

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords: artificial intelligence (AI), upscale, downscale, train, encode, decode

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KR 10-2018-0100976 A (ELECTRONICS & TELECOMMUNICATIONS RES INST) 12 September 2018 See paragraphs [0031]-[0038], [0049]-[0073]; and figures 1-2, 4-9.	1-15
Y	KR 10-2018-0001428 A (KOREA ADVANCED INST SCI & TECH) 04 January 2018 See paragraphs [0151]-[0157]; claims 1-3; and figures 8a-8b.	1-15
A	KR 10-1885855 B1 (UNIV DANKOOK IACF) 07 August 2018 See paragraphs [0023]-[0034]; and figures 2-4.	1-15
A	KR 10-2017-0100045 A (VID SCALE INC.) 01 September 2017 See paragraphs [0012]-[0016]; and figures 3-4.	1-15
A	US 2018-0249158 A1 (MEDIATEK INC.) 30 August 2018 See paragraphs [0040]-[0041]; and figures 5-6.	1-15

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

04 December 2019 (04.12.2019)

Date of mailing of the international search report

04 December 2019 (04.12.2019)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

KIM, Sung Hoon

Telephone No. +82-42-481-8710



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/KR2019/010645

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
KR 10-2018-0100976 A	12/09/2018	None	
KR 10-2018-0001428 A	04/01/2018	CN 109644268 A JP 2019-525544 A KR 10-1974261 B1 US 2019-0230354 A1 WO 2017-222140 A1	16/04/2019 05/09/2019 30/04/2019 25/07/2019 28/12/2017
KR 10-1885855 B1	07/08/2018	None	
KR 10-2017-0100045 A	01/09/2017	CN 104685877 B EP 2901683 A1 JP 2015-534781 A JP 2017-153145 A JP 6141988 B2 KR 10-1774675 B1 KR 10-2008078 B1 TW 201436531 A US 10218971 B2 US 2015-0256828 A1 WO 2014-052740 A1	29/03/2019 05/08/2015 03/12/2015 31/08/2017 07/06/2017 04/09/2017 06/08/2019 16/09/2014 26/02/2019 10/09/2015 03/04/2014
US 2018-0249158 A1	30/08/2018	CA 2997193 A1 CN 107925762 A EP 3342164 A1 KR 10-2018-0052651 A PH 12018500454 A1 WO 2017-036370 A1	09/03/2017 17/04/2018 04/07/2018 18/05/2018 10/09/2018 09/03/2017