



(12) 发明专利

(10) 授权公告号 CN 114631320 B

(45) 授权公告日 2024. 12. 31

(21) 申请号 202080075785.2

(22) 申请日 2020.10.28

(65) 同一申请的已公布的文献号
申请公布号 CN 114631320 A

(43) 申请公布日 2022.06.14

(30) 优先权数据
10-2019-0134514 2019.10.28 KR

(85) PCT国际申请进入国家阶段日
2022.04.28

(86) PCT国际申请的申请数据
PCT/KR2020/014830 2020.10.28

(87) PCT国际申请的公布数据
W02021/086016 EN 2021.05.06

(73) 专利权人 三星电子株式会社
地址 韩国京畿道水原市

(72) 发明人 李钟硕 金载丸 朴永五

(74) 专利代理机构 北京铭硕知识产权代理有限公司 11286
专利代理师 王皎彤 朱志玲

(51) Int.Cl.
H04N 19/70 (2014.01)
H04N 19/124 (2014.01)
H04N 19/59 (2014.01)
H04N 19/91 (2014.01)
H04N 19/93 (2014.01)
G06T 1/20 (2006.01)
G06T 3/4046 (2024.01)
G06T 9/00 (2006.01)
G06N 3/0464 (2023.01)
G06N 3/08 (2023.01)

(56) 对比文件
CN 112840650 A, 2021.05.25

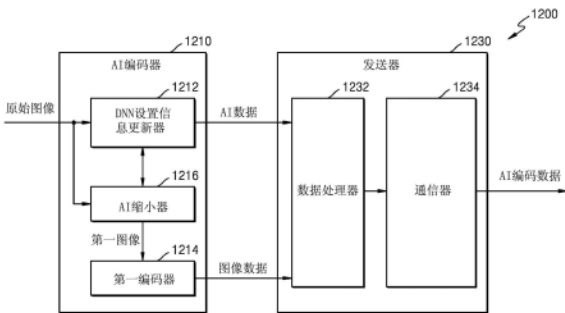
审查员 张楹
权利要求书2页 说明书42页 附图27页

(54) 发明名称

对图像执行人工智能AI编码和AI解码的设备和方法

(57) 摘要

一种人工智能(AI)解码方法,包括:获得通过对第一图像执行第一编码而生成的图像数据以及与至少一个原始图像的AI缩小相关的AI数据,其中,所述至少一个原始图像与第一图像相关;通过对所述图像数据执行第一解码来获得与第一图像相应的第二图像;基于所述AI数据获得用于执行第二图像的AI放大的深度神经网络(DNN)设置信息;以及通过经由根据获得的DNN设置信息操作的放大DNN对第二图像执行AI放大来生成第三图像。所述DNN设置信息是经由放大DNN和用于AI缩小的缩小DNN的联合训练而被更新以用于执行至少一个第二图像的AI放大的DNN信息。



1. 一种人工智能AI解码设备,包括:
存储器,存储一个或更多个指令;以及
处理器,被配置为执行存储在存储器中的所述一个或更多个指令以执行以下操作:
获得通过对第一图像进行第一编码而生成的图像数据;
获得与通过缩小深度神经网络DNN从原始图像到第一图像的AI缩小相关的AI数据;
通过对所述图像数据执行第一解码来获得与第一图像相应的第二图像;
基于所述AI数据更新用于AI放大的预定DNN设置信息;以及
通过经由根据所述更新的DNN设置信息操作的放大DNN对第二图像执行AI放大来生成第三图像,
其中,所述预定DNN设置信息是经由放大DNN和缩小DNN的联合训练而获得的,并且
其中,所述更新的DNN设置信息是经由根据所述预定DNN设置信息操作的放大DNN的单独训练而获得的。
2. 根据权利要求1所述的AI解码设备,其中,所述AI数据包括用于更新所述预定DNN设置信息的滤波器核的权重和偏置。
3. 根据权利要求1所述的AI解码设备,其中,所述预定DNN设置信息是通过以下操作获得的:
通过使用所述原始图像经由缩小DNN生成第一训练图像,
通过使用第一训练图像经由放大DNN生成第二训练图像,以及
基于第一损失信息、第二损失信息和第三损失信息来更新放大DNN和缩小DNN,
其中,第一损失信息和第三损失信息与将第二训练图像与尚未经过AI缩小的所述原始图像进行比较的结果相应,第二损失信息是基于第一训练图像生成的。
4. 根据权利要求3所述的AI解码设备,其中,第一损失信息是基于将第二训练图像的质量参数与所述原始图像的质量参数进行比较的结果而生成的。
5. 根据权利要求3所述的AI解码设备,其中,第三损失信息是基于将第二训练图像的特征相关参数与所述原始图像的特征相关参数进行比较的结果而生成的。
6. 根据权利要求3所述的AI解码设备,其中,第二损失信息与第一训练图像的空间复杂度有关。
7. 根据权利要求1所述的AI解码设备,其中,所述更新的DNN设置信息是通过以下操作获得的:
通过使用所述原始图像经由缩小DNN生成第一训练图像,
对第一训练图像执行第一编码,
通过使用已经经历第一编码的第一训练图像,经由放大DNN生成第二训练图像,以及
基于第一损失信息和第三损失信息更新放大DNN,
其中,第一损失信息和第三损失信息与将第二训练图像与尚未经过AI缩小的所述原始图像进行比较的结果相应。
8. 根据权利要求1所述的AI解码设备,其中,所述AI数据包括权重残差信息和偏置残差信息,其中,权重残差信息和偏置残差信息指示所述预定DNN设置信息的权重和偏置与所述更新的DNN设置信息的权重和偏置之间的差。
9. 根据权利要求8所述的AI解码设备,其中,权重残差信息和偏置残差信息是使用差分

脉冲编码调制DPCM、行程编码RLC和霍夫曼编码方案中的一个编码的信息。

10. 根据权利要求8所述的AI解码设备,其中,权重残差信息和偏置残差信息是关于经由模型压缩生成的权重残差和偏置残差的信息。

11. 根据权利要求8所述的AI解码设备,其中,权重残差信息和偏置残差信息是通过对权重残差和偏置残差执行频率变换而获得的。

12. 根据权利要求1所述的AI解码设备,其中,所述AI数据包括通过基于关于所述预定DNN设置信息的权重和偏置的上下文模型信息,对所述更新的DNN设置信息的权重和偏置进行熵编码而获得的信息。

13. 根据权利要求1所述的AI解码设备,其中,所述AI数据包括标志信息,所述标志信息指示是通过使用所述预定DNN设置信息来执行AI放大,还是通过使用所述更新的DNN设置信息来执行AI放大。

14. 一种用于执行图像的人工智能AI编码的设备,所述设备包括:
存储器,存储一个或更多个指令;以及
处理器,被配置为执行存储在所述存储器中的所述一个或更多个指令以执行以下操作:

通过经由缩小深度神经网络DNN对原始图像执行AI缩小来获得第一图像;
通过对第一图像执行第一编码来生成图像数据;以及
生成包括所述图像数据和AI数据的AI编码数据,其中,所述AI数据包括与AI缩小有关的信息,

其中,所述AI数据包括用于更新存储在AI解码设备中的预定DNN设置信息的信息,
其中,所述预定DNN设置信息是经由放大DNN和缩小DNN的联合训练而获得的,并且
其中,所述更新的DNN设置信息是经由根据所述预定DNN设置信息操作的放大DNN的单独训练而获得的。

15. 一种对图像执行人工智能AI解码的方法,所述方法包括:
获得通过对第一图像进行第一编码而生成的图像数据;
获得与通过缩小深度神经网络DNN从原始图像到第一图像的AI缩小相关的AI数据;
通过对所述图像数据执行第一解码来获得与第一图像相应的第二图像;
基于所述AI数据更新用于AI放大的预定DNN设置信息;以及
通过经由根据所述更新的DNN设置信息操作的放大DNN对第二图像执行AI放大来生成第三图像,

其中,所述预定DNN设置信息是经由放大DNN和缩小DNN的联合训练而获得的,并且
其中,所述更新的DNN设置信息是经由根据所述预定DNN设置信息操作的放大DNN的单独训练而获得的。

对图像执行人工智能AI编码和AI解码的设备和方法

技术领域

[0001] 本公开涉及图像处理领域,并且更具体地,涉及用于基于人工智能(AI)对图像进行编码和解码的方法和设备。

背景技术

[0002] 使用符合预定义数据压缩标准(诸如运动图像专家组(MPEG)标准)的编解码器对图像进行编码,然后将编码后的图像存储在记录介质中或通过通信信道以比特流的形式发送。

[0003] 随着能够再现和存储高分辨率/高清晰度图像的硬件的开发和传播,对能够有效地编码和解码高分辨率/高清晰度图像的编解码器的需求日益增加。

发明内容

[0004] 技术方案

[0005] 根据本公开的实施例,提供了用于对图像执行人工智能(AI)编码和AI解码的方法和设备,由此可以通过基于AI对图像进行编码和解码来实现低比特率。

[0006] 此外,根据本公开的实施例,提供了一种用于对图像执行AI编码和AI解码的方法和设备,由此可通过在周期性地或在必要时更新针对原始图像优化的放大DNN设置信息之后执行放大来改善图像质量。

[0007] 此外,根据本公开的实施例,提供了一种用于对图像执行AI编码和AI解码的方法和设备,由此可通过周期性地或在必要时有效地用信号通知DNN设置信息以用于更新放大DNN的针对原始图像优化的DNN设置信息来有效地减少要编码/解码的信息量。

[0008] 另外的方面将部分地在下面的描述中阐述,并且部分地将从描述中显而易见,或者可以通过实践本公开的所呈现的实施例来学习。

[0009] 根据本公开的实施例,一种AI解码设备包括:存储器,存储一个或更多个指令;以及处理器,被配置为执行存储在存储器中的所述一个或更多个指令以执行以下操作:获得通过对第一图像执行第一编码而生成的图像数据以及与至少一个原始图像的AI缩小相关的AI数据,其中,所述至少一个原始图像与第一图像相关;通过对所述图像数据执行第一解码来获得与第一图像相应的第二图像;基于所述AI数据获得用于执行第二图像的AI放大的DNN设置信息;以及通过经由根据获得的DNN设置信息操作的放大DNN对第二图像执行AI放大来生成第三图像,其中,DNN设置信息是经由放大DNN和用于所述至少一个原始图像的AI缩小的缩小DNN的联合训练而被更新以用于执行与所述至少一个原始图像相应的至少一个第二图像的AI放大的DNN信息,所述联合训练是使用至少一个原始图像执行的。

[0010] 所述获得的DNN设置信息可包括放大DNN的至少一个卷积层中的滤波器核的权重和偏置。

[0011] 所述处理器还可被配置为:通过使用所述至少一个原始图像经由缩小DNN生成第一训练图像,通过使用第一训练图像经由放大DNN生成第二训练图像,并且基于第一损失信

息和第三损失信息、以及第二损失信息更新放大DNN和缩小DNN,其中,第一损失信息和第三损失信息与将第二训练图像与至少一个原始图像中尚未经过AI缩小的原始图像进行比较的结果相应,第二损失信息是基于第一训练图像生成的。

[0012] 可基于将第二训练图像的质量参数与所述至少一个原始图像的质量参数进行比较的结果来生成第一损失信息。

[0013] 可基于将第二训练图像的特征相关参数与所述至少一个原始图像的特征相关参数进行比较的结果来生成第三损失信息。

[0014] 第二损失信息可与第一训练图像的空间复杂度有关。

[0015] 所述处理器还可被配置为:通过使用所述至少一个原始图像经由缩小DNN生成第一训练图像,对第一训练图像执行第一编码,通过使用已经经历第一编码的第一训练图像经由放大DNN生成第二训练图像,以及基于第一损失信息和第三损失信息更新放大DNN,其中,第一损失信息和第三损失信息与将第二训练图像与所述至少一个原始图像中的尚未经过AI缩小的原始图像进行比较的结果相应。

[0016] 放大DNN的更新的DNN设置信息可包括权重残差信息/偏置残差信息,其中,权重残差信息/偏置残差信息指示在更新权重/偏置之前放大DNN中的所有或一些卷积层中的所有或一些滤波器核的权重/偏置与在更新权重/偏置之后放大DNN中的所述所有或一些卷积层中的所述所有或一些滤波器核的权重/偏置之间的差。

[0017] 放大DNN的更新的DNN设置信息可包括关于通过执行频率变换获得的权重残差/偏置残差的信息,关于权重残差/偏置残差的所述信息指示在更新权重/偏置之前放大DNN中的所有或一些卷积层中的所有或一些滤波器核的权重/偏置与在更新权重/偏置之后放大DNN中的所述所有或一些卷积层中的所述所有或一些滤波器核的权重/偏置之间的差。

[0018] 权重残差信息/偏置残差信息可以是使用差分脉冲编码调制(DPCM)、游程长度编码(RLC)和霍夫曼编码技术中的一种编码的信息。

[0019] 权重残差信息/偏置残差信息可以是关于经由模型压缩生成的权重残差/偏置残差的信息。

[0020] 模型压缩可包括修剪或量化中的至少一个。

[0021] 放大DNN的更新的DNN设置信息可以是被更新以用于执行AI放大的信息,所述用于执行AI放大的信息是通过基于与在更新权重/偏置之前放大DNN中的所有或一些卷积层中的所有或一些滤波器核的权重/偏置有关的上下文模型信息,对在更新权重/偏置之后放大DNN中的所有或一些卷积层中的所有或一些滤波器核的权重/偏置进行熵编码而获得的。

[0022] DNN设置信息可包括标志信息,所述标志信息指示是通过使用预定DNN中的卷积层的滤波器核来执行AI放大,还是通过使用经由放大DNN和用于所述至少一个原始图像的AI缩小的缩小DNN的联合训练而被更新以用于执行与所述至少一个原始图像相应的所述至少一个第二图像的AI放大的DNN中的卷积层的滤波器核来执行AI放大,所述联合训练是使用所述至少一个原始图像执行的。

[0023] 根据本公开的另一实施例,一种AI编码设备包括:存储器,存储一个或多个指令;以及处理器,被配置为执行存储在存储器中的一个或多个指令以执行以下操作:通过经由缩小DNN对至少一个原始图像执行AI缩小来获得第一图像;生成包括所述图像数据和AI数据的AI编码数据,其中,所述AI数据包括与所述AI缩小相关的信息,其中,所述AI数据

包括用于对第二图像执行AI放大的放大DNN的DNN设置信息,并且第二图像是通过对所述图像数据执行第一解码而生成的,并且其中所述DNN设置信息是经由放大DNN和用于所述至少一个原始图像的AI缩小的缩小DNN的联合训练而被更新以用于执行与所述至少一个原始图像相应的至少一个第二图像的AI放大的DNN信息,所述联合训练是使用所述至少一个原始图像执行的。

[0024] 根据本公开的另一实施例,一种AI解码方法包括:获得通过对第一图像执行第一编码而生成的图像数据以及与至少一个原始图像的AI缩小相关的AI数据,其中,所述至少一个原始图像与第一图像相关;通过对所述图像数据执行第一解码来获得与第一图像相应的第二图像;基于所述AI数据获得用于执行第二图像的AI放大的DNN设置信息;以及通过经由根据获得的DNN设置信息操作的放大DNN对第二图像执行AI放大来生成第三图像,其中,DNN设置信息是经由放大DNN和用于至少一个原始图像的AI缩小的缩小DNN的联合训练而被更新以用于执行与至少一个原始图像相应的至少一个第二图像的AI放大的DNN信息,所述联合训练是使用至少一个原始图像执行的。

[0025] 根据本公开的另一实施例,一种AI编码方法包括:通过经由缩小DNN对至少一个原始图像执行AI缩小来获得第一图像;通过对第一图像执行第一编码来生成图像数据;以及生成包括所述图像数据和AI数据的AI编码数据,所述AI数据包括与所述AI缩小有关的信息,其中,所述AI数据包括用于对第二图像执行AI放大的放大DNN的DNN设置信息,并且第二图像是通过对所述图像数据执行第一解码而生成的,并且其中所述DNN设置信息是经由放大DNN和用于所述至少一个原始图像的AI缩小的缩小DNN的联合训练而被更新以用于执行与所述至少一个原始图像相应的至少一个第二图像的AI放大的DNN信息,所述联合训练是使用所述至少一个原始图像执行的。

[0026] 根据本公开的另一实施例,一种计算机可读记录介质在其上记录有用于执行上述AI解码方法和AI编码方法的程序。

附图说明

[0027] 通过以下结合附图进行的描述,本公开的特定实施例的上述和其他方面、特征和优点将更加明显,其中:

[0028] 图1是用于描述根据实施例的人工智能(AI)编码处理和AI解码处理的示意图;

[0029] 图2是根据实施例的AI解码设备的框图;

[0030] 图3是示出用于对图像执行AI放大的深度神经网络(DNN)的示意图;

[0031] 图4是用于描述由卷积层进行的卷积操作的示意图;

[0032] 图5是示出图像相关信息和DNN设置信息之间的映射关系的表;

[0033] 图6是示出包括多个帧的图像的示意图;

[0034] 图7是根据实施例的AI编码设备的框图;

[0035] 图8是示出用于对原始图像执行AI缩小的DNN的示意图;

[0036] 图9是用于描述训练第一DNN和第二DNN的方法的示意图;

[0037] 图10是用于描述训练设备对第一DNN和第二DNN的训练处理的示意图;

[0038] 图11是用于对原始图像执行AI缩小的设备和用于对第二图像执行AI放大的设备的示意图;

- [0039] 图12是根据本公开的实施例的AI编码设备的配置的框图；
- [0040] 图13是根据本公开的实施例的AI解码设备的配置的框图；
- [0041] 图14a是根据本公开的实施例的AI编码方法的流程图；
- [0042] 图14b是根据本公开的实施例的经由基于原始图像的DNN联合训练的AI编码方法的流程图；
- [0043] 图14c是根据本公开的实施例的经由基于原始图像的DNN单独训练的AI编码方法的流程图；
- [0044] 图15是根据本公开的实施例的AI解码方法的流程图；
- [0045] 图16a是用于描述根据本公开的实施例的作为联合训练的第一阶段的由AI编码设备执行的通过使用原始图像作为训练图像经由缩小DNN和放大DNN的联合训练来确定缩小DNN和放大DNN的多条最佳设置信息的处理的示意图；
- [0046] 图16b是用于描述根据本公开的实施例的作为联合训练的第二阶段的由AI编码设备执行的选择根据图16a所示的处理确定的缩小DNN的设置信息、经由放大DNN的单独训练确定放大DNN的最佳设置信息、以及生成包括放大DNN的最佳设置信息的AI编码数据的处理的示意图；
- [0047] 图16c是根据本公开的实施例的由AI解码设备执行的基于包括在AI编码数据中的放大DNN的设置信息经由放大DNN对第二图像执行AI放大的处理的示意图；
- [0048] 图17a是用于描述根据本公开的实施例的由AI编码设备执行的选择缩小DNN的设置信息、经由放大DNN的单独训练确定放大DNN的最佳DNN设置信息、以及生成包括放大DNN的最佳DNN设置信息的AI编码数据的处理的示意图；
- [0049] 图17b是用于描述根据本公开的实施例的由AI解码设备执行的基于包括在AI编码数据中的放大DNN的DNN设置信息经由放大DNN对第二图像执行AI放大的处理的示意图；
- [0050] 图18是根据本公开的实施例的由AI解码设备执行的通过更新基于从AI编码数据获得的标志预先确定的放大DNN的DNN设置信息来放大第二图像、或者通过更新放大DNN的针对原始图像优化的DNN设置信息来放大第二图像的处理的流程图；
- [0051] 图19a示出了根据本公开的实施例的默认权重和偏置、放大DNN中的针对原始图像优化的权重和偏置以及放大DNN中的权重差和偏置差的示例；
- [0052] 图19b示出了根据本公开的实施例的放大DNN的针对原始图像优化的权重和偏置以及经由量化和修剪确定的放大DNN中的权重和偏置的示例；
- [0053] 图20a是用于描述根据本公开的实施例的对放大DNN中的针对原始图像优化的权重进行编码的方法的示意图；
- [0054] 图20b是用于描述根据本公开的实施例的对放大DNN中的针对原始图像优化的权重进行编码的方法的示意图；
- [0055] 图21a是用于描述根据本公开的实施例的由AI编码设备执行的基于上下文模型对放大DNN的针对原始图像优化的权重信息进行熵编码的处理的示意图；以及
- [0056] 图21b是用于描述根据本公开的实施例的由AI解码设备执行的基于上下文模型对放大DNN的针对原始图像优化的权重信息进行熵解码的处理的示意图。

具体实施方式

[0057] 由于本公开允许各种改变和许多示例,因此将在附图中示出并在书面描述中详细描述特定实施例。然而,这并不旨在将本公开限制于特定的实践模式,并且应当理解,不脱离本公开的精神和技术范围的所有改变、等同物和替代物都包含在本公开中。

[0058] 在实施例的描述中,当相关技术的某些详细解释可能不必要地模糊本公开的本质时,省略这些解释。另外,在说明书的描述中使用的数字(例如,第一、第二等)仅仅是用于将一个元素与另一个元素区分开的标识性标记。

[0059] 在整个公开内容中,表述“A、B或C中的至少一个”包括仅A、仅B、仅C、A和B两者、A和C两者、B和C两者、A、B和C中的全部或其变体。

[0060] 此外,在本公开中,应当理解,当元件彼此“连接”或“耦接”时,元件可彼此直接连接或耦接,但是可以可选地通过其间的中间元件彼此连接或耦接,除非另有说明。

[0061] 在本发明中,关于表示为“单元”或“模块”的元件,可将两个或多于两个元件组合成一个元件,或可根据细分功能将一个元件划分成两个或多于两个元件。另外,下文描述的每个元件除了其自身的主要功能之外,还可另外执行由另一元件执行的一些或全部功能,并且每个元件的一些主要功能可完全由另一组件执行。

[0062] 此外,在本公开中,“图像”或“画面”可表示静止图像、包括多个连续静止图像(或帧)的运动图像或视频。

[0063] 此外,在本公开中,深度神经网络(DNN)是模拟脑神经的人工神经网络模型的代表性示例,并且不限于使用特定算法的人工神经网络模型。

[0064] 此外,在本公开中,“参数”是在形成神经网络的每个层的运算处理中使用的值,并且例如,可包括当将输入值应用于特定运算表达式时使用的权重。这里,参数可以以矩阵形式表示。该参数是作为神经网络的训练或迭代的结果的数值集合,并且可在必要时通过单独的训练数据来更新。

[0065] 此外,在本公开中,“第一DNN”指示用于人工智能(AI)缩小图像的DNN,并且“第二DNN”指示用于AI放大图像的DNN。

[0066] 此外,在本公开中,“DNN设置信息”包括与构成DNN的元素相关的信息。“DNN设置信息”包括上述作为与构成DNN的元素相关的信息的参数。可通过使用DNN设置信息来设置第一DNN或第二DNN。

[0067] 此外,在本公开中,“原始图像”表示作为AI编码的对象的图像,并且“第一图像”表示作为在AI编码处理期间对原始图像执行AI缩小的结果而获得的图像。此外,“第二图像”表示在AI解码处理期间经由第一解码获得的图像,而“第三图像”表示在AI解码处理期间通过AI放大第二图像获得的图像。

[0068] 此外,在本公开中,“AI缩小”(或缩小)表示基于AI降低图像的分辨率的处理,并且“第一编码”表示根据基于频率变换的图像压缩方法的编码处理。此外,“第一解码”表示根据基于频率变换的图像重构方法的解码处理,而“AI放大”(或放大)表示基于AI提高图像分辨率的处理。

[0069] 图1是根据实施例的用于描述AI编码处理和AI解码处理的示图。

[0070] 当图像的分辨率显著提高时,用于编码和解码图像的信息的吞吐量增大。因此,需要一种用于提高图像的编码和解码效率的方法。

[0071] 如图1所示,根据本公开的实施例,通过对具有高分辨率的原始图像105执行AI缩小110来获得第一图像115。然后,对具有相对低分辨率的第一图像115执行第一编码120和第一解码130,因此与简单地对高分辨率原始图像105执行第一编码和第一解码相比,可显著降低比特率。

[0072] 具体地,在图1中,根据实施例,通过对原始图像105执行AI缩小110来获得第一图像115,并且在AI编码处理期间由编码设备、发送源等对第一图像115执行第一编码120。在AI解码处理期间,接收作为AI编码的结果获得的包括AI数据和图像数据的AI编码数据,经由第一解码130获得第二图像135,并且通过解码设备、接收装置等对第二图像135执行AI放大140获得第三图像145。

[0073] 详细参照AI编码处理,当接收到原始图像105时,对原始图像105执行AI缩小110以获得特定分辨率或特定质量的第一图像115。这里,基于AI执行AI缩小110,并且用于执行AI缩小110的AI模型需要与用于执行第二图像135的AI放大140的AI模型联合训练。这是因为,当分别训练用于AI缩小110的AI模型和用于AI放大140的AI模型时,可能存在由于经受AI编码的原始图像105与通过AI解码重构的第三图像145之间的差异而导致的失真。

[0074] 在本公开的实施例中,AI数据可以是用于在AI编码处理和AI解码处理期间保持这种联合关系的信令。因此,通过AI编码处理获得的AI数据可包括指示放大目标的信息,并且在AI解码处理期间,根据基于AI数据验证的放大目标对第二图像135执行AI放大140。

[0075] 用于AI缩小110的AI模型和用于AI放大140的AI模型可实现为DNN。如稍后将参照图9描述的,因为通过在特定目标下共享损失信息来联合训练第一DNN和第二DNN,所以AI编码设备可将在第一DNN和第二DNN的联合训练期间使用的目标信息提供给AI解码设备,并且AI解码设备可基于所提供的目标信息对第二图像135执行AI放大140以达到目标分辨率。

[0076] 关于图1的第一编码120和第一解码130,可通过第一编码120减少通过对原始图像105执行AI缩小110而获得的第一图像115的信息量。第一编码120可包括:通过预测第一图像115来生成预测数据的处理、生成与第一图像115和预测数据之间的差相应的残差数据的处理、将空间域分量的残差数据变换为频域分量的处理、将变换为频域分量的残差数据量化的处理、以及对量化的残差数据进行熵编码的处理。这样的第一编码120可经由使用频率变换的图像压缩方法(诸如MPEG-2、H.264高级视频编码(AVC)、MPEG-4、H.265/高效视频编码(HEVC)、VC-1、VP8、VP9和AOMediaVideo1(AV1))之一来执行。

[0077] 可通过对图像数据执行第一解码130来重构与第一图像115相应的第二图像135。第一解码130可包括:通过对图像数据进行熵解码来生成量化的残差数据的处理、对量化的残差数据进行逆量化的处理、将频域分量的残差数据变换为空间域分量的处理、生成预测数据的处理、以及通过使用预测数据和残差数据来重构第二图像135的处理。可经由与在第一编码120中使用的使用频率变换的图像压缩方法(诸如MPEG-2、H.264AVC、MPEG-4、H.265/HEVC、VC-1、VP8、VP9和AV1)之一相应的相应图像重构方法来执行这种第一解码130。

[0078] 通过AI编码处理获得的AI编码数据可包括作为对第一图像115执行第一编码120的结果而获得的图像数据、以及与原始图像105的AI缩小110相关的AI数据。可在第一解码130期间使用图像数据,并且可在AI放大140期间使用AI数据。

[0079] 图像数据可以以比特流的形式发送。图像数据可包括基于第一图像115中的像素值获得的数据,例如,作为第一图像115与第一图像115的预测数据之间的差的残差数据。此

外,图像数据包括在对第一图像115执行的第一编码120期间使用的信息。例如,图像数据可以包括在第一编码120期间使用的预测模式信息、运动信息和与量化参数相关的信息。可根据MPEG-2、H.264AVC、MPEG-4、H.265/HEVC、VC-1、VP8、VP9和AV1中的在第一编码120期间使用的图像压缩方法的规则(例如,根据语法)来生成图像数据。

[0080] 在基于第二DNN的AI放大140中使用AI数据。如上所述,因为第一DNN和第二DNN被联合训练,所以AI数据包括使得能够通过第二DNN对第二图像135一致地执行AI放大140的信息。在AI解码处理期间,可基于AI数据对第二图像135执行AI放大140以具有目标分辨率和/或质量。

[0081] AI数据可以以比特流的形式与图像数据一起发送。可选地,根据实施例,AI数据可以以帧或包的形式与图像数据分开发送。作为AI编码的结果获得的AI数据和图像数据可通过相同的网络或通过不同的网络发送。

[0082] 图2是根据实施例的AI解码设备的框图。

[0083] 参照图2,根据实施例的AI解码设备200可包括接收器210和AI解码器230。接收器210可包括通信器212、解析器214和输出器216。AI解码器230可包括第一解码器232和AI放大器234。

[0084] 接收器210接收并解析作为AI编码的结果获得的AI编码数据,并将图像数据和AI数据输出到AI解码器230。

[0085] 具体地,通信器212通过无线网络接收作为AI编码的结果而获得的AI编码数据。作为执行AI编码的结果而获得的AI编码数据包括图像数据和AI数据。可通过相同类型的网络或不同类型的网络接收图像数据和AI数据。

[0086] 解析器214接收通过通信器212接收的AI编码数据,并解析AI编码数据以从AI数据分割图像数据。例如,解析器214可通过读取从通信器212获得的数据的报头来区分图像数据和AI数据,其中,所述报头描述包括图像数据和AI数据的有效载荷或AI数据的配置。根据实施例,解析器214经由通过通信器212接收的数据的报头将图像数据和AI数据发送到输出器216,并且输出器216将图像数据和AI数据分别发送到第一解码器232和AI放大器234。此时,包括在AI编码数据中的图像数据可被验证为是经由特定编解码器(例如,MPEG-2、H.264AVC、MPEG-4、HEVC、VC-1、VP8、VP9或AV1)生成的图像数据。在这种情况下,可通过输出器216将相应的信息发送到第一解码器232,使得经由适当的编解码器处理图像数据。

[0087] 根据实施例,由解析器214解析的AI编码数据可从数据存储介质获得,该数据存储介质包括磁介质(诸如硬盘、软盘或磁带)、光学记录介质(诸如CD-ROM或DVD)或磁光介质(诸如软光盘)。

[0088] 第一解码器232基于图像数据重构与第一图像115相应的第二图像135。由第一解码器232获得的第二图像135被提供给AI放大器234。根据实施例,可将包括在图像数据中的第一解码相关信息(诸如预测模式信息、运动信息、量化参数信息等)进一步提供给AI放大器234。

[0089] 在接收到AI数据时,AI放大器234基于AI数据对第二图像135执行AI放大。根据实施例,可通过进一步使用包括在图像数据中的第一解码相关信息(诸如预测模式信息、量化参数信息等)来执行AI放大。

[0090] 根据实施例的接收器210和AI解码器230被描述为单独的组件或装置,但是可通过

用于控制接收器210和解码器230的联合功能的一个处理器联合实现。在这种情况下,接收器210和AI解码器230可通过专用处理器或通过软件和通用处理器(诸如应用处理器(AP)、中央处理器(CPU)或图形处理器(GPU))的组合来实现。可通过包括用于实现本公开的实施例的存储器或者通过包括用于使用外部存储器的存储器处理器来实现专用处理器。通信器212可包括通信接口(诸如包括无线电和/或天线的无线通信接口或诸如HDMI、以太网等的有线通信接口),以通过网络或从数据存储介质接收AI编码数据,其中,所述数据存储介质包括磁介质(诸如硬盘、软盘或磁带)、光学记录介质(诸如CD-ROM或DVD)或磁光介质(诸如软光盘)。可选地,通信器212可包括专用处理器的一部分或组件,或者软件和通用处理器(诸如应用处理器(AP)、中央处理器(CPU)或图形处理器(GPU))的组合,以接收AI编码数据。

[0091] 此外,接收器210和AI解码器230可由多个处理器配置。在这种情况下,接收器210和AI解码器230可通过专用处理器的组合或通过软件和通用处理器(诸如AP、CPU或GPU)的组合来实现。类似地,AI放大器234及第一解码器232可由不同处理器实施。

[0092] 提供给AI放大器234的AI数据包括使第二图像135能够经由AI放大被处理的信息。这里,放大目标或比率应与第一DNN的缩小目标或比率相应。因此,AI数据包括用于验证第一DNN的缩小目标的信息。

[0093] 包括在AI数据中的信息的示例包括原始图像105的分辨率与第一图像115的分辨率之间的差异信息以及与第一图像115相关的信息。

[0094] 差异信息可表示为关于第一图像115与原始图像105相比的分辨率转换程度的信息(例如,分辨率转换率信息)。此外,因为通过重建的第二图像135的分辨率验证第一图像115的分辨率,并且相应地验证分辨率转换程度,所以差异信息可仅表示为原始图像105的分辨率信息。这里,分辨率信息可表示为垂直/水平尺寸或者表示为比率(16:9、4:3等)和一个轴的尺寸。此外,当存在预设分辨率信息时,分辨率信息可通过一个或多个用信号发送的比特以索引或标志的形式表示。

[0095] 与第一图像115相关的信息可包括关于作为对第一图像115执行第一编码的结果而获得的图像数据的比特率或在第一图像115的第一编码期间使用的编解码器类型中的至少一个的信息。

[0096] AI放大器234可基于包括在AI数据中的差异信息或与第一图像115相关的信息中的至少一个来确定第二图像135的放大目标。放大目标可指示例如将针对第二图像135使分辨率放大到什么程度。当确定了放大目标时,AI放大器234通过第二DNN对第二图像135执行AI放大,以获得与放大目标相应的第三图像145。

[0097] 在描述由AI放大器234执行的根据放大目标对第二图像135执行AI放大的方法之前,将参照图3和图4描述通过第二DNN的AI放大处理。

[0098] 图3是示出用于对第二图像135执行AI放大的DNN 300的示图,并且图4是用于描述图3的第一卷积层310中的卷积操作的示图。

[0099] 如图3所示,第二图像135被输入到第一卷积层310。在图3所示的第一卷积层310中指示的标记 $3 \times 3 \times 4$ 指示通过使用具有 3×3 大小的四个滤波器核对一个输入图像执行卷积处理。作为卷积处理的结果,由四个滤波器核生成四个特征图。每个特征图指示第二图像135的固有特性。例如,每个特征图可表示第二图像135的垂直方向特性、水平方向特性或边

缘特性等。

[0100] 将参照图4详细描述第一卷积层310中的卷积操作。

[0101] 可通过在第一卷积层310中使用的大小为 3×3 的滤波器核430的参数与第二图像135中的相应像素值之间的乘法和加法来生成一个特征图450。因为在第一卷积层310中使用四个滤波器核,所以可使用四个滤波器核通过卷积运算生成四个特征图。

[0102] 在图4中的第二图像135中,I1至I49指示第二图像135中的像素,并且滤波器核430中指示的F1至F9指示滤波器核430的参数。此外,在特征图450中指示的M1至M9指示特征图450的样本。

[0103] 在图4中,第二图像135包括49个像素,但是像素的数量仅是简化示例,并且当第二图像135具有4K的分辨率时,第二图像135可包括例如 3840×2160 个像素。

[0104] 在卷积运算处理期间,将第二图像135的I1、I2、I3、I8、I9、I10、I15、I16和I17的像素值与滤波器核430的F1至F9分别相乘,并且可将相乘的结果值的组合(例如,相加)的值分配为特征图450的M1的值。当卷积运算的步幅为2时,第二图像135的I3、I4、I5、I10、I11、I12、I17、I18和I19的像素值分别与滤波器核430的F1至F9相乘,并且相乘的结果值的组合的值可被分配为特征图450的M2的值。

[0105] 当滤波器核430根据步幅前进到第二图像135的最后一个像素时,在第二图像135中的像素值与滤波器核430的参数之间执行卷积操作,因此可生成具有特定尺寸的特征图450。

[0106] 根据本公开,可通过第一DNN和第二DNN的联合训练来优化第二DNN的参数的值,例如,在第二DNN的卷积层中使用的滤波器核的参数(例如,滤波器核430的F1至F9)的值。如上所述,AI放大器234可基于AI数据确定与第一DNN的缩小目标相应的放大目标,并且将与所确定的放大目标相应的参数确定为在第二DNN的卷积层中使用的滤波器核的参数。

[0107] 包括在第一DNN和第二DNN中的卷积层可根据参照图4描述的卷积运算处理来执行处理,但是参照图4描述的卷积运算处理仅是示例,并且卷积运算处理不限于此。

[0108] 返回参照图3,从第一卷积层310输出的特征图可被输入到第一激活层320。

[0109] 第一激活层320可将非线性特征分配给每个特征图。第一激活层320可包括sigmoid函数、Tanh函数、修正线性单元(ReLU)函数等,但第一激活层320不限于此。

[0110] 第一激活层320分配非线性特征指示:作为第一卷积层310的输出的特征图的至少一个样本值被改变。这里,通过应用非线性特征来执行改变。

[0111] 第一激活层320确定是否将从第一卷积层310输出的特征图的样本值发送到第二卷积层330。例如,特征图的一些样本值由第一激活层320激活并且被发送到第二卷积层330,并且一些样本值由第一激活层320去激活并且不被发送到第二卷积层330。通过第一激活层320强调由特征图表示的第二图像135的固有特性。

[0112] 从第一激活层320输出的特征图325被输入到第二卷积层330。图3中所示的特征图325中的一个特征图325是在第一激活层320中处理参照图4描述的特征图450的结果。

[0113] 在第二卷积层330中指示的标记 $3 \times 3 \times 4$ 指示通过使用具有 3×3 的大小的四个滤波器核对特征图325执行卷积处理。第二卷积层330的输出被输入到第二激活层340。第二激活层340可将非线性特征分配给输入数据。

[0114] 从第二激活层340输出的特征图345被输入到第三卷积层350。在图3所示的第三卷

积层350中指示的标记 $3 \times 3 \times 1$ 指示通过使用大小为 3×3 的一个滤波器核执行卷积处理来生成一个输出图像。第三卷积层350是用于输出最终图像的层,并且通过使用一个滤波器核来生成一个输出。根据本公开的实施例,第三卷积层350可输出第三图像145作为卷积运算的结果。

[0115] 如稍后将描述的,可存在指示第二DNN 300的第一卷积层310、第二卷积层330和第三卷积层350的滤波器核的数量、第二DNN 300的第一卷积层310、第二卷积层330和第三卷积层350的滤波器核的参数等的多条DNN设置信息,并且所述多条DNN设置信息应当与第一DNN的多条DNN设置信息相关联。可经由第一DNN和第二DNN的联合训练来实现第二DNN的多条DNN设置信息与第一DNN的多条DNN设置信息之间的关联。

[0116] 在图3中,第二DNN 300包括三个卷积层(第一卷积层310、第二卷积层330和第三卷积层350)和两个激活层(第一激活层320和第二激活层340),但这仅是示例,并且卷积层和激活层的配置和数量可以根据实施例而变化。此外,根据实施例,第二DNN 300可被实现为递归神经网络(RNN)。在这种情况下,根据本公开的实施例的第二DNN 300的卷积神经网络(CNN)结构被改变为RNN结构。

[0117] 根据实施例,AI放大器234可以包括用于卷积运算和上述激活层的运算的至少一个算术逻辑单元(ALU)。ALU可以被实现为处理器。对于卷积运算,ALU可包括在第二图像135或从先前层输出的特征图的样本值与滤波器核的样本值之间执行乘法的乘法器、以及将乘法的结果值相加的加法器。此外,对于激活层的操作,ALU可包括乘法器和比较器,其中,乘法器将输入样本值乘以在预定的sigmoid函数、Tanh函数或ReLU函数中使用的权重,比较器将乘法结果与特定值进行比较以确定是否将输入样本值发送到下一层。

[0118] 在下文中,将描述由AI放大器234执行的根据放大目标对第二图像135执行AI放大的方法。

[0119] 根据实施例,AI放大器234可存储在第二DNN中可配置的多条DNN设置信息。

[0120] 这里,DNN设置信息可包括关于包括在第二DNN中的卷积层的数量、每个卷积层的滤波器核的数量或每个滤波器核的参数中的至少一个的信息。多条DNN设置信息可分别与各种放大目标相应,并且第二DNN可基于与某个放大目标相应的DNN设置信息进行操作。基于DNN设置信息,第二DNN可具有不同的结构。例如,第二DNN可基于任何一条DNN设置信息而包括三个卷积层,并且可基于另一条DNN设置信息而包括四个卷积层。

[0121] 根据实施例,DNN设置信息可仅包括在第二DNN中使用的滤波器核的参数。在这种情况下,第二DNN的结构不改变,而是仅内部滤波器核的参数可基于DNN设置信息而改变。

[0122] AI放大器234可获得多条DNN设置信息中的用于对第二图像135执行AI放大的DNN设置信息。多条DNN设置信息中的每一条DNN设置信息可以是用于获得预定分辨率和/或预定质量的第三图像145的信息,并且被与第一DNN联合训练。

[0123] 例如,多条DNN设置信息中的一条DNN设置信息可包括用于获得分辨率是第二图像135的分辨率的两倍的第三图像145(例如,两倍于第二图像135的2K(2048×1080)分辨率的4K(4096×2160)分辨率的第三图像145)的信息,并且另一条DNN设置信息可包括用于获得分辨率高于第二图像135的分辨率四倍的第三图像145(例如,比第二图像135的2K(2048×1080)分辨率高四倍的8K(8192×4320)分辨率的第三图像145)的信息。

[0124] 多条DNN设置信息中的每一条DNN设置信息与图6的AI编码设备600的第一DNN的

DNN设置信息联合获得,并且AI放大器234根据与第一DNN的DNN设置信息的缩小率相应的放大率获得所述多条DNN设置信息中的一条DNN设置信息。在这方面,AI放大器234可验证第一DNN的信息。为了使AI放大器234验证第一DNN的信息,根据实施例的AI解码设备200从AI编码设备600接收包括第一DNN的信息的AI数据。

[0125] 换句话说,AI放大器234可通过使用从AI编码设备600接收的信息,来验证用于获得第一图像115的第一DNN的DNN设置信息所针对的信息,并且获得与第一DNN的DNN设置信息联合训练的第二DNN的DNN设置信息。

[0126] 当从所述多条DNN设置信息中获得用于对第二图像135执行AI放大的DNN设置信息时,可基于根据所获得的DNN设置信息操作的第二DNN来处理输入数据。

[0127] 例如,当获得任何一条DNN设置信息时,包括在图3的第二DNN 300的第一卷积层310、第二卷积层330和第三卷积层350中的每一个中的滤波器核的数量以及滤波器核的参数被设置为包括在所获得的DNN设置信息中的值。

[0128] 具体地,在图4的第二DNN的任何一个卷积层中使用的 3×3 的滤波器核的参数被设置为{1,1,1,1,1,1,1,1},并且当之后改变DNN设置信息时,参数被作为包括在改变后的DNN设置信息中的参数的{2,2,2,2,2,2,2,2}替换。

[0129] AI放大器234可基于包括在AI数据中的信息从所述多条DNN设置信息中获得用于AI放大的DNN设置信息,并且现在将描述用于获得DNN设置信息的AI数据。

[0130] 根据实施例,AI放大器234可基于包括在AI数据中的差异信息从所述多条DNN设置信息中获得用于AI放大的DNN设置信息。例如,当原始图像105的分辨率(例如,4K (4096 $\times$ 2160))是第一图像115的分辨率(例如,2K (2048 $\times$ 1080))的两倍时,基于差异信息,AI放大器234可获得用于将第二图像135的分辨率增大两倍的DNN设置信息。

[0131] 根据另一实施例,AI放大器234可基于包括在AI数据中的与第一图像115相关的信息,从所述多条DNN设置信息中获得用于对第二图像135进行AI放大的DNN设置信息。AI放大器234可预先确定图像相关信息和DNN设置信息之间的映射关系,并且获得映射到与第一图像115相关的信息的DNN设置信息。

[0132] 图5是示出图像相关信息和DNN设置信息之间的映射关系的表。

[0133] 根据图5,根据本公开的实施例的AI编码和AI解码处理可不仅仅考虑分辨率的改变。如图5所示,可单独地或共同地考虑分辨率(诸如标准清晰度(SD)、高清晰度(HD)或全HD)、比特率(诸如10Mbps、15Mbps或20Mbps)以及编解码器信息(诸如AV1、H.264或HEVC)来选择DNN设置信息。对于分辨率、比特率和编解码器信息的这种考虑,考虑每个元素的训练应该在AI训练处理期间与编码和解码处理联合执行(参见图9)。

[0134] 因此,当根据训练如图5所示基于包括编解码器类型、图像的分辨率等的图像相关信息提供多条DNN设置信息时,可基于在AI解码处理期间接收的与第一图像115相关的信息来获得用于AI放大第二图像135的DNN设置信息。

[0135] 换句话说,AI放大器234能够通过将图5的表左侧的图像相关信息与表右侧的DNN设置信息相关起来根据图像相关信息使用DNN设置信息。

[0136] 如图5所示,当从与第一图像115相关的信息验证第一图像115的分辨率是SD,作为对第一图像115执行第一编码的结果而获得的图像数据的比特率是10Mbps,并且经由AV1编解码器对第一图像115执行第一编码时,AI放大器234可使用所述多条DNN设置信息中的A

DNN设置信息。

[0137] 此外,当从与第一图像115相关的信息验证第一图像115的分辨率是HD,作为执行第一编码的结果而获得的图像数据的比特率是15Mbps,并且经由H.264编解码器执行第一编码时,AI放大器234可使用所述多条DNN设置信息中的B DNN设置信息。

[0138] 此外,当从与第一图像115相关的信息验证第一图像115的分辨率是全HD,作为执行第一编码的结果而获得的图像数据的比特率是20Mbps,并且经由HEVC编解码器执行第一编码时,AI放大器234可使用所述多条DNN设置信息中的C DNN设置信息。最后,当验证第一图像115的分辨率是全HD,作为执行第一编码的结果获得的图像数据的比特率是15Mbps,并且经由HEVC编解码器执行第一编码时,AI放大器234可使用所述多条DNN设置信息中的D DNN设置信息。可基于作为对第一图像115执行第一编码的结果而获得的图像数据的比特率是20Mbps还是15Mbps来选择C DNN设置信息和D DNN设置信息中的一个。当经由相同编解码器对相同分辨率的第一图像115执行第一编码时获得的图像数据的不同比特率指示重建图像的不同质量。因此,可基于特定图像质量来联合训练第一DNN和第二DNN,并且因此,AI放大器234可根据指示第二图像135的质量的图像数据的比特率来获得DNN设置信息。

[0139] 根据另一实施例,AI放大器234可在考虑从第一解码器232提供的信息(预测模式信息、运动信息、量化参数信息等)和包括在AI数据中的与第一图像115相关的信息两者的情况下,从所述多条DNN设置信息中获得用于对第二图像135执行AI放大的DNN设置信息。例如,AI放大器234可从第一解码器232接收在第一图像115的第一编码处理期间使用的量化参数信息,验证从AI数据获得的作为第一图像115的编码结果的图像数据的比特率,并且获得与量化参数信息和比特率相应的DNN设置信息。即使当比特率相同时,重建图像的质量也可能根据图像的复杂度而变化。比特率是表示对其执行第一编码的整个第一图像115的值,并且即使在第一图像115内,每个帧的质量也可以变化。因此,与仅使用AI数据时相比,当一起考虑来自第一解码器232的针对每个帧可获得的预测模式信息、运动信息和/或量化参数时,可获得更适合于第二图像135的DNN设置信息。

[0140] 此外,根据实施例,AI数据可包括相互商定的DNN设置信息的标识符。DNN设置信息的标识符是用于区分在第一DNN和第二DNN之间联合训练使得对第二图像135执行AI放大到与第一DNN的缩小目标相应的放大目标而得到的一对DNN设置信息的信息。在获得包括在AI数据中的DNN设置信息的标识符之后,AI放大器234可通过使用与DNN设置信息的标识符相应的DNN设置信息对第二图像135执行AI放大。例如,可预先指定指示在第一DNN中可配置或可选择的多个DNN设置信息中的每个DNN设置信息的标识符和指示在第二DNN中可配置或可选择的多个DNN设置信息中的每个DNN设置信息的标识符。在这种情况下,可为在第一DNN和第二DNN中的每一个中可配置或可选择的一对DNN设置信息指定相同的标识符。AI数据可包括在用于原始图像105的AI缩小的第一DNN中设置的DNN设置信息的标识符。接收AI数据的AI放大器234可通过使用多个DNN设置信息中的由包括在AI数据中的标识符指示的DNN设置信息来对第二图像135执行AI放大。

[0141] 此外,根据实施例,AI数据可包括DNN设置信息。AI放大器234可在获得包括在AI数据中的DNN设置信息之后通过使用DNN设置信息对第二图像135执行AI放大。

[0142] 根据实施例,当构成DNN设置信息的多条信息(例如,卷积层的数量、每个卷积层的滤波器核的数量、每个滤波器核的参数等)以查找表的形式存储时,AI放大器234可通过基

于包括在AI数据中的信息将从查找表中的值中选择的一些值进行组合来获得DNN设置信息,并且通过使用获得的DNN设置信息对第二图像135执行AI放大。

[0143] 根据实施例,当确定与放大目标相应的DNN的结构时,AI放大器234可获得与DNN的结构相应的DNN设置信息,例如,滤波器核的参数。

[0144] AI放大器234通过包括与第一DNN有关的信息的AI数据获得第二DNN的DNN设置信息,并且通过基于获得的DNN设置信息设置的第二DNN对第二图像135执行AI放大。因此,与直接分析第二图像135的特征以进行放大时相比,可减少存储器使用和所需的吞吐量。

[0145] 根据实施例,当第二图像135包括多个帧时,AI放大器234可独立地获得特定数量的帧的DNN设置信息,或者可获得全部帧的公共DNN设置信息。

[0146] 图6是示出包括多个帧的图像的示图。

[0147] 如图6中所展示,第二图像135可包括帧 t_0 到 t_n 。

[0148] 根据实施例,AI放大器234可通过AI数据获得第二DNN的DNN设置信息,并且基于DNN设置信息对帧 t_0 至 t_n 执行AI放大。换句话说,可基于公共DNN设置信息经由AI放大来处理帧 t_0 到 t_n 。

[0149] 根据另一实施例,AI放大器234可通过使用从AI数据获得的“A”DNN设置信息对帧 t_0 至 t_n 中的一些(例如,帧 t_0 至 t_a)执行AI放大,并且通过使用从AI数据获得的“B”DNN设置信息对帧 t_{a+1} 至 t_b 执行AI放大。此外,AI放大器234可通过使用从AI数据获得的“C”DNN设置信息来对帧 t_{b+1} 至 t_n 执行AI放大。换句话说,AI放大器234可独立地获得包括多个帧中的特定数量的帧的每个帧组的DNN设置信息,并且通过使用独立获得的DNN设置信息对包括在每个组中的帧执行AI放大。

[0150] 根据另一实施例,AI放大器234可独立地获得形成第二图像135的每个帧的DNN设置信息。换句话说,当第二图像135包括三个帧时,AI放大器234可通过使用关于第一帧获得的DNN设置信息对第一帧执行AI放大,通过使用关于第二帧获得的DNN设置信息对第二帧执行AI放大,并且通过使用关于第三帧获得的DNN设置信息对第三帧执行AI放大。根据基于从第一解码器232提供的信息(预测模式信息、运动信息、量化参数信息等)和包括在上述AI数据中的与第一图像115相关的信息获得DNN设置信息的方法,可针对包括在第二图像135中的每个帧独立地获得DNN设置信息。这是因为可针对包括在第二图像135中的每个帧独立地确定模式信息、量化参数信息等。

[0151] 根据另一实施例,AI数据可包括关于基于AI数据获得的DNN设置信息适用于哪个帧的信息。例如,当AI数据包括指示DNN设置信息直到帧 t_a 有效的信息时,AI放大器234通过使用基于AI数据获得的DNN设置信息对帧 t_0 至 t_a 执行AI放大。此外,当另一条AI数据包括指示DNN设置信息直到帧 t_n 有效的信息时,AI放大器234通过使用基于另一条AI数据获得的DNN设置信息对帧 t_{a+1} 至 t_n 执行AI放大。

[0152] 在下文中,将参照图7描述用于对原始图像105执行AI编码的AI编码设备600。

[0153] 图7是根据实施例的AI编码设备的框图。

[0154] 参照图7,AI编码设备600可包括AI编码器610和发送器630。AI编码器610可包括AI缩小器612和第一编码器614。发送器630可包括数据处理器632和通信器634。

[0155] 在图7中,AI编码器610和发送器630被示出为单独的组件或装置,但是AI编码器610和发送器630可通过用于控制AI编码器610和发送器630的联合功能的一个处理器来联

合实现。在这种情况下, AI编码器610和发送器630可通过专用处理器或通过软件和通用处理器(诸如AP、CPU或图形处理器GPU)的组合来实现。可通过包括用于实现本公开的实施例的存储器或者通过包括用于使用外部存储器的存储器处理器来实现专用处理器。

[0156] 此外, AI编码器610和发送器630可通过多个处理器配置。在这种情况下, AI编码器610和发送器630可通过专用处理器的组合或通过软件和多个通用处理器(诸如AP、CPU或GPU)的组合来实现。AI缩小器612和第一编码器614可通过不同的处理器来实现。

[0157] AI编码器610对原始图像105执行AI缩小并对第一图像115执行第一编码, 并将AI数据和图像数据发送到发送器630。发送器630将AI数据和图像数据发送到AI解码设备200。发送器630可包括通信接口, 诸如包括无线电设备和/或天线的无线通信接口或诸如HDMI、以太网等的有线通信接口。可选地, 发射器630可包括专用处理器的部分或组件, 或通过软件与通用处理器(例如, 应用处理器(AP)、中央处理器(CPU)或图形处理器(GPU))的组合。

[0158] 图像数据包括作为对第一图像115执行第一编码的结果而获得的数据。图像数据可以包括基于第一图像115中的像素值获得的数据, 例如, 作为第一图像115与第一图像115的预测数据之间的差的残差数据。此外, 图像数据包括在第一图像115的第一编码处理期间使用的信息。例如, 图像数据可包括预测模式信息、运动信息、用于对第一图像115执行第一编码的量化参数信息等。

[0159] AI数据包括使得能够对第二图像135执行AI放大到与第一DNN的缩小目标相应的放大目标的信息。根据实施例, AI数据可包括原始图像105和第一图像115之间的差异信息。此外, AI数据可包括与第一图像115相关的信息。与第一图像115相关的信息可包括关于第一图像115的分辨率、作为对第一图像115执行第一编码的结果而获得的图像数据的比特率、或在第一图像115的第一编码期间使用的编解码器类型中的至少一个的信息。

[0160] 根据实施例, AI数据可包括相互商定的使得对第二图像135执行AI放大到与第一DNN的缩小目标相应的放大目标的DNN设置信息的标识符。

[0161] 此外, 根据实施例, AI数据可包括在第二DNN中可配置的DNN设置信息。

[0162] AI缩小器612可获得通过经由第一DNN对原始图像105执行AI缩小而获得的第一图像115。AI缩小器612可基于预定标准来确定原始图像105的缩小目标。

[0163] 为了获得与缩小目标匹配的第一图像115, AI缩小器612可存储在第一DNN中可设置的多条DNN设置信息。AI缩小器612从多条DNN设置信息中获得与缩小目标相应的DNN设置信息, 并且通过按照DNN设置信息设置的第一DNN对原始图像105执行AI缩小。

[0164] 可训练多条DNN设置信息中的每一条DNN设置信息以获得预定分辨率和/或预定质量的第一图像115。例如, 所述多条DNN设置信息中的任何一条DNN设置信息可包括用于获得分辨率为原始图像105的分辨率的一半的第一图像115(例如, 分辨率为原始图像105的4K(4096×2160)的一半的2K(2048×1080)的第一图像115)的信息, 并且另一条DNN设置信息可包括用于获得分辨率为原始图像105的分辨率的四分之一的第一图像115(例如, 分辨率是原始图像105的8K(8192×4320)分辨率的四分之一的2K(2048×1080)的第一图像115)的信息。

[0165] 根据实施例, 当构成DNN设置信息的多条信息(例如, 卷积层的数量、每个卷积层的滤波器核的数量、每个滤波器核的参数等)以查找表的形式存储时, AI缩小器612可通过基于缩小目标对从查找表中的值中选择的一些值进行组合来获得DNN设置信息, 并且通过使

用DNN设置信息对原始图像105执行AI缩小。

[0166] 根据实施例, AI缩小器612可确定与缩小目标相应的DNN的结构, 并且获得与DNN的结构相应的例如滤波器核的参数的DNN设置信息。

[0167] 当第一DNN和第二DNN被联合训练时, 用于对原始图像105执行AI缩小的多条DNN设置信息可具有优化值。这里, 每条DNN设置信息包括第一DNN中包括的卷积层的数量、每个卷积层的滤波器核的数量、或每个滤波器核的参数中的至少一个。

[0168] AI缩小器612可利用获得的用于对原始图像105执行AI缩小的DNN设置信息来设置第一DNN, 以通过第一DNN获得特定分辨率和/或特定质量的第一图像115。当从所述多条DNN设置信息获得用于对原始图像105执行AI缩小的DNN设置信息时, 第一DNN中的每个层可基于包括在DNN设置信息中的信息来处理输入数据。

[0169] 在下文中, 将描述由AI缩小器612执行的确定的缩小目标的方法。缩小目标可指示, 例如, 从原始图像105减小多少分辨率以获得第一图像115。

[0170] 根据实施例, AI缩小器612可基于压缩比 (例如, 原始图像105与第一图像115之间的分辨率差、目标比特率等)、压缩质量 (例如, 比特率的类型)、压缩历史信息或原始图像105的类型中的至少一个来确定缩小目标。

[0171] 例如, AI缩小器612可基于预设或从用户输入的压缩比、压缩质量等来确定缩小目标。

[0172] 作为另一示例, AI缩小器612可通过使用存储在AI编码设备600中的压缩历史信息来确定缩小目标。例如, 根据AI编码设备600可用的压缩历史信息, 可确定用户偏好的编码质量、压缩比等, 并且可根据基于压缩历史信息确定的编码质量来确定缩小目标。例如, 可根据压缩历史信息基于最常使用的编码质量来确定第一图像115的分辨率、质量等。

[0173] 作为另一示例, AI缩小器612可根据压缩历史信息基于比特定阈值更频繁使用的编码质量 (例如, 比特定阈值更频繁使用的编码质量的平均质量) 来确定缩小目标。

[0174] 作为另一示例, AI缩小器612可基于原始图像105的分辨率、类型 (例如, 文件格式) 等来确定缩小目标。

[0175] 根据实施例, 当原始图像105包括多个帧时, AI缩小器612可独立地确定特定数量的帧的缩小目标, 或者可确定全部帧的缩小目标。

[0176] 根据实施例, AI缩小器612可将包括在原始图像105中的帧划分为一定数量的组或子集, 并且独立地确定每个组的缩小目标。可针对每个组确定相同或不同的缩小目标。根据每个组, 包括在组中的帧的数量可相同或不同。

[0177] 根据另一实施例, AI缩小器612可独立地确定包括在原始图像105中的每个帧的缩小目标。可针对每一帧确定相同或不同的缩小目标。

[0178] 在下文中, 将描述AI缩小所基于的第一DNN700的结构示例。

[0179] 图8是示出用于对原始图像执行AI缩小的DNN的示意图。

[0180] 如图8所示, 原始图像105被输入到第一DNN700的第一卷积层710。第一卷积层710通过使用大小为 5×5 的32个滤波器核对原始图像105执行卷积处理。因此, 作为卷积处理的结果而生成的32个特征图被输入到第一激活层720。第一激活层720可将非线性特征分配给32个特征图。

[0181] 第一激活层720确定是否将从第一卷积层710输出的特征图的样本值发送到第二

卷积层730。例如,特征图的一些样本值由第一激活层720激活并且被发送到第二卷积层730,并且一些样本值由第一激活层720去激活并且不被发送到第二卷积层730。通过第一激活层720强调由从第一卷积层710输出的特征图表示的信息。

[0182] 第一激活层720的输出725被输入到第二卷积层730。第二卷积层730通过使用大小为 5×5 的32个滤波器核对输入数据执行卷积处理。因此,作为卷积处理的结果输出的32个特征图被输入到第二激活层740,并且第二激活层740可将非线性特征分配给32个特征图。

[0183] 第二激活层740的输出745被输入到第三卷积层750。第三卷积层750通过使用大小为 5×5 的一个滤波器核对输入数据执行卷积处理。作为卷积处理的结果,可从第三卷积层750输出一个图像。第三卷积层750通过使用所述一个滤波器核作为用于输出最终图像的层来生成一个输出。根据本公开的实施例,第三卷积层750可输出第一图像115作为卷积运算的结果。

[0184] 可存在指示第一DNN 700的第一卷积层710、第二卷积层730和第三卷积层750的滤波器核的数量、第一DNN700的第一卷积层710、第二卷积层730和第三卷积层750的每个滤波器核的参数等的多条DNN设置信息,并且所述多条DNN设置信息可与第二DNN的多条DNN设置信息相关联。可经由第一DNN和第二DNN的联合训练来实现第一DNN的多条DNN设置信息与第二DNN的多条DNN设置信息之间的关联。

[0185] 在图8中,第一DNN700包括三个卷积层(第一卷积层710、第二卷积层730和第三卷积层750)和两个激活层(第一激活层720和第二激活层740),但是这仅是示例配置,并且卷积层和激活层的数量和配置可根据实施例而变化。此外,根据实施例,第一DNN700可被实现为RNN。在这种情况下,根据本公开的实施例的第一DNN700的CNN结构被改变为RNN结构。

[0186] 根据实施例,AI缩小器612可包括用于卷积运算和上述激活层的运算的至少一个ALU。ALU可被实现为处理器。对于卷积运算,ALU可包括乘法器和加法器,乘法器执行原始图像105或从先前层输出的特征图的样本值与滤波器核的样本值之间的乘法,加法器将乘法的结果值相加。此外,对于激活层的操作,ALU可包括乘法器和比较器,乘法器将输入样本值乘以在预定的sigmoid函数、Tanh函数或ReLU函数中使用的权重,比较器将乘法结果与特定值进行比较以确定是否将输入样本值发送到下一层。

[0187] 返回参照图7,在从AI缩小器612接收到第一图像115后,第一编码器614可通过对第一图像115执行第一编码而减小第一图像115的信息量或数据的数量及大小。作为由第一编码器614执行第一编码的结果,可获得与第一图像115相应的图像数据。

[0188] 数据处理器632以某种形式对要发送的AI数据或图像数据中的至少一个进行处理。例如,当要以比特流的形式发送AI数据和图像数据时,数据处理器632可对要以比特流的形式表示的AI数据进行处理,并且通过通信器634以一个比特流的形式发送图像数据和AI数据。作为另一示例,数据处理器632可对要以比特流的形式表示的AI数据进行处理,并且通过通信器634发送与AI数据相应的比特流和与图像数据相应的比特流中的每一个。作为另一示例,数据处理器632可对要以帧或包的形式表示的AI数据进行处理,并且通过通信器634以比特流的形式发送图像数据并以帧或包的形式发送AI数据。

[0189] 通信器634通过网络发送作为执行AI编码的结果而获得的AI编码数据。作为执行AI编码的结果而获得的AI编码数据包括图像数据和AI数据。图像数据和AI数据可通过相同类型的网络或不同类型的网络被发送。

[0190] 根据实施例,作为数据处理器632的处理的结果而获得的AI编码数据可被存储在数据存储介质中,该数据存储介质包括磁介质(诸如硬盘、软盘或磁带)、光学记录介质(诸如CD-ROM或DVD)、或磁光介质(诸如软光盘)。

[0191] 在下文中,将参照图9描述联合训练第一DNN700和第二DNN 300的方法。

[0192] 图9是用于描述训练第一DNN和第二DNN的方法的示意图。

[0193] 在实施例中,经由AI解码处理将通过AI编码处理对其执行AI编码的原始图像105重构为第三图像145,并且为了保持原始图像105和作为AI解码的结果获得的第三图像145之间的一致性,需要AI编码处理和AI解码处理之间的协调。换句话说,需要在AI解码处理期间重建AI编码处理中丢失的信息,并且在这方面,第一DNN700和第二DNN 300需要被联合训练以类似地考虑信息丢失。

[0194] 为了准确的AI解码,最终,应当减少与将图9所示的第三训练图像804和原始训练图像801进行比较的结果相应的质量损失信息830。因此,质量损失信息830用于训练第一DNN 700和第二DNN 300两者。

[0195] 首先,将描述图9所示的训练处理。

[0196] 在图9中,原始训练图像801是要对其执行AI缩小的图像,并且第一训练图像802是通过对原始训练图像801执行AI缩小而获得的图像。此外,第三训练图像804是通过对第一训练图像802执行AI放大而获得的图像。

[0197] 原始训练图像801包括静止图像或包括多个帧的运动图像。根据实施例,原始训练图像801可包括从静止图像或包括多个帧的运动图像中提取的亮度图像。此外,根据实施例,原始训练图像801可包括从静止图像或包括多个帧的运动图像中提取的分块(patch)图像。当原始训练图像801包括多个帧时,第一训练图像802、第二训练图像和第三训练图像804也各自包括多个帧。当原始训练图像801的多个帧被顺序地输入到第一DNN 700时,可通过第一DNN 700和第二DNN 300顺序地获得第一训练图像802、第二训练图像和第三训练图像804的多个帧。

[0198] 对于第一DNN 700和第二DNN 300的联合训练,原始训练图像801被输入到第一DNN700。输入到第一DNN700的原始训练图像801经由AI缩小被输出为第一训练图像802,并且第一训练图像802被输入到第二DNN 300。作为对第一训练图像802执行AI放大的结果,输出第三训练图像804。

[0199] 参照图9,第一训练图像802被输入到第二DNN 300,并且根据实施例,在对第一训练图像802执行第一编码和第一解码时获得的第二训练图像可被输入到第二DNN 300。为了将第二训练图像输入到第二DNN 300,可使用MPEG-2、H.264、MPEG-4、H.265/HEVC、VC-1、VP8、VP9和AV1中的任何一种编解码器。具体地,MPEG-2、H.264、MPEG-4、H.265/HEVC、VC-1、VP8、VP9和AV1中的任何一种编解码器可用于对第一训练图像802执行第一编码,并且用于对与第一训练图像802相应的图像数据执行第一解码。

[0200] 参照图9,与通过第一DNN700输出第一训练图像802分开,获得通过对原始训练图像801执行传统缩小而获得的简化训练图像803。这里,传统缩小可包括双线性缩放、双三次缩放、兰索斯(Lanczos)缩放或阶梯缩放中的至少一个。

[0201] 为了防止第一图像115的结构特征极大地偏离原始图像105的结构特征,获得简化训练图像803以保留原始训练图像801的结构特征。

[0202] 在执行训练之前,可以根据预定的DNN设置信息来配置第一DNN 700和第二DNN 300。当执行训练时,可确定结构损失信息810、复杂度损失信息820和质量损失信息830。

[0203] 可基于比较简化训练图像803和第一训练图像802的结果来确定结构损失信息810。例如,结构损失信息810可与简化训练图像803的结构信息与第一训练图像802的结构信息之间的差异相应。结构信息可包括可从图像提取的各种特征,诸如图像的亮度、对比度、直方图等。结构损失信息810指示原始训练图像801的多少结构信息被保持在第一训练图像802中。当结构损失信息810小时,第一训练图像802的结构信息类似于原始训练图像801的结构信息。

[0204] 可基于第一训练图像802的空间复杂度来确定复杂度损失信息820。例如,第一训练图像802的总方差值可用作空间复杂度。复杂度损失信息820与通过对第一训练图像802执行第一编码而获得的图像数据的比特率相关。定义的是:当复杂度损失信息820小时,图像数据的比特率低。

[0205] 可基于比较原始训练图像801和第三训练图像804的结果来确定质量损失信息830。质量损失信息830可包括关于原始训练图像801和第三训练图像804之间的差异的L1范数值、L2范数值、结构相似度(SSIM)值、峰值信噪比-人类视觉系统(PSNR-HVS)值、多尺度SSIM(MS-SSIM)值、方差膨胀因子(VIF)值或视频多方法评估融合(VMAF)值中的至少一个。质量损失信息830指示第三训练图像804与原始训练图像801有多相似。当质量损失信息830小时,第三训练图像804更类似于原始训练图像801。

[0206] 参照图9,结构损失信息810、复杂度损失信息820和质量损失信息830用于训练第一DNN 700,并且质量损失信息830用于训练第二DNN 300。换句话说,质量损失信息830用于训练第一DNN700和第二DNN 300两者。

[0207] 第一DNN 700可更新参数,使得基于结构损失信息810、复杂度损失信息820和质量损失信息830确定的最终损失信息被减少或最小化。此外,第二DNN 300可更新参数,使得质量损失信息830被减少或最小化。

[0208] 用于训练第一DNN700和第二DNN 300的最终损失信息可被确定为下面的等式1。

[0209] [等式1]

[0210] $LossDS = a \times \text{结构损失信息} + b \times \text{复杂度损失信息} + c \times \text{质量损失信息}$

[0211] $LossUS = d \times \text{质量损失信息}$

[0212] 在等式1中, $LossDS$ 指示要被减小或最小化以训练第一DNN 700的最终损失信息,并且 $LossUS$ 指示要被减少或最小化以训练第二DNN 300的最终损失信息。此外, a 、 b 、 c 和 d 可以是预定的某些权重。

[0213] 换句话说,第一DNN 700更新参数,使得等式1的 $LossDS$ 减小,并且第二DNN 300更新参数,使得 $LossUS$ 减小。当根据在训练期间推导出的 $LossDS$ 更新第一DNN700的参数时,基于更新的参数获得的第一训练图像802变得与基于尚未更新的参数获得的先前的第一训练图像802不同,因此,第三训练图像804也变得与先前的第三训练图像804不同。当第三训练图像804变得与先前的第三训练图像804不同时,还新确定质量损失信息830,并且第二DNN 300相应地更新参数。当新确定了质量损失信息830时,还新确定 $LossDS$,并且第一DNN 700根据新确定的 $LossDS$ 更新参数。换句话说,第一DNN 700的参数的更新导致第二DNN 300的参数的更新,并且第二DNN 300的参数的更新导致第一DNN 700的参数的更新。因此,因为通

过共享质量损失信息830来联合训练第一DNN700和第二DNN 300,所以可联合优化第一DNN700的参数和第二DNN 300的参数。

[0214] 参照等式1,验证了根据质量损失信息830确定LossUS,但是这仅是示例,并且可基于结构损失信息810和复杂度损失信息820中的至少一个以及质量损失信息830来确定损失。

[0215] 在上文中,已经描述了AI解码设备200的AI放大器234和AI编码设备600的AI缩小器612存储多条DNN设置信息,并且现在将描述训练存储在AI放大器234和AI缩小器612中的多条DNN设置信息中的每一条DNN设置信息的方法。

[0216] 如参照等式1所述,第一DNN 700考虑第一训练图像802的结构信息与原始训练图像801的结构信息之间的相似性(结构损失信息810)、作为对第一训练图像802执行第一编码的结果而获得的图像数据的比特率(复杂度损失信息820)、以及第三训练图像804与原始训练图像801之间的差异(质量损失信息830)来更新参数。

[0217] 具体地,可更新第一DNN700的参数,使得获得具有与原始训练图像801相似的结构信息的第一训练图像802,并且当对第一训练图像802执行第一编码时获得具有小比特率的图像数据,并且同时,对第一训练图像802执行AI放大的第二DNN 300获得与原始训练图像801相似的第三训练图像804。

[0218] 优化第一DNN 700的参数的值的方向可通过调整等式1的权重a、b和c而变化。例如,当权重b被确定为高时,可通过使第三训练图像804的低比特率优先于高质量来更新第一DNN 700的参数的值。此外,当权重c被确定为高时,可通过使第三训练图像804的高质量优先于高比特率或保持原始训练图像801的结构信息来更新第一DNN 700的参数的值。

[0219] 此外,优化第一DNN 700的参数的值的方向可根据用于对第一训练图像802执行第一编码的编解码器的类型而变化。这是因为要输入到第二DNN 300的第二训练图像可根据编解码器的类型而变化。

[0220] 换句话说,可基于权重a、b和c以及用于对第一训练图像802执行第一编码的编解码器的类型来联合更新第一DNN 700的参数的值和第二DNN 300的参数的值。因此,当在将权重a、b和c各自确定为特定值并且将编解码器的类型确定为特定类型之后训练第一DNN700和第二DNN 300时,可联合优化第一DNN 700的参数的值和第二DNN 300的参数的相应值。

[0221] 此外,当在改变权重a、b和c以及编解码器的类型之后训练第一DNN 700和第二DNN 300时,可确定联合优化的第一DNN 700的参数的值和第二DNN 300的参数的值。换句话说,当在改变权重a、b和c的值以及编解码器的类型的情况下训练第一DNN700和第二DNN 300时,可在第一DNN 700和第二DNN 300中确定彼此联合训练的多条DNN设置信息。

[0222] 如上面参照图5所述,第一DNN700和第二DNN 300的多条DNN设置信息可被映射到与第一图像相关的信息。为了设置这样的映射关系,可根据特定比特率对经由特定编解码器从第一DNN 700输出的第一训练图像802执行第一编码,并且可将通过对作为执行第一编码的结果而获得的比特流执行第一解码而获得的第二训练图像输入到第二DNN 300。换句话说,通过在设置环境使得根据特定比特率经由特定编解码器对特定分辨率的第一训练图像802执行第一编码之后训练第一DNN 700和第二DNN 300,可确定映射到第一训练图像802的分辨率、用于对第一训练图像802执行第一编码的编解码器的类型、以及作为对第一训练

图像802执行第一编码的结果而获得的比特流的比特率的DNN设置信息对。通过不同地改变第一训练图像802的分辨率、用于对第一训练图像802执行第一编码的编解码器的类型以及根据第一训练图像802的第一编码获得的比特流的比特率,可确定第一DNN700和第二DNN300的多个DNN设置信息与和第一图像相关的多条信息之间的映射关系。

[0223] 图10是用于描述训练设备对第一DNN和第二DNN的训练处理的示意图。

[0224] 参照图9描述的第一DNN700和第二DNN300的训练可由训练设备1000执行。训练设备1000可包括第一DNN700和第二DNN300。训练设备1000可以是例如AI编码设备600或单独的服务器。作为训练结果获得的第二DNN300的DNN设置信息可存储在AI解码设备200中。

[0225] 参照图10,在操作S840和操作S845,训练设备1000初始地设置第一DNN700和第二DNN300的DNN设置信息。因此,第一DNN700和第二DNN300可根据初始化的DNN设置信息进行操作。DNN设置信息可包括关于包括在第一DNN700和第二DNN300中的卷积层的数量、每个卷积层的滤波器核的数量、每个卷积层的滤波器核的大小或每个滤波器核的参数中的至少一个的信息。

[0226] 在操作S850,训练设备1000将原始训练图像801作为输入提供给第一DNN700。原始训练图像801可包括静止图像或包括在运动图像中的至少一个帧。

[0227] 在操作S855,第一DNN700根据初始化的DNN设置信息处理原始训练图像801,并输出通过对原始训练图像801执行AI缩小而获得的第一训练图像802。在图10中,从第一DNN700输出的第一训练图像802被直接输入到第二DNN300,但是从第一DNN700输出的第一训练图像802可经由训练设备1000被输入到第二DNN300。此外,训练设备1000可经由特定编解码器对第一训练图像802执行第一编码和第一解码,并且将第二训练图像作为输入提供给第二DNN300。

[0228] 在操作S860,第二DNN300根据初始化的DNN设置信息处理第一训练图像802或第二训练图像,并输出通过对第一训练图像802或第二训练图像执行AI放大而获得的第三训练图像804。

[0229] 在操作S865,训练设备1000基于第一训练图像802计算复杂度损失信息820。

[0230] 在操作S870,训练设备1000通过比较简化训练图像803和第一训练图像802来计算结构损失信息810。

[0231] 在操作S875,训练设备1000通过比较原始训练图像801和第三训练图像804来计算质量损失信息830。

[0232] 在操作S880,基于最终损失信息经由反馈传播处理来更新初始DNN设置信息。训练设备1000可基于复杂度损失信息820、结构损失信息810和质量损失信息830来计算用于训练第一DNN700的最终损失信息。

[0233] 在操作S885,第二DNN300基于质量损失信息830或最终损失信息经由反馈传播处理来更新DNN设置信息。训练设备1000可基于质量损失信息830计算用于训练第二DNN300的最终损失信息。

[0234] 然后,训练设备1000、第一DNN700和第二DNN300可重复操作S850至操作S885,直到最终损失信息被最小化,以更新DNN设置信息。此时,在每次重复期间,第一DNN700和第二DNN300根据在先前操作中更新的DNN设置信息进行操作。

[0235] 下面的表1示出了当根据本公开的实施例对原始图像105执行AI编码和AI解码时

以及当经由HEVC对原始图像105执行编码和解码时的效果。

[0236] 表1

	内容	分辨率	帧 的 数 量	信息量（比特率） （Mbps）		主观图像质量分数 （VMAF）	
				HEVC	AI 编码 /AI 解码	HEVC	AI 编码 /AI 解码
[0237]	内容_01	8K (7680×4320)	300 帧	46.3	21.4	94.80	93.54
	内容_02			46.3	21.6	98.05	98.98
	内容_03			46.3	21.7	96.08	96.00
	内容_04			46.1	22.1	86.26	92.00
	内容_05			45.4	22.7	93.42	92.98
	内容_06			46.3	23.0	95.99	95.61
			平均	46.11	22.25	94.10	94.85

[0238] 如表1所示,尽管当对包括300个帧的8K分辨率的内容执行AI编码和AI解码时的主观图像质量高于当经由HEVC执行编码和解码时的主观图像质量,但是比特率仍然降低至少50%。

[0239] 图11是用于对原始图像执行AI缩小的设备和用于对第二图像执行AI放大的设备的示意图。

[0240] 设备20接收原始图像105,并通过使用AI缩小器1124和基于变换的编码器1126将图像数据25和AI数据30提供给设备40。根据实施例,图像数据25与图1的图像数据相应,并且AI数据30与图1的AI数据相应。此外,根据实施例,基于变换的编码器1126与图7的第一编码器614相应,并且AI缩小器1124与图7的AI缩小器612相应。

[0241] 设备40接收AI数据30和图像数据25,并且通过使用基于变换的解码器1146和AI放大器1144来获得第三图像145。根据实施例,基于变换的解码器1146与图2的第一解码器232相应,并且AI放大器1144与图2的AI放大器234相应。

[0242] 根据实施例,设备20包括CPU、存储器和包括指令的计算机程序。计算机程序存储在存储器中。根据实施例,设备20根据CPU对计算机程序的执行来执行将参照图11描述的功能。根据实施例,将参照图11描述的功能由专用硬件芯片和/或CPU执行。

[0243] 根据实施例,设备40包括CPU、存储器和包括指令的计算机程序。计算机程序存储在存储器中。根据实施例,设备40根据CPU对计算机程序的执行来执行将参照图11描述的功能。根据实施例,将参照图11描述的功能由专用硬件芯片和/或CPU执行。

[0244] 在图11中,配置控制器1122接收至少一个输入值10。根据实施例,至少一个输入值10可包括AI缩小器1124和AI放大器1144的目标分辨率差、图像数据25的比特率、图像数据25的比特率类型(例如,可变比特率类型、恒定比特率类型或平均比特率类型)或基于变换的编码器1126的编解码器类型中的至少一个。至少一个输入值10可包括预先存储在设备20中的值或从用户输入的值。

[0245] 配置控制器1122基于接收到的输入值10来控制AI缩小器1124和基于变换的编码

器1126的操作。根据实施例,配置控制器1122根据接收到的输入值10获得用于AI缩小器1124的DNN设置信息,并且用DNN设置信息配置AI缩小器1124。根据实施例,配置控制器1122可将输入值10发送到AI缩小器1124,并且AI缩小器1124可基于输入值10获得用于对原始图像105执行AI缩小的DNN设置信息。根据实施例,配置控制器1122可将附加信息(例如,应用AI缩小的颜色格式(亮度分量、色度分量、红色分量、绿色分量或蓝色分量)信息和高动态范围(HDR)的色调映射信息)连同输入值10一起提供给AI缩小器1124,并且AI缩小器1124可考虑输入值10和附加信息来获得DNN设置信息。根据实施例,配置控制器1122将接收到的输入值10的至少一部分发送到基于变换的编码器1126,并且基于变换的编码器1126经由特定值的比特率、特定类型的比特率和特定编解码器对第一图像115执行第一编码。

[0246] AI缩小器1124接收原始图像105并执行参照图1、图7、图8、图9或图10中的至少一个描述的操作以获得第一图像115。

[0247] 根据实施例,AI数据30被提供给设备40。AI数据30可包括原始图像105和第一图像115之间的分辨率差异信息或与第一图像115相关的信息中的至少一个。可基于输入值10中的目标分辨率差来确定分辨率差异信息,并且可基于目标比特率、比特率类型或编解码器类型中的至少一个来确定与第一图像115相关的信息。根据实施例,AI数据30可包括在AI放大期间使用的参数。AI数据30可被从AI缩小器1124提供给设备40。

[0248] 图像数据25在原始图像105由基于变换的编码器1126处理时被获得,并且被发送到设备40。基于变换的编码器1126可根据MPEG-2、H.264AVC、MPEG-4、H.265/HEVC、VC-1、VP8、VP9或VA1来处理第一图像115。

[0249] 配置控制器1142基于AI数据30控制AI放大器1144的操作。根据实施例,配置控制器1142根据接收到的AI数据30获得用于AI放大器1144的DNN设置信息,并且根据DNN设置信息配置AI放大器1144。根据实施例,配置控制器1142可将接收到的AI数据30发送到AI放大器1144,并且AI放大器1144可基于AI数据30获得用于对第二图像135执行AI放大的DNN设置信息。根据实施例,配置控制器1142可将附加信息(例如,应用AI放大的颜色格式(亮度分量、色度分量、红色分量、绿色分量或蓝色分量)信息和HDR的色调映射信息)连同AI数据30一起提供给AI放大器1144,并且AI放大器1144可考虑AI数据30和附加信息来获得DNN设置信息。根据实施例,AI放大器1144可从配置控制器1142接收AI数据30,从基于变换的解码器1146接收预测模式信息、运动信息或量化参数信息中的至少一个,并且基于AI数据30以及预测模式信息、运动信息和量化参数信息中的至少一个来获得DNN设置信息。

[0250] 基于变换的解码器1146可处理图像数据25以重构第二图像135。基于变换的解码器1146可根据MPEG-2、H.264AVC、MPEG-4、H.265/HEVC、VC-1、VP8、VP9或AV1来处理图像数据25。

[0251] AI放大器1144可通过基于设置的DNN设置信息对从基于变换的解码器1146提供的第二图像135执行AI放大来获得第三图像145。

[0252] AI缩小器1124可包括第一DNN,并且AI放大器1144可包括第二DNN,并且根据实施例,根据参照图9和图10描述的训练方法来训练第一DNN和第二DNN的DNN设置信息。

[0253] 上面已经参照图1至图11描述了根据本公开的实施例的AI编码技术和AI解码技术。在下文中,将参照图12至图21详细描述根据本公开的另一实施例的用于通过更新针对原始图像优化的放大DNN设置信息来执行放大的AI编码技术和AI解码技术。DNN设置信息可

包括卷积层的数量、每个卷积层中的滤波器核的数量、每个滤波器核的大小或关于每个滤波器核的参数中的至少一个。

[0254] 在下文中,包括在“DNN设置信息”中的“参数”是在构成神经网络的每个层的数学运算中使用的值,并且可包括例如当将输入值应用于预定义的运算公式时使用的权重和偏置。权重可以是用于执行与输入值的乘法的值,并且偏置可以是用于执行与作为在输入值和权重之间执行乘法的结果而获得的值的加法的值。此外,参数可以以矩阵形式表示。例如,权重可以是在用于AI放大或AI缩小的卷积层中使用的 3×3 滤波器核的参数,并且偏置可以是在将激活函数应用于在执行与用于AI放大或AI缩小的卷积层中的输入通道(深度)的数量相等的数量的卷积操作并将每个卷积操作的结果值相加在一起之后获得的结果之前添加的 1×1 矩阵形式的参数。下面将参照等式9更详细地描述根据本公开的实施例的权重和偏置。参数是作为训练结果的数集,并且可通过单独的训练数据或由原始图像组成的训练数据来更新。

[0255] 在下文中,“模型压缩”是指用于通过减小人工神经网络模型中的参数的数量和大小以降低人工神经网络模型的复杂度而在保持最高可能精度的同时减少数据量的压缩技术。“模型压缩”的示例包括修剪和量化,但是模型压缩不限于此。

[0256] 图12是根据本公开的实施例的AI编码设备的配置的框图。参照图12,AI编码设备1200包括AI编码器1210和发送器1230。AI编码器1210包括DNN设置信息更新器1212、AI缩小器1216和第一编码器1214。

[0257] 如上所述,AI缩小器1216通过经由第一DNN对原始图像(例如,图1的105)执行AI缩小来获得第一图像(例如,图1的115)。第一图像是分辨率低于原始图像的分辨率的图像。因为上面已经描述了由AI缩小器1216进行的AI缩小,所以下面省略了其冗余描述。AI编码设备1200可包括用于控制AI编码器1210和发送器1230的中央处理器。可选地,AI编码器1210和发送器1230由各自的处理器操作,并且当处理器紧密地一起工作时,AI编码设备1200可被整体地操作。可选地,AI编码器1210和发送器1230可由外部处理器控制。

[0258] AI编码设备1200还可包括一个或多个数据存储器或寄存器,用于存储输入到DNN设置信息更新器1212、AI缩小器1216、第一编码器1214和发送器1230或从DNN设置信息更新器1212、AI缩小器1216、第一编码器1214和发送器1230输出的数据。AI编码设备1200还可包括控制存储在数据存储器中的数据输入/输出的存储器控制器。

[0259] 为了对图像进行编码,AI编码设备1200可通过与内置或外部视频编码处理器交互工作来执行包括预测的图像编码操作。根据本公开的实施例,AI编码设备1200的内置视频编码处理器可与包括图像编码处理模块的CPU或GPU以及单独的处理器一起实现基本图像编码操作。

[0260] DNN设置信息更新器1212可更新与用于执行AI缩小的DNN相应的用于执行AI放大的DNN的DNN设置信息。DNN设置信息更新器1212可通过使用至少一个原始图像作为训练图像,经由放大DNN和用于至少一个原始图像的AI缩小的缩小DNN的联合训练或者放大DNN的单独训练,更新用于执行与至少一个原始图像相应的至少一个第二图像的AI放大的设置信息。

[0261] DNN设置信息更新器1212可不选择用于原始图像的AI缩小的缩小DNN的DNN设置信息,而是通过使用原始图像作为训练图像经由放大DNN和缩小DNN的联合训练来确定缩小

DNN的DNN设置信息和放大DNN的DNN设置信息。

[0262] 在这种情况下,可确定包括在DNN设置信息中的卷积层的数量、每个卷积层中的滤波器核的数量、每个滤波器核的大小。然后,可基于卷积层的数量、每个卷积层中的滤波器核的数量和每个滤波器核的大小来执行联合训练。例如,可基于包括在默认放大DNN设置信息中的最复杂的放大DNN的结构来确定卷积层的数量、每个卷积层中的滤波器核的数量和每个滤波器核的大小。换句话说,卷积层的数量可被确定为小于或等于包括在默认放大DNN设置信息中的最复杂的放大DNN中的卷积层的数量,并且每个卷积层中的滤波器核的数量可被确定为小于或等于包括在默认放大DNN设置信息中的每个最复杂的卷积层中的滤波器核的数量。每个滤波器核的大小可被确定为小于或等于包括在默认放大DNN设置信息中的最复杂的放大DNN中的每个滤波器核的大小。

[0263] 此外,在这种情况下,包括在联合训练所需的缩小/放大DNN的DNN设置信息中的初始参数可以是随机初始化的参数。例如,缩小/放大DNN的初始参数可以是基于特定概率分布随机采样的值。例如,缩小/放大DNN的初始参数之间的权重可以是基于高斯概率分布随机采样的值。此外,例如,可将缩小/放大DNN的初始参数之间的偏置初始化为0。

[0264] 可选地,例如,联合训练所需的缩小/放大DNN的初始参数可以是多个预定的缩小/放大DNN参数中的一个。可选地,联合训练所需的缩小/放大DNN的初始参数可以是默认缩小/放大DNN参数。可选地,联合训练所需的缩小/放大DNN的初始参数可以是在紧接在前的图像组(GOP)单元、紧接在前的帧内随机访问点(IRAP)时段单元、紧接在前的序列单元、预设数量的帧的紧接在前的单元等中使用的缩小/放大DNN的参数。

[0265] 在联合训练期间输入到放大DNN的图像可以通过经由缩小DNN缩小原始图像而生成的第一图像,但是该图像不限于此。输入图像可以通过对在缩小之后获得的第一图像执行第一编码和第一解码而生成的第二图像(例如,图1的135)。

[0266] 此后,DNN设置信息更新器1212可通过使用至少一个原始图像作为训练图像经由放大DNN的单独训练来选择确定的缩小DNN的参数并确定放大DNN的参数。在这种情况下,在单独训练之前可用的放大DNN的初始参数可以是在联合训练期间确定的放大DNN的参数,但是初始参数不限于此。放大DNN中的权重可以是基于特定概率分布随机采样的值,并且放大DNN中的偏置可被确定为0。

[0267] 在单独训练期间输入到放大DNN以用于放大DNN的单独训练的图像可以通过对第一图像执行第一编码和第一解码而生成的第二图像,其中,第一图像是通过经由缩小DNN缩小原始图像而生成的。

[0268] 在这种情况下,可经由模型压缩来最终生成在训练期间或之后生成的放大DNN的参数。

[0269] 本文使用的模型压缩是指用于通过减小人工神经网络模型中的参数的数量和大小以降低人工神经网络模型的复杂度,而在保持最高可能准确度的同时减少数据量的压缩技术。在模型压缩期间可能发生信息丢失。模型压缩的代表性示例包括量化和修剪,但是模型压缩不限于此。量化是指将连续变化分成不连续变化的有限数量的等级并为每个等级分配唯一值的处理。量化的典型示例是经由舍入操作(诸如取整舍入、向下舍入和向上舍入)将非整数值转换为整数值的处理。具体地,量化可包括变换放大DNN的参数信息的数据类型,并且例如,可包括将实数数据类型(诸如float32和double64)变换为整数数据类型(诸

如int16和int8)。

[0270] 修剪是指将具有少量数据的值改变为零的处理。修剪的典型示例包括将DNN参数值设置为0的操作(该DNN参数值小于接近0的特定值并且包括在放大DNN的参数信息中),但是修剪不限于此。

[0271] 可选地,代替执行缩小DNN和放大DNN的联合训练,DNN设置信息更新器1212可选择缩小DNN的参数,并且通过使用至少一个原始图像作为训练图像经由单独训练来确定用于放大至少一个原始图像的放大DNN的参数。例如,单独训练所需的缩小DNN的初始参数可以是多个预定缩小DNN参数中的一个预定缩小DNN参数或默认缩小DNN参数。此外,作为在单独训练之前可用的放大DNN的初始参数,放大DNN中的权重可以是基于特定概率分布随机采样的值,并且放大DNN中的偏置可被确定为0。

[0272] 然而,本公开的实施例不限于此,并且单独训练所需的缩小/放大DNN的初始参数可以是在紧接在前的画面组(GOP)单元、紧接在前的瞬时解码刷新(IDR)画面时段(IRAP)时段单元、紧接在前的序列单元、预设数量的帧的紧接在前的单元等中使用的缩小/放大DNN的参数。

[0273] 在这种情况下,输入到放大DNN的用于单独训练放大DNN的图像可以是通过对第一图像执行第一编码和第一解码而生成的第二图像,其中,第一图像是通过经由缩小DNN缩小原始图像而生成的。

[0274] DNN设置信息更新器1212可通过使用至少一个原始图像作为训练图像经由缩小DNN生成第一训练图像,通过使用第一训练图像经由放大DNN生成第三训练图像,并且基于第一损失信息和第三损失信息以及基于第一训练图像生成的第二损失信息来更新放大DNN和缩小DNN,第一损失信息和第三损失信息各自与将第三训练图像与未经缩小的原始图像进行比较的结果相应。这里,第一损失信息可以是基于将第三训练图像的质量参数与原始图像的质量参数进行比较的结果的信息。换句话说,当第一损失信息具有较小的值时,这可指示原始图像的质量与第三训练图像的质量更相似。第一损失信息可以是用于使用原始图像的专用结构信息进行训练的信息。第一损失信息可与图9的质量损失信息830相应。

[0275] 在这种情况下,质量参数的比较结果可以是关于第三训练图像和原始图像之间的差异的质量比较参数,诸如L1范数值、L2范数值、SSIM值、PSNR-HVS值、MS-SSIM值、VIF值或VMAF值,但是质量比较参数不限于此。可以基于质量比较参数的组合来确定质量参数的比较结果。例如,可基于下面的等式2来确定第一损失信息。在等式2中,i、j、k各自表示用于调整每个质量比较参数的相对重要性的权重。

[0276] [等式2]

[0277] 第一损失信息= $i \times \text{L1范数值} + j \times \text{SSIM值} + i \times \text{L2范数值} + j \times \text{VIF值}$ 、或者 $i \times \text{L1范数值} + j \times \text{L2范数值} + k \times \text{SSIM值}$...

[0278] 第三损失信息可以是基于将第三训练图像的特征相关参数与原始图像的特征相关参数进行比较的结果的信息。换句话说,当第三损失信息具有较小的值时,这可以指示原始图像的特征与第三训练图像的特征更相似。第三损失信息可以是用于使用原始图像的专用细节信息(即,特征相关信息)进行训练的信息。

[0279] 例如,比较特征相关参数的结果可以是每个图像的特征的L1范数值或L2范数值,但是比较特征相关参数的结果不限于此。基于Sobel边缘检测器或Canny边缘检测器确定的

边缘信息可以是特征相关参数。

[0280] 第二损失信息可以是与第一训练图像的空间复杂度有关的信息。例如,第二损失信息可表示第一训练图像的总方差。第二损失越小,通过比特流发送的图像数据量越小。第二损失信息可与图9的复杂度损失信息820相应。

[0281] 因此,第一损失信息和第三损失信息是与失真相关的信息,而第二损失信息是与速率相关的信息,并且当通过考虑第一损失信息、第二损失信息和第三损失信息中的全部来执行训练时,DNN设置信息更新器1212可基于速率失真(RD)成本来确定最佳放大DNN参数。

[0282] DNN设置信息更新器1212可基于下面的等式3来确定损失值 $LOSS_{DS}$,并且基于损失值 $LOSS_{DS}$ 来更新放大DNN和缩小DNN的参数,使得损失值 $LOSS_{DS}$ 减小。

[0283] [等式3]

[0284] $LOSS_{DS}=a \times \text{第一损失信息}+b \times \text{第二损失信息}+c \times \text{第三损失信息}$

[0285] 此外,DNN设置信息更新器1212可通过另外考虑第四损失信息来基于下面的等式4确定损失值 $Loss_{DS}$,并且基于损失值 $Loss_{DS}$ 更新放大DNN和缩小DNN的参数,使得损失值 $Loss_{DS}$ 减小。在这种情况下,第四损失信息可与图9的结构损失信息810相应。

[0286] [等式4]

[0287] $Loss_{DS}=a \times \text{第一损失信息}+b \times \text{第二损失信息}+c \times \text{第三损失信息}+d \times \text{第四损失信息}$

[0288] DNN设置信息更新器1212可选择缩小DNN的参数,对通过经由缩小DNN对至少一个原始图像进行缩小而获得的第一训练图像执行第一编码,通过经由放大DNN对通过对已经经历第一编码的第一训练图像执行第一解码而获得的第二训练图像进行放大来生成第三训练图像,并且基于第一损失信息和第三损失信息更新放大DNN,其中,第一损失信息和第三损失信息各自与将第三训练图像与未经历AI缩小的原始图像进行比较的结果相应。第一损失信息和第三损失信息可以是以上描述的第一损失信息和第三损失信息。

[0289] 例如,DNN设置信息更新器1212可基于下面的等式5来确定损失值 $Loss_{US}$,并且基于损失值 $Loss_{US}$ 来更新放大DNN的参数,使得损失值 $Loss_{US}$ 减小。

[0290] [等式5]

[0291] $Loss_{US}=d \times \text{第一损失信息}+e \times \text{第三损失信息}$

[0292] DNN设置信息更新器1212可生成关于经由缩小DNN和放大DNN的联合训练或放大DNN的单独训练获得的放大DNN的参数的信息。

[0293] 此外,DNN设置信息更新器1212可生成关于卷积层的数量、每个卷积层中的滤波器核的数量或每个滤波器核的大小中的至少一个的信息。

[0294] DNN设置信息更新器1212可对放大DNN的DNN设置信息进行编码。例如,DNN设置信息更新器1212可对关于卷积层的数量、每个卷积层中的滤波器核的数量或每个滤波器核的大小中的至少一个的信息进行编码。此外,DNN设置信息更新器1212可对关于放大DNN的参数的信息进行编码。例如,DNN设置信息更新器1212可对放大DNN中的滤波器核的权重残差或偏置残差进行编码。在这种情况下,权重残差或偏置残差可以是在紧接在前的时段单元中编码的权重或偏置与针对当前时段单元确定的权重或偏置之间的差。

[0295] 在这种情况下,可使用差分脉冲编码调制(DPCM)、游程长度编码(RLC)和霍夫曼编

码来执行编码,但是编码不限于此。可使用各种其它熵编码技术来执行编码。

[0296] 本公开的实施例不限于此,并且DNN设置信息更新器1212可基于上下文模型对放大DNN的DNN设置信息执行熵编码。例如,DNN设置信息更新器1212可以基于放大DNN的用于先前时段单元的DNN设置信息来生成上下文模型,更新放大DNN的用于当前时段单元的DNN设置信息,并且通过使用上下文模型对放大DNN的更新的DNN设置信息进行熵编码。在这种情况下,上下文模型是指表示基于周围环境的各种上下文信息估计的符号的出现概率的模型。

[0297] 每当对当前时段单元的放大DNN中的滤波器核的权重或偏置进行编码时,可更新上下文模型,并且可基于更新的上下文模型对关于当前时段单元的放大DNN中的滤波器核的下一权重或偏置的信息进行熵编码。可针对整个放大DNN确定上下文模型,但是该确定不限于此,并且可针对每个卷积层、针对与卷积层中的一个输出通道相应的至少一个滤波器核、以及针对卷积层中的每个滤波器核来确定上下文模型。基于上下文的熵编码的示例可包括范围编码、算术编码和霍夫曼编码,但是熵编码不限于此。

[0298] DNN设置信息更新器1212可更新放大DNN的所有层中的滤波器核的权重或偏置,并且对关于所有层中的滤波器核的更新的权重或偏置的信息进行编码,但是更新不限于此。DNN设置信息更新器1212可更新一些层中的权重或偏置,并且对关于相应层中的更新的权重或偏置的信息进行编码。可选地,DNN设置信息更新器1212可更新层中的一些滤波器核的权重或偏置,并且对关于一些滤波器核的更新的权重或偏置的信息进行编码。例如,DNN设置信息更新器1212可更新放大DNN的最后一层中的权重或偏置,并且对关于最后一层中的更新的权重或偏置的信息进行编码。DNN设置信息更新器1212可对放大DNN的包括权重残差信息或偏置残差信息的更新的DNN设置信息进行编码。权重残差信息或偏置残差信息指示在更新权重或偏置之前放大DNN中的所有或一些卷积层中的所有或一些滤波器核的权重或偏置与在更新权重或偏置之后放大DNN中的所有或一些卷积层中的所有或一些滤波器核的权重或偏置之间的差。

[0299] DNN设置信息更新器1212可对权重残差执行频率变换,并且对放大DNN的更新的DNN设置信息进行编码,其中,更新的DNN设置信息包括关于已经经历频率变换的权重残差的信息。

[0300] DNN设置信息更新器1212可基于上下文模型对被更新以在当前时段单元中使用的权重或偏置执行熵编码,但是熵编码不限于此。DNN设置信息更新器1212可基于上下文模型对被更新以在当前时段单元中使用的权重或偏置与在紧接在前的时段单元中使用的权重或偏置之间的差进行熵编码。在这种情况下,上下文模型可基于关于在紧接在前的时段单元或紧接在前的时段单元之前的时段单元中的至少一个中被确定用于在放大DNN中使用的权重或偏置之间的差异的信息。

[0301] DNN设置信息更新器1212可结合模型压缩技术(例如,修剪或量化)对权重信息或偏置信息进行熵编码。

[0302] DNN设置信息更新器1212可周期性地更新放大DNN的DNN设置信息。例如,DNN设置信息更新器1212可针对每个GOP、每个IRAP时段(例如,瞬时解码刷新(IDR)画面时段)、每个序列或预设数量的帧的每个单元更新DNN设置信息。

[0303] 例如,AI编码设备1200可在每个IDR画面时段对图像进行编码的同时执行场景变

化检测,并且当在帧中检测到场景变化时,将该帧确定为IDR画面并在每个IDR画面时段对后续图像进行编码。在这种情况下,DNN设置信息更新器1212可针对检测到场景变化的每个IDR图像时段更新DNN设置信息。

[0304] DNN设置信息更新器1212可在必要时更新DNN设置信息。例如,DNN设置信息更新器1212可确定针对每个预设时段是否更新DNN设置信息,并且当DNN设置信息更新器1212确定更新DNN设置信息时更新DNN设置信息。DNN设置信息更新器1212可基于至少一个原始图像中的至少一些来确定放大DNN的DNN设置信息,并且通过基于RD成本将基于确定的DNN设置信息执行AI缩小和AI放大的结果与基于尚未更新的DNN设置信息执行AI缩小和AI放大的结果进行比较来确定是否更新DNN设置信息。

[0305] 在这种情况下,DNN设置信息更新器1212可对关于是否更新DNN设置信息的标志信息进行编码。当DNN设置信息更新器1212确定更新DNN设置信息时,DNN设置信息更新器1212可将标志信息编码为第一值。另一方面,当DNN设置信息更新器1212确定不更新DNN设置信息时,DNN设置信息更新器1212可将标志信息编码为第二值。

[0306] 当DNN设置信息更新器1212确定不更新DNN设置信息时,可不对标志信息进行编码。第一值或第二值可以是0或1,但是该值不限于此。

[0307] DNN设置信息更新器1212可确定是将通过使用AI放大DNN的预定DNN设置信息,还是通过使用AI放大DNN的使用原始图像作为训练图像更新的DNN设置信息来放大第二图像。在本文中,术语“AI放大DNN”和“放大DNN”或“AI缩小DNN”和“缩小DNN”可互换使用。DNN设置信息更新器1212可基于至少一个原始图像中的至少一些来确定DNN设置信息,并且通过基于RD成本比较基于确定的DNN设置信息执行AI缩小和AI放大的结果与基于预定的DNN设置信息执行AI缩小和AI放大的结果,来确定是将基于AI放大DNN的预定DNN设置信息,还是基于AI放大DNN的使用原始图像作为训练图像更新的DNN设置信息来放大第二图像。

[0308] DNN设置信息更新器1212可将标志信息编码为指示是将基于AI放大DNN的预定DNN设置信息,还是基于AI放大DNN的使用原始图像作为训练图像更新的DNN设置信息来放大第二图像。当标志信息具有第一值时,标志信息可指示将通过使用AI放大DNN的预定设置信息来放大第二图像,并且当标志信息具有第二值时,标志信息可指示将通过使用AI放大DNN的使用原始图像作为训练图像更新的DNN设置信息来放大第二图像。当将通过使用AI放大DNN的预定设置信息来放大第二图像时,DNN设置信息更新器1212可不对标志信息进行编码。

[0309] DNN设置信息更新器1212可基于原始图像或第一图像的分辨率或图像数据的比特率中的至少一个来更新AI放大DNN的DNN设置信息。

[0310] 例如,DNN设置信息更新器1212可通过基于原始图像或第一图像的分辨率或图像数据的比特率中的至少一个改变卷积层的数量、每个卷积层中的滤波器核的数量或每个滤波器核的大小中的至少一个来改变DNN结构。AI编码设备1200可通过基于改变的DNN结构执行放大DNN的训练来更新DNN设置参数。例如,可将DNN结构改变为具有3个卷积层、 3×3 大小的滤波器核和4个通道,并且可基于配置的DNN结构来更新包括AI放大DNN的设置参数的DNN设置信息。DNN结构还可被改变为包括4个卷积层、 3×3 大小的滤波器核和8个通道,并且可基于改变的DNN结构来更新AI放大DNN的设置参数。此外,DNN结构可被配置为包括5个卷积层、 5×5 大小的滤波器核和8个通道,并且可基于改变的DNN结构来更新AI放大DNN的设置参数。当通过考虑图像数据的相同比特率以及原始图像和第一图像的相同分辨率来执行放大

DNN和缩小DNN的联合训练时,待编码的图像数据可随着DNN结构改变而相应地变化。然而,在单独训练放大DNN的情况下,图像数据可保持完整,其中,在单独训练放大DNN的期间,即使在DNN结构改变时,也保持选择已经经历缩小的第一图像。此外,在单独训练放大DNN的情况下,当更新DNN设置信息时,可进一步提高第三图像(图1的145)的质量,第三图像是在执行AI放大之后由AI解码设备(图13的1300)获得的图像。

[0311] 此外,DNN设置信息更新器1212可通过以各种方式改变DNN结构来生成用于多个DNN结构的多条DNN设置信息,并且可对用于所述多个DNN结构的多条DNN设置信息进行编码。在这种情况下,所述多个DNN结构可以是与图像数据的相同比特率以及原始图像和第一图像的相同分辨率相应的DNN结构。然而,本公开的实施例不限于此,并且所述多个DNN结构可以是与图像数据的各种比特率以及原始图像和第一图像的各种分辨率相应的DNN结构。

[0312] 尽管已经参照图12描述了DNN设置信息更新器1212对放大DNN的DNN设置信息进行编码,但是本公开的实施例不限于此,并且本领域普通技术人员将理解,DNN设置信息更新器1212可将放大DNN的DNN设置信息发送到数据处理器1232,并且数据处理器1232可执行DNN设置信息更新器1212的上述编码操作。

[0313] 第一编码器1214对由执行缩小的AI缩小器1216获得的第一图像进行编码。

[0314] 如上所述,编码可包括通过预测第一图像来生成预测数据的处理、生成与第一图像和预测数据之间的差相应的残差数据的处理、将空间域中的残差数据变换为频域分量的处理、将已经经历变换为频域分量的残差数据进行量化的处理、以及对量化的残差数据进行熵编码的处理。

[0315] 数据处理器1232对要以预定义格式发送的AI数据或图像数据中的至少一个进行处理。例如,当需要以比特流的形式发送AI数据和图像数据时,数据处理器1232可对AI数据进行处理,使得AI数据以比特流的形式表示,并且经由通信器1234以一个比特流的形式发送AI数据和图像数据。作为另一示例,数据处理器1232可对AI数据进行处理,使得AI数据以比特流的形式表示,并且经由通信器1234分别发送与AI数据相应的比特流和与图像数据相应的比特流。作为另一示例,数据处理器1232可对AI数据进行处理,使得AI数据被表示为帧或包,并且经由通信器1234发送比特流形式的图像数据和帧或包形式的AI数据。

[0316] 通信器1234经由网络发送作为AI编码的结果而生成的AI编码数据。作为AI编码的结果而生成的AI编码数据包括图像数据和AI数据。

[0317] 图像数据和AI数据可经由同构或异构网络发送。

[0318] 图像数据包括作为第一图像的第一编码的结果而生成的数据。图像数据可包括基于第一图像中的像素值生成的数据,诸如作为第一图像与预测数据之间的差的残差数据。此外,图像数据包括在第一图像的第一编码期间使用的信息。例如,图像数据可包括用于对第一图像执行第一编码的模式信息和与量化参数有关的信息。

[0319] AI数据包括使得AI解码设备1300能够根据与第一DNN(缩小DNN)的缩小目标相应的放大目标对第二图像执行AI放大的信息。例如,AI数据可包括原始图像和第一图像之间的差异信息。此外,AI数据可包括与第一图像有关的信息。与第一图像相关的信息可包括关于第一图像的分辨率、作为第一图像的第一编码的结果而生成的图像数据的比特率或在第一图像的第一编码中使用的编解码器类型中的至少一个的信息。

[0320] 图13是根据本公开的实施例的AI解码设备的配置的框图。参照图13,AI解码设备

1300包括接收器1310和AI解码器1330。AI解码器1330包括第一解码器1332、AI放大器1334和DNN设置信息更新器1336。

[0321] AI解码设备1300可包括控制接收器1310和AI解码器1330的中央处理器。可选地,接收器1310和AI解码器1330由各自的处理器操作,并且当处理器紧密地一起工作时,AI解码设备1300可被完整地操作。可选地,接收器1310和AI解码器1330也可由外部处理器控制。

[0322] AI解码设备1300还可包括一个或更多个数据存储器,用于存储输入到DNN设置信息更新器1336、AI放大器1334、第一解码器1332和接收器1310或从DNN设置信息更新器1336、AI放大器1334、第一解码器1332和接收器1310输出的数据。AI解码设备1300还可包括控制存储在数据存储器中的数据的输入/输出的存储器控制器。

[0323] 为了对图像进行解码,AI解码设备1300可通过与内置或外部视频解码处理器交互工作来执行包括预测的图像解码操作。根据本公开的实施例,AI解码设备1300的内置视频解码处理器可与包括图像解码处理模块的CPU或GPU以及单独的处理器一起实现基本图像解码操作。

[0324] 通信器1312经由网络接收包括图像数据和AI数据的AI编码数据。图像数据包括作为第一图像的第一编码的结果而生成的信息,并且AI数据包括放大DNN的DNN设置信息。

[0325] 解析器1314将经由通信器1312接收的AI编码数据划分为图像数据和AI数据,并经由输出器1316将图像数据发送到第一解码器1332并将AI数据发送到AI放大器1334。

[0326] 因为图13的AI解码设备1300的通信器1312、解析器1314和输出器1316的操作分别与参照图2描述的AI解码设备200的通信器212、解析器214和输出器216的操作相应,所以省略了其冗余描述。

[0327] AI放大器1334通过经由第二DNN(放大DNN)对第二图像执行AI放大来获得第三图像(例如,图1的145)。第三图像是具有比第二图像的分辨率更高的分辨率的图像。因为上面已经描述了由AI放大器1334进行的AI放大,所以省略了其冗余描述。

[0328] DNN设置信息更新器1336可被与AI放大器1334分开提供,但是DNN设置信息更新器1336不限于此,并且DNN设置信息更新器1336可包括在AI放大器1334中。

[0329] DNN设置信息更新器1336可基于DNN设置信息来更新与AI缩小DNN相应的AI放大DNN的DNN设置信息。

[0330] DNN设置信息更新器1336可获得与至少一个原始图像到第一图像的AI缩小相关的AI数据。DNN设置信息更新器1336可基于AI数据获得用于执行AI放大的DNN设置信息。DNN设置信息更新器1336可基于根据AI数据获得的DNN设置信息来更新DNN设置信息。

[0331] 在这种情况下,DNN设置信息可以是这样的DNN信息,其中,该DNN信息通过使用至少一个原始图像作为训练图像,经由放大DNN和用于至少一个原始图像的AI缩小的缩小DNN的联合训练、或者放大DNN的单独训练而被更新以用于执行与至少一个原始图像相应的至少一个第二图像的AI放大。

[0332] 上面已经参照图12描述了AI编码设备1200通过使用至少一个原始图像作为训练图像,经由放大DNN和用于至少一个原始图像的AI缩小的缩小DNN的联合训练、或者放大DNN的单独训练来更新用于执行与至少一个原始图像相应的至少一个第二图像的AI放大的DNN设置信息。因为AI解码设备1300使用AI放大DNN的使用原始图像作为训练图像更新的参数,而不是先前通过使用单独提供的图像作为训练图像确定的AI放大DNN的参数,所以AI解码

设备1300可对第二图像最佳地执行AI放大。

[0333] DNN设置信息可包括卷积层的数量、每个卷积层中的滤波器核的数量或关于每个滤波器核的大小的信息中的至少一个。

[0334] 此外,DNN设置信息可包括构成放大DNN的至少一个卷积层中的至少一个滤波器核的权重。本公开的实施例不限于此,并且DNN设置信息可包括至少一个输出通道的偏置。

[0335] DNN设置信息更新器1336可基于AI数据获得第一标志,并且基于第一标志确定是否将通过使用AI放大DNN的预定DNN设置信息,还是通过使用AI放大DNN的使用原始图像作为训练图像更新的DNN设置信息来放大第二图像。第一标志可以是指示将通过使用AI放大DNN的预定DNN设置信息或基于AI放大DNN的使用原始图像作为训练图像更新的DNN设置信息来放大第二图像的标志。当第一标志具有第一值时,第一标志可指示将通过使用如以上参照图9所述训练的AI放大DNN的预定DNN设置信息来放大第二图像,其中,预定DNN设置信息存储在AI解码设备1300中。

[0336] 在这种情况下,DNN设置信息更新器1336可从多个AI放大DNN的多条预定DNN设置信息中确定AI放大DNN的DNN设置信息,并且将AI放大DNN的所述DNN设置信息确定为用于对第二图像执行AI放大的DNN设置信息。

[0337] 当第一标志具有第二值时,第一标志可指示将基于AI放大DNN的使用原始图像作为训练图像更新的DNN设置信息来放大第二图像。

[0338] 在该配置中,DNN设置信息更新器1336可将AI放大DNN的使用原始图像作为训练图像更新的DNN设置信息确定为用于对第二图像执行AI放大的DNN设置信息。

[0339] 第一值或第二值可以是0或1,但是该值不限于此。

[0340] 在该配置中,可经由模型压缩生成AI放大DNN的使用原始图像作为训练图像更新的DNN设置信息。AI编码设备1200可对通过使用原始图像作为训练图像经由缩小DNN和放大DNN的联合训练或放大DNN的单独训练获得的放大DNN的DNN设置信息进行编码,并且DNN设置信息更新器1336可获得关于编码的DNN设置信息的信息。

[0341] 例如,DNN设置信息更新器1336可获得关于放大DNN的结构的信息(例如,卷积层的数量、每个卷积层中的滤波器核的数量以及关于每个滤波器核的大小的信息)。DNN设置信息更新器1336可获得关于放大DNN中的卷积层的数量、每个卷积层中的滤波器核的数量和每个滤波器核的大小的信息,并且基于获得的信息更新放大DNN中的卷积层的数量、每个卷积层中的滤波器核的数量和每个滤波器核的大小。

[0342] DNN设置信息更新器1336可获得用于多个放大DNN结构的多条DNN设置信息,基于图像数据的比特率或原始图像或第一图像的分辨率中的至少一个来从所述多个放大DNN结构中确定DNN结构,并且基于与确定的DNN结构相应的DNN设置信息来更新放大DNN中的卷积层的数量、每个卷积层中的滤波器核的数量和每个滤波器核的大小。

[0343] 可选地,DNN设置信息更新器1336可获得用于多个DNN结构的多条DNN设置信息,基于至少一个可用DNN结构从所述多个DNN结构中确定一个DNN结构,并且基于与所确定的DNN结构相应的DNN设置信息来更新放大DNN中的卷积层的数量、每个卷积层中的滤波器核的数量和每个滤波器核的大小。在这种情况下,可根据在AI解码设备1300中的实现AI放大器1334的处理器的能力来确定至少一个可用DNN结构。例如,DNN设置信息更新器1336可获得关于放大DNN中的权重残差或偏置残差的信息。在这种情况下,权重残差或偏置残差可表示

用于在紧接在前的时段单元中经由放大DNN执行放大的权重或偏置与用于在当前时段单元中经由放大DNN执行放大的权重或偏置之间的差。在这种情况下,关于放大DNN中的权重残差或偏置残差的编码信息是使用各种熵编码技术(诸如差分脉冲编码调制(DPCM)、游程长度编码(RLC)和霍夫曼编码)编码的信息。DNN设置信息更新器1336可通过执行逆编码操作来获得关于放大DNN中的权重残差或偏置残差的信息。

[0344] DNN设置信息更新器1336可通过将放大DNN中的权重残差或偏置残差与用于在紧接在前的时段单元中经由放大DNN执行放大的权重或偏置相加,来更新用于在当前时段单元中经由放大DNN执行放大的权重或偏置。本公开的实施例不限于此,并且DNN设置信息更新器1336可基于上下文模型对编码的DNN设置信息执行熵解码。例如,DNN设置信息更新器1336可基于用于先前时段单元的放大DNN的DNN设置信息来生成上下文模型,并且通过使用生成的上下文模型对放大DNN的编码的DNN设置信息执行熵解码。

[0345] 每当对用于当前时段单元的放大DNN中的滤波器核的权重或偏置进行解码时,可更新上下文模型,并且可基于更新的上下文模型对关于用于当前时段单元的放大DNN中的滤波器核的下一权重或偏置的信息进行熵解码。基于上下文的熵解码的示例可包括范围解码、算术解码和霍夫曼解码,但是熵解码不限于此。可针对整个放大DNN确定上下文模型,但是上下文模型不限于此,并且可针对每个卷积层、针对与卷积层中的一个输出通道相应的至少一个滤波器核、以及针对卷积层中的每个滤波器核来确定上下文模型。

[0346] DNN设置信息更新器1336可对用于放大DNN的所有层的权重信息或偏置信息进行解码,并且基于解码的权重信息或偏置信息更新所有层中的权重或偏置,但是DNN设置信息更新器1336不限于此。DNN设置信息更新器1336可对一些层的权重信息或偏置信息进行解码,并更新相应层中的权重或偏置。可选地,DNN设置信息更新器1336可对层中的一些滤波器核的权重信息或偏置信息进行解码,并更新该层中的相应滤波器核中的权重或偏置。例如,DNN设置信息更新器1336可对放大DNN的最后一层中的权重信息/偏置信息进行解码,并更新最后一层中的权重或偏置。

[0347] DNN设置信息更新器1336可对权重残差信息或偏置残差信息进行解码,然后基于解码的权重残差信息或偏置残差信息更新所有或一些卷积层中的所有或一些滤波器核的权重或偏置,其中,所述权重残差信息或偏置残差信息指示更新权重或偏置之前的放大DNN中的所有或一些卷积层中的所有或一些滤波器核的权重或偏置与更新权重或偏置之后的放大DNN中的所有或一些卷积层中的所有或一些滤波器核的权重或偏置之间的差。DNN设置信息更新器1336可对权重残差信息进行解码,并通过对解码的权重残差信息执行逆频率变换来获得权重残差。

[0348] 虽然DNN设置信息更新器1336可基于上下文模型对关于用于对第二图像执行AI放大的放大DNN中的权重或偏置的信息执行熵解码,但是DNN设置信息更新器1336不限于此。DNN设置信息更新器1336可基于上下文模型对针对当前时段单元更新的权重或偏置与用于在紧接在前的时段单元中进行放大的权重或偏置之间的差进行熵解码。在这种情况下,上下文模型可基于关于用于放大紧接在前的图像或紧接在前的图像之前的图像中的至少一个而确定的权重或偏置之间的差异的信息。可通过结合模型压缩技术(例如,修剪或量化)对权重信息或偏置信息进行熵编码来生成编码的DNN设置信息。

[0349] DNN设置信息更新器1336可周期性地更新放大DNN的DNN设置信息。例如,DNN设置

信息更新器1336可针对每个GOP、每个IRAP时段(例如,每个IDR画面时段)、每个序列或预设数量的帧的每个单元对编码的DNN设置信息进行解码,并且基于解码的DNN设置信息来更新放大DNN的DNN设置信息。

[0350] DNN设置信息更新器1336可在必要时更新DNN设置信息。例如,DNN设置信息更新器1336可对针对每个预设时段编码的第一标志信息进行解码,基于解码的第一标志信息确定是否更新放大信息的DNN设置信息,并且当DNN设置信息更新器1336确定更新DNN设置信息时更新DNN设置信息。在这种情况下,编码的第一标志信息可以是关于是否要更新放大DNN的DNN设置信息的标志信息。

[0351] 当第一标志信息具有第一值时,DNN设置信息更新器1336可确定不更新放大DNN的DNN设置信息。

[0352] 当第一标志信息具有第二值时,DNN设置信息更新器1336可确定更新放大DNN的DNN设置信息。

[0353] 第一值或第二值可以是0或1,但是该值不限于此。

[0354] DNN设置信息更新器1336可针对每个预设时段解析第一标志信息,并且当针对每个预设时段未解析出第一标志信息时,确定不更新放大DNN的DNN设置信息。

[0355] DNN设置信息更新器1336可对编码的第二标志信息进行解码,并且基于解码的第二标志信息来确定是将通过使用AI放大DNN的预定DNN设置信息还是通过使用AI放大DNN的使用原始图像作为训练图像更新的DNN设置信息来放大第二图像。在这种情况下,编码的第二标志信息可以是指示是将基于AI放大DNN的预定DNN设置信息还是基于AI放大DNN的使用原始图像作为训练图像更新的DNN设置信息来放大第二图像的标志信息。

[0356] 当第二标志信息具有第一值时,DNN设置信息更新器1336可通过使用放大DNN的预定DNN设置信息来确定要放大第二图像。

[0357] 当第二标志信息具有第二值时,DNN设置信息更新器1336可确定将通过使用放大DNN的使用原始图像作为训练图像更新的DNN设置信息来放大第二图像。

[0358] DNN设置信息更新器1336可针对每个预设时段解析第二标志信息,并且当针对每个预设时段未解析出第二标志信息时,确定将通过使用放大DNN的预定DNN设置信息来放大第二图像。

[0359] 尽管上面已经描述了DNN设置信息更新器1336对放大DNN的编码的DNN设置信息进行解码,但是本公开的实施例不限于此,并且本领域普通技术人员将理解,解析器1314可执行DNN设置信息更新器1336的上述解码操作,并经由输出器1316将解码的DNN设置信息发送到DNN设置信息更新器1336。

[0360] AI放大器1334可通过经由根据DNN设置信息操作的放大DNN对第二图像执行放大来生成第三图像。可从DNN设置信息更新器1336接收DNN设置信息。

[0361] 图14a是根据本公开的实施例的AI编码方法的流程图。

[0362] AI编码设备1200生成用于执行与AI缩小相应的AI放大的放大DNN设置信息,该放大DNN设置信息是经由使用原始图像的DNN联合或单独训练而更新的DNN信息(操作S1405)。下面将参照图16a、图16b和图17a详细描述由AI编码设备1200执行的生成使用原始图像经由DNN联合或单独训练更新的DNN设置信息的处理。

[0363] AI编码设备1200可获得通过对第一图像执行第一编码而生成的图像数据,其中,

第一图像是通过对原始图像执行AI缩小而获得的(操作S1410)。

[0364] AI编码设备1200发送图像数据和包括放大DNN设置信息的AI数据(操作S1415)。

[0365] 图14b是根据本公开的实施例的经由基于原始图像的DNN联合训练的AI编码方法的流程图。

[0366] AI编码设备1200可通过经由缩小DNN对原始图像执行AI缩小来获得第一训练图像,并且通过经由放大DNN对第一训练图像执行AI放大来获得第三训练图像(操作S1420)。

[0367] AI编码设备1200可基于原始图像、第一训练图像和第三训练图像经由DNN联合训练来更新DNN信息(操作S1425)。

[0368] AI编码设备1200可通过基于更新的DNN信息经由缩小DNN对原始图像执行AI缩小来获得第一图像,并且通过经由放大DNN对第二训练图像执行放大来获得第三训练图像,其中,第二训练图像是通过对第一训练图像执行第一编码和第一解码而生成的(操作S1430)。

[0369] AI编码设备1200可通过经由基于原始图像和第三训练图像的对放大DNN的训练更新DNN信息,来生成用于执行与AI缩小相应的AI放大的放大DNN设置信息(操作S1435)。

[0370] AI编码设备1200可获得通过对第一图像执行第一编码而生成的图像数据,其中,第一图像是通过对原始图像执行缩小而生成的(操作S1440)。在这种情况下,可基于经由使用原始图像的DNN联合训练设置的DNN信息来执行AI缩小。

[0371] AI编码设备1200可生成包括图像数据和包含放大DNN设置信息的AI数据的AI编码数据(操作S1445)。

[0372] 下面将参照图16a和图16b详细描述由AI编码设备1200执行的生成经由使用原始图像的DNN联合训练更新的DNN设置信息的处理。

[0373] 图14c是根据本公开的实施例的经由基于原始图像的DNN单独训练的AI编码方法的流程图。

[0374] AI编码设备1200可通过经由缩小DNN对原始图像执行AI缩小来获得第一训练图像,并且通过经由放大DNN对通过对第一训练图像执行第一编码和第一解码而生成的第二图像执行AI放大来获得第三训练图像(操作1450)。

[0375] AI编码设备1200可经由基于原始图像和第三训练图像的对放大DNN的单独训练来更新DNN信息,并且生成用于执行与AI缩小相应的AI放大的放大DNN设置信息(操作S1455)。

[0376] AI编码设备1200可获得通过对第一图像执行第一编码而生成的图像数据,其中,第一图像是通过对原始图像执行缩小而生成的(操作S1460)。

[0377] AI编码设备1200可生成包括图像数据和包含放大DNN设置信息的AI数据的AI编码数据(操作S1465)。

[0378] 下面将参照图17a详细描述由AI编码设备1200执行的生成经由使用原始图像的DNN单独训练更新的DNN设置信息的处理。

[0379] 图15是根据本公开的实施例的AI解码方法的流程图。

[0380] AI解码设备1300获得图像数据和AI数据(操作S1510)。

[0381] AI解码设备1300基于图像数据获得第二图像(操作S1520)。AI解码设备1300可通过对图像数据执行第一解码来获得第二图像。

[0382] AI解码设备1300从AI数据获得用于执行AI放大的放大DNN设置信息,该放大DNN设置信息是经由使用原始图像的DNN联合或单独训练而更新的DNN信息(操作S1530)。

[0383] 下面将参照图16c和图17b详细描述由AI解码设备1300执行的从AI数据获得用于执行AI放大的放大DNN设置信息的处理,该放大DNN设置信息是经由使用原始图像的DNN联合或单独训练而更新的DNN信息。

[0384] AI解码设备1300通过基于放大DNN设置信息对第二图像执行放大来获得第三图像(操作S1540)。

[0385] 图16a是用于描述根据本公开的另一实施例的作为联合训练的第一阶段的由AI编码设备1200执行的通过使用原始图像作为训练图像经由缩小DNN和放大DNN的联合训练来确定缩小DNN和放大DNN的多条最佳DNN设置信息的处理的示意图。

[0386] 参照图16a,AI编码设备1200可生成第一训练图像1602,作为通过使用原始图像1601作为训练图像基于第一DNN700执行AI缩小的结果。在这种情况下,原始图像1601可以是要被编码为一个或更多个帧的当前GOP单元、当前IRAP时段单元、当前序列单元或预定数量的帧的当前单元中的帧,但是原始图像1601不限于此。此外,第一DNN700的初始DNN设置信息和第二DNN 300的初始DNN设置信息可分别用于在紧接在前的时段单元中对图像进行缩小和放大。可选地,第一DNN700或第二DNN 300的初始DNN设置信息可以是多条预定DNN设置信息中的一条或设置为默认的DNN设置信息。

[0387] 然而,本公开的实施例不限于此,并且第一DNN700和第二DNN300中的初始权重可以是利用遵循特定概率分布的随机采样获得的值。第一DNN700和第二DNN300中的初始偏置可被确定为零。

[0388] AI编码设备1200可基于第一训练图像1602的复杂度来生成第二损失信息1620。在这种情况下,第二损失信息1620可与上面参照图12描述的第二损失信息相应。

[0389] AI编码设备1200可生成第三训练图像1604,作为通过使用第一训练图像1602基于第二DNN 300执行AI放大的结果。

[0390] AI编码设备1200可生成第一损失信息1610和第三损失信息1630,第一损失信息1610和第三损失信息1630各自与将与原始图像1601有关的信息和与第三训练图像1604有关的信息进行比较的结果相应。在这种情况下,第一损失信息1610和第三损失信息1630可分别与上面参照图12描述的第一损失信息和第三损失信息相应。

[0391] AI编码设备1200可根据下面的等式6基于第一损失信息1610、第二损失信息1620和第三损失信息1630来确定损失值 $Loss_{DS}$ 。

[0392] [等式6]

[0393] $Loss_{DS} = a \times \text{第一损失信息} + b \times \text{第二损失信息} + c \times \text{第三损失信息}$

[0394] AI编码设备1200可基于损失值 $Loss_{DS}$ 更新第一DNN700和第二DNN 300的多条DNN设置信息。

[0395] AI编码设备1200可通过迭代地重复上述处理来执行第一DNN700和第二DNN 300的联合训练,并且经由基于损失值 $Loss_{DS}$ 的联合训练来确定第一DNN 700和第二DNN 300的最佳DNN设置信息。

[0396] 图16b是用于描述根据本公开的实施例的作为联合训练的第二阶段的AI编码设备1200选择根据图16a所示的处理确定的缩小DNN的DNN设置信息,经由放大DNN的单独训练确定放大DNN的最佳DNN设置信息,并通过比特流发送放大DNN的最佳DNN设置信息的处理的示意图。

[0397] 参照图16b, AI编码设备1200选择经由如上参照图16a所述的第一DNN 700和第二DNN 300的联合训练确定的最佳DNN设置信息。在这种情况下, 可基于经由第一DNN700和第二DNN 300的联合训练确定的第二DNN 300的最佳DNN设置信息来确定第二DNN 300的初始DNN设置信息, 但是初始DNN设置信息不限于此。第二DNN 300中的初始权重可以是基于特定概率分布随机采样的值, 并且第二DNN 300中的初始偏置可以是0。

[0398] AI编码设备1200可生成第一训练图像1602, 作为通过使用原始图像1601作为训练图像基于第一DNN700执行AI缩小的结果。AI编码设备1200可对第一训练图像1602执行第一编码1603和第一解码1605。AI编码设备1200可通过基于第二DNN 300对经由第一解码1605生成的图像执行AI放大来生成第三训练图像1604。AI编码设备1200可生成第一损失信息1610和第三损失信息1630, 第一损失信息1610和第三损失信息1630各自与将与原始图像1601有关的图像信息和与第三训练图像1604有关的图像信息进行比较的结果相应。

[0399] AI编码设备1200可根据下面的等式7基于第一损失信息1610和第三损失信息1630来确定损失值 $Loss_{US}$ 。

[0400] [等式7]

[0401] $Loss_{US} = d \times \text{第一损失信息} + e \times \text{第三损失信息}$

[0402] AI编码设备1200可基于损失值 $Loss_{US}$ 更新第二DNN 300的DNN设置信息。在这种情况下, 可经由模型压缩来更新第二DNN 300的DNN设置信息。

[0403] AI编码设备1200可通过重复上述处理来执行第二DNN 300的单独训练, 并且经由基于损失值 $Loss_{US}$ 的单独训练来确定第二DNN 300的最佳DNN设置信息。

[0404] AI编码设备1200可对所确定的第二DNN 300的最佳DNN设置信息进行编码, 并且AI编码设备1200的发送器1230可生成包括编码的DNN设置信息和经由第一编码1603生成的图像数据的AI编码数据。在这种情况下, DNN设置信息可包括第二DNN 300的权重残差信息或偏置残差信息。此外, AI编码数据可以以比特流的形式表示。

[0405] 图16c是用于描述根据本公开的实施例的AI解码设备1300基于包括在AI编码数据中的放大DNN的DNN设置信息经由放大DNN对第二图像执行AI放大的处理的示图。

[0406] 如上面参照图16b所述, AI编码设备1200可生成AI编码数据。

[0407] 参照图16c, AI解码设备1300的接收器1310接收AI编码数据。

[0408] 如上文所描述, AI编码数据可表示为比特流。AI解码设备1300的接收器1310从AI编码数据获得图像数据和DNN设置信息。AI解码设备1300的DNN设置信息更新器1336可基于DNN设置信息来配置第二DNN 300。AI解码设备1300可从AI编码数据获得图像数据, 并通过对图像数据执行第一解码1660来生成第二图像1670。

[0409] AI解码设备1300可通过基于DNN设置信息经由第二DNN 300放大第二图像1670来生成第三图像1680。在这种情况下, 因为DNN设置信息是如上参照图16a和16b所述的使用原始图像作为训练图像优化的第二DNN 300的DNN设置信息, 所以与通过基于第二DNN的基于由AI编码设备1200和AI解码设备1300单独提供的训练图像预先确定的DNN设置信息放大第二图像而生成的第三图像的质量相比, 第三图像1680的质量可得到改善。

[0410] 图17a是用于描述根据本公开的实施例的AI编码设备选择缩小DNN的DNN设置信息、经由放大DNN的单独训练确定放大DNN的最佳DNN设置信息、并以比特流发送放大DNN的最佳DNN设置信息的处理的示图。

[0411] 参照图17a, AI编码设备1200选择第一DNN700的权重。在这种情况下, 第一DNN700的所选择的DNN设置信息可以是用于在紧接在前的时段单元中缩小原始图像的设置信息。然而, 本公开的实施例不限于此, 并且第一DNN700的DNN设置信息可以是多条预定DNN设置信息中的一条或被设置为默认的DNN设置信息。

[0412] 此外, 第二DNN 300的初始DNN设置信息可以是用于在紧接在前的单元中放大原始图像的DNN设置信息。然而, 本公开的实施例不限于此, 并且第二DNN 300中的初始权重可以是基于特定概率分布随机采样的值, 并且第二DNN 300中的初始偏置可以是0。

[0413] AI编码设备1200可生成第一训练图像1702, 作为通过使用原始图像1701作为训练图像基于第一DNN700执行AI缩小的结果。AI编码设备1200可对第一训练图像1702执行第一编码1703和第一解码1705。AI编码设备1200可通过基于第二DNN 300对经由第一解码1705生成的图像执行AI放大来生成第三训练图像1704。AI编码设备1200可生成第一损失信息1710和第三损失信息1730, 第一损失信息1710和第三损失信息1730各自与将与原始图像1701相关的图像信息和与第三训练图像1704相关的图像信息进行比较的结果相应。

[0414] AI编码设备1200可根据下面的等式8基于第一损失信息1710和第三损失信息1730来确定损失值 $Loss_{US}$ 。

[0415] [等式8]

[0416] $Loss_{US} = d \times \text{第一损失信息} + e \times \text{第三损失信息}$

[0417] AI编码设备1200可基于损失值 $Loss_{US}$ 更新第二DNN 300的DNN设置信息。在这种情况下, 可经由模型压缩来更新第二DNN 300的DNN设置信息。

[0418] AI编码设备1200可通过重复上述处理来执行第二DNN 300的单独训练, 并且经由基于损失值 $Loss_{US}$ 的单独训练来确定第二DNN 300的最佳DNN设置信息。

[0419] AI编码设备1200可对所确定的第二DNN 300的最佳DNN设置信息进行编码, 并且AI编码设备1200的发送器1230可生成包括编码的DNN设置信息和经由第一编码1703生成的图像数据的AI编码数据。在这种情况下, DNN设置信息可以包括第二DNN 300的权重残差信息或偏置残差信息。

[0420] 图17b是用于描述根据本公开的实施例的由AI解码设备1300执行的基于包括在AI编码数据中的放大DNN的DNN设置信息经由放大DNN对第二图像执行AI放大的处理的示图。

[0421] 如上面参照图17a所述, AI编码设备1200可生成AI编码数据。

[0422] 参照图17b, AI解码设备1300的接收器1310接收AI编码数据。如上文所描述, AI编码数据可表示为比特流。AI解码设备1300的接收器1310从AI编码数据获得图像数据和DNN设置信息。AI解码设备1300的DNN设置信息更新器1336可基于DNN设置信息确定第二DNN 300。

[0423] AI解码设备1300可通过对图像数据执行第一解码1760来生成第二图像1770。AI解码设备1300可通过基于DNN设置信息经由第二DNN 300放大第二图像1770来生成第三图像1780。在这种情况下, 因为DNN设置信息是如上面参照图17a所述的使用原始图像作为训练图像优化的第二DNN 300的DNN设置信息, 所以与通过基于第二DNN的基于单独提供的训练图像预先确定的DNN设置信息放大第二图像而生成的第三图像的质量相比, 第三图像1780的质量可被改善。

[0424] 图18是根据本公开实施例的由AI解码设备1300执行的通过更新基于从AI编码数

据获得的标志预先确定的放大DNN的DNN设置信息来放大第二图像、或者通过更新放大DNN的针对原始图像优化的DNN设置信息来放大第二图像的处理的流程图。

[0425] 参照图18, AI解码设备1300可基于第一标志的值来确定第一标志是否指示是否使用放大DNN的预定DNN设置信息(操作S1805)。

[0426] 当第一标志指示将使用放大DNN的预定DNN设置信息时, AI解码设备1300可通过使用放大DNN的预定DNN设置信息来放大图像(操作S1810)。在这种情况下, 可将放大DNN的预定DNN设置信息存储在AI解码设备1300中。

[0427] AI解码设备1300可周期性地获得第二标志, 并且基于第二标志的值来确定第二标志是否指示是否使用放大DNN的多条预定DNN设置信息中的除了先前DNN设置信息之外的放大DNN的DNN设置信息(操作S1815)。

[0428] 当第二标志指示将使用放大DNN的多条预定DNN设置信息中的除了先前DNN设置信息之外的放大DNN的DNN设置信息时, AI解码设备1300可将用于放大DNN的DNN设置信息改变为多条预定DNN设置信息中的除了在先前时段中使用的DNN设置信息之外的放大DNN的DNN设置信息(操作S1820)。

[0429] 当第二标志指示将不使用放大DNN的多条预定DNN设置信息中的除了先前DNN设置信息之外的放大DNN的DNN设置信息时, AI解码设备1300可确定在先前时段中使用的DNN设置信息被保持作为将用于放大DNN的DNN设置信息(操作S1825)。

[0430] AI解码设备1300可通过使用放大DNN的预定DNN设置信息来放大图像(操作S1830)。在这种情况下, 放大DNN的预定DNN设置信息可以是在操作S1820中改变的DNN设置信息或在操作S1825中保持的DNN设置信息。

[0431] 当第一标志指示将不使用放大DNN的预定DNN设置信息时, AI解码设备1300可通过使用放大DNN的针对原始图像优化的DNN设置信息来放大图像(操作S1835)。上面已经参照图16a至图17b提供了关于放大DNN的针对原始图像优化的DNN设置信息的描述, 因此, 下面不再重复冗余描述。

[0432] AI解码设备1300可周期性地获得第三标志, 并且基于第三标志的值来确定第三标志是否指示是否用放大DNN的针对当前时段优化的DNN设置信息来更新DNN设置信息(操作S1840)。

[0433] 当第三标志指示将更新DNN设置信息时, AI解码设备1300可将要用于放大DNN的DNN设置信息改变为针对原始图像的当前时段优化的放大DNN的DNN设置信息(操作S1845)。

[0434] 当第三标志指示将不更新DNN设置信息时, AI解码设备1300可确定保持在先前时段中使用的放大DNN的DNN设置信息(操作S1850)。AI解码设备1300可通过使用放大DNN的针对原始图像优化的DNN设置信息来放大图像(操作S1855)。在这种情况下, 针对原始图像优化的DNN设置信息可以是在S1845中改变的DNN设置信息或在S1850中保持的DNN设置信息。

[0435] 图19a示出了根据本公开的实施例的默认权重和偏置、放大DNN中的针对原始图像优化的权重和偏置以及放大DNN中的权重差和偏置差的示例。

[0436] 参照图19a, AI编码设备1200可确定放大DNN的卷积层中的滤波器核的大小为 3×3 的默认权重矩阵1910和大小为 1×1 的默认偏置矩阵1920。在这种情况下, 默认权重矩阵1910和默认偏置矩阵1920可以是通过在AI编码设备1200中使用单独的训练图像预先确定的权重和偏置, 但是权重和偏置不限于此。权重和偏置可以是在紧接在前的时段单元中通

过使用原始图像作为训练图像而确定的权重和偏置。

[0437] 根据下面的等式9,可以使用输入矩阵 $X_{L,t,d_{in}}$ 、默认权重矩阵 $W_{L,t,d_{in},d_{out}}$ (例如,1910)和默认偏置矩阵 $B_{L,t,d_{out}}$ (例如,1920)来生成输出矩阵 $Y_{L,t,d_{out}}$ 。在等式9中,activation()和convolution()分别表示激活函数和卷积运算函数, L 、 d_{in} 和 d_{out} 分别表示卷积层、层 L 的输入通道(深度)和层 L 的输出通道,并且 t 表示时间。

[0438] [等式9]

$$[0439] \quad Y_{L,t,d_{out}} = \text{activation}\left(\sum_{d_{in}} \text{convolution}(X_{L,t,d_{in}}, W_{L,t,d_{in},d_{out}}) + B_{L,t,d_{out}}\right)$$

[0440] 参照图19a, AI编码设备1200可通过使用当前原始图像作为训练图像来确定大小为 3×3 的最优权重矩阵1930和大小为 1×1 的最优偏置矩阵1940。最佳权重矩阵1930和最佳偏置矩阵1940中的元素的值可分别与默认权重矩阵1910和默认偏置矩阵1920的元素的值不同。

[0441] AI编码设备1200可基于上下文模型对最优权重矩阵1930和最优偏置矩阵1940进行熵编码以进行发送。

[0442] AI编码设备1200可基于默认权重矩阵1910、默认偏置矩阵1920、最优权重矩阵1930和最优偏置矩阵1940来确定大小为 3×3 的权重差矩阵1950和大小为 1×1 的偏置差矩阵1960。

[0443] 每个权重差矩阵1950中的元素的值可以是默认权重矩阵1910中的相应的一个默认权重矩阵中的元素的值与相应的最优权重矩阵1930中的元素的值之间的差,并且每个偏置差矩阵1960中的元素的值可以是默认偏置矩阵1920中的相应的一个默认偏置矩阵中的元素的值与相应的最优偏置矩阵1940中的元素的值之间的差。

[0444] AI编码设备1200可对权重差矩阵1950和偏置差矩阵1960进行熵编码以进行发送。

[0445] 图19b示出了根据本公开的实施例的放大DNN中的针对原始图像优化的权重和偏置以及经由量化和修剪确定的放大DNN中的权重和偏置的示例。

[0446] 参照图19b, AI编码设备1200可通过使用当前原始图像作为训练图像来确定大小为 3×3 的最优权重矩阵1970和大小为 1×1 的最优偏置矩阵1975。

[0447] AI编码设备1200可通过对最优权重矩阵1970执行修剪来确定大小为 3×3 的权重矩阵1980。换句话说,参照图19b, AI编码设备1200可通过将最佳权重矩阵1970中的小于0.02的绝对值的元素的所有值改变为零来确定权重矩阵1980。然而,用于将元素的值改变为零的绝对值不限于0.02,并且本领域普通技术人员将理解,小于接近0的值的元素的值可以改变为0。

[0448] AI编码设备1200可通过分别对权重矩阵1980和最优偏置矩阵1975执行量化来确定大小为 3×3 的权重矩阵1990和大小为 1×1 的偏置矩阵1995。换句话说,参照图19b, AI编码设备1200可通过执行权重矩阵1980中的元素的值乘以128的乘法运算和对元素乘积的舍入运算来确定权重矩阵1990,并且通过对最优偏置矩阵1975中的元素的值执行舍入运算来确定偏置矩阵1995。然而,本公开的实施例不限于此,并且本领域普通技术人员将理解,可以理解,可对权重/偏置矩阵执行本公开的各种实施例的量化。

[0449] 图20a是用于描述根据本公开的实施例的对放大DNN中的针对原始图像优化的权重进行编码的方法的示意图。

[0450] 参照图20a, AI编码设备1200可确定 4×4 滤波器核的权重差矩阵2010。在这种情况下, 权重差矩阵2010中的权重差值是放大DNN中的针对原始图像优化的权重差值, 如上面参照图19a所述, 并且下面将省略其详细描述。

[0451] AI编码设备1200可通过对 4×4 滤波器核的权重差矩阵2010执行修剪来确定权重矩阵2015。

[0452] AI编码设备1200可通过对权重矩阵2015执行量化来确定权重矩阵2020。上面已经参照图19b提供了关于修剪和量化的描述, 因此将不再重复。

[0453] 然后, AI编码设备1200可通过对权重矩阵2020中的权重执行RLC 2030来生成AI数据2040。在这种情况下, AI数据2040可以以比特流的形式表示。

[0454] 图20b是用于描述根据本公开的另一实施例的对放大DNN中的针对原始图像优化的权重进行编码的方法的示意图。

[0455] 参照图20b, AI编码设备1200可确定 4×4 滤波器核的权重差矩阵2055。在这种情况下, 权重差矩阵2055中的权重差值是放大DNN中的针对原始图像优化的权重差值, 如上面参照图19a所述, 并且下面省略其冗余描述。

[0456] AI编码设备1200可通过对 4×4 滤波器核的权重差矩阵2055执行变换(例如, 离散余弦变换)来确定权重矩阵2060。

[0457] AI编码设备1200可通过对权重矩阵2060执行修剪来确定权重矩阵2065。

[0458] AI编码设备1200可通过对权重矩阵2065执行量化来确定权重矩阵2070。

[0459] AI编码设备1200可通过对权重矩阵2070的权重执行RLC 2080来生成AI数据2090。

[0460] AI解码设备1300可对AI数据2090执行游程长度解码以生成二维(2D)权重矩阵, 并经由逆量化和逆变换来恢复权重差矩阵。在这种情况下, 恢复的权重差矩阵可以与权重差矩阵2055不同。失配是通过由于修剪和量化导致的数据丢失而引起的。

[0461] 根据图20b的方法, AI编码设备1200可对权重差矩阵2055执行变换, 使得数据值可集中在低频处, 而高频处的数据值可更接近零。因此, 与使用参照图20a描述的方法时相比, 在后续修剪和量化期间可更有效地丢弃不必要的信息。AI解码设备1300可执行与AI编码设备1200执行的变换相应的逆变换。

[0462] 在这种情况下, AI解码设备1300可通过对AI数据2090执行游程长度解码来生成2D权重矩阵, 并通过对2D权重矩阵执行逆量化和逆变换来恢复权重差矩阵。

[0463] 图21a是用于描述根据本公开实施例的由AI编码设备1200执行的基于上下文模型对放大DNN的针对原始图像优化的权重信息进行熵编码的处理的示意图, 并且图21b是用于描述根据本公开实施例的由AI解码设备1300执行的基于上下文模型对放大DNN的针对原始图像优化的权重信息进行熵解码的处理的示意图。

[0464] 参照图21a, 在AI编码设备1200中, 上下文模型生成器2110可基于权重 w_{t-1} (其是先前时间 $t-1$ 的图像的权重)生成上下文模型 M_{t-1} , 并且熵编码器2120可通过基于上下文模型 M_{t-1} 对权重 w_t (其是当前时间 t 的图像的权重)进行熵编码来生成权重信息。在这种情况下, 可结合诸如修剪和量化的模型压缩技术来执行熵编码。在这种情况下, 先前时间 $t-1$ 的图像可表示紧接在时间 t 的当前帧之前的帧, 但是该时间不限于此, 并且该时间可表示紧接在包括时间 t 的帧的当前时段单元之前的时段单元。在这种情况下, 时段单元可以是以下项之一: 预设数量的帧、GOP中的帧、序列中的帧和IRAP时段中的帧, 但不限于此。

[0465] 参照图21b,在AI解码设备1300中,上下文模型生成器2130可以以与AI编码设备1200中相同的方式基于权重 w_{t-1} (权重 w_{t-1} 是先前时间 $t-1$ 的图像的恢复的权重)生成上下文模型 M_{t-1} ,并且熵解码器2140可通过基于上下文模型 M_{t-1} 对从AI编码设备1200接收的权重信息进行熵解码来获得权重 w_t ,其中,权重 w_t 是当前时间 t 的图像的权重。在这种情况下,当熵编码是结合模型压缩技术(诸如修剪和量化)被执行的时,AI解码设备1300可结合熵解码操作执行与模型压缩技术相应的操作。例如,当熵编码是结合量化操作被执行的时,AI解码设备1300可结合熵解码操作执行逆量化操作。

[0466] 此外,由AI编码设备1200生成并且与至少一个质量(例如,基于分辨率或比特率中的至少一个的质量)相应的图像数据和DNN设置信息通过流传输系统被提供给AI解码设备1300。流传输系统是包括至少一个服务器(例如,服务服务器或内容服务器)和终端的系统,并且是指服务器根据经由网络来自终端的请求提供与请求相应的图像相关数据的系统。在这种情况下,所生成的图像数据和DNN设置信息可存储在与服务服务器分开的内容服务器中,并且终端可根据请求从服务服务器接收至少一种质量的图像数据的位置信息,然后从与至少一种质量的图像数据的位置信息相应内容服务器接收图像数据和DNN设置信息。在这种情况下,内容服务器是用于存储图像数据和DNN设置信息的服务器,并且可以是AI编码设备1200,但是内容服务器不限于此。内容服务器可与AI编码设备1200分离,并且在这种情况下,内容服务器可从AI编码设备1200接收图像数据和DNN设置信息以进行存储。

[0467] 同时,上述本公开的实施例可被编写为可存储在介质中的计算机可执行程序或指令。

[0468] 介质可永久地存储计算机可执行程序或指令,或者存储用于执行或下载的计算机可执行程序或指令。此外,介质可以是其中组合了单件或多件硬件的各种记录介质或存储介质中的任何一种,并且介质不限于直接连接到计算机系统的介质,而是可以分布在网络上。介质的示例包括被配置为存储程序指令的磁介质(诸如硬盘、软盘和磁带)、光学记录介质(诸如CD-ROM和DVD)、磁光介质(诸如软光盘)以及ROM、RAM和闪存。介质的其他示例包括由发布应用的应用商店或由提供或发布各种类型的软件的网站、服务器等管理的记录介质和存储介质。

[0469] 同时,可经由软件来实现与上述DNN相关的模型。当DNN模型经由软件(例如,包括指令的程序模块)实现时,DNN模型可存储在计算机可读记录介质中。

[0470] 此外,DNN模型可通过以硬件芯片的形式被集成而成为上述AI解码设备200或AI编码设备600的一部分。例如,DNN模型可以以用于AI的专用硬件芯片的形式被制造,或者可被制造为现有通用处理器(例如,CPU或应用处理器)或图形专用处理器(例如,GPU)的一部分。

[0471] 此外,DNN模型可以以可下载软件的形式提供。计算机程序产品可包括通过制造商或电子市场电子发布的软件程序形式的产品(例如,可下载的应用)。对于电子发布,软件程序的至少一部分可存储在存储介质中或者可以临时生成。在这种情况下,存储介质可以是制造商或电子市场的服务器,或者是中继服务器的存储介质。

[0472] 根据本公开的实施例,用于执行图像的AI编码和AI解码的方法和设备能够经由基于AI的图像编码和解码以低比特率处理图像。

[0473] 此外,根据本公开的实施例,用于执行图像的AI编码和AI解码的方法和设备可通过在周期性地或在必要时更新针对原始图像优化的放大DNN设置信息之后执行放大来改善

图像质量。

[0474] 另外,根据本公开的实施例,用于执行图像的AI编码和AI解码的方法和设备可通过周期性地或在必要时有效地用信号通知DNN设置信息以更新放大DNN的针对原始图像优化的DNN设置信息,有效地减少要编码和解码的信息量。

[0475] 然而,本领域普通技术人员将理解,通过根据本公开的实施例的用于执行图像的AI编码和AI解码的方法和设备可实现的效果不限于上文描述的那些效果,并且从以下描述中将更清楚地理解本文未描述的本公开的其他效果。

[0476] 虽然已经参照附图描述了本公开的一个或更多个实施例,但是本领域普通技术人员将理解,在不脱离由下面的权利要求限定的精神和范围的情况下,可在其中进行形式和细节上的各种改变。

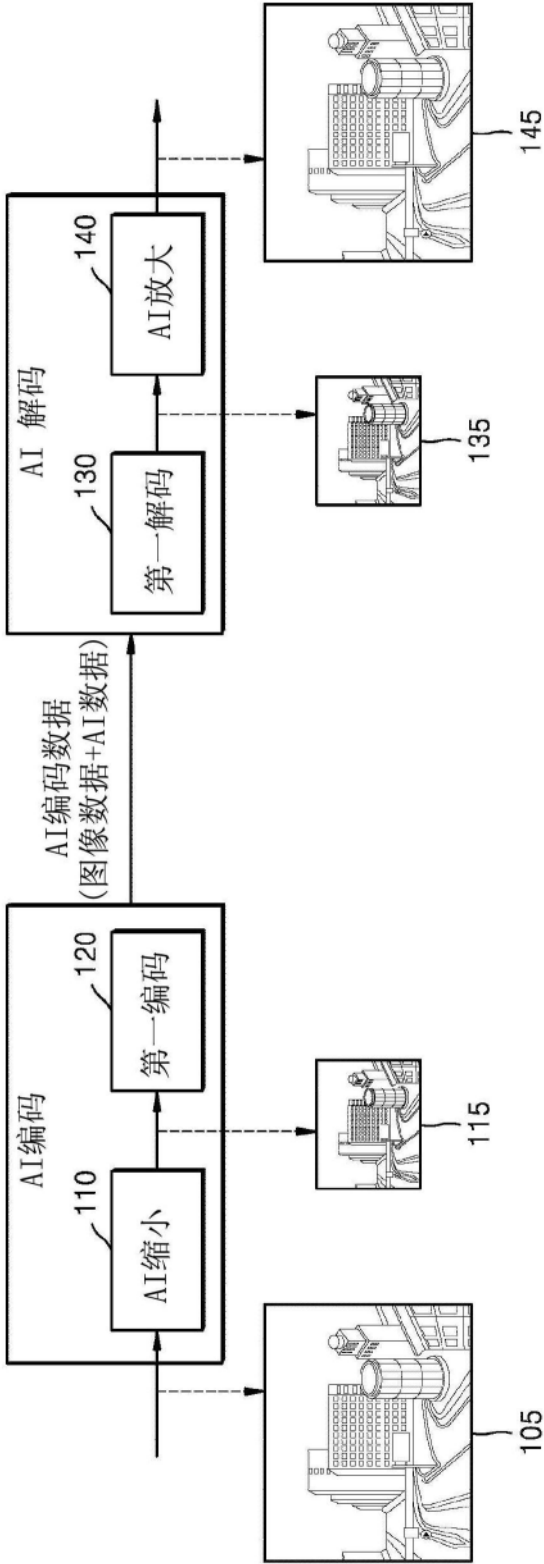


图1

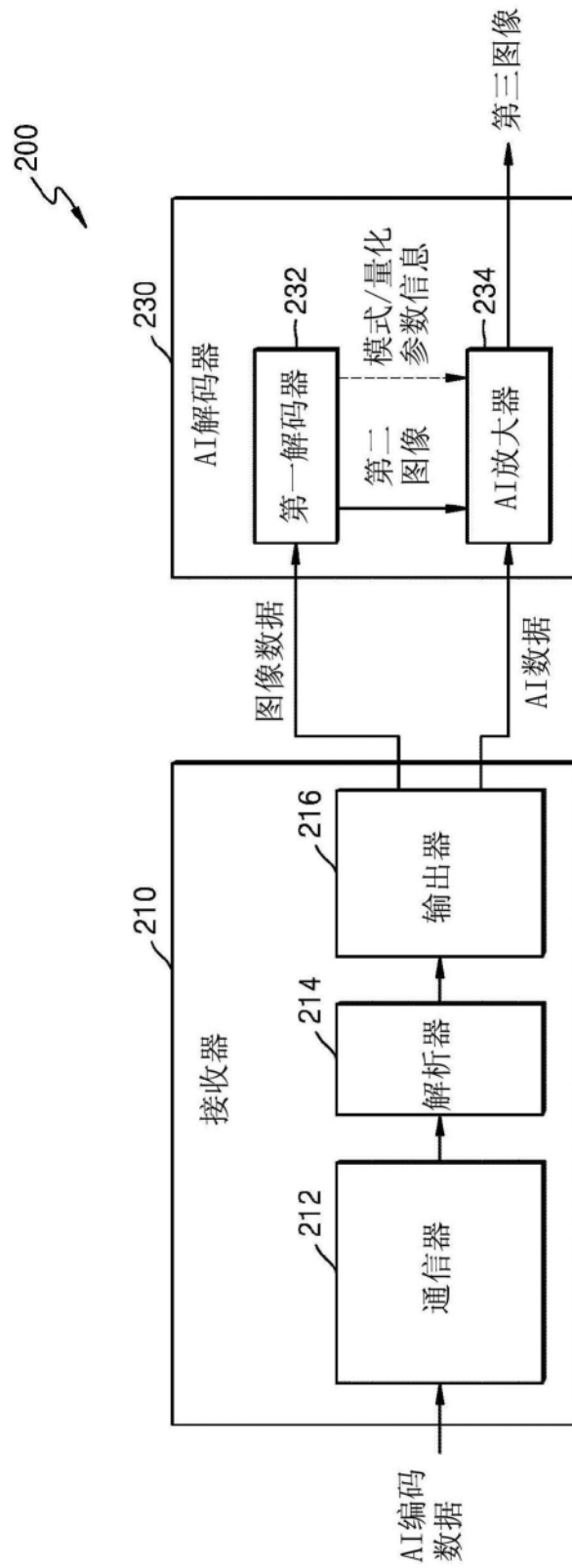


图2

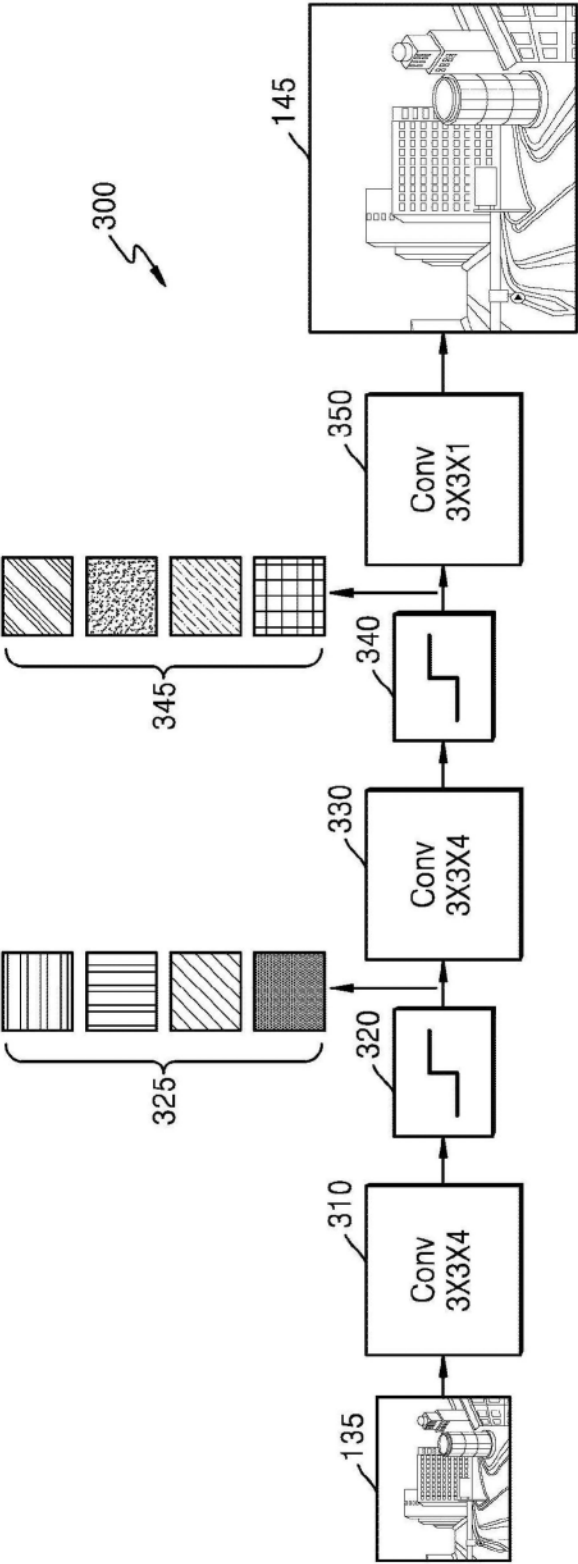


图3

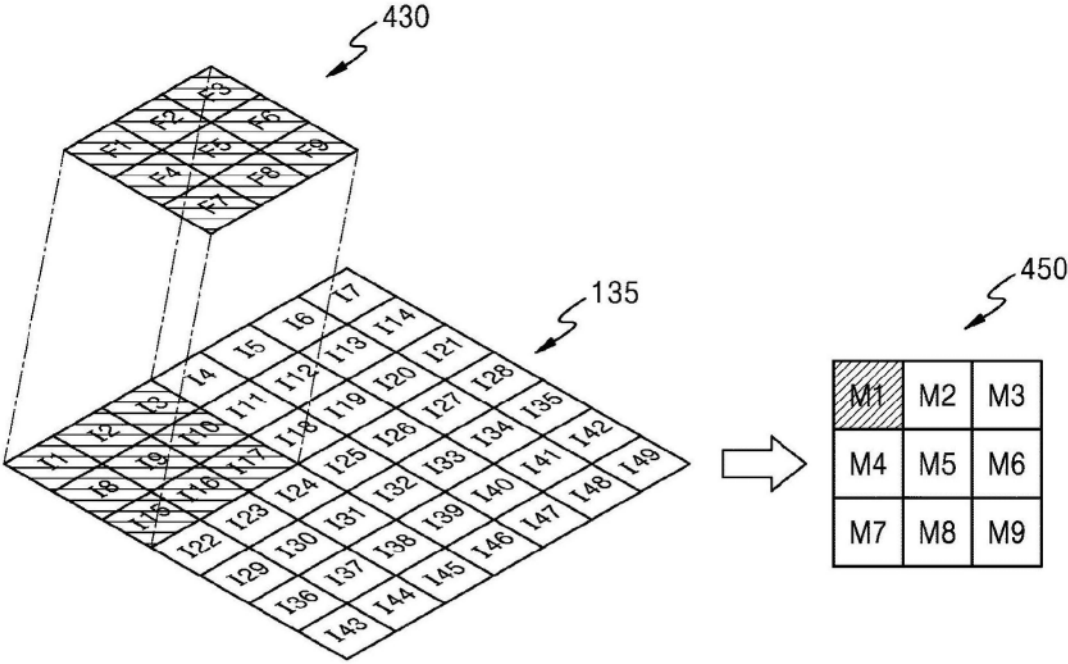


图4

图像相关信息	DNN设置信息
SD, 10Mbps, AV1	A DNN 设置信息
HD, 15Mbps, H.264	B DNN 设置信息
全 HD, 20Mbps, HEVC	C DNN 设置信息
全 HD, 15Mbps, HEVC	D DNN 设置信息

图5

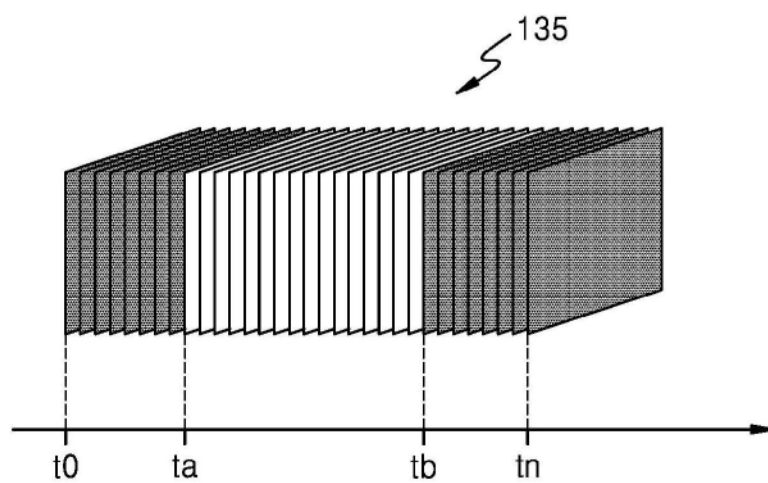


图6

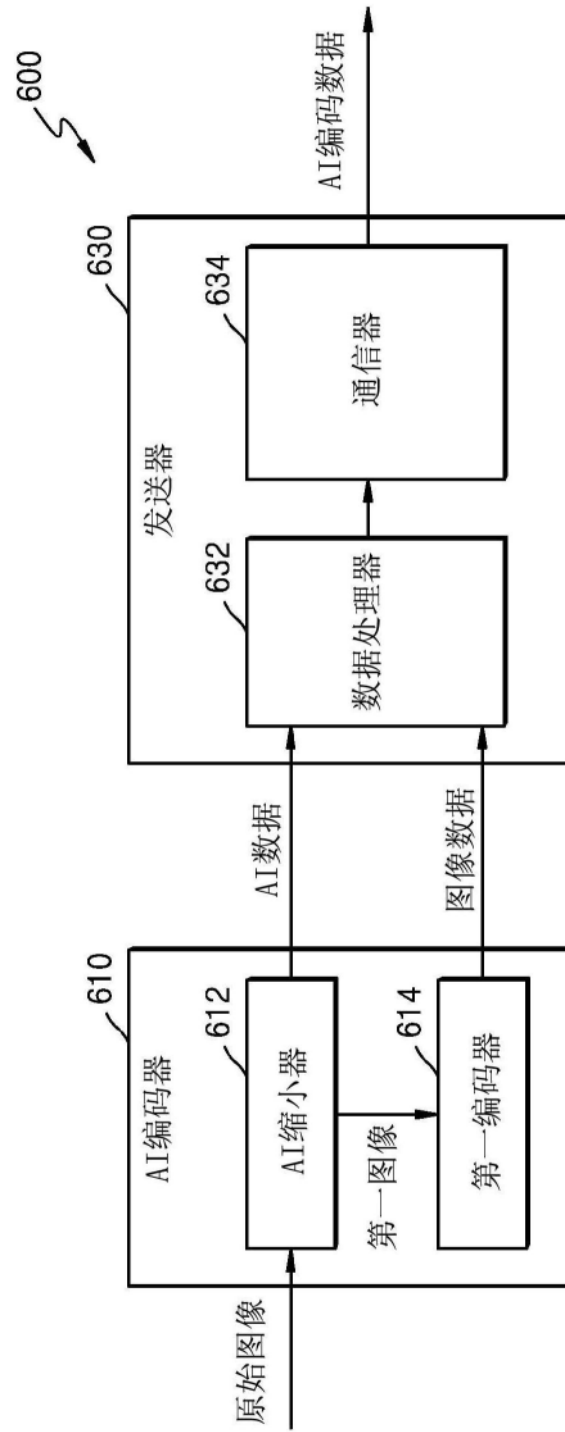


图7

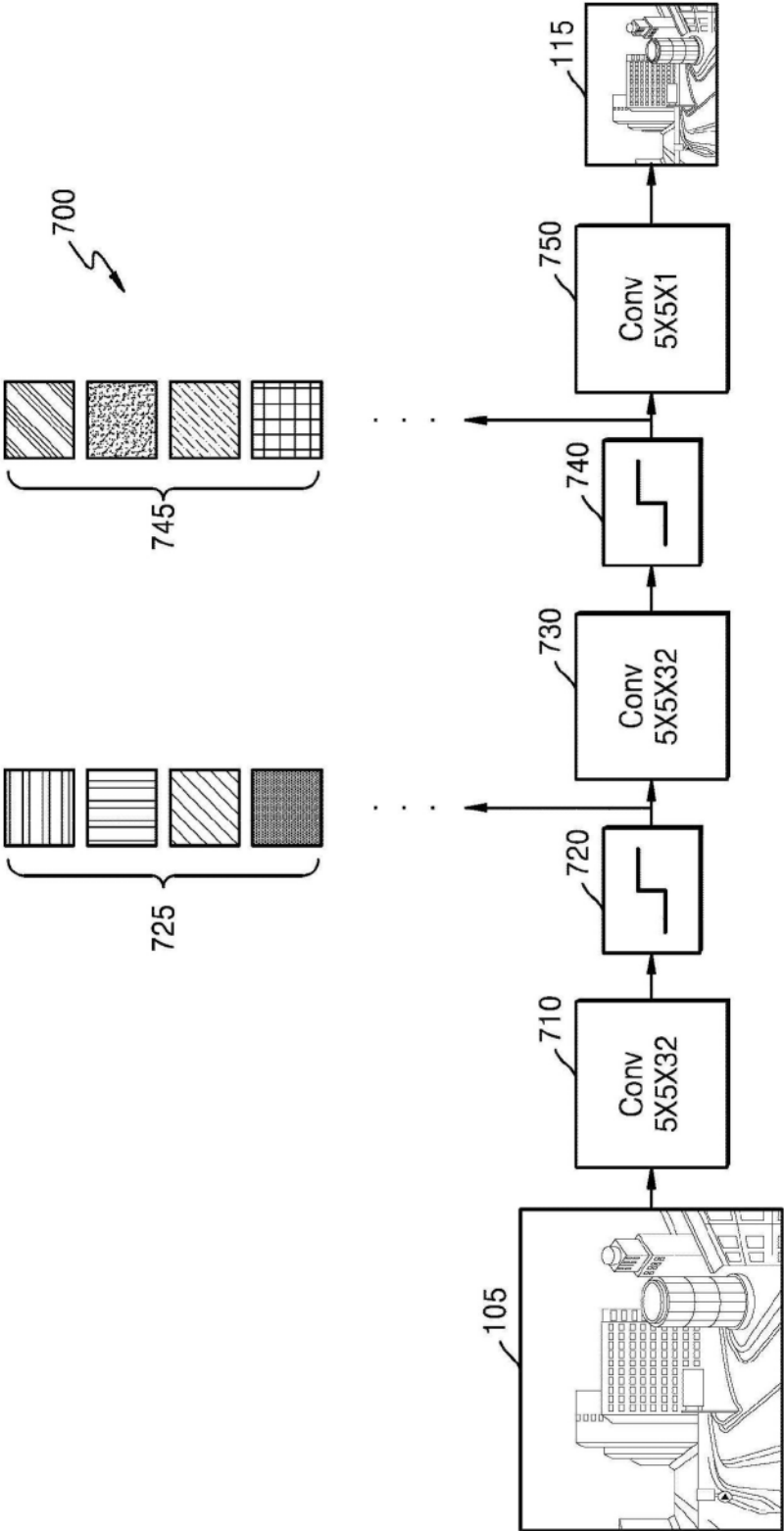


图8

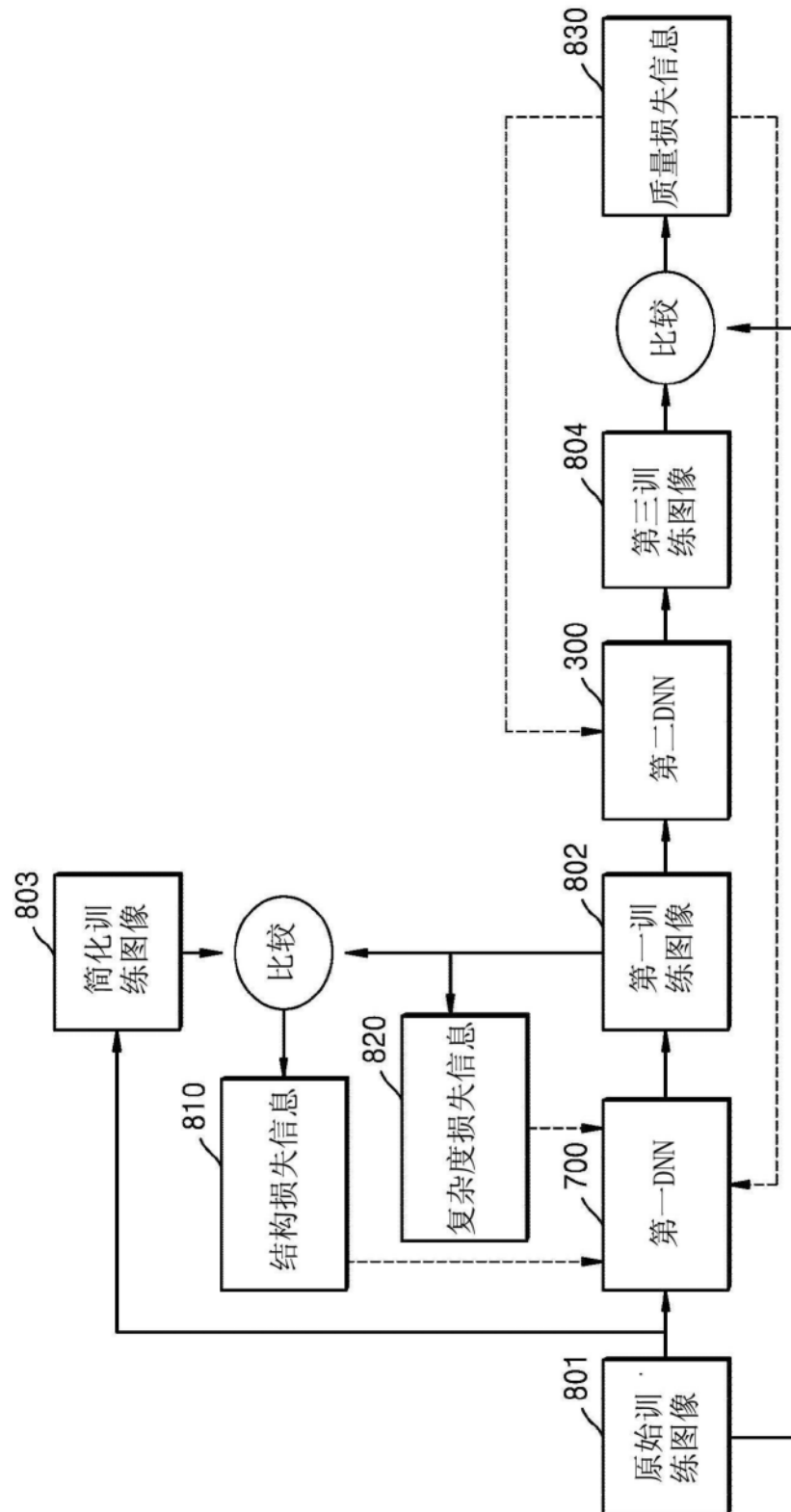


图9

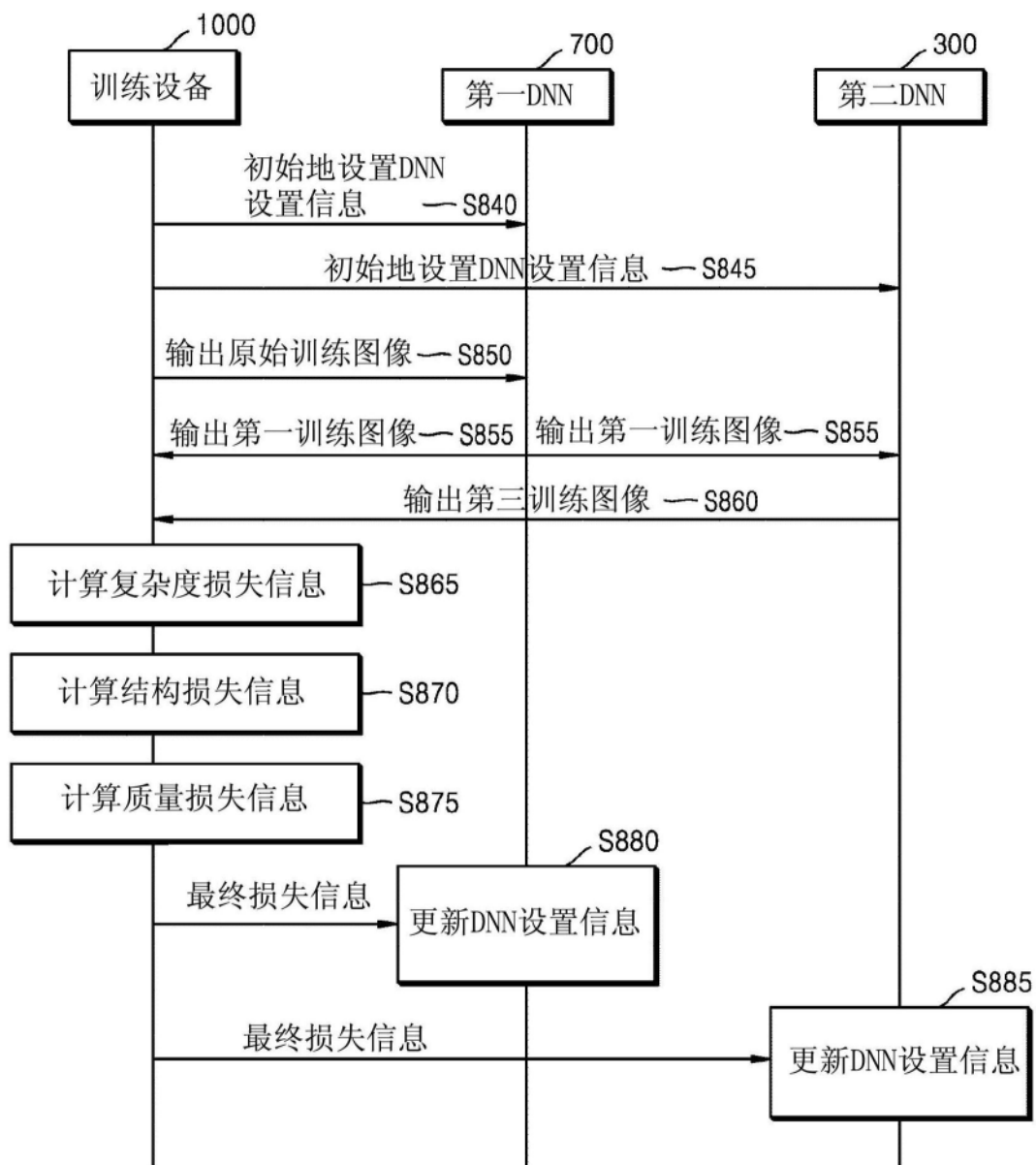


图10

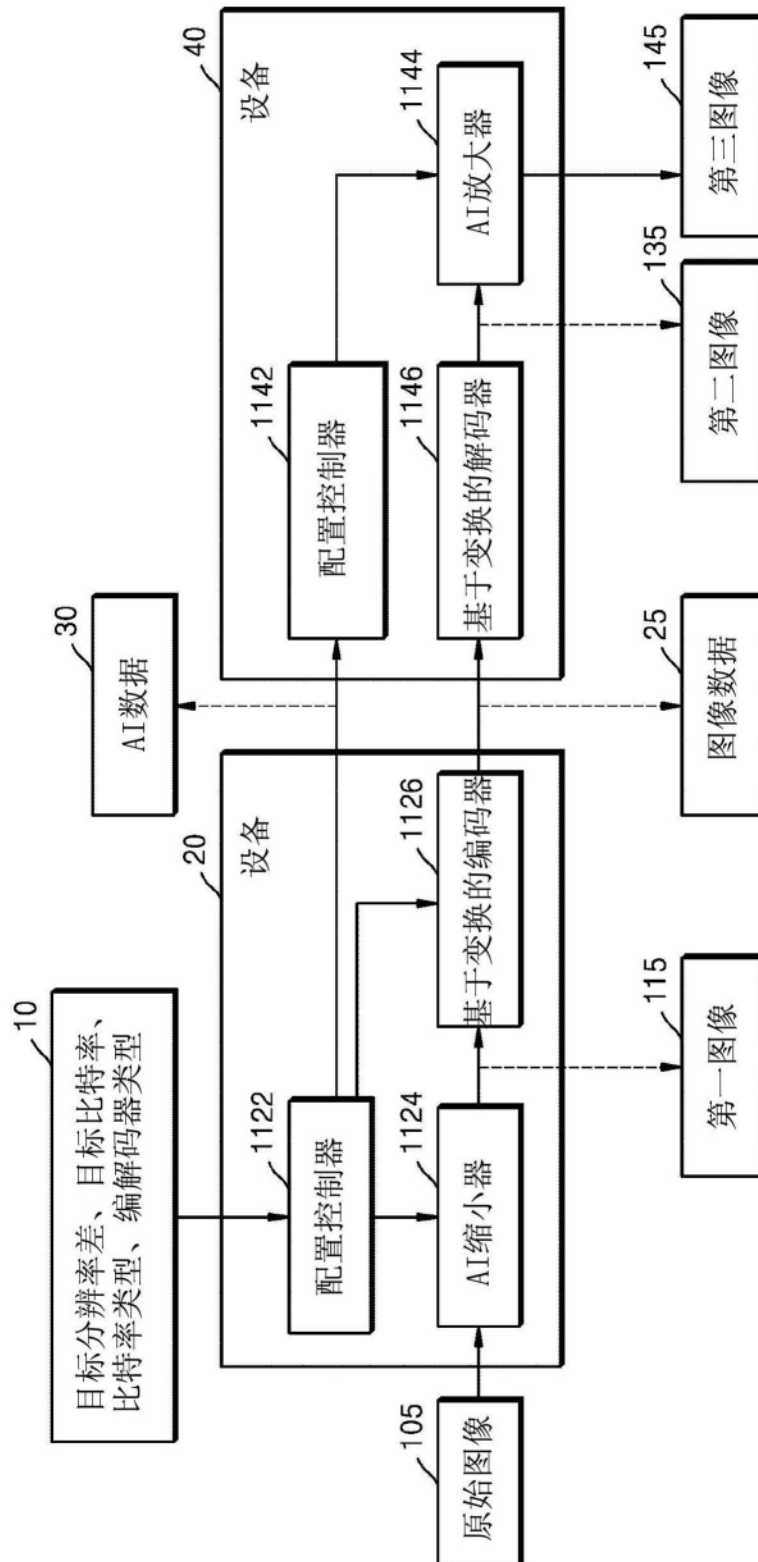


图11

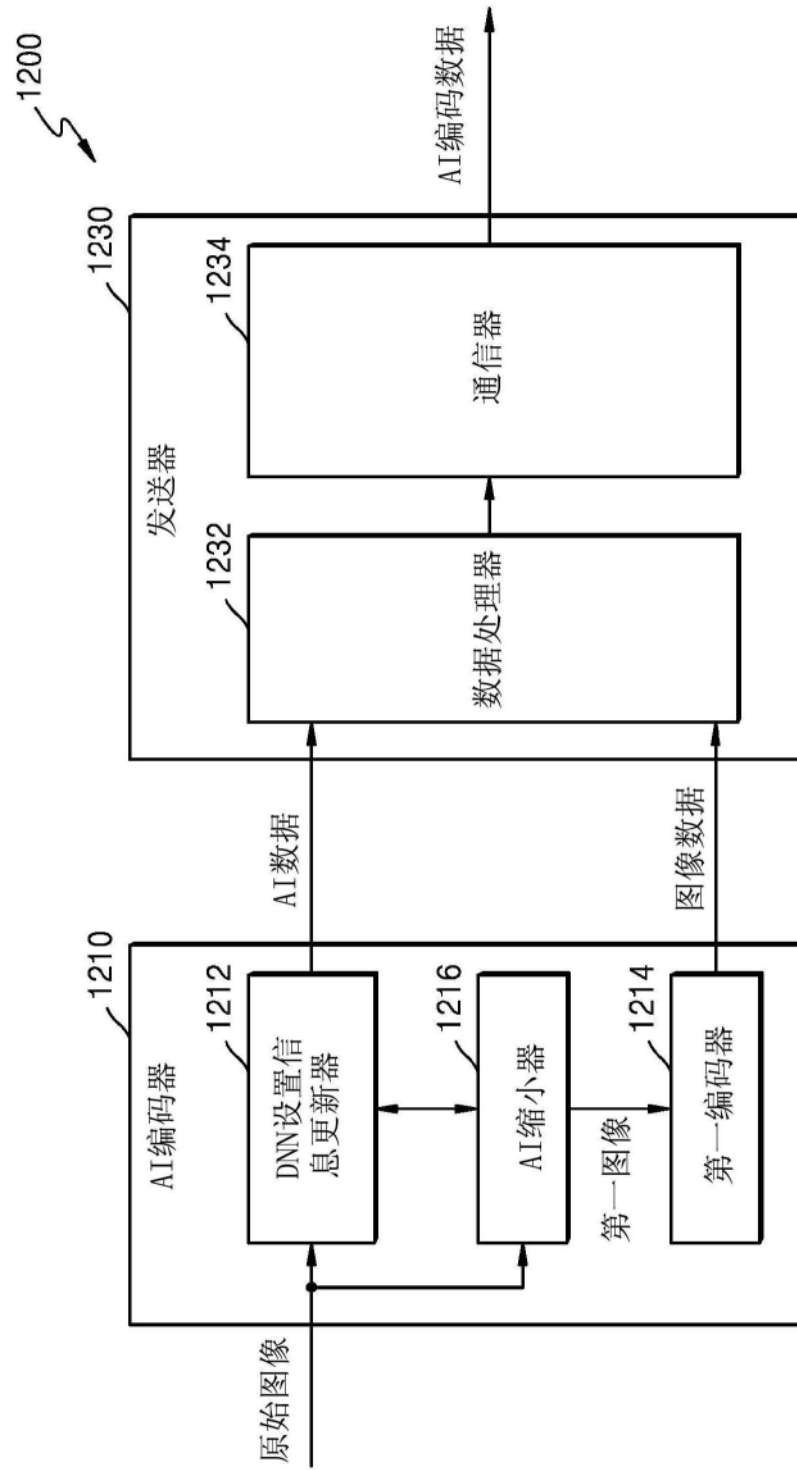


图12

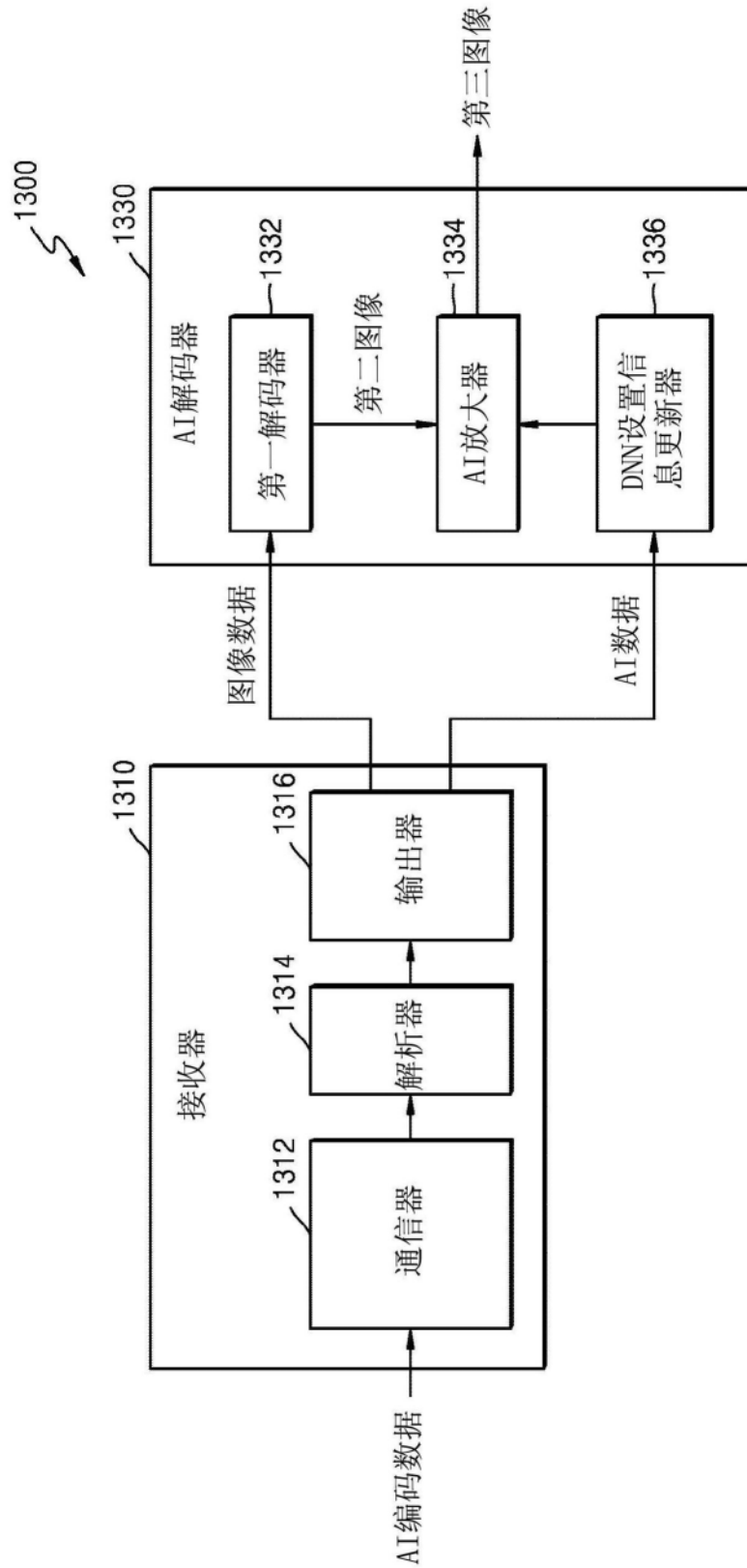


图13

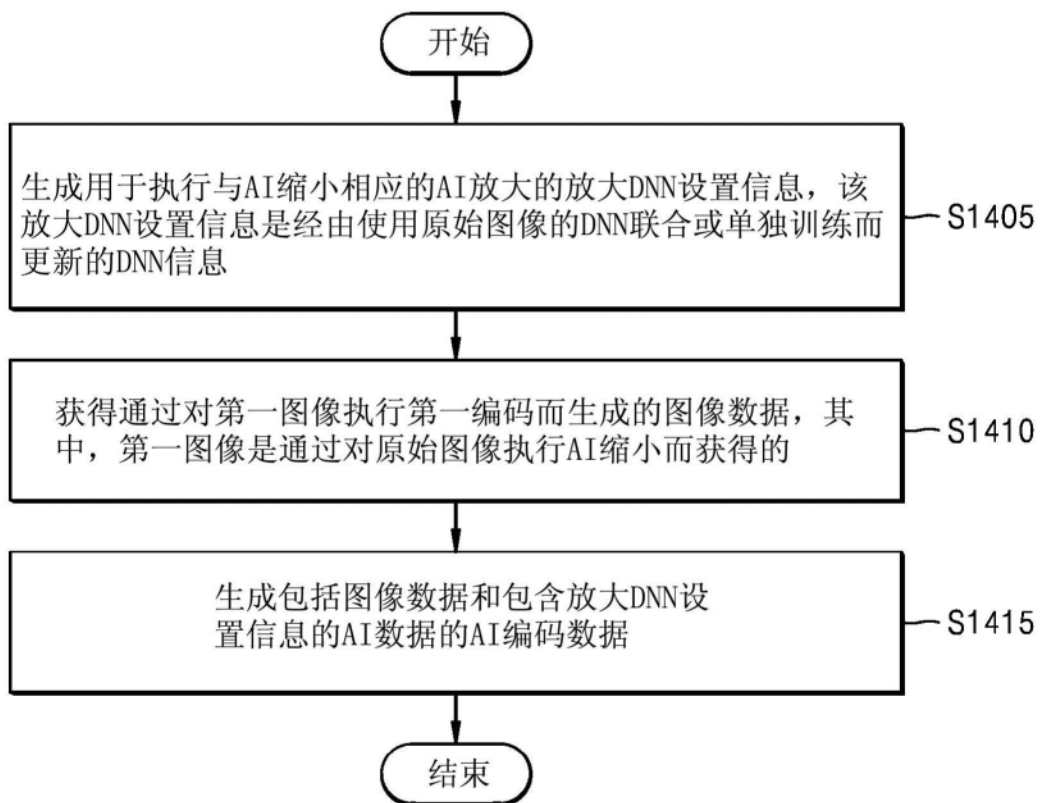


图14a

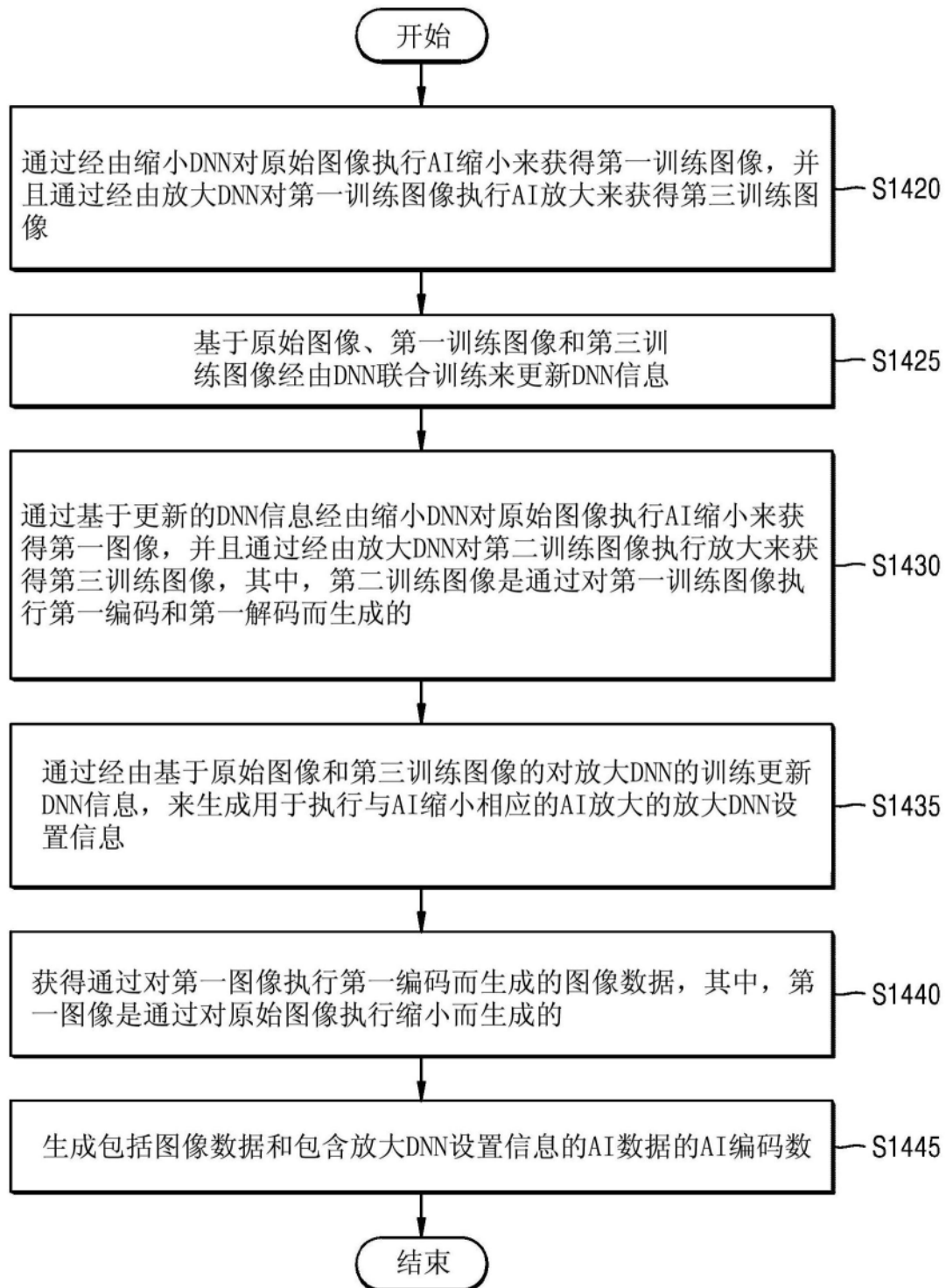


图14b

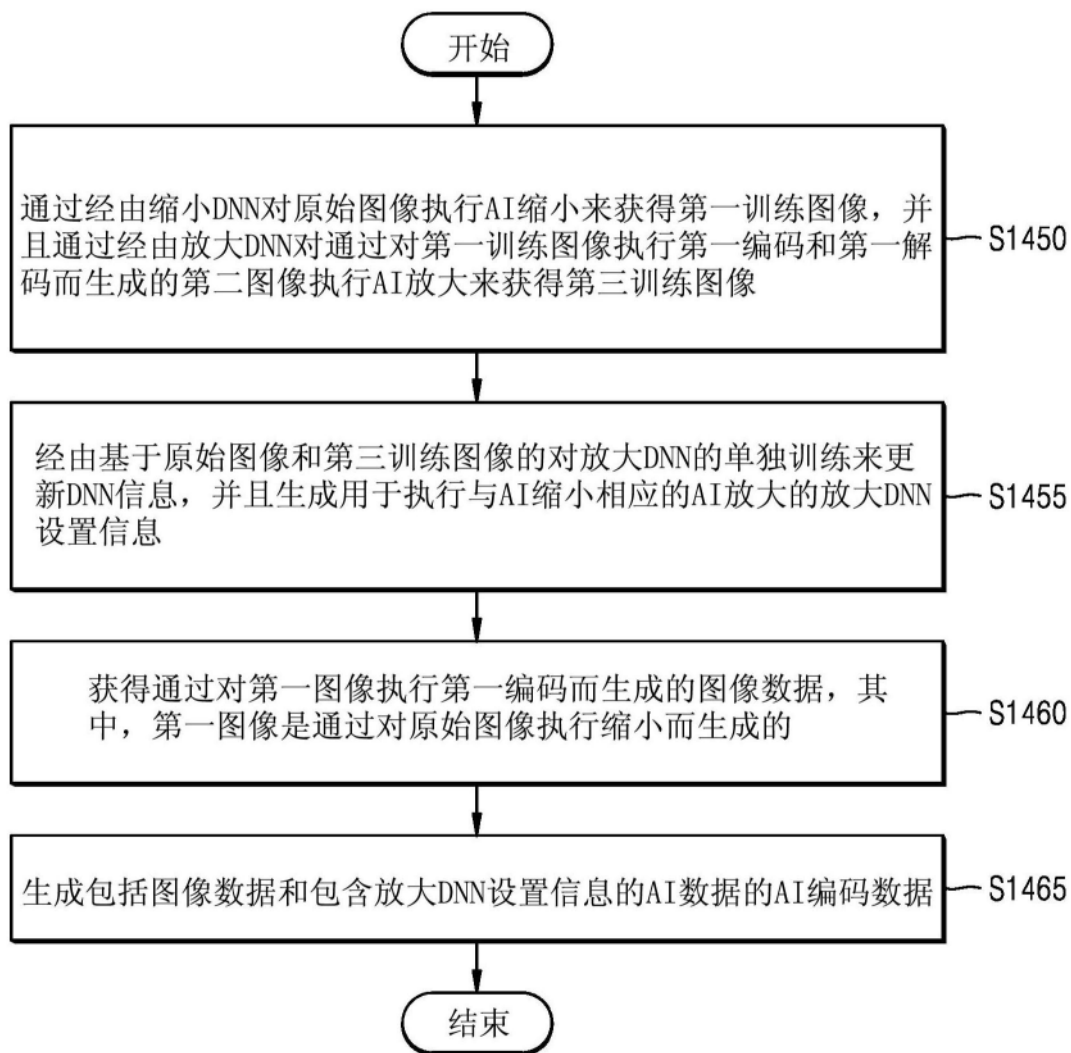


图14c

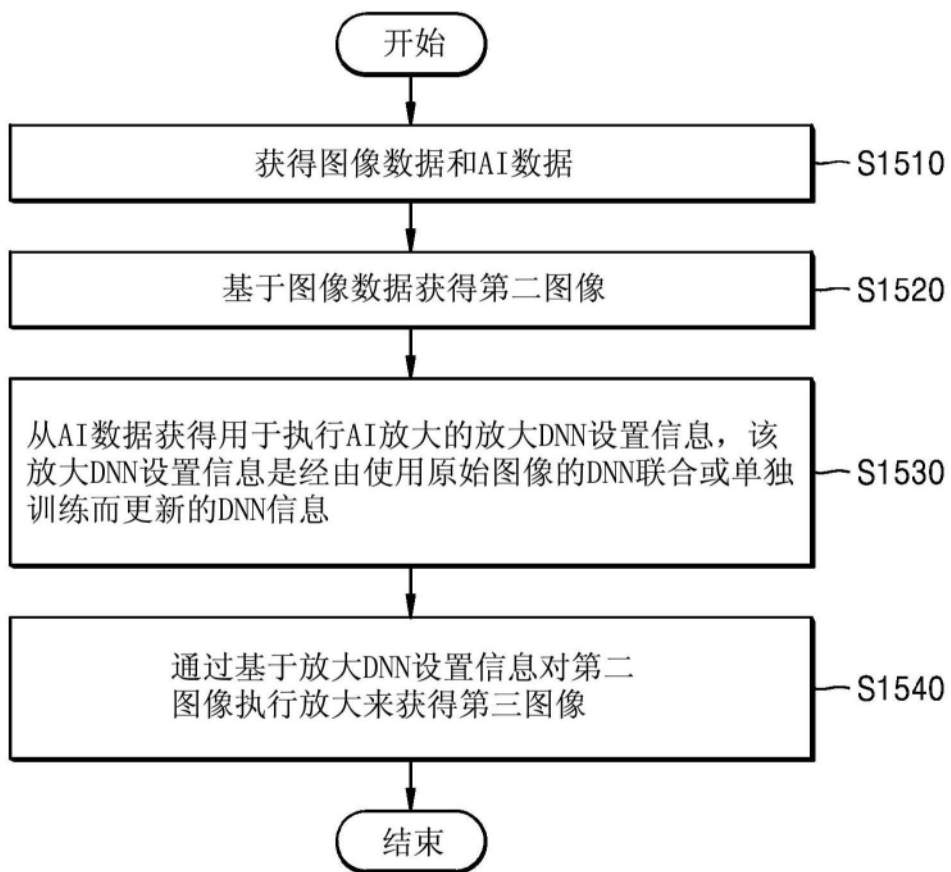


图15

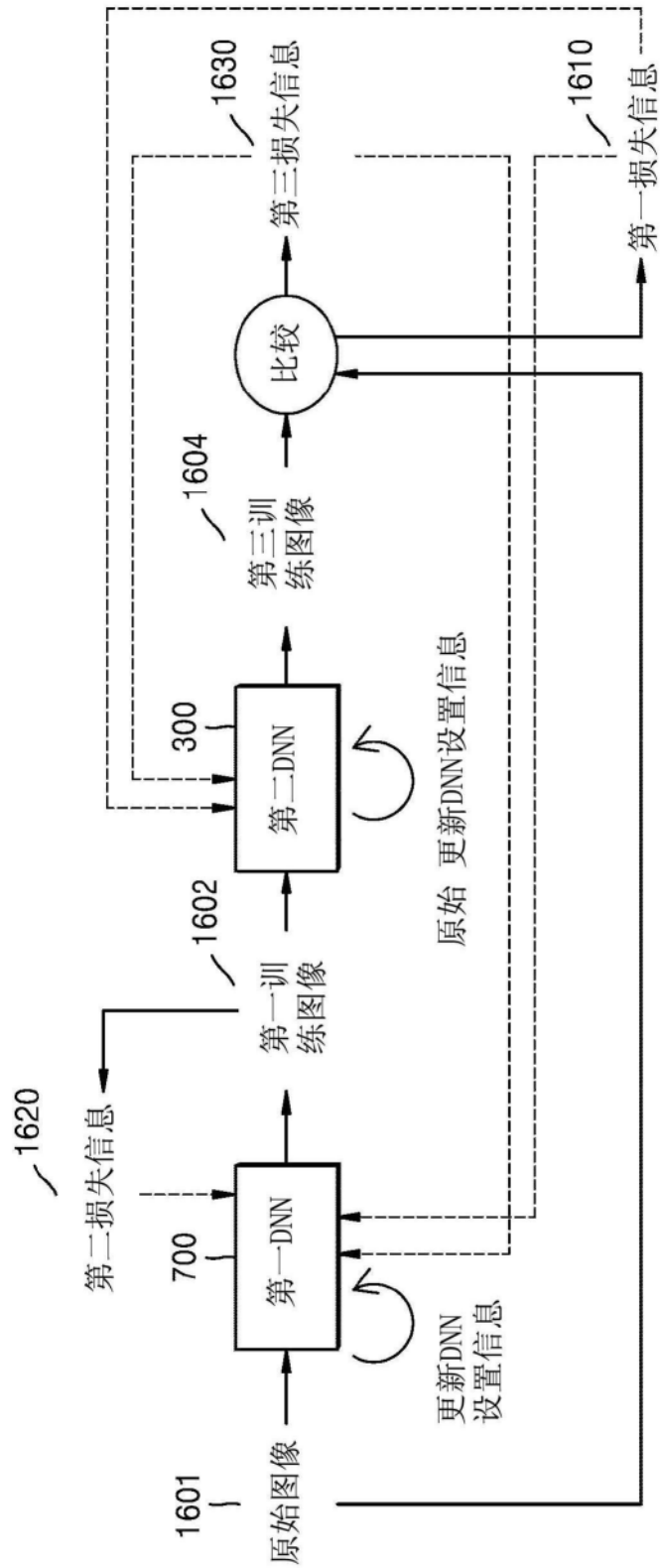


图16a

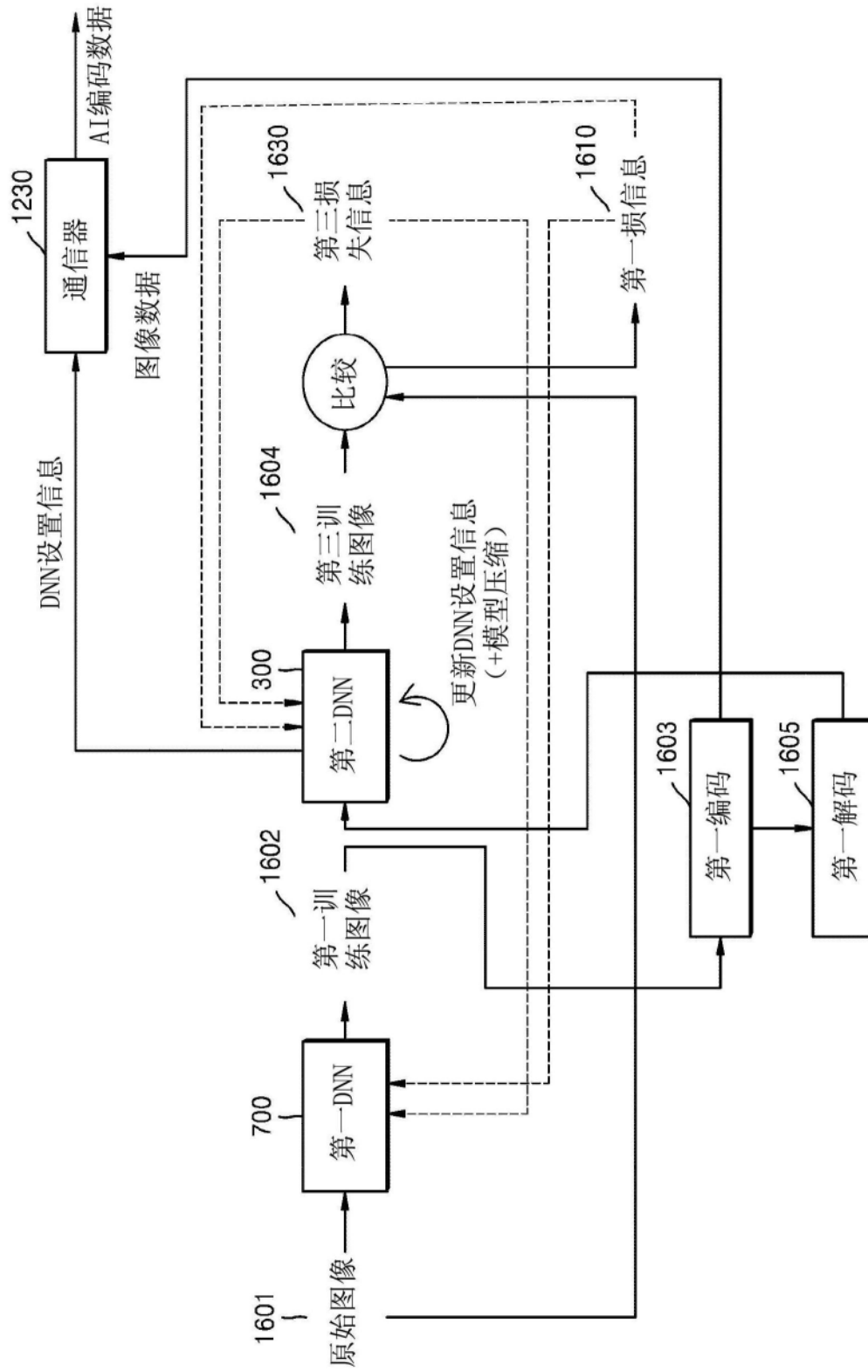


图16b

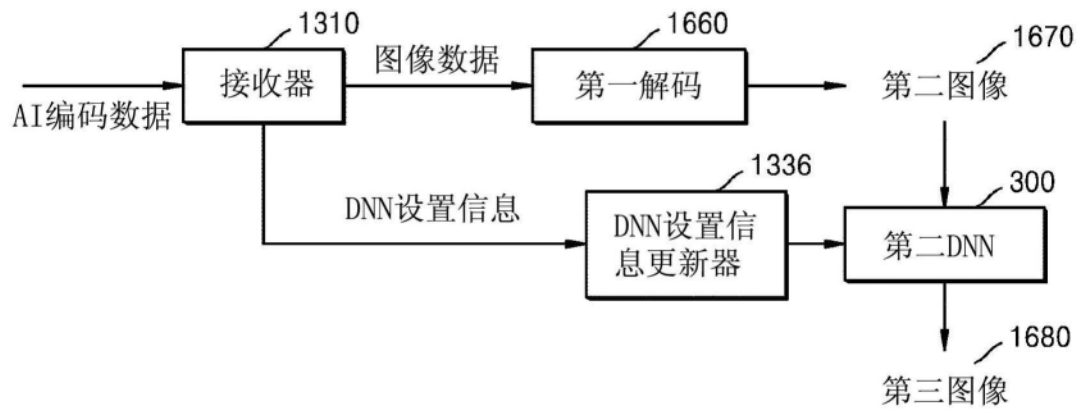


图16c

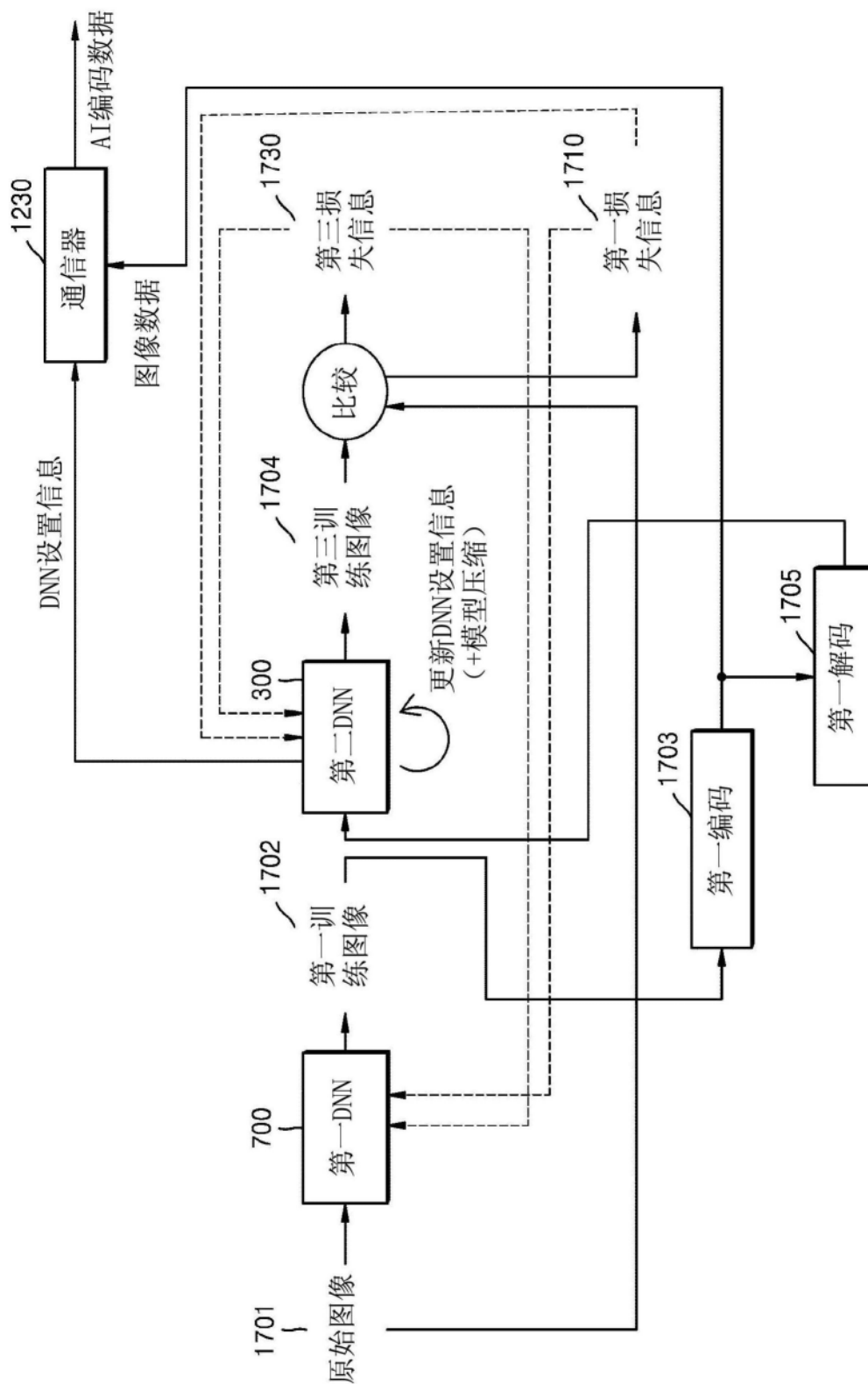


图17a

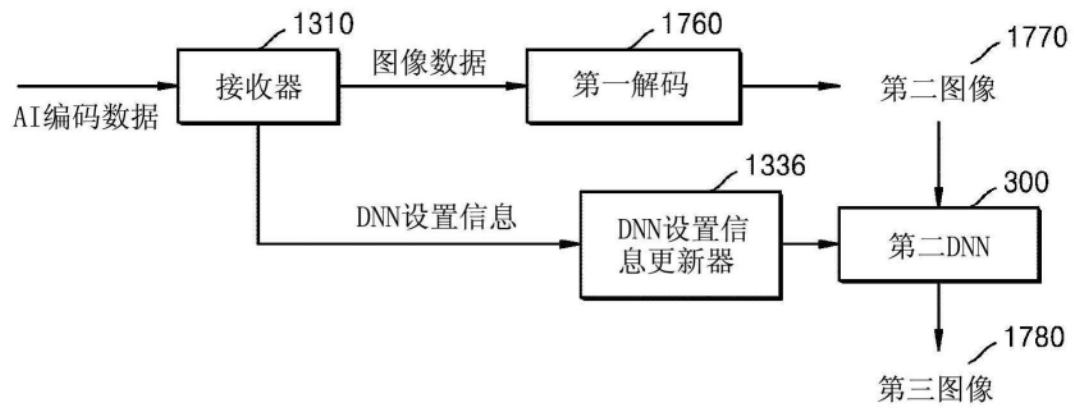


图17b

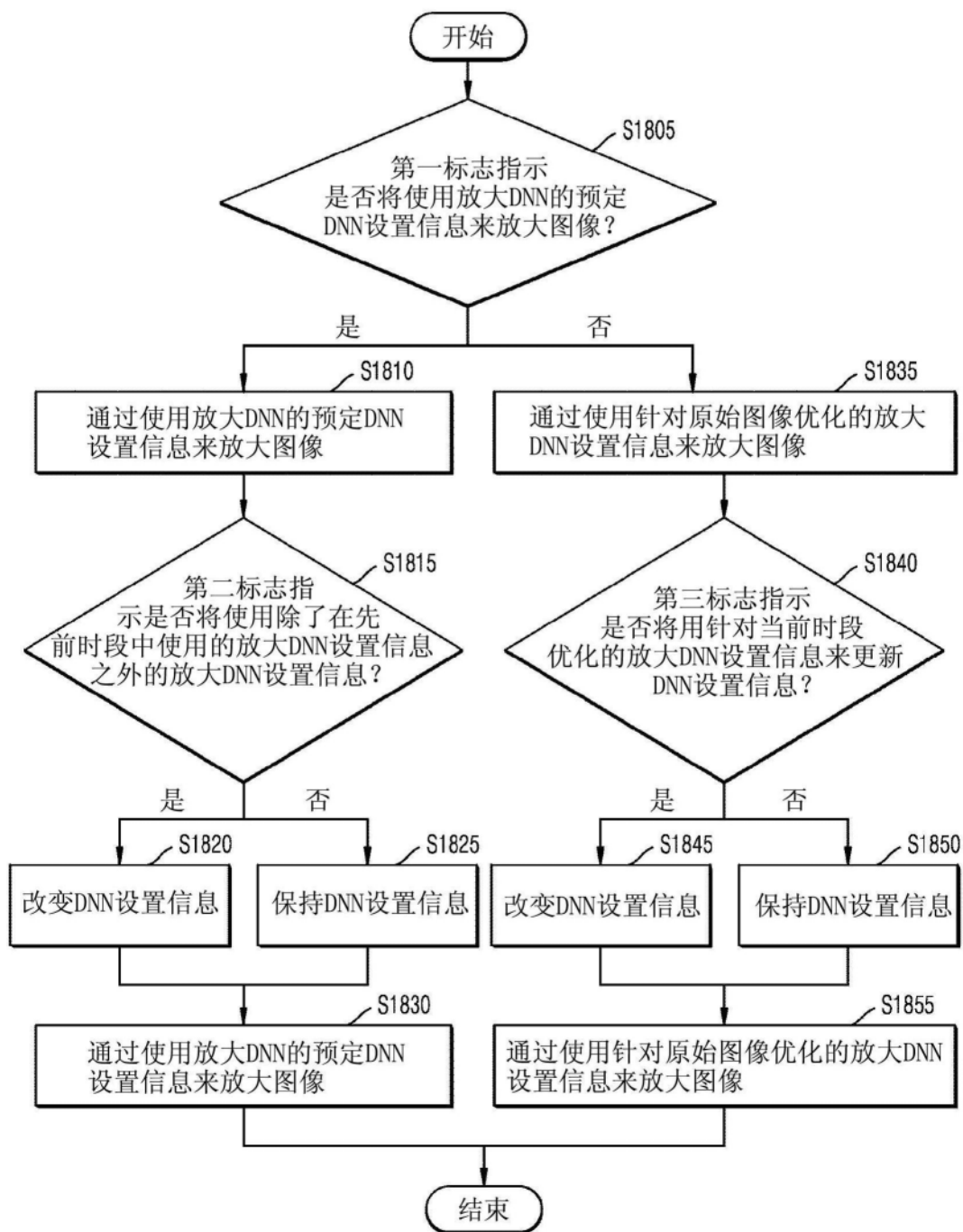


图18

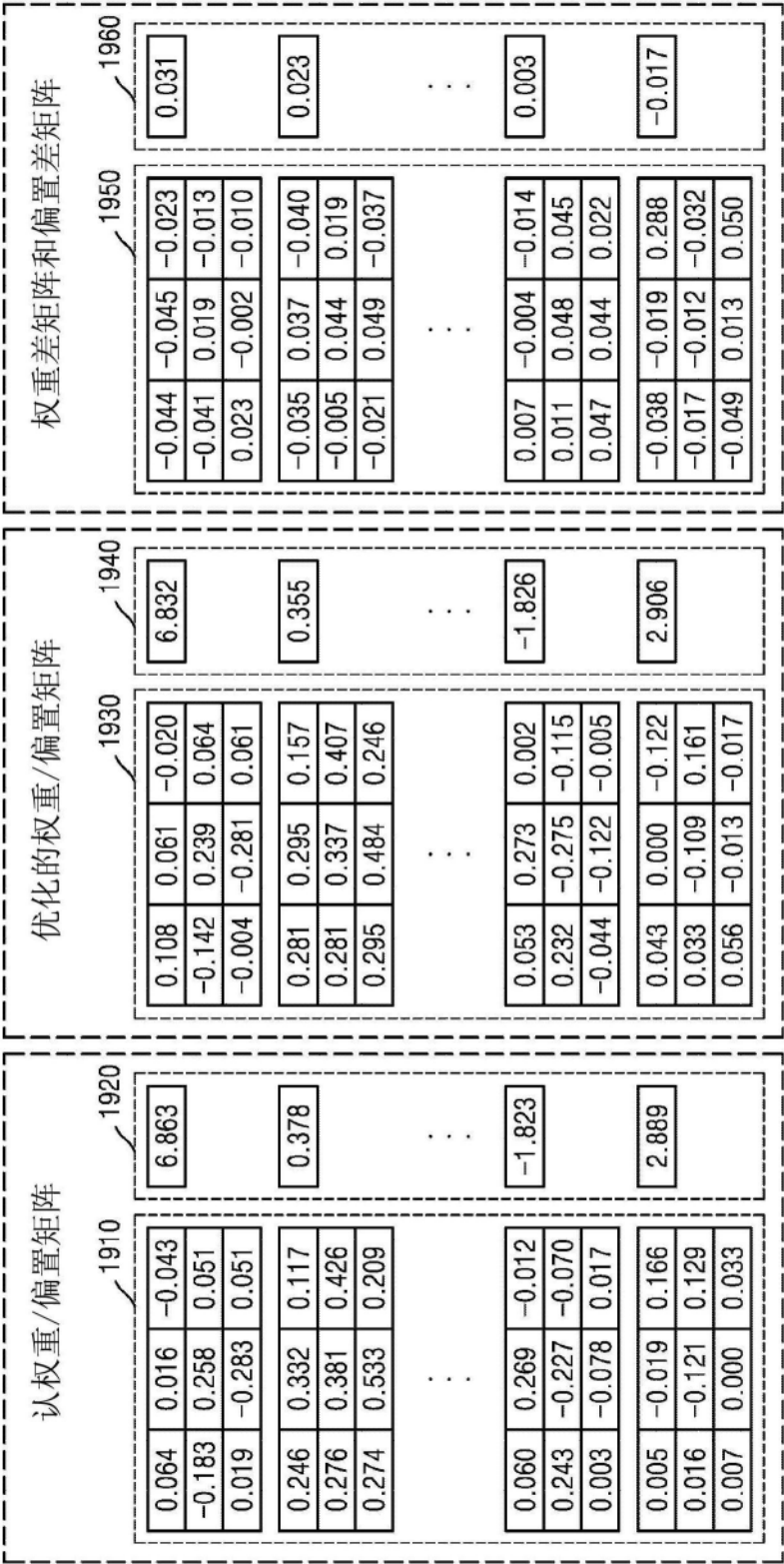


图19a

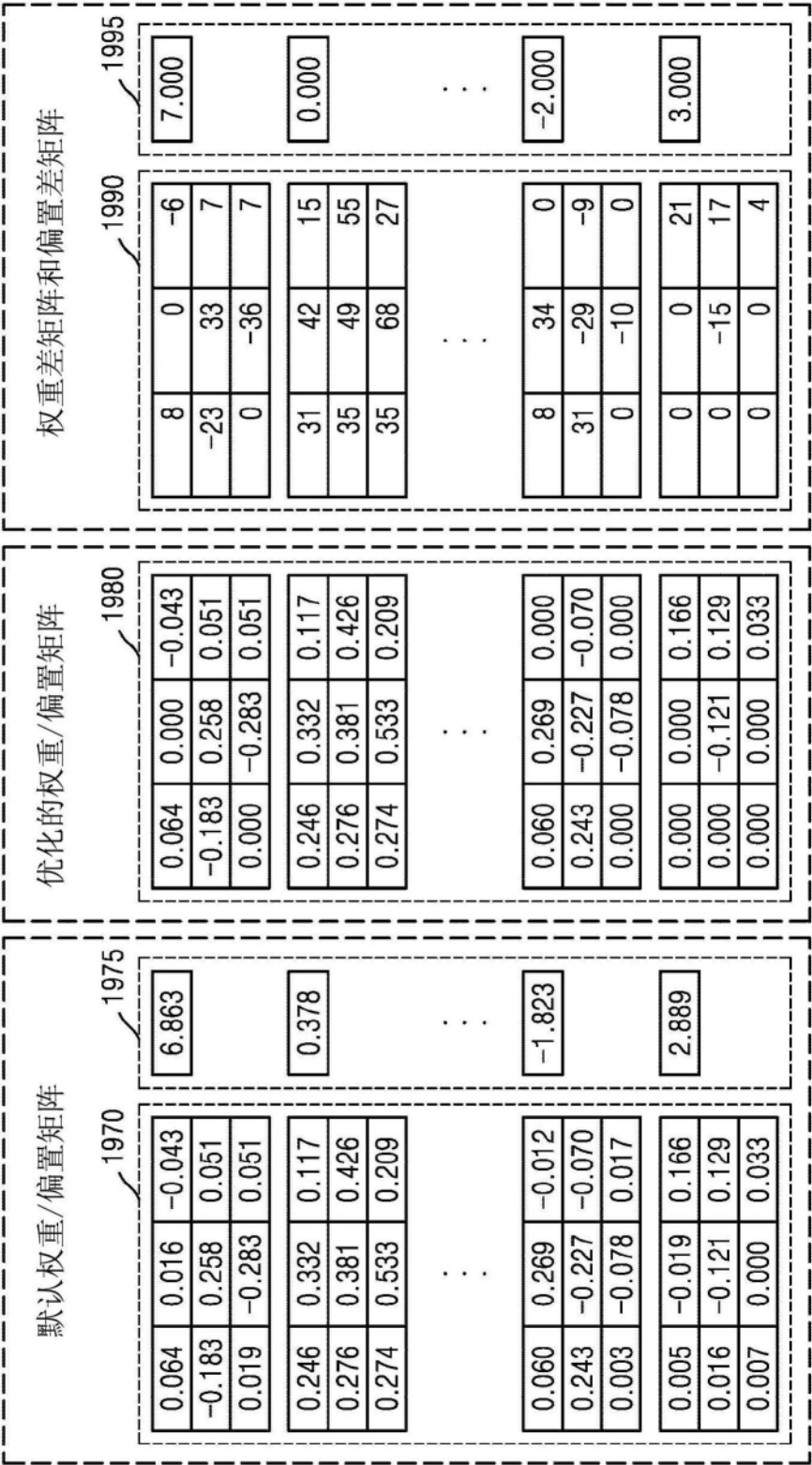


图19b

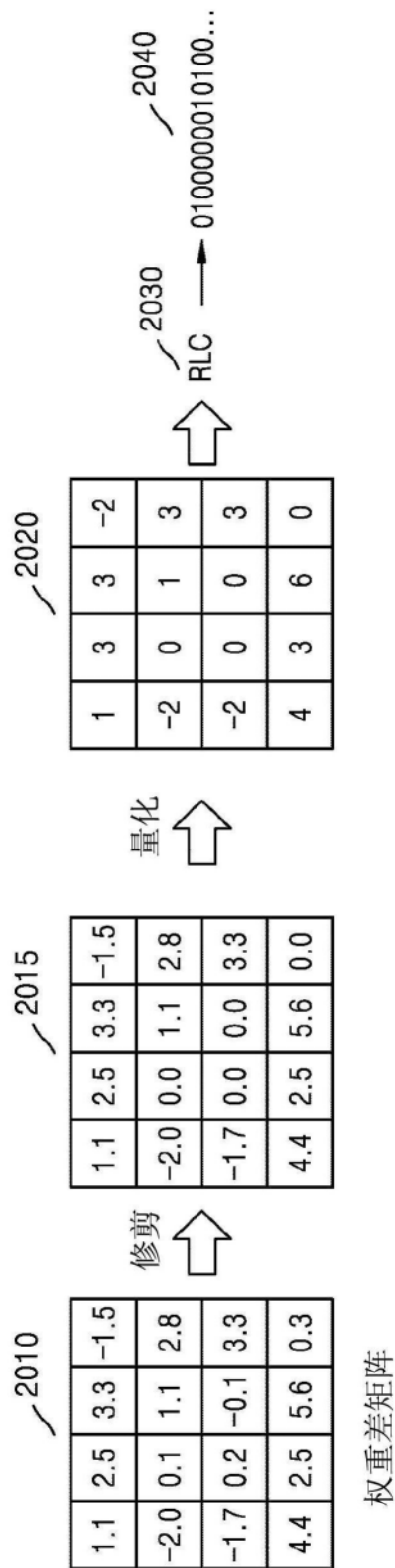


图20a

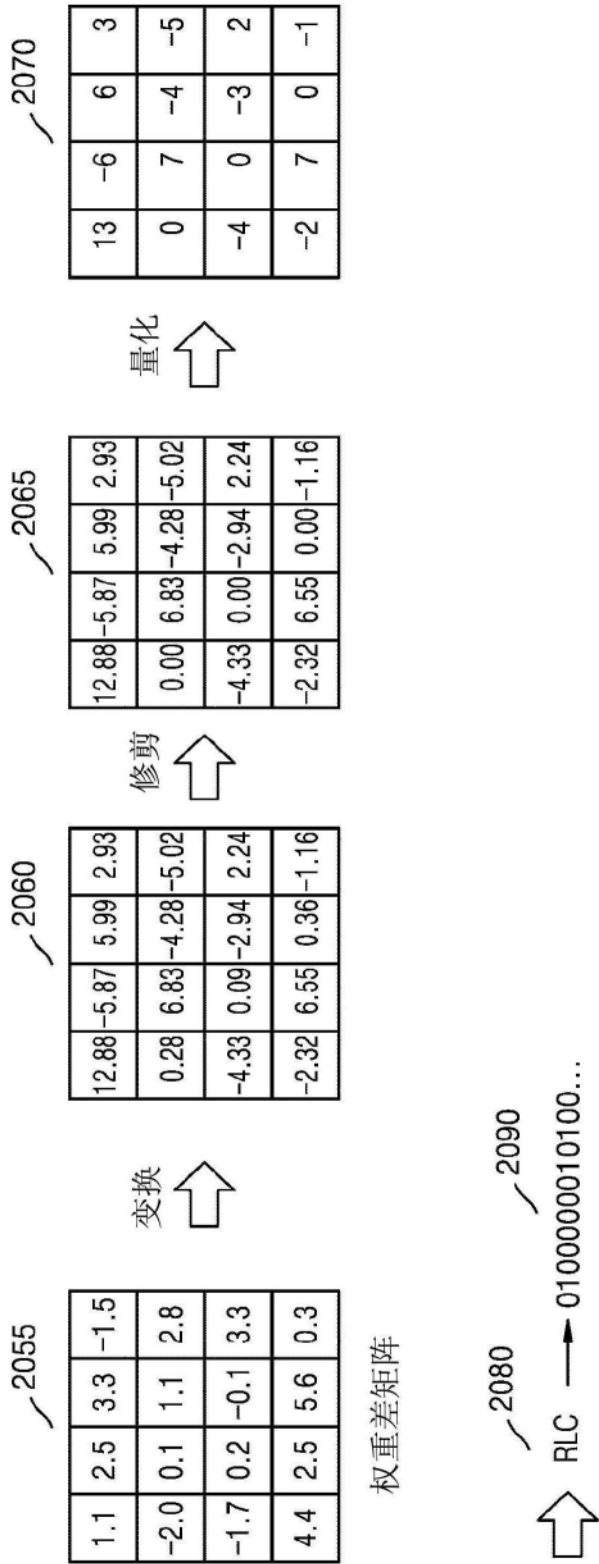


图20b

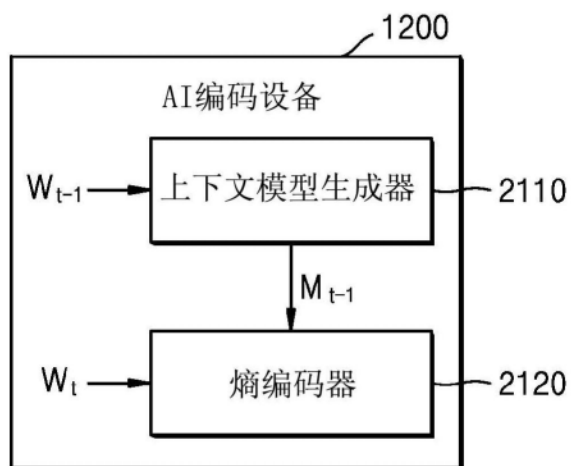


图21a

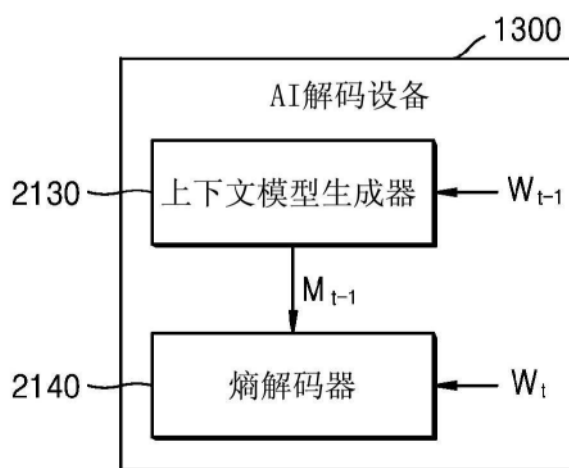


图21b