(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118872269 A
(43)申请公布日 2024.10.29

(21)申请号 202380027336.4

(22)申请日 2023.04.11

(30)优先权数据
2022202470 2022.04.13 AU

(85)PCT国际申请进入国家阶段日
2024.09.12

(86)PCT国际申请的申请数据
PCT/AU2023/050289 2023.04.11

(87)PCT国际申请的公布数据
WO2023/197029 EN 2023.10.19

(71)申请人 佳能株式会社
地址 日本

(72)发明人 克里斯托弗·詹姆斯·罗斯沃恩
陈红绒·阮

(74)专利代理机构 北京魏启学律师事务所 11398
专利代理师 陈涛

(51)Int.Cl.
H04N 19/169(2006.01)
H04N 19/103(2006.01)
H04N 19/159(2006.01)
H04N 19/172(2006.01)
H04N 19/18(2006.01)
H04N 19/50(2006.01)
H04N 19/91(2006.01)
H04N 19/96(2006.01)

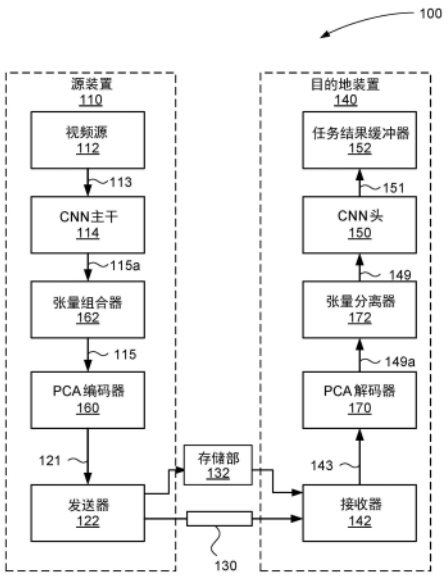
权利要求书4页 说明书52页 附图27页

(54)发明名称

用于对张量进行编码和解码的方法、设备和系统

(57)摘要

一种从位流中解码张量的方法包括:从位流中解码第一单元,该第一单元与跨张量的通道的平均张量值相对应;从位流中解码第二单元,该第二单元与张量的分解的分量相对应;从位流中解码第三单元,该第三单元与同第一单元相关的系数相对应;以及从位流中解码第四单元,该第四单元与同第二单元相关的系数相对应。该方法还包括:通过将针对各个通道的系数与平均张量值相乘并将张量的分解的分量与同张量的分解相关联的系数相乘来逐个通道地产生张量,其中,这些单元中的至少一个单元相对于其他单元中的至少一个单元可独立地解码,并且各单元布置在多个二维样本阵列中。



CN 118872269 A

1.一种从位流中解码张量的方法,所述方法包括:

从所述位流中解码第一信息单元,所述第一信息单元与跨所述张量的通道的平均张量值相对应;

从所述位流中解码第二信息单元,所述第二信息单元与所述张量的分解的分量相对应;

从所述位流中解码第三信息单元,所述第三信息单元与所述张量的同所述第一信息单元相关的系数相对应;

从所述位流中解码第四信息单元,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应;以及

通过将针对各个通道的系数与所述平均张量值相乘并将所述张量的分解的分量与同所述张量的分解相关联的系数相乘来逐个通道地产生所述张量,其中,来自所述位流的所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

2.根据权利要求1所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的全部信息单元能够相对于彼此独立地解码。

3.根据权利要求1所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元是水平布置的用于定义图片的宽度的子图片。

4.根据权利要求1所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元是水平布置的用于定义图片的宽度的子图片,以及所述第四信息单元是位于所述第一信息单元、所述第二信息单元和所述第三信息单元下方的跨越所述图片的宽度并延伸到所述图片的底部的子图片。

5.根据权利要求1所述的方法,其中,所述方法应用于多个张量,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的各信息单元包含针对所述多个张量中的各张量的相应信息。

6.根据权利要求1所述的方法,其中,所述第三信息单元布置在二维样本阵列中,所述样本阵列的纵横比基于所述信息被调整为1:1或尽可能接近1:1。

7.根据权利要求1所述的方法,其中,所述第三信息单元布置在二维样本阵列中,所述样本阵列的纵横比基于所述信息被调整为2:1或尽可能接近2:1。

8.根据权利要求1所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的各信息单元被布置为通用视频编码所用的子图片。

9.根据权利要求1所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的各信息单元被布置为通用视频编码或高效视频编码所用的图块。

10.根据权利要求1所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的各信息单元被布置为通用视频编码或高效视频编码所用的切片。

11.一种将张量编码到位流中的方法,所述方法包括:

对第一信息单元进行编码,所述第一信息单元与跨所述张量的通道的平均张量值相对

应;

对第二信息单元进行编码,所述第二信息单元与所述张量的分解的分量相对应;

对第三信息单元进行编码,所述第三信息单元与所述张量的同所述第一信息单元相关的系数相对应;以及

对第四信息单元进行编码,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应,

其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

12.根据权利要求11所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的全部信息单元能够相对于彼此独立地解码。

13.根据权利要求11所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元是水平布置的用于定义图片的宽度的子图片。

14.根据权利要求11所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元是水平布置的用于定义图片的宽度的子图片,以及所述第四信息单元是位于所述第一信息单元、所述第二信息单元和所述第三信息单元下方的跨越所述图片的宽度并延伸到所述图片的底部的子图片。

15.根据权利要求11所述的方法,其中,所述方法应用于多个张量,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的各信息单元包含针对所述多个张量中的各张量的相应信息。

16.根据权利要求11所述的方法,其中,所述第三信息单元布置在二维样本阵列中,所述样本阵列的纵横比基于所述信息被调整为1:1或尽可能接近1:1。

17.根据权利要求11所述的方法,其中,所述第三信息单元布置在二维样本阵列中,所述样本阵列的纵横比基于所述信息被调整为2:1或尽可能接近2:1。

18.根据权利要求11所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的各信息单元被布置为通用视频编码所用的子图片。

19.根据权利要求11所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的各信息单元被布置为通用视频编码或高效视频编码所用的图块。

20.根据权利要求11所述的方法,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的各信息单元被布置为通用视频编码或高效视频编码所用的切片。

21.一种非暂态计算机可读存储介质,其存储用于执行从位流中解码张量的方法的程序,所述方法包括:

从所述位流中解码第一信息单元,所述第一信息单元与跨所述张量的通道的平均张量值相对应;

从所述位流中解码第二信息单元,所述第二信息单元与所述张量的分解的分量相对应;

从所述位流中解码第三信息单元,所述第三信息单元与所述张量的同所述第一信息单

元相关的系数相对应;

从所述位流中解码第四信息单元,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应;以及

通过将针对各个通道的系数与所述平均张量值相乘并将所述张量的分解的分量与同所述张量的分解相关联的系数相乘来逐个通道地产生所述张量,其中,来自所述位流的所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

22.一种用于从位流中解码张量的解码器,所述解码器被配置为:

从所述位流中解码第一信息单元,所述第一信息单元与跨所述张量的通道的平均张量值相对应;

从所述位流中解码第二信息单元,所述第二信息单元与所述张量的分解的分量相对应;

从所述位流中解码第三信息单元,所述第三信息单元与所述张量的同所述第一信息单元相关的系数相对应;

从所述位流中解码第四信息单元,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应;以及

通过将针对各个通道的系数与所述平均张量值相乘并将所述张量的分解的分量与同所述张量的分解相关联的系数相乘来逐个通道地产生所述张量,其中,来自所述位流的所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

23.一种非暂态计算机可读存储介质,其存储用于执行将张量编码到位流中的方法的程序,所述方法包括:

对第一信息单元进行编码,所述第一信息单元与跨所述张量的通道的平均张量值相对应;

对第二信息单元进行编码,所述第二信息单元与所述张量的分解的分量相对应;

对第三信息单元进行编码,所述第三信息单元与所述张量的同所述第一信息单元相关的系数相对应;以及

对第四信息单元进行编码,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应,

其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

24.一种用于将张量编码到位流中的编码器,所述编码器被配置为:

对第一信息单元进行编码,所述第一信息单元与跨所述张量的通道的平均张量值相对应;

对第二信息单元进行编码,所述第二信息单元与所述张量的分解的分量相对应;

对第三信息单元进行编码,所述第三信息单元与所述张量的同所述第一信息单元相关

的系数相对应;以及

对第四信息单元进行编码,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应,

其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

用于对张量进行编码和解码的方法、设备和系统

[0001] 相关申请的交叉引用

[0002] 本申请根据35 U.S.C. § 119要求2022年4月13日提交的澳大利亚专利申请2022202470的权益,该申请通过整体引用而并入本文,如同在本文中全部阐述一样。

技术领域

[0003] 本发明通常涉及数字视频信号处理,特别地,涉及用于对来自卷积神经网络的张量进行编码和解码的方法、设备和系统。本发明还涉及包括记录有用于使用视频压缩技术对来自卷积神经网络的张量进行编码和解码的计算机程序的计算机可读介质的计算机程序产品。

背景技术

[0004] 视频压缩是用于支持许多应用(包括用于发送和存储视频数据的应用)的普遍存在的技术。已开发了许多视频编码标准并且其他视频编码标准当前正在开发中。视频编码标准化的最新进展已导致形成被称为“联合视频专家组”(JVET)的组。该联合视频专家组(JVET)包括两个标准设置组织(SSO)的成员,即:国际电信联盟(ITU)的电信标准化部门(ITU-T)的研究组16、问题6(SG16/Q6),其还已知为“视频编码专家组”(VCEG);以及国际标准化组织/国际电工协会联合技术协会1/小组协会29/工作组11(ISO/IEC JTC1/SC29/WG11),其还已知为“运动图片专家组”(MPEG)。

[0005] 联合视频专家组(JVET)已开发了名为“通用视频编码”(VVC)的视频压缩标准。

[0006] 卷积神经网络(CNN)是用于应对涉及机器视觉的用例等的新兴技术。CNN典型地包括许多层,诸如卷积层(包括全连接层)等,其中数据以“张量”的形式从一个层传递到下一层。在训练阶段中确定各层的权重,其中在训练阶段中,非常大量的训练数据被传递通过CNN,并且将所确定的结果与同训练数据相关联的地面真值(ground truth)进行比较。应用用于更新网络权重的处理(诸如随机梯度下降等),以迭代细化网络权重,直到网络以期望的准确度水平来进行。当卷积阶段具有大于1的“步幅(stride)”时,来自卷积的输出张量具有比相应的输入张量更低的空间分辨率。与输入张量相比,诸如“最大池化”(或“Maxpool”)等的操作还减少了输出张量的空间大小。最大池化通过将输入张量分割成数据样本组(例如,数据样本的2×2组),并且从各组中选择最大值作为输出张量中相应值的输出,来产生输出张量。利用输入来执行CNN并且逐步地将输入变换为输出的处理通常被称为“推断”。

[0007] 通常,张量具有四个维度,即:批量、通道、高度和宽度。当对视频数据进行推断时,大小为“一”的第一维度“批量”指示一次一个帧被传递通过CNN。当训练网络时,批量维度的值可以增加,使得根据预定的“批量大小”在更新网络权重之前多个帧被传递通过网络。多帧视频可以作为批量维度的大小根据给定视频的帧的数量而增加的单个张量被传递通过。然而,出于与存储器消耗和访问相关的实际考虑,对视频数据的推断通常基于帧来进行。“通道”维度指示给定张量的并发“特征图”的数量。高度和宽度维度指示CNN特定阶段的特征图的大小。通过CNN的通道计数根据网络架构而变化。取决于特定网络层中发生的子采

样,特征图的大小也变化。

[0008] 针对CNN的第一层的输入是图像或视频帧,该图像或视频帧通常被调整大小以与输入到第一层的张量的维度相符合。张量的维度取决于CNN架构,CNN架构通常具有与输入宽度和高度相关的一些维度以及另外的“通道”维度。

[0009] 基于通道维度对张量进行切片会得到“特征图”集,之所以这么说,是因为张量的各切片与相应的输入图像具有某种关系,从而捕获诸如各种边缘类型等的属性。在离网络的输入更远的层,属性可能更抽象。通过将CNN使用特定输入进行任务的结果与所提供的地面真值进行比较来度量CNN的“任务性能”,其中地面真值通常由人类准备且被认为指示“正确”的结果。

[0010] 一旦决定了网络拓扑,随着更多的训练数据变得可用,网络权重可以随时间而更新。也可以重新训练CNN的一部分,从而使网络的(一个或多于一个)其他部分的权重保持不变。在进行相对大量的乘法累加运算并且相对于存储器写入和读取大量的中间张量的情况下,CNN的整体复杂性往往很高。在一些应用中,CNN完全在“云”中被实现,这导致需要高且昂贵的处理能力。在其他应用中,CNN在边缘装置(诸如照相机或移动电话等)中被实现,这导致灵活性更低,但处理负荷更分散。新兴架构涉及将网络拆分成部分,其中一个部分在边缘装置中运行,另一个部分在云中运行。这种分布式网络架构可以称为“协作智能”。协作智能提供了诸如将来自网络的第一部分的部分结果对于若干不同的第二部分重用等的益处,也许各部分针对不同的任务进行了优化。协作智能架构引入了对张量数据的高效压缩的需求,以通过诸如WAN等的网络进行发送。张量数据可以以压缩形式存储在云中的服务器上,或者作为部分处理后的数据存储在装置中,以稍后用于各种任务。

[0011] 预料VVC将特别是随着视频格式的能力的增加(例如,具有更高的分辨率和更高的帧频)而解决针对甚至更高的压缩性能的持续需求、以及解决针对通过WAN的服务递送(其中,带宽成本相对较高)的日益增长的市场需求。VVC可以在当代硅工艺中实现,并且在所实现的性能与实现成本之间提供可接受的折衷。实现成本可以被认为是例如硅面积、CPU处理器负荷、存储器利用率和带宽方面中的一个或多于一个方面。VVC标准的通用性的一部分在于可用于压缩视频数据的工具的广泛选择、以及VVC适合的应用的广泛范围。

[0012] 视频数据包括图像数据的帧序列,各帧包括一个或多于一个颜色通道。通常,需要一个主颜色通道和两个辅颜色通道。主颜色通道通常被称为“亮度”通道,并且(一个或多于一个)辅颜色通道通常被称为“色度”通道。尽管视频数据通常在RGB(红-绿-蓝)颜色空间中显示,但该RGB颜色空间在三个相应分量之间具有高度相关性。编码器或解码器所看到的视频数据表示通常使用诸如YCbCr等的颜色空间。YCbCr将(根据传递函数映射到“亮度”的)辉度集中在Y(主)通道中,并且将色度集中在Cb和Cr(辅)通道中。由于使用了去相关的YCbCr信号,亮度通道的统计信息与色度通道的统计信息明显不同。主要不同在于,在量化之后,与相应亮度通道块的系数相比,色度通道包含给定块的相对较少的有效系数。此外,可以以与亮度通道相比更低的速率(子采样)(例如,在水平方向上为一半且在垂直方向上为一半(已知为“4:2:0色度格式”))对Cb和Cr通道进行空间采样。4:2:0色度格式通常在(诸如互联网视频流式传输、广播电视和Blu-ray™盘上的存储等的)“消费者”应用中使用。当仅亮度样本存在时,所得到的单色帧被称为使用“4:0:0色度格式”。

[0013] VVC标准指定“基于块”的架构,在该架构中,帧首先被分割成已知为“编码树单元”

(CTU)的正方形区阵列。CTU通常占据相对大的区域,诸如128×128个亮度样本等。在使用VV标准时的其他可能的CTU大小是32×32和64×64。然而,各帧的右侧和下方边缘处的CTU的面积可能较小,其中发生隐式拆分以确保CB保持在帧中。与各CTU相关联的是用于亮度通道和色度通道这两者的“编码树”(“共享树”),或者各自用于亮度通道和色度通道的单独的树。编码树定义CTU的区域向一组块(也被称作“编码块”(CB))的分解。当使用共享树时,单个编码树指定用于亮度通道和色度通道这两者的块,在这种情况下,并置编码块的集合被称为“编码单元”(CU)(即,各CU具有用于各颜色通道的编码块)。以特定顺序处理CB以用于编码或解码。由于使用4:2:0色度格式,具有用于128×128个亮度样本区域的亮度编码树的CTU具有用于(与128×128个亮度样本区域并置的)64×64个色度样本区域的相应色度编码树。当单个编码树用于亮度通道和色度通道时,给定区域的并置块的集合通常被称为“单元”,例如上述CU、以及“预测单元”(PU)和“变换单元”(TU)。具有跨越4:2:0色度格式视频数据的颜色通道的CU的单个树得到相应亮度块的一半宽度和高度的色度块。当针对给定区域使用单独的编码树时,使用上述CB以及“预测块”(PB)和“变换块”(TB)。

[0014] 尽管在“单元”和“块”之间存在上述区别,但是术语“块”可以用作帧的区域(area)或区(region)的通用术语,对于该区域或区,操作被应用于所有颜色通道。

[0015] 对于各CU,生成帧数据的相应区域的内容(样本值)的预测单元(PU)(“预测单元”)。此外,形成了预测与在编码器的输入处看到的区域的内容之间的差(或“空间域”残差)的表示。各颜色通道中的差可以被变换和编码为残差系数序列,从而形成给定CU的一个或多于一个TU。所应用的变换可以是应用于残差值的各块的离散余弦变换(DCT)或其他变换。变换被单独地应用(即,分两次进行二维变换)。首先通过对块中的各行样本应用一维变换来变换块。然后,通过对部分结果的各列应用一维变换来变换该部分结果,以产生基本上对残差样本去相关的变换系数的最终块。VVC标准支持各种大小的变换,包括各边的尺寸是2的幂的矩形块的变换。变换系数被量化用于熵编码到位流中。

[0016] VVC表征为帧内预测和帧间预测。帧内预测涉及使用正使用的帧中先前处理的样本以生成该帧中当前数据样本块的预测。帧间预测涉及使用从先前解码的帧中获得的样本块来生成帧中的当前样本块的预测。从先前解码的帧中获得的样本块根据运动向量而相对于当前块的空间位置偏移,该运动向量通常被应用滤波。帧内预测块可以是:(i)均匀样本值(“DC帧内预测”),(ii)具有偏移以及水平和垂直梯度的平面(“平面帧内预测”),(iii)具有在特定方向上应用的相邻样本的块的总体(“角度帧内预测”),或者(iv)使用相邻样本和所选择的矩阵系数的矩阵乘法的结果。通过将“残差”编码到位流中,可以在一定程度上校正预测块和相应输入样本之间的进一步差异。残差通常从空间域变换到频域,以在“主变换域”中形成残差系数,该残差系数可以通过应用“二次变换(secondary transform)”进一步变换,以在“二次变换域”中产生残差系数。根据量化参数对残差系数进行量化,导致解码器处产生的样本的重建准确度损失,但位流中的位率降低。图片序列可以根据使用帧内预测的图片以及使用帧内或帧间预测的图片的指定结构、以及对编码顺序中的在前图片的指定依赖性进行编码,该编码顺序可以与显示或递送顺序不同。“随机访问”配置得到周期性帧内图片,从而形成解码器开始解码位流的进入点。随机访问配置中的其他图片通常使用帧间预测,以根据指定深度的层级结构从显示或递送顺序中在当前图片之前和之后的图片预测其内容。使用显示顺序中在当前图片之后的图片来预测当前图片需要在给定图片的解码

和给定图片的显示(以及从缓冲器的移除)之间进行一定程度的图片缓冲和延迟。

发明内容

[0017] 本发明的目的是基本上克服或至少改善现有布置的一个或多于一个缺点。

[0018] 本公开的一方面提供了一种从位流中解码张量的方法,所述方法包括:从所述位流中解码第一信息单元,所述第一信息单元与跨所述张量的通道的平均张量值相对应;从所述位流中解码第二信息单元,所述第二信息单元与所述张量的分解的分量相对应;从所述位流中解码第三信息单元,所述第三信息单元与所述张量的同所述第一信息单元相关的系数相对应;从所述位流中解码第四信息单元,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应;以及通过将针对各个通道的系数与所述平均张量值相乘并将所述张量的分解的分量与同所述张量的分解相关联的系数相乘来逐个通道地产生所述张量,其中,来自所述位流的所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

[0019] 本公开的另一方面提供了一种将张量编码到位流中的方法,所述方法包括:对第一信息单元进行编码,所述第一信息单元与跨所述张量的通道的平均张量值相对应;对第二信息单元进行编码,所述第二信息单元与所述张量的分解的分量相对应;对第三信息单元进行编码,所述第三信息单元与所述张量的同所述第一信息单元相关的系数相对应;以及对第四信息单元进行编码,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

[0020] 本公开的另一方面提供了一种非暂态计算机可读存储介质,其存储用于执行从位流中解码张量的方法的程序,所述方法包括:从所述位流中解码第一信息单元,所述第一信息单元与跨所述张量的通道的平均张量值相对应;从所述位流中解码第二信息单元,所述第二信息单元与所述张量的分解的分量相对应;从所述位流中解码第三信息单元,所述第三信息单元与所述张量的同所述第一信息单元相关的系数相对应;从所述位流中解码第四信息单元,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应;以及通过将针对各个通道的系数与所述平均张量值相乘并将所述张量的分解的分量与同所述张量的分解相关联的系数相乘来逐个通道地产生所述张量,其中,来自所述位流的所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

[0021] 本公开的另一方面提供了一种用于从位流中解码张量的解码器,所述解码器被配置为:从所述位流中解码第一信息单元,所述第一信息单元与跨所述张量的通道的平均张量值相对应;从所述位流中解码第二信息单元,所述第二信息单元与所述张量的分解的分量相对应;从所述位流中解码第三信息单元,所述第三信息单元与所述张量的同所述第一信息单元相关的系数相对应;从所述位流中解码第四信息单元,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应;以及通过将针对各个通道的系数与所述平

均张量值相乘并将所述张量的分解的分量与同所述张量的分解相关联的系数相乘来逐个通道地产生所述张量,其中,来自所述位流的所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

[0022] 本公开的另一方面提供了一种非暂态计算机可读存储介质,其存储用于执行将张量编码到位流中的方法的程序,所述方法包括:对第一信息单元进行编码,所述第一信息单元与跨所述张量的通道的平均张量值相对应;对第二信息单元进行编码,所述第二信息单元与所述张量的分解的分量相对应;对第三信息单元进行编码,所述第三信息单元与所述张量的同所述第一信息单元相关的系数相对应;以及对第四信息单元进行编码,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

[0023] 本公开的另一方面提供了一种用于将张量编码到位流中的编码器,所述编码器被配置为:对第一信息单元进行编码,所述第一信息单元与跨所述张量的通道的平均张量值相对应;对第二信息单元进行编码,所述第二信息单元与所述张量的分解的分量相对应;对第三信息单元进行编码,所述第三信息单元与所述张量的同所述第一信息单元相关的系数相对应;以及对第四信息单元进行编码,所述第四信息单元与所述张量的同所述第二信息单元相关的系数相对应,其中,所述第一信息单元、所述第二信息单元、所述第三信息单元和所述第四信息单元中的至少一个信息单元能够相对于其他信息单元中的至少一个信息单元独立地解码,并且各信息单元布置在多个二维样本阵列中。

[0024] 还公开了其他方面。

附图说明

[0025] 现在将参考以下的附图和附录来说明本发明的至少一个实施例,其中:

[0026] 图1是示出分布式机器任务系统的示意框图;

[0027] 图2A和图2B形成可以实践图1的分布式机器任务系统的通用计算机系统的示意框图;

[0028] 图3A是示出CNN的主干部的功能模块的示意框图;

[0029] 图3B是示出图3A的残差块的示意框图;

[0030] 图3C是示出图3A的残差单元的示意框图;

[0031] 图3D是示出图3A的CBL模块的示意框图;

[0032] 图4是示出CNN的可替代主干部的功能模块的示意框图;

[0033] 图5是示出作为分布式机器任务系统的一部分的基于通道间去相关的张量编码器的示意框图;

[0034] 图6是示出作为基于通道间去相关的张量编码器的一部分的特征图打包器、打包帧编码器和解包器的示意框图;

[0035] 图7是示出视频编码器的功能模块的示意框图;

[0036] 图8是示出作为分布式机器任务系统的一部分的基于通道间去相关的张量解码器

的示意框图;

[0037] 图9是示出视频解码器的功能模块的示意框图;

[0038] 图10A是示出CNN的头部的示意框图;

[0039] 图10B是示出图10A的升级器模块的示意框图;

[0040] 图10C是示出图10A的检测模块的示意框图;

[0041] 图11是示出CNN的可替代头部的示意框图;

[0042] 图12A和图12B是示出用于保持通道间去相关张量数据的子图片的布置的示意框图;

[0043] 图13是示出保持编码的通道间去相关特征图和相关联的元数据的位流的示意框图;

[0044] 图14示出用于进行CNN的第一部分并对所得的特征图进行去相关和编码的方法;

[0045] 图15示出用于对位流进行解码、重建去相关特征图并进行CNN的第二部分的方法;

[0046] 图16A和图16B示出两个连续的数据集和相关联的特征向量;

[0047] 图17示出用于调整由对从CNN的第一部分产生的张量的连续去相关操作得到的基向量的方法;

[0048] 图18A是示出用于在使用分解进行压缩之前组合来自FPN中不同层的张量的张量减少模块的示意框图;

[0049] 图18B是示出用于在解压缩后从组合张量中提取不同FPN层的张量的张量分离器模块的示意框图;

[0050] 图19示出用于对来自CNN的多层第一部分的张量进行重采样以与来自CNN的多层第一部分的另一张量进行去相关的方法;

[0051] 图20示出用于将从位流重建的张量重采样到CNN的第二部分的原始分辨率的方法;

[0052] 图21示出用于确定用于对来自CNN的第一部分的张量进行编码的基向量的方法;

[0053] 图22示出用于对为了对来自CNN的第一部分的张量进行编码而使用的基向量进行解码的方法;

[0054] 图23是示出网络中的拆分点处的不同层的张量之间的关系以及所得到的用于从张量组合器再现编码张量的选项的示意框图;以及

[0055] 附录A示出位流中的用于表示与基向量打包相关联的元数据的SEI消息格式和相关语义。

具体实施方式

[0056] 在任一个或多于一个附图中参考具有相同附图标记的步骤和/或特征的情况下,除非出现相反意图,否则这些步骤和/或特征为了本说明书的目的而具有相同的(一个或多于一个)功能或(一个或多于一个)操作。

[0057] 分布式机器任务系统可以包括边缘装置,诸如产生中间压缩数据的网络照相机或智能电话等。分布式机器任务系统还可以包括最终装置,诸如对中间压缩数据进行操作以产生任务结果的基于服务器场(“云”)的应用。任务结果的示例包括用于示出对象检测和实例分段任务的性能的mAP(平均精度均值)以及用于度量在多个视频帧上跟踪对象的准确度

的MOTA(多对象跟踪准确度)。其他任务结果的示例包括人姿态估计和动作识别。另外,边缘装置功能性可以体现在云中,并且中间压缩数据可以被存储以供以后处理(根据需要可能用于多个不同的任务)。

[0058] 中间压缩数据的方便形式是压缩视频位流,这归因于高性能压缩标准及其实现的可用性。视频压缩标准通常对一些给定位深度(诸如布置在平面阵列中的10位)的整数样本进行操作。彩色视频具有三个平面阵列,根据应用,例如对应于颜色分量Y、Cb、Cr或R、G、B。CNN通常对张量形式的浮点数据进行操作。与CNN操作的传入视频数据相比,张量通常具有相对较小的空间维度。张量通常具有比彩色视频数据典型的三个通道多的通道。

[0059] 张量通常具有以下维度:帧、通道、高度和宽度。例如,维度[1,256,76,136]的张量将被认为包含二百五十六(256)个特征图(通道),各特征图的大小为136×76。对于视频数据,通常一次对一个帧进行推断(帧值为1),而不是使用包含多个帧的张量。

[0060] VVC支持将图片分割为多个子图片,各子图片可以独立地编码和独立地解码。在一种方法中,各子图片被编码为一个“切片”、或连续的编码CTU序列。还可以利用“图块”机制以将图片分割为多个可独立解码的区。子图片可以以某种灵活的方式指定,其中各种矩形CTU集被编码为相应的子图片。子图片尺寸的灵活定义允许在一个图片中高效地保持需要不同区域的数据类型,避免大的“未使用”区域,即未用于重建张量数据的帧区域。

[0061] 图1是示出分布式机器任务系统100的功能模块的示意框图。将机器任务跨多个系统分布的概念有时被称为“协作智能”(CI)。将特定神经网络分割为两个部分需要在网络中指定“拆分点”。网络中从输入层直到拆分点的层在第一装置中进行,并且压缩所得的(一个或多于一个)中间张量。使用来自第一装置的解压缩的(一个或多于一个)张量进行从拆分点直到网络末端的层。在拆分点处可能有一个或多于一个张量需要压缩。张量的维度可以相同,或者可以不同。在使用“特征金字塔网络”(FPN)的情况下,层的宽度和高度通常相关,使得给定层是先前层的宽度和高度的一半。FPN架构还可以定义在每个交替层上发生的宽度和高度减半。在一些架构中,可以看到相同宽度和高度的多个张量。层之间的通道计数可以相同,或者可以不同。因此,适用于当代CNN中所使用的各种网络拓扑的压缩方法有益于应用于广泛的场景。

[0062] 系统100可以用于实现用于将特征图去相关、打包和量化到平面帧中以对特征图进行编码并从编码数据中解码特征图的方法。系统100可以用于使得相关联的开销数据的计算负担是可接受的,并且关于解码特征图的任务性能对位流的位率变化具有弹性。系统100还可以用于使得张量的量化表示不会不必要地消耗位,这些位在任务性能方面不会提供相称的益处。

[0063] 系统100包括用于以编码视频位流121的形式从CNN主干114生成编码张量数据115a的源装置110。系统100还包括用于对编码视频位流143形式的张量数据进行解码的目的地装置140。通信通道130用于将编码视频位流121从源装置110通信到目的地装置140。在一些布置中,源装置110和目的地装置140可以其中一个或两者包括各自的移动电话手机(例如,“智能电话”)或网络照相机和云应用。通信通道130可以是诸如以太网等的有线连接或者诸如WiFi或5G等的无线连接,包括跨广域网(WAN)或跨自组织连接的连接。此外,源装置110和目的地装置140可以包括在一些计算机可读存储介质(诸如文件服务器中的硬盘驱动器或存储器等)上捕获编码视频数据的应用。

[0064] 如图1所示,源装置110包括视频源112、CNN主干114、张量组合器162、主分量分析(PCA)编码器160和发送器122。视频源112通常包括所捕获的视频帧数据(表示为113)的源,诸如摄像传感器、存储在非暂态记录介质上的先前捕获到的视频序列、或者来自远程摄像传感器的视频馈送等。视频源112还可以是计算机显卡的输出,例如,显示操作系统和在计算装置(例如,平板计算机)上执行的各种应用的视频输出。可以包括摄像传感器作为视频源112的源装置110的示例包括智能电话、视频摄录机、专业摄像机和网络视频照相机。视频源112可以产生独立的图像,或者可以产生时间上连续的图像,即视频。

[0065] CNN主干114接收视频帧数据113并进行整个CNN的特定层,诸如与CNN的“主干”相对应的层,从而输出张量115a。CNN的主干层可以产生多个张量作为例如与视频帧数据113所表示的输入图像的不同空间尺度相对应的输出。不同的空间尺度可以称为“特征金字塔网络”(FPN)架构。例如,如果系统100进行(用于对象检测的)“YOLOv3”网络,则FPN可以得到从主干114输出作为张量115a的与三个层相对应的三个张量,这三个张量具有不同的空间分辨率和通道计数。当系统100正在进行诸如(用于对象检测的)“Faster RCNN X101-FPN”或(用于实例分段的)“Mask RCNN X101-FPN”等的网络时,张量115a包括五层的张量P2～P5。张量组合器162可以通过对一层的张量进行重采样并将所得的张量与另一层(例如,FPN中的相邻层)的张量级联来组合两层,从而产生组合张量115。PCA编码器160接收从张量组合器162输出的张量115。当存在足够的层间相关性以使用比单独去相关层所需的基向量更少的基向量来表示组合层时,层的组合是合适的。层间相关程度是网络本身和所提供的输入数据的属性。层间相关程度许可组合张量的基向量总数与各单独去相关层的张量的基向量数量之和相比有所减少。例如,如果通常每层使用25个基向量,并且级联两层,则级联层所需的基向量数量可以设置为小于50。使用的基向量数量适配于基向量的解释方差的布置能够通过利用这种层间相关性来实现基向量计数的减少,并参考图21和图22进行讨论。在层间相关性受输入数据影响的范围内,张量组合器162可以选择是否组合层,以使基向量面积最小化。在网络被理解为具有一定程度的层间相关性的情况下,可以应用固定配置,以例如对FPN的最小分解层进行下采样,并将结果与FPN的第二最小分解层相组合以进行去相关步骤(即,1480)。如果层的组合不合适或不需要,则张量组合器162不进行任何操作,并且张量115等于张量115a。

[0066] PCA编码器160用于对整个CNN的一个或多于一个内部层进行编码,这些内部层由CNN主干114输出。PCA编码器160首先使用诸如主分量分析(PCA)等的方法对跨各个张量的通道的张量的DC归一化版本进行去相关。通过从张量的各通道中去除“均值特征”来生成张量的DC归一化版本。可以从张量的所有通道中均匀地去除(减去)均值特征,或者可以通过应用正向变换以通道相关方式去除均值特征,从而针对张量的各通道得到相应的“均值系数”。由PCA得到的特征向量的子集用作“基向量”(或者,可互换地,术语“分量”可用于指代基向量)以表示正被编码的张量的通道。张量的各通道使用基向量进行变换以产生系数。基向量、均值特征图和相关联的系数被量化并打包到视频帧中。视频帧使用视频编码器进行编码并输出为位流121。位流121被供给到发送器122以通过通信通道130进行发送,或者位流121被写入存储部132以供以后使用。

[0067] 源装置110支持针对CNN主干114的特定网络。然而,目的地装置140可以针对CNN头150使用若干网络之一。在针对CNN头150使用若干网络之一时,可以存储打包特征图形式的

部分处理的数据,以供以后在进行各种任务时使用,而不需要再次进行CNN主干114的操作。

[0068] 位流121由发送器122通过通信通道130作为编码视频数据(或“编码视频信息”)被发送。在一些实现中,位流121可以被存储在存储用存储器132中,直到以后通过通信通道130被发送(或者代替通过通信通道130的发送),其中存储部132是非暂态存储装置,诸如“闪速”存储器或硬盘驱动器等。例如,可以根据需要通过用于视频分析应用的广域网(WAN)将编码视频数据提供给客户。

[0069] 目的地装置140包括接收器142、PCA解码器170、张量分离器172、CNN头150和CNN任务结果缓冲器152。接收器142从通信通道130接收编码视频数据,并将视频位流143传递到PCA解码器170。PCA解码器170输出解码的张量149a,解码的张量149a被供给到张量分离器172。张量分离器进行张量组合器162的逆操作,以产生提取的张量149。提取的张量149被传递到CNN头150。在层组合不适合或不需要的架构中,张量分离器172不进行任何操作,并且张量149等于张量149a。CNN头150进行以CNN主干114开始的任务的稍后层以产生任务结果151。任务结果151存储在任务结果缓冲器152中。任务结果缓冲器152的内容可以例如经由图形用户接口呈现给用户,或者被提供给基于任务结果来决定某个动作的分析应用,其可以包括向用户的聚合任务结果的概要级呈现。源装置110和目的地装置140各自的功能性也可以被体现在单个装置中,单个装置的示例包括移动电话手机和平板计算机以及云应用。

[0070] 尽管以上说明了示例装置,但源装置110和目的地装置140各自通常经由硬件组件和软件组件的组合可以配置在通用计算机系统内。图2A例示这种计算机系统200,该计算机系统200包括:计算机模块201;诸如键盘202、鼠标指示器装置203、扫描器226、可被配置为视频源112的照相机227、以及麦克风280等的输入装置;以及包括打印机215、显示装置214、以及扬声器217的输出装置。计算机模块201可以使用外部调制器-解调器(调制解调器)收发器装置216来经由连接部221与通信网络220进行通信。可以表示通信通道130的通信网络220可以是WAN,诸如因特网、蜂窝电信网络或私有WAN等。在连接部221是电话线的情况下,调制解调器216可以是传统的“拨号上网”调制解调器。可替代地,在连接部221是高容量(例如,线缆或光学的)连接部的情况下,调制解调器216可以是宽带调制解调器。还可以使用无线调制解调器来进行向通信网络220的无线连接。收发器装置216可以提供发送器122和接收器142的功能性,并且通信通道130可以体现在连接部221中。

[0071] 计算机模块201通常包括至少一个处理器单元205和存储器单元206。例如,存储器单元206可以具有半导体随机存取存储器(RAM)和半导体只读存储器(ROM)。计算机模块201还包括多个输入/输出(I/O)接口,其中这多个输入/输出(I/O)接口包括:音频-视频接口207,其耦接至视频显示器214、扬声器217和麦克风280;I/O接口213,其耦接至键盘202、鼠标203、扫描器226、照相机227以及可选的操纵杆或其他人机接口装置(未例示);以及外部调制解调器216和打印机215所用的接口208。从音频-视频接口207向计算机监视器214的信号通常是计算机显卡的输出。在一些实现中,调制解调器216可以内置于计算机模块201内,例如内置于接口208内。计算机模块201还具有本地网络接口211,其中该本地网络接口211允许计算机系统200经由连接部223耦接至已知为局域网(LAN)的局域通信网络222。如图2A所例示,局域通信网络222还可以经由连接部224连接至广域网220,其中该局域通信网络222通常将包括所谓的“防火墙”装置或具有相似功能性的装置。本地网络接口211可以包括以太网(Ethernet™)电路卡、蓝牙(Bluetooth™)无线布置或IEEE 802.11无线布置;然而,

对于接口211,可以实践多种其他类型的接口。本地网络接口211还可以提供发送器122和接收器142的功能性,并且通信通道130也可以体现在局域通信网络222中。

[0072] I/O接口208和213可以提供串行连接性和并行连接性中的任一个或这两者,其中前者通常根据通用串行总线(USB)标准来实现并且具有相应的USB连接器(未例示)。设置有存储装置209,并且存储装置209通常包括硬盘驱动器(HDD)210。还可以使用诸如软盘驱动器和磁带驱动器等的其他存储装置(未例示)。通常设置有光盘驱动器212以用作数据的非易失性源。可以使用诸如光盘(例如,CD-ROM、DVD、蓝光盘(Blu ray Disc™))、USB-RAM、便携式外部硬盘驱动器和软盘等的便携式存储器装置作为针对计算机系统200的数据的适当源。通常,HDD 210、光盘驱动器212、网络220和222中的任意者还可被配置为作为视频源112进行操作、或者作为为了经由显示器214进行再现所要存储的解码视频数据的目的地而进行操作。系统100的源装置110和目的地装置140可以体现在计算机系统200中。

[0073] 计算机模块201的组件205～213通常经由互连总线204并且以得到相关领域技术人员已知的计算机系统200的传统操作模式的方式进行通信。例如,处理器205使用连接部218耦接至系统总线204。同样,存储器206和光盘驱动器212通过连接部219耦接至系统总线204。可以实践所述布置的计算机的示例包括IBM-PC和兼容机、Sun SPARCstation、Apple Mac™或相似的计算机系统。

[0074] 在适当或期望的情况下,可以使用计算机系统200来实现PCA编码器160和PCA解码器170以及以下所述的方法。特别地,可以将PCA编码器160、PCA解码器170和要说明的方法作为在计算机系统200内可执行的一个或多于一个软件应用程序233来实现。特别地,利用软件233中的在计算机系统200内执行的指令231(参见图2B)来实现PCA编码器160、PCA解码器170和所述方法的步骤。可以将软件指令231形成为各自用于进行一个或多于一个特定任务的一个或多于一个代码模块。还可以将软件分割成两个单独部分,其中第一部分和相应的代码模块进行所述方法,并且第二部分和相应的代码模块管理第一部分和用户之间的用户接口。

[0075] 例如,可以将软件存储在包括以下所述的存储装置的计算机可读介质中。软件从计算机可读介质载入计算机系统200,然后由计算机系统200来执行。具有这样的软件的计算机可读介质或者该计算机可读介质上所记录的计算机程序是计算机程序产品。在计算机系统200中使用该计算机程序产品优选地实现了用于实施视源装置110和目的地装置140以及所述方法的有利设备。

[0076] 通常将软件233存储在HDD 210或存储器206中。该软件从计算机可读介质载入计算机系统200,并且由计算机系统200来执行。因而,例如,可以将软件233存储在光盘驱动器212所读取的光学可读盘存储介质(例如,CD-ROM)225上。

[0077] 在一些实例中,将应用程序233以编码在一个或多于一个CD-ROM 225上并且经由相应的驱动器212进行读取的方式供给至用户,或者可替代地,可以由用户从网络220或222读取应用程序233。更进一步地,还可以将软件从其他计算机可读介质载入计算机系统200。计算机可读存储介质是指将所记录的指令和/或数据提供至计算机系统200以供执行和/或处理的任何非暂态有形存储介质。这种存储介质的示例包括软盘、磁带、CD-ROM、DVD、蓝光盘、硬盘驱动器、ROM或集成电路、USB存储器、磁光盘、或者诸如PCMCIA卡等的计算机可读卡等,而与这些装置在计算机模块201的内部还是外部无关。还可以参与软件、应用程序、指令

和/或视频数据或编码视频数据向计算机模块201的提供的暂态或非有形计算机可读传输介质的示例包括:无线电或红外发送通道以及向着另一计算机或联网装置的网络连接部、以及包括电子邮件发送和网站上所记录的信息等的因特网或内联网。

[0078] 可以执行上述的应用程序233的第二部分和相应的代码模块来实现要渲染或以其他方式表示在显示器214上的一个或多于一个图形用户接口(GUI)。通过典型地对键盘202和鼠标203进行操纵,计算机系统200的用户和应用可以以在功能上可适用的方式对接口进行操纵,以将控制命令和/或输入提供至与这些(一个或多于一个)GUI相关联的应用。还可以实现在功能上可适用的其他形式的用户接口,诸如利用经由扬声器217所输出的语音提示和经由麦克风280所输入的用户声音命令的音频接口等。

[0079] 图2B是处理器205和“存储器”234的详细示意框图。存储器234表示图2A中的计算机模块201可以访问的(包括存储装置209和半导体存储器206的)所有存储器模块的逻辑聚合。

[0080] 在初始对计算机模块201通电的情况下,执行上电自检(power-on self-test,POST)程序250。通常将POST程序250存储在图2A的半导体存储器206的ROM 249中。有时将诸如存储有软件的ROM 249等的硬件装置称为固件。POST程序250检查计算机模块201内的硬件以确保适当工作,并且通常检查处理器205、存储器234(209、206)和通常还存储在ROM 249中的基本输入-输出系统软件(BIOS)模块251,以进行正确操作。一旦POST程序250成功运行,BIOS 251激活图2A的硬盘驱动器210。激活硬盘驱动器210使得经由处理器205执行驻留在硬盘驱动器210上的引导装入程序252。这样将操作系统253载入RAM存储器206,其中在该RAM存储器206上,操作系统253开始操作。操作系统253是处理器205可执行的系统级应用,以实现包括处理器管理、存储器管理、装置管理、存储管理、软件应用接口和通用用户接口等的各种高级功能。

[0081] 操作系统253管理存储器234(209、206),以确保计算机模块201上运行的各处理或应用具有在不与分配至另一处理的存储器冲突的情况下执行的充足存储器。此外,需要适当使用图2A的计算机系统200中可用的不同类型的存储器,以使得各处理可以高效地运行。因此,聚合存储器234并不意图例示如何分配存储器的特定分段(除非另外说明),而是提供计算机系统200可访问的存储器的概述图以及如何使用这种存储器。

[0082] 如图2B所示,处理器205包括多个功能模块,其中这多个功能模块包括控制单元239、算术逻辑单元(ALU)240和有时称为高速缓冲存储器的本地或内部存储器248。高速缓冲存储器248在寄存器区段中通常包括多个存储用寄存器244～246。一个或多于一个内部总线241从功能上使这些功能模块相互连接。处理器205通常还具有用于使用连接部218经由系统总线204与外部装置进行通信的一个或多于一个接口242。存储器234使用连接部219耦接至总线204。

[0083] 应用程序233包括可以包含条件分支指令和循环指令的指令序列231。程序233还可以包括执行程序233时所使用的数据232。将指令231和数据232分别存储在存储器位置228、229、230和235、236、237中。根据指令231和存储器位置228～230的相对大小,如存储器位置230中示出的指令所描述的,可以将特定指令存储在单个存储器位置中。可替代地,如存储器位置228和229中示出的指令段所描述的,可以将指令分段成各自被存储在单独的存储器位置的多个部分。

[0084] 通常,向处理器205给予指令集,其中在该处理器205内执行该指令集。处理器205等待后续输入,其中处理器205通过执行另一指令集来对该后续输入作出反应。可以从多个源中的一个或多于一个源提供各输入,其中该输入包括输入装置202、203中的一个或多于一个所生成的数据、跨网络220、202其中之一从外部源接收到的数据、从存储装置206、209其中之一检索到的数据或者从插入相应的读取器212内的存储介质225检索到的数据(所有这些均在图2A中示出)。执行指令集在一些情况下可能会导致输出数据。执行还可能涉及将数据或变量存储至存储器234。

[0085] PCA编码器160、PCA解码器170和所述方法可以使用存储器234内的相应存储器位置255、256、257中所存储的输入变量254。PCA编码器160、PCA解码器170和所述方法产生存储器234内的相应存储器位置262、263、264中所存储的输出变量261。可以将中间变量258存储在存储器位置259、260、266和267中。

[0086] 参考图2B的处理器205,寄存器244、245、246、算术逻辑单元(ALU)240和控制单元239一起工作以进行微操作序列,其中这些微操作序列是针对构成程序233的指令集中的每个指令进行“提取、解码和执行”周期所需的。各提取、解码和执行周期包括:

[0087] 提取操作,用于从存储器位置228、229、230提取或读取指令231;

[0088] 解码操作,其中在该解码操作中,控制单元239判断提取了哪个指令;以及

[0089] 执行操作,其中在该执行操作中,控制单元239和/或ALU 240执行该指令。

[0090] 之后,可以执行针对下一指令的进一步提取、解码和执行周期。类似地,可以进行存储周期,其中通过该存储周期,控制单元239将值存储至或写入存储器位置232。

[0091] 要说明的图15至图22的方法中的各步骤或子处理与程序233的一个或多于一个分段相关联,并且通常通过处理器205中的寄存器区段244、245、246、ALU 240和控制单元239一起工作以针对程序233的所述分段的指令集中的每个指令进行提取、解码和执行周期,来进行该各步骤或子处理。

[0092] 图3A是示出可以用作CNN主干114的实现的CNN的主干部310的功能模块的示意框图300。主干部114有时被称为“DarkNet-53”,尽管不同的主干也是可能的,从而得到针对各帧的张量115的层的不同数量和维度。

[0093] 如图3A中所示,视频数据113被传递到调整大小器模块304。该调整大小器模块304将视频数据113的各帧调整到适合于CNN主干310处理的分辨率,从而产生调整大小后的帧数据312。如果视频数据113的分辨率已经适合于CNN主干310,则不需要调整大小器模块304的操作。调整大小后的帧数据312被传递到卷积批量归一化泄漏修正线性(convolutional batch normalisation leaky rectified linear(CBL))模块314以产生张量316。CBL 314包含参考图3D所示的CBL模块360所描述的模块。

[0094] CBL模块360将调整大小后的帧数据312的张量361作为输入。该张量361被传递到卷积层362以产生张量363。如果卷积层362具有1的步幅,则张量363具有与张量361相同的空间维度。如果卷积层362具有较大的步幅(诸如2等),则张量363与张量361相比具有较小的空间维度,例如,对于2的步幅,张量363的宽度和高度被减半。无论步幅如何,对于特定的CBL块,张量363的通道维度的大小可以与张量361的通道维度相比而变化。张量363被传递到批量归一化模块364,该模块输出张量365。批量归一化模块364对输入张量363进行归一化,并且应用缩放因子和偏移值来产生输出张量365。缩放因子和偏移值从训练处理中导

出。张量365被传递到泄漏修正线性激活(“LeakyReLU”)模块366以产生张量367。模块366提供“leaky”激活函数,由此张量中的正值被传递通过,并且负值在幅度上被严重降低,例如降低到其先前值的0.1倍。

[0095] 张量316从CBL块314被传递到残差块模块320,诸如在内部包含1个残差单元、2个残差单元和8个残差单元的级联的1+2+8模块等。

[0096] 参考如图3B所示的ResBlock 340描述残差块。ResBlock 340接收张量341。该张量341被零填充模块342进行零填充以产生张量343。张量343被传递到CBL模块344以产生张量345。张量345被传递到残差单元346,。残差单元346包含一系列级联的残差单元。残差单元346中的最后残差单元输出张量347。

[0097] 参考如图3C中所示的ResUnit 350来描述残差单元。ResUnit 350将张量351作为输入。张量351被传递到CBL模块352以产生张量353。张量353被传递到第二CBL单元354以产生张量355。加法模块356将张量355与张量351求和以产生张量357。加法模块356也可以被称为“快捷方式”,因为输入张量351实质上影响输出张量357。对于未经训练的网络,ResUnit 350进行动作以传递通过张量。当进行训练时,CBL模块352和354进行动作以根据训练数据和地面真值数据使张量357偏离张量351。

[0098] 返回到图3A,Res11模块320输出张量322。张量322作为层之一从主干模块310输出,并且还被提供给Res8模块324。Res8模块324是包括八个残差单元(即350)的残差块(即340)。Res8模块324产生张量326。张量326被传递到Res4模块328,并且作为层之一从主干模块310输出。Res4模块是包括四个残差单元(即350)的残差块(即340)。Res4模块328产生张量329。张量329作为层之一从主干模块310输出。总体地,层张量322、326和329作为张量115输出。主干CNN 310可以将分辨率为1088×608的视频帧作为输入,并产生对应于三层的三个张量,这些张量具有以下维度:[1,256,76,136]、[1,512,38,68]、[1,1024,19,34]。对应于三层的三个张量115的另一示例可以是[1,512,34,19]、[1,256,68,38]、[1,128,136,76],其分别在CNN 310中的第75个特征图、第90个特征图和第105个特征图处分离。分离点取决于CNN 310。

[0099] Res11 320、Res8 324和Res4 328各自以与ResBlock 340类似的方式操作。CBL 314、CBL 344和CBL 354各自以与CBL 360类似的方式运行。

[0100] 图4是示出可以用作CNN主干114的实现的、CNN的可替代主干部400的功能模块的示意框图。主干部400实现具有特征金字塔网络的残差网络(“ResNet FPN”),并且是CNN主干114的可替代版本。帧数据113被输入并经由张量409、413、417、421、425被传递通过stem网络408、res2模块412、res3模块416、res4模块420、res5模块424和最大池模块428,其中最大池模块428产生P6张量429作为输出。

[0101] stem网络408包括步幅为二(2)的7×7卷积和最大池化操作。res2模块412、res3模块416、res4模块420和res5模块424进行卷积运算,诸如LeakyReLU激活等。各模块412、416、420和424也经由2的步幅设置将处理后的张量的分辨率减半。张量413、417、421和425各自被分别传递到1×1横向卷积模块446、444、442和440中之一。模块446、444、442和440分别产生张量447、445、443、441。张量441被传递到3×3输出卷积模块470,该模块产生输出张量P5471。张量441也被传递到上采样器模块450以产生上采样张量451。求和模块460对张量443和451求和以产生张量461,张量461被传递到上采样器模块452和3×3横向卷积模块

472。模块472输出P4张量473。上采样器模块452产生上采样张量453。求和模块462对张量445和453求和以产生张量463,该张量463被传递到3×3横向卷积模块474和上采样器模块454。模块474输出P3张量475。上采样器模块454输出上采样张量455。求和模块464对张量447和455求和以产生张量465,张量465被传递到3×3横向卷积模块476。模块476输出P2张量477。上采样器模块450、452和454使用最近邻插值来降低计算复杂性。张量429、471、473、475和477形成CNN主干400的输出张量115。尽管图4示出Faster RCNN网络架构的特定主干部("P层"拆分点),但可以进行向主干和头的不同分割。在张量413、417、421和425处拆分网络被称为"res"拆分点。在张量409处拆分网络被称为"stem"拆分点。在张量447、445、443和441处拆分网络被称为"C层"拆分点。

[0102] 图5是示出作为分布式机器任务系统100的一部分的、与编码器160(或"PCA编码器")的实现相对应的基于通道间去相关的张量编码器的示意框图500。PCA编码器160进行操作以降低输入张量115的维度,使得PCA解码器170可以以最小的保真度损失产生输入张量的重建版本。与帧数据113相比,张量115具有较低的空间分辨率,但具有多得多的通道。例如,RGB或YCbCr格式的帧数据113具有3个通道,而张量可能具有256个通道。因此,直接将张量115的通道打包到视频帧中导致需要编码的帧的大小相对较大。张量的通道虽然内容各异,但表现出一定程度的通道间相关性。当被打包到帧中时,使用视频编码器很难利用通道间相关性,因为打包格式会导致特征图之间存在特定的空间偏移,而视频编码器通常无法利用这种偏移。此外,即使视频编码器发现了通道间相关性以供利用,使用帧间预测或帧内块复制来做到这一点也相当于从一个通道预测另一个通道。通道之间的对应关系将重复到如下的程度:在测试预测模式和搜索运动向量或块向量时可以在视频编码器中发现通道对之间的相关性。

[0103] 在PCA编码器160中使用利用主分量分析(PCA)或等效方法的通道间去相关方法。PCA产生定义"特征空间"的向量("主分量"或"特征向量")集。PCA的输出是一维向量。在所描述的布置中,输出的一维向量被理解为转换为与正进行PCA的张量的宽度和高度的维度相对应的二维形式。各特征向量相对于定义特征空间的其他特征向量是正交的。正交性意味着各特征向量在表达特征图时所做的贡献没有冗余。如果向量集小于输入张量,则从输入张量导出这样的向量集可以称为"分解"或降维。特征空间是张量的通道可以表达为主分量的标量倍数的空间。主分量的排序使得分量中的最大方差量捕获张量中的尚未被在前分量捕获的最大剩余未解释方差量。因此,选择向量集的前N个分量(其中N小于张量的通道计数)使得能够以与原始张量相比、在解码器处重建(或"投影")张量时的精度损失最小的方式在降维空间中表达张量。使用点积运算实现将张量从一个空间变换到另一个空间,其中输入适当地被布置为一维阵列并根据需要进行转置。

[0104] 主分量分析的方法包括"奇异值分解"(SVD),并且可以用作PCA编码器160的(以下关于模块548所述的)分解函数。在548处可以在PCA编码器160中使用的其他分解方法包括QR分解,其可以使用Gram-Schmidt处理、Householder变换或Givens旋转、或LU(下上)分解、或Cholesky分解来进行。各分解方法中所使用的值N可以大大地小于张量115中的通道数。例如,在张量115中有256个通道的情况下,仅使用前25个主分量,PCA解码器170就可以产生足够高保真度的重建张量,以使CNN头150几乎无损地操作,即,就好像CNN主干114和CNN头150之间没有或可以忽略不计的有损压缩一样。

[0105] 主分量或基向量使用有损压缩进行编码,这导致在一些实施例中编码器需要使用基向量的有损表示。在编码器中使用基向量的有损表示确保了用于生成系数的正向变换相对于用于将基向量投影回解码张量的通道中的反向变换更接近可逆。分解模块548通常还需要或受益于对零中心数据的操作,从而产生更准确的基向量作为结果。由于主分量或基向量相对于原点,因此正被分解的张量也必须以原点为中心,即零中心。使用零中心张量允许生成基向量,该基向量是正交的并且具有各连续基向量准确地指示张量中的最大剩余未解释方差量的属性。当连续基向量各自解释张量中的最大剩余未解释方差量时,在给定程度上解释张量所需的基向量数量就会减少。换句话说,零中心张量的分解允许最大程度地减少维度,从而实现最大的压缩效率。

[0106] 张量编码器500接收张量115并将张量115输入到提取均值模块510。为了允许分解模块548对零中心数据进行操作,提取均值模块510进行操作以通过跨张量的通道维度进行平均来生成均值特征图511。均值特征图511的版本(即有损均值特征图522)通过点积模块540使用导出的有损系数538进行正向变换以产生每通道均值特征图542。在模块240处实现的点积是在特征图的“平坦化”版本上进行的,即,从二维阵列重新成形为一维阵列。在对特征图和基向量进行点积时,使特征图和基向量两者平坦化,一个水平平坦化,另一个垂直平坦化。对于本文的其余部分,在提到对(二维)特征图或其他数据进行点积时,重新成形为一维表示以及从一维表示重新成形被理解为隐式的。减法器544从张量中的各通道减去每通道均值特征图542,以产生DC归一化张量546。

[0107] 为了解决浮点域相对于整数样本域的量化以及样本域中可能使用有损编码所导致的差异,PCA编码器160对均值特征图511进行量化、编码、解码和逆量化,以产生有损均值特征图522。均值特征图511由量化器模块512量化,以产生整数均值特征图513。整数均值特征图513被传递到子图片编码器514。子图片编码器514将特征图513打包到子图片中。子图片被编码为位流部515。位流部515对应于跨张量的通道进行平均的均值特征图511的压缩表示。子图片编码器514还输出重建的整数均值特征图516,该重建的整数均值特征图516对应于PCA解码器170中产生的未打包解码均值特征图844,两者都是均值特征图511的有损表示。重建的整数均值特征图516由逆量化器模块520转换回浮点域以产生有损均值特征图522。有损均值特征图522提供整数均值特征图516的浮点表示。

[0108] 由于张量115中各个通道的振幅不同,因此可以使用“均值系数”来改变均值特征图522被去除的程度,该“均值系数”由点积模块524对有损均值特征图522和张量115的各相应通道进行点积而产生。点积模块生成均值系数的集合526,其包括针对张量115的每个通道的一个均值系数。均值系数526由量化器模块528量化到样本域中以产生整数均值系数530。均值系数530被打包到子图片中并由子图片编码器532编码以产生位流部533和有损子图片表示534。位流部533对应于与平均张量值511相关的系数526。表示解码的均值系数的有损子图片表示534被供给到逆量化器模块536。逆量化器模块536输出重建的均值系数538,即均值系数526的有损浮点版本。因此,零平均张量546对应于PCA解码器170中导出的零平均张量852。

[0109] 损失来源包括来自量化和来自有损编码。由于每通道均值特征图542的生成对与均值系数有关的任何损失的敏感性,因此子图片编码器532以比子图片编码器514所需的更高的质量级别(更小的量化步长大小或更低的量化参数)操作。子图片编码器532可以在无

损模式下操作以获得更高的质量,代价是在位流部533中花费更多位。诸如HEVC等的标准为CU提供了“变换量化旁路”模式,作为用于通过将变换和量化处理旁路来在本地提供无损操作的手段。作为代替,诸如VVC等的标准使用足够低的QP,使得量化处理变为无损处理,并且可以独立选择变换跳过模式以避免来自变换的损失。这种无损编码的本地应用允许“混合”有损/无损编码,其中无损编码用于图片的高敏感部,诸如用于包含打包系数的子图片区等。均值或DC去除的结果是更适合分解模块548中的基向量导出的零中心张量,即DC归一化张量546。

[0110] 分解模块进行操作以跨张量546的通道进行分解函数以产生基向量,基向量是能够表示张量的分量,并且分量的数量和表示的准确性之间具有可接受的折衷。分解模块548中可以使用各种分解机制。奇异值分解(SVD)是一种适用于分解模块548的流行的、计算速度快的分解方法。尽管有损编码可用于需要从源装置110传输到目的地装置140的不同类型的数据(均值特征、系数、基向量),但使用数据的样本表示需要从浮点域量化到整数域,整数域的范围受视频编码和解码处理的位深度约束。为了针对量化数据产生单个相干容器,图像帧可以由多个子图片组成,各子图片是被独立编码的子图片,并且重建的版本(即,在相应解码器中产生的子图片的版本)在PCA编码器160中可用。下文将参考图13A和图13B描述子图片的布置。由于使用了逆量化模块520和536,因此在系统500所实现的张量编码处理中考虑了由于有损编码导致的压缩损失。

[0111] 分解模块548产生基向量集550。基向量集550的大小通常是张量115的通道数的一小部分。基向量550定义新的子空间,在该子空间中,张量115可以以最小的保真度损失近似地表达。分解张量时产生的基向量比通道少被称为“降维”。降维提供了一种通过丢弃或不生成对表达张量115的通道贡献最小的基向量来实现有损压缩的手段。例如,分解模块548可以产生255个或25个基向量的输入通道计数的约10%。基向量形成特征空间,该特征空间能够表示张量115中的特征图,同时显著减少维度,从而减少图片或子图片中的编码区域。基向量550由量化器552量化以产生整数基向量张量554。量化器552进行操作以使用基向量550生成张量的(整数)系数。整数基向量张量554被打包到子图片中并由子图片编码器556编码。子图片编码器556输出位流部557,并且作为解包重建子图片的结果,输出重建整数张量558。位流部557对应于由模块548分解张量546而生成的分量550。逆量化器模块560对重建整数张量558进行逆量化以产生重建基向量张量562。

[0112] 点积模块564进行正向变换,生成零平均张量546的各通道与重建基向量张量562中的各基向量的点积,以产生系数的集合566。在零平均张量546中有256个通道并且张量562中有25(二十五)个分量的情况下,有256×25=6400个系数作为集合566输出。量化器模块568将系数566从浮点域量化到整数域。量化器模块568输出一组系数样本570。系数样本570被打包到子图片中,并由子图片编码器572编码成位流部573。位流部573对应于与分量550相关的系数566。位流部515、533、557和573对应于四个子图片,并由子图片位流组合器580组合以产生编码位流121。由于任务性能对PCA解码器170中所看到的重建系数818的保真度的敏感性,子图片编码器572中使用的质量级别相对高于用于编码基向量(即,由子图片编码器556编码基向量)的质量级别。

[0113] 在PCA编码器160的布置中,分解模块548可操作以输出可变数量的基向量550,使得至少张量115的最小方差量由基向量550表达,直至最大分量数M的限制,例如,对于具有

256个通道的张量,最大分量数M为25。“特征值”与基向量550中的各向量相关联,从而指示通过使用相应基向量容纳或“解释”了多少方差。由于各连续基向量解释了张量115中的最大剩余未解释方差量,因此基向量550的特征值的幅度在随着基向量550中的向量前进时减小。可以建立最小方差阈值,诸如85%、90%、95%、98%或其他值等。在前N个特征值的累积和小于最小方差阈值并且N小于或等于分解模块548所生成的基向量的数量的情况下,可以选择最大值N。可替代地,在前N个特征值的累积和大于最小方差阈值并且N小于或等于分解模块548所生成的基向量的数量的情况下,可以选择最小值N。N个基向量的子集作为基向量550输出。用于将基向量550打包到(如关于图12A所述的)子图片1216中并编码到位流部557中的模块552、556在一些布置中可以对前N个基向量进行编码,省略剩余向量(如果有)用于正向变换,打包到子图片1216中。对于方差较小的张量115,所选择的值N可以低于固定阈值,诸如25等。由于选择了N个基向量,因此可以减少要打包在子图片1216中的基向量的面积和要打包在(如关于图12A所述的)子图片1214中的系数的数量,从而减少位流121的编码大小。在网络的拆分点包含多个张量的情况下,例如,在使用FPN的情况下,可以针对FPN中的各层独立地导出值N,从而允许适配于从CNN主干114针对FPN的各层产生的张量中所包含的复杂度(以及因此方差)的差异。PCA编码器160将针对正被压缩的各张量确定的值N编码在位流121中,以使得PCA解码器170能够从解码的图片中解包正确数量的基向量。PCA解码器170使用解包的基向量的数量N来确定基向量张量计数,并使用最坏情况基向量计数M来确定子图片1216中的打包位置。PCA解码器170可以使用当前帧和后续帧的解码值N,在解码出新值N时更新N,以将前N个位置处的基向量解包为来自位流143的解码基向量。

[0114] 为了减少损失,可以无损地编码诸如均值系数530和系数样本570等的敏感数据。可以使用与VVC不同的编码对均值系数530和系数样本570进行编码。代替将530和570打包到相应的子图片中以编码为视频样本,530和570可以使用用于数字列表的可替代通用方案进行编码。一种方案是“DeepCABAC”,其中使用上下文自适应二进制算术编码以及与诸如VVC或HEVC等的标准中用于残差样本的二值化方案类似的二值化方案来对具有聚集在低幅度的幅度和许多零值数字(例如,高斯或类似分布)的数字序列进行高效编码。无论对整数域中的系数使用无损编码还是有损编码,从浮点域到整数域的转换会引入一些损失,这些损失是以其他方式不可避免的(除非浮点值在位流121中表示,这是较不高效的)。在系数530和570在位流121中使用与VVC不同的方案(诸如DeepCABAC等)表示的情况下,DeepCABAC位流可以使用SEI消息(诸如“用户数据SEI消息”或其他专用SEI消息等)与帧相关联。对于各图片编码基向量,相关联的SEI消息伴随图片,以使得能够从系数和基向量重建特征图。

[0115] 量化器512、528、552和568通常都将浮点值量化为与视频帧的位深度所提供的整数样本范围相对应的整数样本范围(例如,在使用10位编码时为0～1023)。量化为样本意味着量化为封闭的值范围,而如果使用VVC的可替代压缩方案,则可以量化为开放(不受限制)的范围。例如,如果子图片编码器532和572使用诸如DeepCABAC等的算法,则无需将整数值拟合到值范围中。当将浮点值拟合到样本范围中时,量化器512、528、552和568各自使用相应的“量化范围”进行操作,该量化范围可以设置为各种粒度。每次量化张量时,可以确定各相应量化器的量化范围。所确定的各量化范围需要在位流121中用信号通知,因此PCA解码器170可以将整数值逆量化回浮点域中的正确范围。

[0116] 位流115是从CNN主干114和CNN头150所形成的神经网络中的阶段所获得的。因此,

张量546也来自CNN主干114和CNN头150所形成的神经网络中的阶段。尽管图5示出对一个张量115的基于PCA的压缩,但是当CNN在存在特征金字塔网络的点处分离时,存在需要压缩的多个张量。当使用FPN时,PCA编码器160对FPN的各张量进行操作。子图片编码器514、532、556和572可以在FPN的张量之间共享,并且打包和解包能够在各个子图片中为与FPN中的各张量相对应的数据保留区域。在使用FPN的情况下,在处理FPN的各层时,量化器512、528、552和568以及逆量化器520、536、560可以使用单独的量化范围。单独的量化范围在位流121中编码,以供PCA解码器170在参考图13所描述的SEI消息中使用。

[0117] 在PCA编码器160的另一布置中,量化器512、528、552和568所确定的量化范围被设置为各个量化器的先前量化范围,并且根据需要增加以适应针对当前PCA编码操作正量化的数据(数据511、526、550和566)的范围。量化范围的设置导致图片之间的记忆效应,这导致量化范围在张量压缩操作序列上“增长”。在张量序列上增长的量化范围导致对特定张量中看到的异常值的敏感性降低,这些异常值将改变诸如基向量等的大部的数据的样本值。在张量序列上增长的量化范围可以抑制在将连续张量压缩操作的打包图片视为视频序列时可以观察到的“闪烁”效应。通过减少亮度偏移(其对于编码器来说难以检测并以压缩形式高效表示),抑制打包图片中的闪烁便于使用视频编码器(例如,编码器514、532、556和572)进行压缩。在视频编码器使用周期性刷新(诸如随机访问配置中的帧内图片等)的情况下,可以仅根据这种图片的当前张量来确定量化范围。对于随机访问序列,非帧内图片可以使用帧间预测编码工具来预测来自先前图片的内容(诸如基向量等),从而由于用于表示连续帧中的基向量的样本值的分布相对稳定,因此可以实现较高的压缩效率。

[0118] 在PCA编码器160和PCA解码器170的另一布置中,当使用FPN时,在层之间的基向量系数的量化范围(QR)之间建立关系。在CNN主干114和CNN头150形成Faster RCNN或Mask RCNN网络的情况下,层N的基向量系数的QR被设置为层N+1的QR的两倍,依此类推。将QR设置为先前层的QR的两倍包括如下的情况:两层级联成单个层以进行张量压缩,例如,下采样的P2和P3级联以形成层N并且P4形成层N+1。如果用信号通知基向量系数的一个QR,则基于层索引导出各层的QR。当CNN主干114和CNN头150形成YOLOv3网络或类似网络时,基于分辨率的排序被反转,使得层N+1的QR是层N的两倍。位流121中的信号通知指示在使用FPN时,在PCA解码器170中要使用哪种排序来对基向量的系数进行逆量化。

[0119] 图6是编码器架构的示意框图600。编码器架构600包括特征图打包器610、打包帧编码器614和解包器620。子图片编码器514、532、556和572被实现为架构600的实例。打包器610接收具有给定通道计数、宽度和高度尺寸并包含整数值(即已经量化)的张量608,例如,子图片编码器514接收整数均值特征图513。打包器610将接收到的张量打包成2D平面样本阵列。通常,在所描述的布置中,张量608的各个通道的特征图以从左到右和从上到下的方式存储在子图片帧612中。子图片帧612需要具有足够的大小以保持张量608的通道,包括对由于特征图大小和子图片帧612尺寸之间不匹配而导致的打包间隙的允许。参考图7描述了通常实现为VVC编码器的打包帧编码器614。编码器614产生对应于相应子图片的编码位流部616。例如,子图片编码器514输出编码位流部515。编码器614还输出重建帧618,该重建帧618对应于在解码位流部616时再现的有损版本。重建帧618表示均值特征图513的重建,其中对由于编码而引起的损失进行建模或表示。损失反映了编码损失,诸如编码器614处使用的特定编码方法所导致的损失等。在图7中描述的示例中,编码器614是VVC编码器。在另一

实现中,编码器614可以是HEVC编码器,这将导致不同的编码损失。无损编解码器可用于编码器614,诸如JPEG LS或DeepCABAC等,在这种情况下,重建帧618等于子图片帧612。如果HEVC或VVC用于编码器614,则它们可以被配置为对系数进行无损编码。

[0120] 重建帧618被传递到解包器620。解包器620提取特征图以产生具有与张量608相同维度的重建张量622。作为模块600的操作的结果,PCA编码器160被允许使用与PCA解码器170中看到的版本相对应的特征图(或系数)的版本,因此以比不考虑损失源时更高级别的保真度进行操作。子图片编码器514、532、556和572被配置为在内部和跨子图片边界禁用环路滤波,因为环路滤波通常针对人类对解码图片的消耗进行了优化。系统100使用视频压缩作为一种用以高效地表示对需要从CNN主干114传播到CNN头150的中间张量数据进行降维所得到的数据的手段。

[0121] 图7是示出视频编码器614的功能模块的示意框图700。视频编码器614对定义整个图片的子图片集合中的一个子图片进行编码。虽然可以在一次编码中将所有子图片编码为一个图片,但使用一次编码意味着PCA编码器160中的反馈无法解决流水线中的有损编码。通常,数据以样本或系数的组(诸如块向固定大小的子块的分割等)或者作为阵列在视频编码器614内的功能模块之间传递。如图2A和图2B所示,可以使用通用计算机系统200来实现视频编码器614,其中可以利用计算机系统200内的专用硬件、利用计算机系统200内可执行的软件(诸如驻留在硬盘驱动器205上并且由处理器205控制其执行的软件应用程序233的一个或多于一个软件代码模块等),来实现各种功能模块。可替代地,可以利用在计算机系统200内可执行的专用硬件和软件的组合来实现视频编码器614。可以可替代地在诸如进行所述方法的功能或子功能的一个或多于一个集成电路等的专用硬件中实现视频编码器614和所述方法。这种专用硬件可以包括图形处理单元(GPU)、数字信号处理器(DSP)、专用标准产品(ASSP)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)或者一个或多于一个微处理器和关联存储器。特别地,视频编码器614包括模块710～790,其中这些模块各自可被实现为软件应用程序233的一个或多于一个软件代码模块。

[0122] 尽管图7的视频编码器614是通用视频编码(VVC)视频编码流水线的示例,但是也可以使用其他视频编解码器来进行本文描述的处理阶段。例如,可以使用HEVC。所述示例生成编码数据的位流。如果使用其他编解码器,则一些实现可以将数据打包成不同的格式,诸如帧格式等。视频编码器614接收子图片帧数据612,诸如子图片的帧序列等,各帧包括一个或多于一个颜色通道。帧数据612可以是所使用的档次支持的任何色度格式和位深度,例如,VVC标准的“主10”档次的4:0:0、4:2:0,样本精度为八(8)至十(10)位。块分区器710首先将帧数据612分割成CTU,通常形状为正方形且被配置使得使用CTU的特定大小。CTU的最大有效大小例如可以是32×32、64×64或128×128亮度样本,由存在于“序列参数集”中的“sps_log2_ctu_size_minus5”句法元素配置。CTU大小还提供最大CU大小,因为没有进一步拆分的CTU将包含一个CU。块分区器710还根据亮度编码树和色度编码树将各CTU分割成一个或多于一个CB。亮度通道也可以被称为主颜色通道。各色度通道也可以被称为辅颜色通道。CB具有各种大小,并且可以包括正方形和非正方形纵横比这两者。然而,在VVC标准中,CB、CU、PU和TU总是具有为二的幂的边长。因此,根据CTU的亮度编码树和色度编码树,从块分区器710输出表示为712的当前CB(根据CTU的一个或多于一个块上的迭代来前进)。

[0123] 可以按光栅扫描顺序扫描由帧数据614的第一分割得到的CTU,并且可以将该CTU

分组成一个或多于一个“切片(slice)”。切片可以是“帧内”(或“I”)切片。帧内切片(I切片)指示切片中的每个CU是帧内预测的。通常,编码层视频序列(CLVS)中的第一图片仅包含I切片,并且被称为“帧内图片”。CLVS可以包含形成“随机访问点”(即,视频序列中可以开始被解码的中间帧)的周期性帧内图片。可替代地,切片可以是单预测或双预测的(分别为“P”或“B”切片),分别指示切片中单预测和双预测的附加可用性。

[0124] 视频编码器614根据图片结构对图片序列进行编码。一种图片结构是“低延迟”,在这种情况下,使用帧间预测的图片可能仅参考序列中先前出现的图片。低延迟使得各图片一被解码就被输出,此外还被存储以供后续图片可能参考用。另一种图片结构是“随机访问”,其中图片的编码顺序与显示顺序不同。随机访问允许帧间预测图片参考其他图片,这些图片尽管已被解码但尚未被输出。需要一定程度的图片缓冲,以便将来按显示顺序的参考图片存在于解码图片缓冲器中,从而导致多帧时延。

[0125] 当使用除4:0:0之外的色度格式时,在I切片中,各CTU的编码树可以在64×64水平以下发散成两个单独编码树,一个用于亮度,且另一个用于色度。使用单独树允许在CTU的亮度64×64区域内在亮度和色度之间存在不同的块结构。例如,大的色度CB可以与许多较小的亮度CB并置,并且反之亦然。在P或B切片中,CTU的单个编码树定义了亮度和色度共同的块结构。单个树的所得块可以是帧内预测的或帧间预测的。

[0126] 除了将图片分割为切片之外,图片还可以分割为“图块”。图块是覆盖图片的矩形区域的CTU序列。CTU扫描以光栅扫描的方式在各图块内进行,并从一个图块前进到下一个图块。切片可以是整数个图块,或给定图块内整数个连续的CTU行。

[0127] 对于各CTU,视频编码器614在两个阶段中操作。在第一阶段(称为“搜索”阶段)中,块分区器710测试编码树的各种潜在配置。编码树的各潜在配置具有相关联的“候选”CB。第一阶段涉及测试各种候选CB以选择提供相对高的压缩效率和相对低的失真的CB。测试通常涉及拉格朗日优化,由此基于速率(即,编码成本)和失真(即,相对于输入帧数据612的误差)的加权组合来评估候选CB。选择“最佳”候选CB(即,具有最低评估速率/失真的CB)用于随后编码到位流616中。包括在候选CB的评估中的是如下选项,该选项用于针对给定区域使用CB,或者根据各种拆分选项进一步拆分区域并用进一步的CB对各较小的所得区域进行编码,或者甚至进一步拆分区域。因此,在搜索阶段中选择编码树和CB这两者本身。

[0128] 视频编码器614针对各CB(例如,CB 712)产生由箭头720指示的预测块(PB)。PB 720是关联的CB 712的内容的预测。减法器模块722产生PB 720和CB 712之间的表示为724的差(或“残差”,指差在空间域中)。差724是PB 720和CB 712中的相应样本之间的块大小差。差724被变换、量化和表示为由箭头736指示的变换块(TB)。PB 720以及关联的TB 736通常例如基于所评估的成本或失真而选自多个可能的候选CB其中之一。

[0129] 候选编码块(CB)是从视频编码器614针对关联的PB和所得到的残差可用的预测模式其中之一得到的CB。当与视频编码器614中的预测PB组合时,TB 736以位流中的附加信号通知为代价来减少经解码CB和原始CB 712之间的差。

[0130] 因而,各候选编码块(CB)(即,与变换块(TB)相结合的预测块(PB))均具有关联的编码成本(或“速率”)和关联的差(或“失真”)。CB的失真通常被估计为样本值的差,诸如绝对差和(SAD)、平方差和(SSD)或应用于差的哈达玛(Hadamard)变换等。模式选择器786使用差724可以确定从各候选PB得到的估计,以确定预测模式787。预测模式787指示针对当前CB

使用特定预测模式(例如,帧内预测或帧间预测)的决定。可以以与残差的熵编码相比明显更低的成本来进行与各候选预测模式相关联的编码成本和相应的残差编码的估计。因此,即使在实时视频编码器中,也可以评估多个候选模式,以确定率失真意义上的最佳模式。

[0131] 根据率失真确定最佳模式通常是使用拉格朗日优化的变型来实现的。

[0132] 可以采用拉格朗日或类似的优化处理来选择(利用块分区器710的)CTU向CB的最佳分区以及从多个可能性中的最佳预测模式的选择这两者。通过在模式选择器模块786中应用候选模式的拉格朗日优化处理,选择具有最低成本度量的帧内预测模式作为“最佳”模式。最低成本模式包括所选择的二次变换索引788,该索引也由熵编码器738编码到位流614中。

[0133] 在视频编码器614的操作的第二阶段(称为“编码”阶段),在视频编码器614中对各CTU的所确定的(一个或多于一个)编码树进行迭代。对于使用单独树的CTU,针对CTU的各64×64亮度区,首先对亮度编码树进行编码,接着对色度编码树进行编码。在亮度编码树内,仅对亮度CB进行编码,并且在色度编码树内,仅对色度CB进行编码。对于使用共享树的CTU,单个树根据共享树的共同块结构描述CU(即,亮度CB和色度CB)。

[0134] 熵编码器738支持使用可变长度和固定长度码字的句法元素的按位编码、以及句法元素的算术编码模式。位流的诸如“参数集”(例如序列参数集(SPS)和图片参数集(PPS))的部分使用固定长度码字和可变长度码字的组合。切片(也称为连续部分)具有使用可变长度编码的切片头,随后是使用算术编码的切片数据。切片头定义特定于当前切片的参数,诸如切片级量化参数偏移等。切片数据包括切片中各CTU的句法元素。可变长度编码和算术编码的使用需要位流的各部分内的顺次解析。这些部分可以用起始代码来描述,以形成“网络抽象层单元”或“NAL单元”。使用上下文自适应二进制算术编码处理来支持算术编码。

[0135] 经算术编码的句法元素由一个或多于一个“bin”(二进制文件)的序列组成。与位一样,bin的值为“0”或“1”。然而,bin未作为离散位被编码在位流614中。bin具有关联预测(或“可能”或“最大概率”)值和关联概率(已知为“上下文”)。当要编码的实际bin与预测值匹配时,对“最大概率符号”(MPS)进行编码。对最大概率符号进行编码在位流614中的消耗位方面相对便宜,包括总计少于一个离散位的成本。当要编码的实际bin与可能值不匹配时,对“最小概率符号”(LPS)进行编码。对最小概率符号进行编码在消耗位方面具有相对高的成本。bin编码技术使得能够使“0”vs“1”的概率歪斜的bin进行高效编码。对于具有两个可能值的句法元素(即,“flag(标志)”),单个bin就足够了。对于具有许多可能值的句法元素,需要bin的序列。

[0136] 可以基于序列中较早bin的值来确定序列中较晚bin的存在。另外,各bin可以与多于一个上下文相关联。特定上下文的选择可以取决于句法元素中的较早bin、相邻句法元素(即,来自相邻块的相邻句法元素)的bin值等。每次对上下文编码bin进行编码时,针对该bin(如果存在)选择的上下文以反映新bin值的方式更新。因此,二进制算术编码方案被称为自适应的。

[0137] 熵编码器738还支持缺少上下文的bin(被称为“旁路bin”)。假定“0”和“1”之间的等概率分布对旁路bin进行编码。因此,各bin在位流614中具有一位的编码成本。不存在上下文节省了存储器并降低了复杂性,因而使用特定bin的值的分布未歪斜的旁路bin。采用上下文和自适应的熵编码器的一个示例在本领域中已知为CABAC(上下文自适应二进制算

术编码器),并且在视频编码中已经采用了该编码器的许多变型。

[0138] QP控制器790确定用于建立量化器734和去量化器740所使用的量化步长大小的量化参数792。较大的量化步长大小导致主变换系数728被量化为较小的值,从而以降低逆变换系数746的保真度为代价来降低位流121的位率。

[0139] 熵编码器738使用上下文编码的和旁路编码的bin的组合来对量化参数792进行编码,并且如果用于当前CB,则对LFNST索引788进行编码。量化参数792在各切片开始时被编码,并且切片内的量化参数792的变化使用“增量QP”句法元素进行编码。在已知为“量化组”的各区域中,最多用信号通知增量QP句法元素一次。量化参数792被应用于亮度CB的残差系数。调整后的量化参数被应用于并置色度CB的残差系数。调整后的量化参数可以包括根据映射表和从偏移列表中选择的CU级偏移从亮度量化参数792进行映射。当与变换块相关联的残差仅在通过应用二次变换而变换成主要系数的这些系数位置中包括有效残差系数时,用信号通知二次变换索引788。

[0140] 使用残差句法对与CB相关联的各TB的残差系数进行编码。残差句法被设计为高效地对具有低幅度的系数进行编码,主要使用算术编码的bin来指示系数的有效性以及较低值的幅度,并为较高幅度的残差系数保留旁路bin。因此,包括非常低的幅度值和有效系数的稀疏放置的残差块被高效压缩。此外,存在两种残差编码方案。如应用变换时所见,常规残差编码方案针对有效系数主要位于TB的左上角的TB进行了优化。变换跳过残差编码方案可用于未进行变换的TB,并且能够高效地对残差系数进行编码,而不管它们在整个TB中的分布如何。

[0141] 多路复用器模块784根据从各候选CB的测试预测模式中选择的所确定的最佳帧内预测模式从帧内预测模块764输出PB 720。候选预测模式不需要包括视频编码器614所支持的每个可设想的预测模式。帧内预测分为三个类型:第一,“DC帧内预测”,其涉及用表示附近重建样本的平均值的单个值填入PB;第二,“平面帧内预测”,其涉及根据平面、利用DC偏移以及从附近重建的相邻样本导出的垂直和水平梯度用样本填入PB。附近重建样本通常包括在当前PB上方的一行重建样本(延伸到PB右侧一定程度)以及在当前PB左侧的一列重建样本(向下延伸超过PB一定程度);以及第三,“角度帧内预测”,其涉及用在特定方向(或“角度”)上跨PB滤波和传播的重建相邻样本填入PB。在VVC中,支持六十五(65)个角度,其中矩形块能够利用正方形块不可用的附加角度,以产生总共八十七(87)个角度。

[0142] 第四类型的帧内预测可用于色度PB,由此根据“交叉分量线性模型”(CCLM)模式从并置亮度重建样本生成PB。三种不同的CCLM模式是可用的,各模式使用从相邻亮度和色度样本导出的不同模型。所导出的模型用于从并置亮度样本生成色度PB的样本块。可以使用参考样本的矩阵乘法,使用从预定义的矩阵集合中选择的一个矩阵来对亮度块进行帧内预测。这种矩阵帧内预测(MIP)通过使用针对视频数据的大集合训练的矩阵来实现增益,其中矩阵表示在角度、平面或DC帧内预测模式下不容易捕获的参考样本和预测块之间的关系。

[0143] 模块764还可以通过使用“块内复制”(IBC)方法从当前帧附近复制块来生成预测单元。参考块的位置被约束为等效于一个CTU的区域(该CTU被分割为已知为VPDU的64×64个区),该区域覆盖各行CTU内以及各切片或图块内的当前CTU的已处理VPDU和先前的(一个或多于一个)CTU的VPDU,直到与一个128×128亮度样本相对应的区域限制,而不管位流的配置CTU大小如何。该区域称为“IBC虚拟缓冲器”,并且限制了IBC参考区域,从而限制所需

的存储部。IBC缓冲器填入有重建样本754(即在环路滤波之前),因此需要与帧缓冲器772分开的缓冲器。当CTU大小为128×128时,虚拟缓冲器包括仅来自当前CTU左侧相邻的CTU的样本。当CTU大小为32×32或64×64时,虚拟缓冲器包括来自当前CTU左侧最多四个或十六个CTU的CTU。无论CTU大小如何,访问相邻CTU以获得IBC参考块的样本都会受到边界(诸如图片、切片或图块的边缘等)的约束。特别是对于具有较小尺寸的FPN层的特征图,使用诸如32×32或64×64等的CTU大小得到更加对准的参考区域以覆盖先前特征图的集合。在基于SAD、SSE或其他差异度量对特征图放置进行排序的情况下,访问类似的特征图以进行IBC预测提供了编码高效的优势。

[0144] 当对特征图数据进行编码时的预测块的残差不同于针对自然视频所见的残差。自然视频通常由图像传感器或屏幕内容捕获,如通常在操作系统用户接口等中看到的那样。特征图残差往往包含许多细节。特征图残差中的细节水平相比于各种变换的显著低频系数,更适合变换跳过编码。发明人为了度量残差编码的益处使用DCT-2、MTS(DST-7、DCT-8水平和垂直组合)以及LFNST(各种经训练的不可分离的变换)进行的实验表明,特征图残差具有足够的局部相似性,以受益于变换编码。然而,特征图残差系数的分布不朝向变换块的DC(左上)系数聚集。换句话说,当对特征图数据进行编码时,存在足够的相关性以使变换表明增益。当使用块内复制来产生特征图数据的预测块时,也存在足够的相关性。因此,当对特征图数据进行编码时,当评估由用于块内复制的候选块向量产生的残差时,可以使用哈达玛成本估计,而不是仅仅依赖于SAD或SSD成本估计。SAD或SSD成本估计倾向于选择具有更适合变换跳过编码的残差的块向量,并且可能错过具有将使用变换进行紧凑编码的残差的块向量。当对特征图数据进行编码时,可以使用VVC标准的多重变换选择(MTS)工具,使得除了DCT-2变换之外,DST-7和DCT-8变换的组合也可水平和垂直地用于残差编码。

[0145] 可以垂直或水平地将帧内预测亮度编码块分区成大小相等的预测块集合,各块具有十六(16)个亮度样本的最小区域。这种帧内子分区(ISP)方法使得单独变换块有助于来自亮度编码块中的一个子分区到下一子分区的预测块生成,从而提高了压缩效率。

[0146] 在先前重建的相邻样本不可用的情况下,例如在帧的边缘,使用样本范围一半的默认半色调值。例如,对于10位视频,使用五百一十二(512)的值。由于针对位于帧的左上位置的CB,没有先前的样本可用,角度和平面帧内预测模式产生与DC预测模式的输出相同的输出(即,以半色调值作为幅度的样本的平坦面)。

[0147] 对于帧间预测,由运动补偿模块780使用来自位流中编码顺序帧中当前帧之前的一个或两个帧的样本产生预测块782,并由多路复用器模块784输出该预测块782作为PB720。此外,对于帧间预测,单个编码树通常用于亮度通道和色度通道这两者。位流中的编码帧的顺序可以与捕获或显示时的帧的顺序不同。当一个帧用于预测时,块被称为“单预测”并且具有一个关联的运动向量。当两个帧用于预测时,块被称为“双预测”并且具有两个关联的运动向量。对于P切片,各CU可以是帧内预测的或单预测的。对于B切片,各CU可以是帧内预测的、单预测的或双预测的。

[0148] 通常使用“图片组”结构对帧进行编码,从而实现帧的时间层级。帧可以被分割成多个切片,各切片对帧的一部分进行编码。帧的时间层级允许帧按照显示帧的顺序参考前一和后一图片。以必要的顺序对图像进行编码,以确保满足对各帧进行解码的相关性。仿射帧间预测模式是可用的,其中不是使用一个或两个运动向量来选择和滤波预测单元的参考

样本块,而是将预测单元分割成多个较小的块,并且产生运动场,使得各较小的块具有不同的运动向量。运动场使用预测单元的附近点的运动向量作为“控制点”。仿射预测允许对不同于平移的运动进行编码,而较少需要使用深度拆分的编码树。VVC可用的双预测模式沿着所选择的轴进行两个参考块的几何混合,并且用信号通知相对于块的中心的角度和偏移。这种几何分区模式(“GPM”)允许沿着两个对象之间的边界使用更大的编码单元,并且针对编码单元的编码的边界的几何形状作为角度和中心偏移。运动向量差可以被编码为方向(上/下/左/右)和距离(支持2的幂距离的集合),而不是使用笛卡尔(x,y)偏移。从相邻块(“合并模式”)获得运动向量预测子,就好像没有应用偏移一样。当前块将与所选择的相邻块共享相同的运动向量。

[0149] 根据运动向量778和参考图片索引来选择样本。运动向量778和参考图片索引应用于所有颜色通道,因此主要在对PU而不是PB的操作方面描述帧间预测。用单个编码树描述各CTU到一个或多于一个帧间预测块的分解。帧间预测方法可以在运动参数的数量及其精度方面变化。运动参数通常包括参考帧索引,该索引指示将使用参考帧列表中的哪个(哪些)参考帧加上各参考帧的空间平移,但是可以包括更多帧、特定帧或复杂仿射参数(诸如缩放和旋转等)。另外,可以应用预定的运动细化处理来基于参考样本块生成密集运动估计。

[0150] 已经确定和选择了PB 720,并且在减法器722处从原始样本块中减去PB 720,获得具有最低编码成本的残差(表示为724),并且对该残差进行有损压缩。有损压缩处理包括变换、量化和熵编码的步骤。正向主变换模块726对差724应用正向变换,将差724从空间域转换到频域,并产生由箭头728表示的主变换系数。一维中的最大主变换大小是由序列参数集中的“sps_max_luma_transform_size_64_flag”配置的32点DCT-2或64点DCT-2变换。如果正被编码的CB大于表示为块大小的最大支持的主变换大小(例如64×64或32×32),则以平铺方式应用主变换726来变换差724的所有样本。在使用非正方形CB的情况下,也使用CB的各维度中的最大可用变换大小来进行平铺。例如,当使用三十二(32)的最大变换大小时,64×16CB使用两个以平铺方式布置的32×16主变换(primary transform)。当CB的大小大于最大支持的变换大小时,以平铺方式用TB填充CB。例如,用处于2×2布置的四个64×64TB填充具有64-pt变换最大大小的128×128CB。用处于2×4布置的八个32×32的TB变换填充具有32-pt最大大小的64×128CB。

[0151] 变换726的应用得到针对CB的多个TB。当变换的各应用对大于32×32(例如64×64)的差724的TB进行操作时,TB的左上32×32区域之外的所有得到的主变换系数728被设置为零(即,被丢弃)。剩余的主变换系数728被传递到量化器模块734。根据与CB相关联的量化参数792对主变换系数728进行量化,以产生主变换系数732。除了量化参数792之外,量化器模块734还可以应用“缩放列表”以通过根据残差系数在TB内的空间位置进一步缩放残差系数来允许TB内的非均匀量化。对于亮度CB与各色度CB,量化参数792可以不同。主变换系数732被传递到正向二次变换模块730,以通过进行不可分离的二次变换(NSST)操作或将二次变换旁路掉来产生由箭头736表示的变换系数。正向主变换726通常是可分离的,从而变换各TB的行集合然后列集合。对于宽度和高度不超过16个样本的亮度TB,正向主变换模块726在水平和垂直方向上使用类型II离散余弦变换(DCT-2),或者在水平和垂直方向上使用变换的旁路,或者在水平或垂直方向上使用类型VII离散正弦变换(DST-7)和类型VIII离散

余弦变换(DCT-8)的组合。在VVC标准中,使用DST-7和DCT-8的组合被称为“多变换选择集”(MTS)。

[0152] 模块730的正向二次变换通常是不可分离的变换,其仅被应用于帧内预测CU的残差,并且尽管如此也可以被旁路掉。正向二次变换对十六(16)个样本(布置为主变换系数728的左上4×4子块)或四十八(48)个样本(布置为主变换系数728的左上8×8系数中的三个4×4子块)进行操作,以产生二次变换系数集合。二次变换系数集合在数量上可以少于从中导出它们的主变换系数集合。由于二次变换仅应用于彼此相邻且包括DC系数的系数集合,因而二次变换被称为“低频不可分离的二次变换”(LFNST)。这种二次变换可通过训练处理获得,并且由于其不可分离性质和经训练的原点,可利用残差信号中的不能通过可分离变换(诸如DCT和DST的变型等)捕获的附加冗余。此外,当应用LFNST时,TB中的所有剩余系数在主变换域和二次变换域这两者中都为零。

[0153] 量化参数792对于给定TB是恒定的,并且因此得到针对TB的在主变换域中的残差系数的产生的均匀缩放。量化参数792可以随着用信号通知的“增量量化参数”而定期地变化。对于包含在给定区域内的CU(称为“量化组”),用信号通知增量量化参数(增量QP)一次。如果CU大于量化组大小,则利用CU的TB之一,用信号通知增量QP一次。即,通过熵编码器738针对CU的第一量化组用信号通知增量QP一次,而不针对CU的任何后续量化组用信号通知增量QP。通过应用“量化矩阵”也可以实现非均匀缩放,由此针对各残差系数所应用的缩放因子从量化参数792与缩放矩阵中的相应条目的组合导出。缩放矩阵可以具有小于TB的大小的大小,并且当应用于TB时,最近邻方法用于从大小小于TB大小的缩放矩阵提供针对各残差系数的缩放值。残差系数736被供给至熵编码器738以在位流614中编码。通常,根据扫描图案,扫描TU的具有至少一个有效残差系数的各TB的残差系数以产生值的有序列表。扫描图案通常将TB扫描为4×4“子块”的序列,从而以4×4组残差系数的粒度提供常规扫描操作,其中子块的布置取决于TB的大小。各子块内的扫描和从一个子块到下一子块的进展通常遵循反向对角线扫描模式。另外,使用增量QP句法元素将量化参数792编码到位流614中,并且给定切片或子图片中的初始值的切片QP和二次变换索引788被编码到位流614中。

[0154] 如上所述,视频编码器614需要访问与视频解码器中看到的解码帧表示相对应的帧表示。因此,残差系数736被传递通过逆二次变换模块744,该逆二次变换模块744根据二次变换索引788操作以产生由箭头742表示的中间逆变换系数。由去量化器模块740根据量化参数792对中间逆变换系数742进行逆量化,以产生由箭头746表示的逆变换系数。去量化器模块740还可以使用缩放列表进行残差系数的逆非均匀缩放,这对应于在量化器模块734中进行的正向缩放。逆变换系数746被传递到逆主变换模块748,以产生TU的残差样本(由箭头750表示)。逆主变换模块748水平和垂直地应用DCT-2变换,这受参考正向主变换模块726描述的最大可用变换大小的约束。由逆二次变换模块744进行的逆变换的类型与由正向二次变换模块730进行的正向变换的类型相对应。由逆主变换模块748进行的逆变换的类型与由主变换模块726进行的主变换的类型相对应。求和模块752将残差样本750和PU 720相加以产生CU的重建样本(由箭头754指示)。

[0155] 重建样本754被传递到参考样本高速缓冲存储器756和环内滤波器模块768。通常使用ASIC上的静态RAM实现以避免昂贵的片外存储器访问的参考样本高速缓冲存储器756提供了满足用于为帧中的后续CU生成帧内PB的依赖关系所需的最小样本存储。最小依赖关

系通常包括沿着一行CTU的下方的样本的“线缓冲器”,以供下一行CTU以及列缓冲(其范围由CTU的高度设置)使用。参考样本高速缓冲存储器756将参考样本(由箭头758表示)供给至参考样本滤波器760。样本滤波器760应用平滑操作以产生滤波参考样本(由箭头762指示)。滤波参考样本762由帧内预测模块764使用以产生由箭头766表示的样本的帧内预测块。对于各候选帧内预测模式,帧内预测模块764产生样本块(即766)。通过模块764使用诸如DC、平面或角度帧内预测等的技术来生成样本块766。也可以使用矩阵乘法方法来产生样本块766,其中相邻参考样本作为输入,并且由视频编码器614从矩阵集合中选择矩阵,使用索引在位流614中用信号通知所选择的矩阵来标识矩阵集合中的哪个矩阵将被视频解码器使用。

[0156] 环内滤波器模块768对重建样本754应用数个滤波阶段。滤波阶段包括“去块滤波器”(DBF),该DBF应用与CU边界对准的平滑化,以减少由不连续而产生的伪影。环内滤波器模块768中存在的另一滤波阶段是“自适应环路滤波器”(ALF),该ALF应用基于Wiener的自适应滤波器以进一步降低失真。环内滤波器模块768中的另一可用滤波阶段是“样本自适应偏移”(SAO)滤波器。SAO滤波器通过首先将重建样本分类为一个或多于一个类别、并且根据所分配的类别在样本级应用偏移来操作。

[0157] 从环内滤波器模块768输出由箭头770表示的滤波样本。滤波样本770被存储在帧缓冲器772中。帧缓冲器772通常具有存储数个(例如,多达十六(16)个)图片的容量,因而存储在存储器206中。由于所需的大存储器消耗,通常不使用片上存储器来存储帧缓冲器772。如此,对帧缓冲器772的访问在存储器带宽方面是昂贵的。帧缓冲器772将参考帧(由箭头774表示)提供至运动估计模块776和运动补偿模块780。参考帧774作为相应子图片编码器模块600(514、532、556、572)的重建帧618输出并提供给解包器模块620。在图7的示例中,重建帧是有损VVC编码操作的结果,即由于模块726至748和752至754的操作。

[0158] 运动估计模块776估计多个“运动向量”(表示为778),其各自是相对于当前CB的位置的笛卡尔空间偏移,从而参考帧缓冲器772中的参考帧其中之一中的块。针对各运动向量产生参考样本的滤波块(表示为782)。滤波参考样本782形成可供模式选择器786的潜在选择用的进一步候选模式。此外,对于给定CU,PB 720可以使用一个参考块(“单预测”)形成,或者可以使用两个参考块(“双预测”)形成。对于所选择的运动向量,运动补偿模块780根据支持运动向量中的子像素精度的滤波处理来产生PB 720。如此,运动估计模块776(其对许多候选运动向量进行操作)与运动补偿模块780(其仅对所选择的候选进行操作)相比可以进行简化滤波处理,以实现降低的计算复杂性。当视频编码器614针对CU选择帧间预测时,运动向量778被编码到位流616中。

[0159] 尽管参考通用视频编码(VVC)说明了图7的视频编码器614,但其它视频编码标准或实现也可以采用模块710～790的处理阶段。帧数据612(和位流616)也可以从存储器206、硬盘驱动器210、CD-ROM、蓝光盘或其它计算机可读存储介质中读取(或者被写入存储器206、硬盘驱动器210、CD-ROM、蓝光盘或其它计算机可读存储介质)。另外,帧数据612(和位流616)可以从外部源(诸如连接至通信网络220的服务器或者射频接收器等)接收(或者被发送至该外部源)。通信网络220可能提供有限的带宽,从而需要在视频编码器120中使用速率控制,以避免有时在帧数据612难以压缩时使网络饱和。

[0160] 位流616可以由一个或多于一个切片构建,该切片表示帧数据612的空间区段(CTU

的集合)、由视频编码器614的一个或多于一个实例产生、在处理器205的控制下以协调的方式操作。位流616还可以包含与一个子图片相对应的一个切片,以作为形成一个图片的子图片集合输出,各子图片相对于图片中的任何其他切片或子图片可独立编码和独立解码。独立编码和解码图片中的任何子图片的能力允许通过在张量压缩的稍后阶段使用特征图或系数的有损版本,来在PCA编码器160中考虑到有损压缩对任何给定子图片中所包含的打包特征图或系数的影响。

[0161] 图8是示出作为分布式机器任务系统100的一部分的基于通道间去相关的张量解码器170的实现的示意框图800。视频位流143被传递到图片解码器804,该图片解码器804实现VVC视频解码器并解码由PCA编码器160生成的位流。图片解码器804解码视频位流143中存在的子图片,各子图片对应于重建张量149所需的不同数据类型。各解码的子图片为解码的张量提供信息单元,该单元对应于平均张量值511、系数526、基向量550和系数566其中之一。均值特征子图片840由图片解码器804输出并传递到解包器842。解包器842基于张量149的宽度和高度来提取整数均值特征图844。整数均值特征图844被传递到逆量化器846,其中使用例如从位流143获得的合适量化范围进行从样本域到浮点域的转换,从而得到解码的均值特征图848。

[0162] 均值系数子图片830也由图片解码器804输出并传递到解包器832。解包器832基于张量149的通道计数提取整数均值系数834。整数均值系数834被传递到逆量化器836。逆量化器836使用从位流143获得的量化范围将均值系数834从整数域转换到浮点域,以输出浮点均值系数838。点积模块850通过对均值特征848与均值系数838中的相应系数进行点积来产生张量149的各通道的均值特征图852。逆量化器846、836、824和816是在实现500中用于对位流进行编码的相应量化函数(512、528、552和568)的逆。

[0163] 包含在量化和有损压缩之前由分解模块548(例如,进行PCA)生成的打包基向量的子图片820也由图片解码器804输出并传递到解包器822。解包器822将整数基向量823提取为以不重叠方式放置在子图片820中的一系列特征图。整数基向量823被传递到逆量化器824。逆量化器824构建浮点基向量826,从而应用从位流143获得的量化范围。

[0164] 包含系数(针对张量143的每个特征图具有一个系数)的子图片810也由图片解码器804输出并传递到解包器812。解包器812基于张量149的维度从子图片810中提取各系数。每个特征图或通道的每个基向量需要一个系数,并且该系数表示各基向量在重建给定特征图或通道中的贡献,从而共同形成从解包器812输出的整数系数814。逆量化器816根据从位流143获得的量化范围将整数系数814从整数域转换为浮点域,从而输出浮点系数818。零中心张量854由点积模块828产生,该点积模块828对系数818和基向量826进行点积。求和模块856将零中心张量854与均值特征图848相加,以产生张量149作为来自PCA解码器170的输出。PCA解码器170中的操作流程将压缩张量的维度恢复回原始维度,从而允许将张量149传递到CNN头150以进行给定的机器任务。

[0165] 如后面关于图12A和图12B所述,各解码图片840、830、820和810中的信息布置在多个二维样本阵列中。在以上关于图8所述的示例中,各解码图片840、830、820和810可独立解码。在其他实现中,根据量化器512、528、552和568的输出如何打包并编码成位流,解码图片840、830、820和810中的至少一个相对于其他可独立解码。例如,可以进行解码图片840、830、820和810的独立和/或并发解码,以便于以比单个解码器顺序解码解码图片840、830、

820和810所能实现的帧速率更高的帧速率进行操作。例如,当目的地装置140在云中的服务器中实现时,解码性能可能受益于分布在处理器205中的多个核上,尤其是在没有硬件模块可用于减轻计算要求更高的操作时。

[0166] 采用各种基向量(820)和相关联的系数(830和810)的形式的张量149的降维表示使得在将向量和系数按通道打包到图片或子图片中时能够减少面积。与尝试直接压缩包含张量115的打包特征图或通道的帧相比,打包数据的面积减少会产生更高效的压缩表示。

[0167] 图9中示出图片解码器804(也称为视频解码器)的示例实现900。尽管图9的视频解码器804是通用视频编码(VVC)视频解码流水线的示例,但其它视频编解码器也可用于进行本文所述的处理阶段,例如HEVC等。如图9所示,位流910(其对应于位流143)被输入到视频解码器804。位流143可以从存储器206、硬盘驱动器210、CD-ROM、蓝光盘或其它非暂态计算机可读存储介质中读取,并且作为位流910被提供到实现900。可替代地,位流910可以从外部源(诸如连接至通信网络220的服务器或者射频接收器等)接收。位流143包含表示要解码的所捕获帧数据的编码句法元素。在子图片被独立解码的情况下,可以将与各子图片相对应的位流143的部分供给到实现900的单独实例。针对各子图片的实现910的单独实例允许并行解码子图片以提高吞吐量。

[0168] 位流910被输入到熵解码器模块920。熵解码器模块920通过对“bin”序列进行解码来从位流910中提取句法元素,并将句法元素的值传递给视频解码器804中的其他模块。熵解码器模块920使用可变长度和固定长度解码来解码SPS、PPS或切片头,使用算术解码引擎将切片数据的句法元素解码为一个或多于一个bin的序列。各bin可以使用一个或多于一个“上下文”,其中上下文描述了要用于编码bin的“一”和“零”值的概率等级。在多个上下文可用于给定bin的情况下,进行“上下文建模”或“上下文选择”步骤以选择可用上下文中的一个对bin进行解码。解码bin的处理形成顺次反馈环路,因此各切片可以由给定的熵解码器920实例被解码在切片的整体中。单个(或几个)高性能熵解码器920实例可以从位流910中解码帧或图片的所有切片或子图片,多个较低性能熵解码器920实例可以并发地从位流910中解码帧的切片。

[0169] 熵解码器模块920应用算术编码算法,例如“上下文自适应二进制算术编码”(CABAC),以从位流910中解码句法元素。解码的句法元素被用于重建视频解码器804内的参数。参数包括残差系数(由箭头924表示)、量化参数974、二次变换索引970和诸如帧内预测模式等的模式选择信息(由箭头958表示)。模式选择信息还包括诸如运动向量以及将各CTU分区为一个或多于一个CB等的信息。通常与来自先前解码的CB的样本数据相结合,参数被用于生成PB。

[0170] 残差系数924被传递到逆二次变换模块936,在逆二次变换模块936处,根据二次变换索引应用二次变换或者不进行操作(旁路)。逆二次变换模块936从二次变换域系数产生重建变换系数932,即主变换域系数。重建变换系数932被输入到去量化器模块928。去量化器模块928对残差系数932(即在主变换系数域中)进行逆量化(或“缩放”),以根据量化参数974创建由箭头940表示的重建中间变换系数。去量化器模块928还可以应用缩放矩阵来提供TB内的非均匀去量化,这对应于去量化器模块740的操作。如果在位流143中指示使用非均匀逆量化矩阵,则视频解码器804从位流143中读取量化矩阵作为缩放因子序列,并将缩放因子布置成矩阵。逆缩放将量化矩阵与量化参数组合使用以创建重建中间变换系数940。

[0171] 重建变换系数940被传递到逆主变换模块944。模块944将系数940从频域变换回空间域。逆主变换模块944水平和垂直地应用逆DCT-2变换,这受如参考正向主变换模块726所述的最大可用变换大小的约束。模块944的操作结果是由箭头948表示的残差样本块。残差样本块948在大小上等于相应的CB。残差样本948被供给至求和模块950。

[0172] 在求和模块950处,残差样本948被加到解码PB(表示为952),以产生由箭头956表示的重建样本块。重建样本956被供给至重建样本高速缓冲存储器960和环内滤波模块988。环内滤波模块988产生表示为992的帧样本的重建块。帧样本992被写入帧缓冲器996。帧缓冲器996输出与图1的张量149相对应的图像或视频帧914。

[0173] 重建样本高速缓冲存储器960以类似于视频编码器614的重建样本高速缓冲存储器756的方式操作。重建样本高速缓冲存储器960在无存储器206的情况下(例如,通过作为代替使用通常是片上存储器的数据232)为对后续CB进行帧内预测所需的重建样本提供存储。由箭头964表示的参考样本是从重建样本高速缓冲存储器960获得的,并被供给至参考样本滤波器968以产生由箭头972表示的滤波参考样本。滤波参考样本972被供给至帧内预测模块976。模块976根据在位流910中用信号通知的并由熵解码器920解码的帧内预测模式参数958,产生由箭头980表示的帧内预测样本的块。帧内预测模块976支持编码器侧模块764的模式,包括IBC和MIP。使用诸如DC、平面或角度帧内预测等的模式来生成样本块980。

[0174] 当CB的预测模式在位流143中被指示使用帧内预测时,帧内预测样本980经由多路复用器模块984形成解码PB 952。帧内预测产生样本的预测块(PB),PB是一个颜色分量中的块,其是使用同一颜色分量中的“相邻样本”导出的。相邻样本是与当前块邻近的样本,并且由于处于块解码顺序中的前面,因此已经被重建。在亮度块和色度块并置的情况下,亮度块和色度块可以使用不同的帧内预测模式。然而,两个色度CB共享相同的帧内预测模式。

[0175] 当CB的预测模式在位流910中被指示为帧间预测时,运动补偿模块934产生表示为938的帧间预测样本的块。使用由熵解码器920从位流143中解码的运动向量和参考帧索引从帧缓冲器996中选择和过滤样本块998,来产生帧间预测样本块938。样本块998是从帧缓冲器996中所存储的先前解码帧获得的。为了进行双预测,产生两个样本块并将这两个样本块混合在一起以产生用于解码PB 952的样本。帧缓冲器996填入有来自环内滤波模块988的滤波块数据992。与视频编码器614的环内滤波模块768一样,环内滤波模块988应用DBF、ALF和SAO滤波操作中的任意者。一般地,运动向量应用于亮度和色度通道这两者,尽管用于亮度和色度通道中的子样本插值的滤波处理不同。

[0176] 图7和图9中未示出如下的模块,该模块用于在编码之前对视频进行预处理和在解码之后对视频进行后处理以使样本值移位,从而实现各色度通道内样本值的范围的更均匀使用。在视频编码器614中导出多段线性模型,并且在位流中用信号通知多段线性模型,以供视频解码器804使用,从而撤销样本移位。线性模型色度缩放(LMCS)工具为特定颜色空间和内容提供压缩益处,该特定颜色空间和内容在其样本空间的使用中具有一些不均匀性,尤其是有限范围的利用,这可能导致来自量化的应用的更高质量损失。

[0177] 图10A是示出用于对象检测的CNN的头部150的示例实现1000的示意框图,其中CNN的头部150对应于“YOLOv3”网络中除“DarkNet-53”主干部之外的部分。当CNN主干例如如图3A中实现时,可以使用实现1000。根据要在目的地装置140中进行的任务,不同的网络可以替代CNN头150。传入张量149被分成各层的张量(即,张量1010、1020和1034)。张量1010被传

递到CBL模块1012以产生张量1014。张量1014被传递到检测模块1016和升级器模块1022。检测模块以检测张量的形式输出边界框1018。边界框1018被传递到非最大抑制(NMS)模块1048。

[0178] 为了产生寻址原始视频数据113中的坐标的边界框,在针对网络114的主干部调整大小之前,在升级器模块1022处进行原始视频宽度和高度的缩放。升级器模块1022接收张量1014和张量1020并产生升级的张量1024,该张量1024被传递到CBL模块1026。CBL模块1026产生张量1028作为输出。张量1028被传递到检测模块1030和升级器模块1036。检测模块1030产生检测张量1032,该检测张量1032被供给到NMS模块1048。升级器模块1036是模块1020的另一实例。升级器模块1036接收张量1028和张量1034并输出升级的张量1038。升级的张量1038被传递到CBL模块1040,该模块1040将张量1042输出到检测模块1044。CBL模块1012、1026和1040各自包含五个CBL模块的级联,例如图3D中所示的CBL模型360。升级器模块1022和1036各自是图10B所示的升级器模块1060的实例。模块1048接收张量1018、1032和1036并输出任务结果151。

[0179] 如图10B所示,升级器模块1060接受张量1062(例如图10A中的张量1014)作为输入。张量1062被传递到CBL模块1066以产生张量1068。张量1068被传递到上采样器1070以产生上采样张量1072。级联模块1074通过将上采样张量1072与第二输入张量1064(例如图10A中输入到升级器1022的张量1020)级联来产生张量1076。

[0180] 检测模块1016、1030和1044是图10C中所示的检测模块1080的实例。检测模块1080接收张量1082。张量1082输入到CBL模块1084。CBL模块1084生成张量1086。张量1086被传递到卷积模块1088,该卷积模块1088实现检测核。在一些布置中,检测核应用1×1核以在张量的三层中的各层上产生关于特征图的输出。检测核是1×1×(B×(5+C)),其中B是特定单元可以预测的边界框的数量,通常是三(3)个,并且C是类的数量,其可以是八十(80)个,使得核大小为二百五十五(255)个检测属性(即张量1090)。常数“5”表示四个边界框属性(框中心x,y和大小比例x,y)和一个对象置信度级别(“对象性”)。检测核的结果具有与输入特征图相同的空间维度,但是输出的深度对应于检测属性。检测核被应用在各个层,通常是三个层,这得到大量候选边界框。NMS模块1048对得到的边界框应用非最大抑制处理,以丢弃冗余框,诸如相似比例的重叠预测,从而得到边界框的最终集合作为对象检测的输出。

[0181] 图11是示出CNN的可替代头部1100的示意框图。头部1100可以实现为CNN头150,其中CNN主干114例如实现为主干400。头部1100形成被称为“faster RCNN”的整体网络的一部分,并且包括特征网络(即,主干部400)、区建议网络和检测网络。头部1100的输入是张量149,其包括P2至P6层张量1110、1112、1114、1116和1118。P2至P6层张量1110、1112、1114、1116和1118对应于图4的P2至P6输出477、475、473、471和429。P2至P6张量1110、1112、1114、1116和1118输入到区建议网络(RPN)头模块1120。RPN头模块1120对输入张量进行卷积,从而生成中间张量。中间张量被馈送到两个后续的同级层中,(i)一个层用于分类,以及(ii)一个层用于边界框或“感兴趣区”(ROI)回归。结果输出是分类和边界框1122。分类和边界框1122被传递到NMS模块1124。NMS模块1124通过去除得分较低的重叠框来修剪冗余边界框,以产生修剪后的边界框1126。边界框1126被输入到感兴趣区(ROI)池化器1128。ROI池化器1128使用(下文将进一步描述的)张量149的一些层张量和边界框1126,以使用最大池化操作从各种输入大小图中产生固定大小的特征图。在最大池化操作中,子采样取各组输入值

中的最大值,以在输出张量中产生一个输出值。

[0182] 在CNN主干400和CNN头1100的布置中,从(作为张量149接收的)输出张量115中省略“P6”层张量429。在省略P6张量的布置中,CNN头1100通过对P5张量1116进行步幅等于2的“Maxpool”操作来产生P6输入张量1118。由于P6层可以从P5层重建,因此无需在PCA编码器160和PCA解码器170中单独编码和解码作为显式FPN层的P6层。

[0183] ROI池化器1128的输入是P2～P5特征图1110、1112、1114和1116、以及感兴趣区建议1126。来自1126的各建议(ROI)与特征图(1110～1116)的一部分相关联,以产生固定大小的图。固定大小的图的大小独立于特征图1110～1116的底部。例如,根据以下规则,选择特征图1110～1116中的一个,使得所得到的裁剪图具有足够的细节:floor(4+log2(sqrt(box_area)/224)),其中224是规范的框大小。ROI池化器1128进行操作以根据建议1126裁剪传入的特征图,从而产生张量1130。

[0184] 张量1130被馈送到全连接(FC)神经网络头1132。FC头1132进行两个全连接层以产生类得分和边界框预测子增量张量1134。类得分通常是80个元素的张量,各元素对应于相应对象类别的预测得分。边界框预测增量张量是80×4＝320个元素的张量,包含相应对象类别的边界框。最终处理由输出层模块1136进行,输出层模块1136接收张量1134并进行滤波操作以产生滤波后的张量1138。低得分(低分类)对象不再被进一步考虑。非最大抑制模块1140接收滤波后的张量1138,并通过去除具有较低分类得分的重叠框来去除重叠的边界框,从而得到与张量151相对应的推断输出张量1140。

[0185] 图12A和图12B是示出例如由子图片位流组合器580实现的图片1200向子图片1210、1212、1214和1216的分割的示意框图。各子图片由编码器514、532、556和572中的相应编码器的打包器610打包。各子图片包括以二维样本阵列布置的信息。参考图12B,子图片1210保持均值特征图,诸如FPN的张量中的张量的均值特征图1220等。均值特征图1220对应于图5的布置中的子图片编码器514的位流515。子图片1212保持对应于图5的533的均值系数1222。均值系数1222以值行布置,每个编码层或FPN层对应一行,相应张量的每个通道对应一个值。如参考图18A、图18B、图19和图20所述,编码层的数量和维度可能基于张量重采样和/或级联与FPN层的数量和维度不同。子图片1216保持对应于图5的557的基向量。在图12B的示例中,基向量包括基向量1226等,其中基向量以不重叠的方式打包到子图片1216的区域中。基向量可以相邻地打包,或者可以包括各基向量集之间的例如一个或两个样本的间隔。各编码层或FPN层的基向量被打包到子图片1216中。图片1200中不用于存储任何数据的区域(诸如1208)可以由在应用逆量化以将样本值转换回浮点域之后对应于值“0”的样本值占据。类似地,子图片1210、1212、1214和1216中不用于存储任何数据的区域可以由在应用逆量化以将样本值转换回浮点域之后对应于值“0”的样本值占据。子图片1214保持对应于图5的572的系数,其中针对FPN的张量中的张量的各通道的各基向量都有一个系数。例如,在张量中有256个通道、25个基向量的情况下,一个张量的系数被布置为256×25的样本阵列。

[0186] 在图片1200中编码多个张量的情况下,相应的样本阵列堆叠在子图片1214中。由于量化特征图是二维样本阵列,并且系数也映射到一维或二维样本阵列中,因此针对图片1200使用单色格式就足够了。在CNN主干114为帧113产生了张量115的情况下,PCA编码器160可以基于张量115的维度确定子图片1210、1212、1214和1216的大小。子图片1210、1212

和1216的宽度和高度可以被指定为整数个CTU,即128个亮度样本的倍数。子图片1216的宽度等于水平布置的子图片1210、1212和1214的宽度之和。位于子图片1210、1212和1214下方的子图片1216的高度被设置为针对打包基向量(例如基向量1226)提供足够的区域。

[0187] 在图12A和图12B的示例中,分别对应于量化系数513、530和568的子图片1210、1212、1214水平布置,从而定义图片1200的宽度。对应于量化系数554的子图片1216位于第一单元、第二单元和第三单元下方,跨越子图片1210、1212、1214的宽度并延伸到图片1200的底部。在另一实现中,量化系数可以以不同于子图片的格式布置。例如,对于VVC或HEVC,量化系数可以布置或打包为切片或图块。在又一些实现中,量化系数可以以不同的方式布置,这取决于诸如编码类型(例如VVC或HEVC)和预期数据特性等的因素。例如,量化系数可以在应用了变换跳过的情况下按小块的扫描顺序布置,使得当不使用变换时,残差编码应用于有序的量化系数集。例如,当使用HEVC压缩量化系数时,可以对4×4块应用变换跳过,并且量化系数可以按4×4块中的对角扫描图案布置,以利用残差编码,而无需对4×4块应用变换。在基向量的数量为16或少于16个时,各特征图的量化系数可以存储在单独的4×4块中。在各4×4块内,量化系数的压缩效率受益于在4×4块内应用自适应Rice参数编码。在其他实现中,可以使用单个图片来代替单独的子图片,其中单独的切片对应于各组量化系数或者各子图片可以以与图片1200不同的方式布置。

[0188] 在图片1200的布置中,系数1222向子图片1212中的打包将阵列从N×1阵列(其中N是通道数)修改为W×H阵列(其中W×H等于N),并且选择系数以实现将子图片的纵横比调整为尽可能接近正方形(1:1)或矩形(诸如2:1比率),但避免N×1的极度拉长比率。更接近正方形的纵横比会导致与打包系数样本的边缘相邻的样本更少,这些样本不会被解包器使用,而是仍以相对高的保真度进行编码,因此会增加残差编码成本,而不会对张量重建有任何益处。

[0189] 在图片1200的另一布置中,进行系数1222或1224向子图片1212或1214中的打包,使得各系数被打包到多于一个样本中。例如,打包器610可以将系数打包到2×2样本集中。解包器620或832或812读取2×2样本集并进行滤波操作以恢复系数。打包到多于一个样本中提供了一种用以补偿由于视频编码器614中使用有损编码而引入系数值中的误差的机制。样本集的其他布置也是可能的,包括布置为2×4、4×2、4×4。

[0190] 图13是示出保持编码的打包特征图和相关联的元数据的位流1300的示意框图。位流1300对应于PCA编码器160所产生的位流121或PCA解码器170所解码的位流143。位流1300包含以“网络抽象层”单元头开头的句法组。例如,NAL单元头1308位于序列参数集(SPS)1310之前。SPS1310可以包括“档次水平等级”(profile level tier(PLT))句法单元,其指定档次(编码工具集)以及水平和等级(联合指定操作点,例如,就最大采样率和压缩位率而言)。PLT句法单元可以包括“通用约束信息”(GCI)句法单元,其与所选档次相比进一步约束可用工具集。SPS1310指定图片1200的布局,包括子图片1210、1212、1214和1216的位置和大小。GCI包括标志的集合,其中各标志约束特定编码工具不用于位流1300中。PLT可以用信号通知位流1300中可以使用的特定工具集,该特定工具集被称为“档次”。档次的示例是“Main10”,其提供色度格式为4:0:0或4:2:0并且目标是广泛部署的8到10位视频。GCI可以指示将档次的工具集进一步约束为工具的子集(称为“子档次”)。在PCA编码器160对打包成帧的特征图进行编码(即,根据图5操作)时,VVC标准的某些工具不提供压缩益处。视频编码

器614无需尝试未针对打包特征图提供压缩益处的工具,并且可以在GCI中将该工具用信号通知为未在位流1300中使用。SPS1310还指示位流1300所表示的帧数据的色度格式、位深度、分辨率。

[0191] PCA编码器160所使用的打包格式可以使用索引编码在SEI消息1313中,以从所有可用特征打包格式的枚举中选择一个特征打包格式。也可以在SEI消息1313中使用索引来指示用于产生特征图的特定CNN主干,以从预定CNN主干集合(其一些或全部可用于源装置110)的枚举中选择一个CNN主干。根据CNN主干类型索引,可以确定层的数量和各层中通道的数量以及各层中的各特征图的分辨率。基向量550的大小(即,宽度和高度)、基向量550的最大数量(即,FPN的各层的最大数量、或由张量组合器162得到的FPN的简化形式)以及所使用的基向量550的数量(即,由方法2100得到)被编码为参考附录A所述的SEI消息1313中的基向量打包信息1391。在张量提取器172中要进行以恢复具有与张量115a相同的数量和维度的张量149的操作(由张量组合器162确定)被描述为层映射1390。层映射1390存在于SEI消息1313中。层映射1390包括参考附录A所述的layer_update_mapping_info、layer_extraction_flag、layer_upsample_flag、layer_downsampling_flag、resampling_filter_idx、resample_vertical_only_flag、resample_horizontal_only_flag、src_layer_idx、src_channel_offset、dst_layer_idx。

[0192] 图片1314被编码在位流1300中。各图片包括一个或多于一个子图片,诸如对子图片1210进行编码的编码子图片1320等。对于位流的第一图片并且通常对于“随机访问点”访问单元,使用帧内切片来避免对位流1300中的其他访问单元的任何预测依赖性。对子图片1212进行编码的编码子图片1322包括切片头1330,接着是切片数据1340。切片数据1340包括CTU序列,从而提供帧数据的编码表示。CTU是正方形的,并且通常大小为128×128,这与典型的特征图大小不是良好地对准。将特征图对准到最小块大小(诸如4×4网格)部分地改善了这种未对准。为了能够从整数或样本域转换回浮点域,量化范围1392存在于位流1300中,并且定义浮点最小值和最大值,根据该浮点最小值和最大值,在逆量化处理中映射样本(例如,如在520、536、560、846、836、824、816中进行的)。量化范围在量化器模块(即,512、528、552和568)中确定,并在位流121中被编码为SEI消息1313中的量化范围1392。编码子图片1324对子图片1214进行编码,并且编码子图片1326对子图片1216进行编码。编码子图片1220、1222、1224和1226分别对应于位流部515、533、557和573。代替单独的子图片,单个图片还可以与对应于子图片1320、1322、1324和1326的单独切片一起使用。

[0193] 图14示出用于进行CNN的第一部分并对视频数据帧的所得特征图进行编码的方法1400。在对特征图进行编码时,张量被编码到位流中。方法1400可以使用诸如配置的FPGA、ASIC或ASSP等的设备来实现。可替代地,如下所述,方法1400可以由源装置110在处理器205的执行下实现为应用程序233的一个或多于一个软件代码模块。实现方法1400的应用程序233的软件代码模块可以例如驻留在硬盘驱动器210和/或存储器206中。对视频源112所产生的各视频数据帧重复方法1400。方法1400可以存储在计算机可读存储介质上和/或存储器206中。

[0194] 方法1400以进行CNN第一部分步骤1410开始。在步骤1410,CNN主干114在处理器205的执行下进行特定CNN的层的子集,以将输入帧113转换为中间张量115a。中间张量115a可以例如存储在存储器206和/或硬盘驱动器210中。如图4所示,示例CNN是“Faster R-CNN”

或“Mask R-CNN”,并且层的子集对应于直到“P层”拆分点的所有层。当从CNN主干114提取多个张量时,例如由于FPN的使用,张量115a可以包含多个张量。方法1400进行操作以对与来自视频源112的一个视频数据帧相对应的张量进行编码。处理器205中的控制从步骤1410进行到进行张量减少步骤1415。

[0195] 在步骤1415,张量组合器162在处理器205的执行下进行操作以对张量的数量进行重采样。图19示出在步骤1415实现的方法1900。作为步骤1415的操作的结果,来自FPN的张量115a可被(i)下采样(数量减少)、(ii)上采样(数量增加)或者(iii)直接传递通过以生成张量115。进行上采样或下采样使得来自两层的张量以相兼容的宽度和高度级联。对张量115a的给定张量进行上采样或下采样、或者保持所有张量115a不变的决定被编码在位流121中。处理器205中的控制从步骤1415进行到确定均值特征图步骤1420。

[0196] 在确定均值特征图步骤1420,模块510在处理器205的执行下跨通道维度对张量115进行平均以产生均值特征511。在步骤1420,均值特征511由量化器512在处理器205的执行下进行量化以产生整数(量化)均值特征图513。处理器205中的控制从步骤1420进行到编码均值特征步骤1430。

[0197] 在编码均值特征步骤1430,子图片编码器514在处理器205的执行下将整数均值特征图513打包到子图片1210中,并对子图片1210进行编码以产生位流部515。处理器205中的控制从步骤1430进行到恢复重建均值特征步骤1440。

[0198] 在恢复重建均值特征步骤1440,子图片编码器514在处理器205的执行下输出重建图片(例如,编码器514的实现内的618),其对应于用于视频压缩的子图片输入的有损版本(例如,612)。处理器205中的控制从步骤1440进行到确定均值系数步骤1450。

[0199] 在确定均值系数步骤1450,逆量化器520在处理器205的执行下进行操作以对整数均值特征图516进行逆量化,从而将图516转换回浮点域作为特征图522。点积524在处理器205的执行下使用张量115和特征图522,以使用点积函数生成均值系数的集合526,其中针对张量115中的每个通道生成一个均值系数。处理器205中的控制从步骤1450进行到编码均值系数步骤1460。

[0200] 在编码均值系数步骤1460,量化器模块528在处理器205的执行下对均值系数526进行量化,以生成或产生整数均值系数530。子图片编码器532在处理器205的执行下将整数均值系数530打包到子图片1212中,并对子图片1212进行编码以产生位流部533。处理器205中的控制从步骤1460进行到恢复重建均值系数步骤1470。

[0201] 在恢复重建均值系数步骤1470,子图片编码器532在处理器205的执行下输出子图片1212的重建版本,从该子图片1212的重建版本生成重建均值系数534。重建均值系数534由模块536进行逆量化,以产生恢复均值系数538。在步骤1470中,图片的重建版本对应于编码器532的实现内的618。处理器205中的控制从步骤1470进行到确定基向量步骤1480。

[0202] 在确定基向量步骤1480,逆量化器536、点积模块540、减法器544和分解模块548在处理器205的执行下进行操作以产生张量115的基向量集。逆量化器536输出浮点均值系数538。点积模块540对恢复均值特征图522和系数538进行点积运算以产生偏移张量542。张量542具有与张量115相同的维度,并且包含张量115中各通道的特征图的预测,该预测是仅使用均值(平均)特征图信息进行预测的。因此,张量542通过减去该每通道均值特征图数据,能够产生张量115的零平均版本。模块544从张量115中减去偏移特征图542,以产生零中心

张量546。分解模块548接收张量546作为输入,并生成基向量集550。集合550中的各基向量可被称为“分量”。参考图21描述分解模块548在产生基向量550时的操作。向量550包含的基向量比张量115中的通道少,这对应于张量115的维度降低。550中的基向量表示子空间中的张量115,该子空间解释了针对基向量550中分量数的张量115的最大方差量。换句话说,基向量550使得能够以针对给定分量数的质量下降最小的方式表示张量115。处理器205中的控制从步骤1480进行到编码基向量步骤1490。

[0203] 在编码基向量步骤1490,量化器模块552在处理器205的执行下进行操作以将基向量550量化到整数域中。子图片编码器556在处理器205的执行下将所得的整数基向量554打包到子图片中并对其进行编码以产生位流部557。处理器205中的控制从步骤1490进行到恢复重建基向量步骤14100。

[0204] 在步骤14100,子图片编码器556在处理器205的执行下生成重建整数张量558,并进行操作以获得子图片1316的重建版本(例如,618)。逆量化器560在处理器205的执行下将基向量558从重建子图片中解包并逆量化回浮点域(作为重建基向量562)。处理器205中的控制从步骤14100进行到确定系数步骤14110。

[0205] 步骤1440、1470和14100中的各步骤恢复重建系数。在各步骤生成的重建系数提供了在步骤1430、140和1490中量化和编码的特征或向量的重建。各重建允许对由于子图片编码器所执行的编码操作引起的损失进行建模和说明,使得在解码器侧看到的特征或向量的版本反映了在编码位流121时使用的特征或向量。还说明了在实现500中用于生成编码位流的(一个或多于一个)量化函数(例如512、528和552处)的操作所产生的损失。

[0206] 在步骤14110,点积模块564在处理器205的执行下,对张量546中的各通道与重建基向量562中的各向量进行点积,以产生系数的集合(即,系数566)。系数566表示各基向量在再现各特征图的内容时的贡献。处理器205中的控制从步骤14110进行到编码系数步骤14120。

[0207] 在步骤14120,量化器模块568在处理器205的执行下量化系数566以产生整数系数570。整数系数570被传递到子图片编码器572以打包到子图片1214中。子图片1214由模块572编码以产生位流部573。处理器205中的控制从步骤14120进行到组合子图片步骤14130。在张量115包含多于一个张量的情况下,对张量115中的各张量重复步骤1420～14120,其中在步骤14130将最终压缩表示布置成单个图片。

[0208] 如上所述,模块552至572以与模块512至532类似的方式进行操作以生成位流。在步骤14130,子图片位流组合器580在处理器205的执行下将分别对应于子图片1210、1212、1216和1214的位流部515、533、557和573组合成单个位流,即121。

[0209] 方法1400完成并且处理器205中的处理继续调用方法1400用于下一帧。

[0210] 图15示出由用于进行拆分CNN的第二部分的PCA解码器800实现的、用于解码位流143以产生解码张量149的方法1500。方法1500可以使用诸如配置的FPGA、ASIC或ASSP等的设备来实现。可替代地,如下所述,方法1500可以由源装置110在处理器205的执行下实现为应用程序233的一个或多于一个软件代码模块。实现方法1500的应用程序233的软件代码模块可以例如驻留在硬盘驱动器210和/或存储器206中。方法1500可以被存储在计算机可读存储介质上和/或存储器206中。方法1500以解码均值特征步骤1510开始。

[0211] 在步骤1510,图片解码器804在处理器205的执行下从位流143解码子图片1210(对

应于子图片840)。处理器205中的控制从步骤1510进行到解码均值系数步骤1520。

[0212] 在步骤1520,图片解码器804在处理器205的执行下从位流143解码子图片1212(对应于子图片830)。处理器205中的控制从步骤1520进行到确定DC分量步骤1530。

[0213] 在步骤1530,解包器842和832、逆量化器846和836以及点积模块850在处理器205的执行下生成张量852。解包器842和832解包子图片,各子图片包含如使用实现500的(一个或多于一个)量化函数进行量化的、与均值特征511相关的量化特征图。逆量化器846对由解包器842从子图片840解包的均值特征图进行逆量化以导出均值特征图848。逆量化器836使用位流143中用信号通知的相应量化范围,将由解包器832从子图片830解包的均值系数从整数域逆量化到浮点域,以产生均值系数838。点积模块850对均值特征848和均值系数838进行点积以产生张量852。张量852表示从零中心张量重建非零中心张量所用的偏移。处理器205中的控制从步骤1530进行到解码基向量步骤1540。

[0214] 在步骤1540,图片解码器804在处理器205的执行下从位流143解码子图片1216(对应于子图片820)。子图片1216作为解码的子图片820被传递到解包器822。解包器822然后提取解码的基向量(例如1226),以在步骤1560进行逆量化。参考图22描述了目的地装置140在从位流143解码基向量时的操作。处理器205中的控制从步骤1540进行到解码系数步骤1550。

[0215] 在步骤1550,图片解码器804在处理器205的执行下从位流143解码子图片1214(对应于子图片810)。子图片1214作为解码的子图片810被传递到解包器812。解包器812提取解码的系数(例如1224),以在步骤1560进行逆量化。处理器205中的控制从步骤1550进行到确定AC分量步骤1560。

[0216] 在步骤1560,解包器822和812、逆量化器824和816以及点积模块828在处理器205的执行下进行操作以产生零中心张量854。解包器822从子图片820中提取基向量并将基向量作为整数特征图822传递到逆量化器824。逆量化器根据从位流143获得的量化范围将基向量从整数域转换为浮点域,从而产生基向量826。解包器812从子图片810中提取系数并将系数作为整数系数传递到逆量化器816。逆量化器816产生浮点域系数818。点积模块828对系数818中的各系数与向量826中的相应基向量进行点积以产生零中心重建张量,即张量854。处理器205中的控制从步骤1560进行到产生张量步骤1570。

[0217] 在步骤1570,求和模块856在处理器205的执行下将零中心重建张量854与张量852求和以产生重建张量149。由于PCA处理的最佳或接近最佳的降维,张量149与张量115基本相似,因此适合由CNN头150用于进行给定的机器任务。如果正在使用FPN,则针对由方法1900得到的FPN层的表示的各张量进行步骤1510～1570。处理器205中的控制从步骤1570进行到提取张量步骤1575。

[0218] 在步骤1575中,处理器205通过张量提取器172的操作从任何复合张量中提取单独的FPN层张量。张量提取器172进行图20所示的方法2000,以从张量149a的任何复合张量中提取单独的FPN层张量,从而将结果作为张量149传递。处理器205中的控制从步骤1575进行到进行神经网络第二部分步骤1580。

[0219] 在步骤1580,CNN头150在处理器205的执行下,使用张量149作为输入来进行神经网络的后期阶段。例如,可以进行使用参考图10A或图11描述的网络的分类,这取决于张量149的维度(以及其中包含的张量的数量,如果使用FPN)。此外,在数个网络在主干部中具有

匹配的架构和权重的情况下,可以使用从这样的主干部编码的张量以通过运行CNN头150的不同实例来进行这些不同网络的完成。步骤1580执行以输出任务结果151。

[0220] 图16A和图16B示出两个连续的数据集和相关联的特征向量,以例示特征向量“不稳定性”的概念,当对相似的数据集重复进行分解时,就会发生特征向量“不稳定性”。当从视频生成去相关张量时,连续帧高度相关,因此所得到的基向量也非常相似。图16A中的第一数据集1600示出二维数据集,图16B中的第二数据集1640示出二维数据集的稍后版本。数据集1600和1640相似但不相同,并且可以通过对采样处理(诸如来自视频装置112的帧捕获等)进行CNN主干114操作来导出。特征图中的各样本形成单独的维度以为了在系统100中进行分解的目的,然而为了说明目的,图16A和图16B中示出二维数据集。产生单个特征向量,从而将维度从二维减少到一维。对于数据集1600,产生特征向量1610,并且对于数据集1640,产生特征向量1650。由于数据集1600和1640相似,因此特征向量1610和1650具有相似的绝对斜率,其中“斜率”反映二维向量情况下的向量的方向,注意主分量的截距是通过原点的。然而,由于数据集1600和1640的差异小,因而用于进行分解的处理可能使特征向量的方向“翻转”或逆转,从而导致向量1610和1650具有相反的方向。换句话说,1650的逆转表示各元素(即,x和y分量)是向量1610的相应元素的负。基向量的方向的逆转对应于向量的各元素具有与先前产生的基向量的相应元素相同(或相似)的幅度,但符号相反。对于基向量,存在比图16A和图16B的示例所示多得多的维度。例如,特征图大小为136×72指示基向量具有9792个元素或9792个维度。因此,将向量1650乘以负一得到具有与向量1610基本相同的值的向量。

[0221] 当重复进行跨张量115的特征图的分解时,会出现相同的问题,即所得到的基向量逐元素逆转。即使基向量的内容以其他方式基本相似,各个基向量有时也可能逆转其方向。任何基向量可能会发生这种极性逆转,并且给定基向量的发生独立于其他基向量的发生。当压缩包含打包基向量的连续帧时,逆转会降低帧间预测实现高编码效率的能力,因为视频编码器614不易识别与逆转基向量相对应的逆转区,从而无法进行高效的工具选择。

[0222] 图17示出用于对来自CNN第一部分的给定层的连续特征图进行编码的方法1700,这些连续特征图具有相同的维度。方法1700进行操作,使得在PCA的调用之间,降维空间中的表示在方向性方面是稳定的。方法1700特别适合于处理视频数据,其中降维表示需要多次重新生成,可能每帧一次,并且连续视频帧之间存在相关性,这导致通过对来自连续视频帧的张量进行PCA而产生的基向量具有相关性。在这种情况下,即使基向量内的整体和逐个样本的幅度相对不变,从一次调用到下一次调用,所得基向量的方向也可能跨基向量的所有样本逆转。这种“极性逆转”可能归因于基向量生成处理中的不稳定性、或对输入张量的小变化的敏感性。方法1700可以通过诸如配置的FPGA、ASIC或ASSP等的设备来实现。方法1700可以由目的地装置140在处理器205的执行下实现例如作为PCA编码器160的功能、作为应用程序233的一个或多于一个软件代码模块。方法1700针对位流143中编码的各视频数据帧重复。实现方法1700的应用程序233的软件代码模块可以存储在例如硬盘驱动器210上和/或存储器206中。方法1700从进行第一PCA步骤1710开始。

[0223] 在步骤1710,PCA编码器160在处理器205的执行下为第一张量生成第一基向量集,该第一张量是由CNN主干114处理来自视频源112的给定帧而产生的。例如,分解模块548可以接收张量546。处理器205中的控制从步骤1710进行到存储基向量步骤1720。

[0224] 在步骤1720,PCA编码器160在处理器205的执行下将第一基向量集存储在存储器206中,以供随后调用分解模块548使用。处理器205中的控制从步骤1720进行到进行第二PCA步骤1730。

[0225] 在步骤1730,PCA编码器160在处理器205的执行下为第二张量生成第二基向量集,第二张量是由CNN主干114处理来自视频源112的步骤1710的给定帧的后续帧而产生的。步骤1730以与步骤1720类似的方式(即,使用分解模块548)进行操作。后续帧可以是视频源112所提供的下一帧,或者可以是较晚的帧。在步骤1730生成的基向量提供当前张量的分量,而在步骤1720存储的基向量与前一张量相关。如果基向量的生成以低于每帧一次的频率进行,则生成可以是周期性的(每N帧一次)或者由与帧的内容相关的某个标准驱动,诸如在帧内容被认为包含显著的新信息时驱动。例如,由于与视频源112所提供的较早帧相比检测到大运动或其他显著变化,因此可以进一步生成基向量。由于来自视频源112的视频帧中的相关性,第一基向量集和第二基向量集具有一定程度的相关性,使得基向量适合使用图12A和图12B的打包表示中的帧间预测工具进行压缩。然而,第一帧和第二帧之间的噪声或其他小变化可能导致第二基向量集的各个基向量以与第一基向量集的相应基向量相比相反的“方向”(或逆转)而生成。处理器205中的控制从步骤1730进行到基向量逆转检查步骤1740。

[0226] 在步骤1730生成的基向量提供当前张量的分量,而在步骤1710生成的基向量提供前一张量的分量,从而表示视频数据113中的较早帧。在步骤1740,测试当前基向量以检查极性逆转。检查极性逆转涉及测试当前(步骤1730)和较早或前一(步骤1710)生成的给定编码层或FPN层的基向量,以确定除了交换或逆转基向量中各元素的符号之外,向量是否保持基本相似。对于给定的基向量(从步骤1710的第一或前一向量开始),确定总和。总和是首先对来自第一基向量集和第二基向量集的相应向量进行逐元素相加的结果。针对各基向量来确定逐元素相加所得和的绝对值之和,以产生相似性得分,相似性得分是每个基向量的单个标量值。

[0227] 给定基向量的相似性得分的较低值指示第二基向量集与第一基向量集中的对应基向量基本相似但极性相反。因此,相似性得分的较低值指示连续进行分解之间的基向量的极性逆转。在总和低于阈值的情况下,检测到极性逆转(步骤1740处为“是”),并且处理器205中的控制从步骤1740进行到逆转基向量步骤1750。在总和不低于阈值的情况下,未检测到极性逆转(步骤1740处为“否”),并且处理器205中的控制从步骤1740进行到最后基向量测试步骤1760。相似性阈值通常是预定的。在一些实施例中,相似性阈值与当前张量的宽度和高度中的至少一个线性相关,例如是当前张量的宽度和/或高度的0.8或0.9。在其他布置中,相似性阈值可以等于张量的一个通道的面积的log2。相似性阈值可以基于代表非逆转情况的诸如滤波后的相似性得分等的因素来确定。可以将中值滤波器应用于相似性得分以排除异常值(例如,指示极性逆转的低得分)。然后可以缩放中值滤波后的相似性得分并将其与给定帧的相似性得分进行比较以检测极性逆转。可以针对每个基向量和每个层粒度进行滤波,以考虑基向量和层之间的不同统计数据。可以允许由于预期内容和内容方差的变化而导致的相似性得分的变化。设置阈值的允许也可以基于所使用的分解方法等。

[0228] 在逆转基向量步骤1750,PCA编码器160在处理器205的执行下使当前正在处理的基向量的极性逆转,例如,通过将向量乘以-1(负1)以产生更新的基向量。实际上,生成从当

前基向量更新的分量集,使得更新的分量集中的分量是在步骤1730生成的基向量中的相应分量的逆。由于这种逆转,当前基向量变得与来自第一基向量集的相应基向量基本相似。一旦发生逆转,只要在下一个分解操作中没有发生进一步的极性逆转,就需要继续逆转来自后续分解操作的基向量,以保持与当前分解操作的逆转基向量的极性匹配。标志集(每个基向量一个标志)用于指示各基向量的逆转状态,以用于将来的基向量生成操作,例如,为后续帧提供基向量逆转步骤的持久性。在标志集中切换或更改当前基向量的标志以指示极性逆转已发生,使得能够处置在第一次检测到极性逆转时开始逆转并在检测到第二次极性逆转时停止逆转的情况。该标志集提供了是否要逆转使用分解函数针对任何后续张量生成的分量的基础。处理器205中的控制从步骤1750进行到最后基向量测试步骤1760。

[0229] 在步骤1760,处理器205测试正在处理的当前向量是否是基向量集中的最后一个。如果不是(步骤1760为“否”),则处理进行到基向量集中的下一个基向量,并且处理器205中的控制从步骤1760进行到步骤1740。如果是(步骤1760为“是”),则处理从步骤1760进行到编码基向量步骤1770。

[0230] 在步骤1770,编码器160在处理器205的执行下对最终基向量进行操作,该最终基向量包括针对在步骤1750中已进行逆转的向量的更新基向量、以及来自第二基向量集的针对尚未进行逆转的向量的基向量。该最终基向量集被存储作为方法1700的后续调用的参考点,即作为“第一基向量集”。最终基向量集作为基向量550被传递到量化器模块552,并由子图片编码器556编码。此外,在进行一次逆转的情况下,各分量的标志指示在方法1700的后续调用中需要针对相应分量进行逆转,以确保极性与方法1700的上一次调用的极性相匹配。在方法1700的后续调用中第二次检测到逆转的情况下,切换标志,从而导致在方法1700的进一步调用中停止相应分量的逆转。然后,方法1700针对当前帧终止,并且可以再次调用用于下一帧或较晚帧,其中PCA编码器160以低于每帧一次的频率进行。

[0231] 图18A是示出用于使用上采样或下采样操作来组合来自不同FPN层的张量的模块162的示例实现1800的示意框图。图18A的示例使用来自CNN主干部400的P2～P5层。由于PCA解码器170能够从解码的P5层重建P6层,因此不用考虑层P6 429。图18A中所示的示例涉及对P2和P3之一进行重采样,并在以下参考图19进行描述。

[0232] 图18B是示出用于使用上采样或下采样操作将来自组合表示的张量提取回与FPN层相对应的单独张量的模块172的实现1890的示意框图。图18B的示例为CNN头1100提供了示例P2～P5层。层P6 1118是通过对层P5 1116进行最大池化操作而导出的,因此不需要压缩或解压缩。图18B中所示的示例涉及图18A中所示的P2和P3已组合的情况,并且在以下参考图20进行描述。

[0233] 图19示出用于对来自CNN的多层第一部分的张量进行重采样以与来自CNN的多层第一部分的另一张量进行去相关的方法1900。方法1900可以通过诸如配置的FPGA、ASIC或ASSP等的设备来实现。可替代地,如下所述,方法1900可以由源装置110在处理器205的执行下实现为应用程序233的一个或多于一个软件代码模块。方法1900针对位流143中编码的各视频数据帧重复。实现方法1900的应用程序233的软件代码模块可以例如存储在硬盘驱动器210上和/或存储器206中。方法1900从确定张量映射步骤1910开始。

[0234] 在步骤1910,张量组合器162在处理器205的执行下确定是否要组合来自FPN的两个不同层的张量(张量115a中的张量)。如果张量115a至少具有第一特征图集(具有第一大

小)和第二特征图集(具有不同于第一大小的第二大小),则通常进行组合。如果未组合,则张量115a将不经修改地被传递。如果要组合张量,则张量组合器162确定是要通过执行较大张量的下采样或较小张量的上采样来组合张量,还是要在不重采样的情况下组合张量。当张量115a包括至少两个张量时,可以使用上采样和下采样的选项,其中两个张量的宽度和高度以2为因子相关,即在包括步幅等于2的操作的情况下从分层网络生成。当张量115a包括至少两个张量时,可以使用在没有重采样的情况下级联的选项,其中两个张量的宽度和高度相同。由于张量115a的维度由网络架构和拆分点设置,因此张量组合器162可用的选项受到网络架构和拆分点的选项约束。在张量115a仅包含一个张量的情况下,没有组合张量的机会,并且张量组合器162的唯一选项是将张量115a作为张量115无改变地传递。在要评估下采样或上采样到宽度相同但高度为一半或两倍的张量的情况下,resample_vertical_only_flag句法元素指示仅在垂直方向上应用重采样。在要评估下采样或上采样到高度相同但宽度为一半或两倍的张量的情况下,resample_horizontal_only_flag句法元素指示仅在水平方向上应用重采样。仅垂直或仅水平重采样的选项允许在不同层中进行张量组合,其中仅在垂直或水平方向上应用等于二的步幅,如在某些CNN架构中可能遇到的那样。

[0235] 如果对较大张量进行下采样,则较大张量中的细节会丢失。因此,下采样更适合于缺乏精细细节的数据,诸如主要包含大对象的帧数据113等。如果确定帧数据113包含更小的细节,则处理器205不会选择对较大张量进行下采样。无论是下采样还是上采样,在应用基于跨通道分解的压缩之前组合来自两个不同层的张量的结果是发现跨层的冗余的能力。如果各层由PCA编码器160独立地编码,尽管各张量的跨通道冗余会被发现,也无法发现跨层的冗余。如果识别出跨层的冗余,则对组合两个不同FPN层的张量进行编码所需的总基向量数量可能低于分配给各FPN层的张量的基向量数量,其中张量不会组合成单个层。在要对较小的张量进行上采样的情况下,看到利用层间冗余的益处,导致很少或不需要增加用于组合张量的基向量数量。对于两层之间宽度和高度相同的分层网络(通道计数可能不同),张量组合器162可以尝试发现跨层冗余,而无需对来自两个层的具有相同宽度和高度的张量进行重采样。处理器205可以尝试组合而不进行重采样、上采样或下采样、并且保持张量未经修改的选项,并选择导致所得张量115的编码成本最低的选项。在其他布置中,步骤1910可以设置是否要实现重采样的预定默认操作。在步骤1910确定的映射可以是预定映射,或者可以更定期地确定,诸如每帧确定一次。当映射是预定时,映射仅需在位流开始时用信号通知。SEI消息1313中的layer_update_mapping_info句法元素许可用信号通知SEI消息1313的实例,其中映射将被通信给目的地装置140。对于预定映射,利用CLVS的第一图片1200用信号通知一次映射的实例就足够了。处理器205中的控制从步骤1920进行到编码张量映射步骤1920。

[0236] 在步骤1920中,张量组合器162在处理器205的执行下将步骤1920的操作所得到的确定编码到位流121中。编码的确定表示需要在张量分离器172中进行以使张量组合器162中所进行的操作反转的操作。使用标志或可变长度码字将上采样、下采样或保持FPN层未经修改的决定编码在SEI消息1313中。参考附录A描述了用于将步骤1910中做出的决定编码到位流121中的句法。该句法对要在张量分离器172中进行的操作进行编码,使得张量149具有与张量115a相同的数量、维度和通道计数,即在张量组合器162的操作之前对要在张量分离器172中进行的操作进行编码。在张量被上采样并通过张量组合器162与另一张量组合的情

况下,信号通知指示张量分离器172要从解码的组合张量中提取两个张量。所提取的张量之一被输出作为张量149之一,并且所提取的张量被下采样以产生包括在张量149中的附加张量。在张量被下采样并通过张量组合器162与另一张量组合的情况下,信号通知指示张量分离器172要从解码的组合张量中提取两个张量。提取的张量之一被输出为张量149之一,并且另一张量被上采样以产生包括在张量149中的附加张量。在张量未被重采样但通过张量组合器162与另一张量组合的情况下,信号通知指示张量分离器172要从解码的组合张量中提取两个张量,这两个张量都输出为张量149中的张量。索引以获得组合张量,即src_layer_idx,并插入重采样张量,即dst_layer_idx。源层索引和目的地层索引的信号通知是灵活的,以例如支持级联层在张量115a中不相邻的情况、或者具有另一层一半空间分辨率的层在张量115a中较早出现或在张量115a中较晚出现的情况。实现上采样的编码指定了在编码时被上采样的特征图将在解码时被下采样,并且反之亦然。处理器205中的控制从步骤1920进行到张量编码器映射测试步骤1930。

[0237] 在步骤1930,如果在步骤1910做出的决定是不组合来自张量115a的不同FPN层的张量,则方法1900终止。处理器205中的控制返回到方法1400的确定均值特征步骤1420,其中针对剩余步骤,所有张量115a都作为张量115传递。如果在步骤1910处做出的决定是在张量组合器162中对来自张量115a的FPN层的张量进行上采样,则处理器205中的控制从步骤1930进行到上采样选择张量步骤1940。如果在步骤1910处做出的决定是在张量组合器162中对来自张量115a的FPN层的张量进行下采样,则处理器205中的控制从步骤1930进行到下采样选择张量步骤1950。如果在步骤1910处做出的决定是在张量组合器162中对张量115a之间的张量进行级联而不应用重采样滤波器,则处理器205中的控制从步骤1958进行。

[0238] 参考图18A,实现1800接收输入P3～P5(分别为1810～1813)。实现1800包括张量上采样器1822、张量下采样器1831、级联模块1818和1836以及多路复用器1844。

[0239] 在第二特征图集(或张量)的第二大小大于第一大小时,对第一张量进行上采样,或者反之亦然。在步骤1940,张量上采样器1822在处理器205的执行下,对来自张量115a的张量(例如,P3张量1811)进行上采样,以产生上采样的P3张量1824。上采样的张量1824也可以称为上采样的特征图集1824。P3张量1811具有宽度w和高度h。上采样的P3张量1824具有宽度2d和高度2h。可以使用插值滤波器、最近邻滤波器或其他滤波器来进行上采样,以产生上采样的P3张量1824所需的中间样本。处理器205中的控制从步骤1940进行到级联张量步骤1945。

[0240] 在步骤1945,级联模块1818在处理器205的执行下,使上采样的P3张量1824和P2张量1810沿通道维度级联以产生级联张量1842。级联张量1842具有宽度2w和高度2h,并且通道计数等于张量1824和1810的通道计数之和。多路复用器1844选择张量1842作为级联输出张量1846。针对剩余的编码步骤,输出张量1846、1812和1813作为张量115。然后方法1900终止,并且处理器205中的控制在方法1400的步骤1420处恢复。

[0241] 在步骤1950,张量下采样器1831在处理器205的执行下,从张量115a中下采样张量(例如,P2张量1810),以产生下采样的P2张量1832。P2张量1810具有宽度2w和高度2h。下采样的P2张量1832在空间上被下采样,以具有P2张量1810的一半宽度和一半高度,从而得到尺寸w、h。可以使用子采样的双线性、双三次或其他滤波操作来进行下采样。在P2张量1810的通道维度中不应用下采样或其他滤波,因此下采样的P2张量1832的通道计数不变。处理

器205中的控制从步骤1950进行到级联张量步骤1955。

[0242] 在步骤1955,级联模块1836在处理器205的执行下,使P3张量1811和下采样的P2张量1832沿通道维度级联,以产生级联张量1839。级联张量1839具有宽度w和高度h,并且通道计数等于张量1811和1832的通道计数之和。多路复用器1844选择张量1839作为级联输出张量1846。针对剩余编码步骤,输出张量1846、1812和1813作为张量115。如果实现了重采样,则如图18A中关于P2和P3以及级联器1818和1836的操作所述,输出张量1846包括重采样的特征图集和尚未重采样的特征图集。然后,方法1900终止,并且处理器205中的控制在方法1400的步骤1420处恢复。

[0243] 在步骤1958,在处理器205的执行下,级联模块1823使张量1811和张量1832(在这种情况下,两者具有相同的宽度和高度)级联,以产生级联张量1838。多路复用器1844将张量1838作为级联的输出张量1846传递。针对剩余的编码步骤,张量1846、1812和1813作为张量115传递。然后,方法1900终止,并且处理器205中的控制在方法1400的步骤1420处恢复。

[0244] 方法1400继续执行步骤1420至14120,从而例如在步骤1480中使用步骤2070输出的张量确定基向量,在步骤14110中使用基向量导出级联张量中针对两个特征图相关联的系数,并在步骤1490和14120对系数和基向量进行编码。

[0245] 作为方法1900的操作结果,组合张量作为组合FPN的两个单独层的张量149之一传递。使用诸如PCA等的分解方法压缩组合张量。分解方法还能够沿着发生上采样/下采样的FPN的两个层进行去相关,而不是将去相关限制到FPN的每个层。此外,在需要保留细节的情况下,源装置110具有从较小FPN层(例如P3)上采样张量的选项,以用于基向量导出和压缩的目的。与纯粹为了P2层生成的基向量的数量相比,为了上采样的P3和P2层的级联生成的基向量的数量不需要增加。因此,不需要图片1200中用于P3层的附加基向量区域。

[0246] 图20示出用于从解码的组合表示(即,张量149a)中重采样张量以产生用于提供给CNN头150的输出张量149的方法2000。方法2000用于从编码数据中解码张量。方法2000可以通过诸如配置的FPGA、ASIC或ASSP等的设备来实现。可替代地,如下所述,方法2000可以由目的地装置130在处理器205的执行下实现为应用程序233的一个或多于一个软件代码模块。方法2000针对位流143中编码的各视频数据帧重复。实现方法2000的应用程序233的软件代码模块可以存储在例如硬盘驱动器210上和/或存储器206中。方法2000从解码张量映射步骤2010开始。

[0247] 在步骤2010,熵解码器920在处理器205的执行下从位流143中所包含的SEI消息1313解码句法元素,从而指示张量149a是否包括组合的重采样张量。解码器例如基于指定重采样是否在编码时发生的解码信息或基于默认设置,来确定是否要执行重采样。也就是说,熵解码器解码张量之一是否是如图19中的编码所实现的、来自一个FPN层的张量与来自另一FPN层的张量的重采样(上采样或下采样)版本的级联。当步骤1540和1550中对基向量和系数的解码已经发生并且在步骤1570中导出张量时,执行方法2000。组合张量可以是具有相同最终宽度和高度的张量的级联,或者张量149a中可能不存在级联的张量。

[0248] SEI消息1313中的信号通知指示张量分离器172需要进行以恢复张量149(即,产生具有与适合CNN头150使用的张量115a相同的数量和维度的张量)的操作。在张量组合器162已对张量进行下采样以产生级联张量的情况下,位流143中的信号通知指示张量分离器172对级联张量的一部分进行上采样以恢复张量。在张量组合器162已对张量进行上采样以产

生级联张量的情况下,位流143中的信号通知指示张量分离器172对级联张量的一部分进行下采样以恢复张量。在张量组合器162尚未对张量进行重采样以产生级联张量的情况下,位流143中的信号通知指示张量分离器172提取级联张量的一部分以恢复张量,而不应用重采样滤波器。从级联张量中提取张量是根据解码的layer_extraction_flag句法元素确定的。经由layer_downsample_flag句法元素用信号通知向提取的张量应用下采样滤波器。经由layer_upsample_flag用信号通知向提取的张量应用上采样滤波器。经由src_layer_idx句法元素用信号通知张量149a中的哪个张量是级联张量。经由dst_layer_idx句法元素用信号通知表示要插入提取的张量以产生张量149的偏移。经由src_channel_offset句法元素用信号通知用于在两个层之间的级联张量内划定边界的通道索引。处理器205中的控制从步骤2010进行到张量解码器映射测试步骤2020。

[0249] 在步骤2020,如果步骤2010的结果指示不应用重映射,则处理器205中的控制从步骤2010进行到输出张量步骤2070。如果步骤2010的结果指示要应用张量下采样,则处理器205中的控制从步骤2020进行到提取张量步骤2030。如果步骤2010的结果指示要应用张量上采样,则处理器205中的控制从步骤2020进行到提取张量步骤2050。如果步骤2010的结果指示要应用无重采样的张量提取,则处理器205中的控制从步骤2020进行到提取张量步骤2035。

[0250] 图18B的实现1890接收输入P2/P3、P4和P5(分别为1850、1852和1854)。输入P2/P3是通过编码侧的上采样或下采样(例如通过方法1900的操作)生成的组合张量。实现1890包括张量切片器1856、下采样器1862、上采样器1864以及选择器1870和1872。

[0251] 在步骤2030,张量切片器1856在处理器205的执行下,沿通道维度(在根据src_channel_offset的索引处)将组合的P2/P3张量1850切片成两个张量,各张量具有与P2和P3层的通道计数相对应的通道计数,即256个通道。张量切片器1856输出第一切片张量1858和第二切片张量1860。处理器205中的控制从步骤2030进行到下采样提取张量步骤2040。

[0252] 在步骤2035,张量切片器1856在处理器205的执行下,沿通道维度(在根据src_channel_offset的索引处)将组合的P2/P3张量1850切片成两个张量。张量切片器1856将第一切片张量1858输出到选择器1870,并将第二切片张量1860输出到选择器1872。处理器205中的控制从步骤2035进行到输出张量步骤2070。

[0253] 在步骤2040,下采样器1862在处理器205的执行下,对张量1858进行下采样以导出重新缩放的张量1866。重新缩放的张量提供了用于对组合张量进行解码的特征图集。下采样器1862可以使用抽取、或子采样的双线性或双三次或其他滤波器来产生重新缩放的张量1866。处理器205中的控制从步骤2040进行到输出张量步骤2070。

[0254] 在步骤2050,张量切片器1856在处理器205的执行下,将组合的P2/P3张量1850沿通道维度切片成两半,每半具有与P2和P3层的通道计数相对应的通道计数,即256个通道。张量切片器1856输出第一切片张量1858和第二切片张量1860。处理器205中的控制从步骤2050进行到上采样提取张量步骤2060。

[0255] 在步骤2060,上采样器1864在处理器205的执行下对张量1860进行上采样以产生重新缩放的张量1868。上采样可以使用双线性、双三次、最近邻或其他滤波方法来导出填充重新缩放的张量1868所需的插值样本。处理器205中的控制从步骤2060进行到输出张量步骤2070。

[0256] 步骤2040和2060进行操作以通过对步骤2030或2050中导出的张量的第一部分进行重采样(分别进行下采样和上采样)来解码张量149a,从而导出第一特征图集。

[0257] 在步骤2070,提供输出张量149以供CNN头150使用。P5张量1854和P4张量1852从张量149a传递到张量149。如果没有发生重映射或重采样操作,则将对应于P2和P3层的张量从张量149a传递到张量149。如果发生了重映射操作,则输出P3张量1882和P2张量1880作为张量149的相应FPN层。如果发生了下采样,则选择器1870输出下采样的张量1866作为张量1882,并且选择器1872输出张量1858作为张量1880。如果发生了上采样,选择器1870输出张量1860作为张量1882,并且选择器1872输出张量1868作为张量1880。如果实现了重采样,则如图18B中关于P2和P3以及选择器1870和1872的操作所述,输出张量1880和1882之一包括重采样的特征图集和尚未重采样的特征图集。如果没有发生重采样,则选择器1870输出张量1860作为张量1882,并且选择器1872输出张量1858作为张量1880。方法2000终止,并且处理器205中的控制返回到方法1500。

[0258] 在方法1900和2000的布置中,步骤1910还决定P2～P5中哪些层要进行上采样或下采样,从而选择一对相邻层,即P2和P3、P3和P4或者P4和P5。在步骤1920,将所选择的一对相邻层编码到位流121中。在步骤2010,从位流143中解码所选择的一对相邻层,并从级联的张量中恢复所指示的张量。

[0259] 图21示出由分解模块548确定用于对来自CNN主干114的张量115进行编码的基向量的方法2100。方法2100可以通过诸如配置的FPGA、ASIC或ASSP等的设备来实现。方法2100是参考附录A中呈现的基向量打包格式信息SEI消息句法来描述的。可替代地,如下所述,方法2100可以由源装置110在处理器205的执行下实现为应用程序233的一个或多于一个软件代码模块。方法2100针对位流121中编码的各视频数据帧重复。实现方法2100的应用程序233的软件代码模块可以存储在例如硬盘驱动器210上和/或存储器206中。方法2100从编码最大基向量计数步骤2110开始。

[0260] 在步骤2110,熵编码器738在处理器205的执行下将值bv_max_cnt编码到SEI消息1313中。值bv_max_cnt作为基向量打包信息1391的一部分被编码到SEI消息1313中,该基向量打包信息1391对应于分解模块548要生成到位流121中的基向量的数量。值bv_max_cnt指示在张量115中分解张量时能够使用的最大基向量数量。如果在张量组合器162中没有发生层减少,则为张量115的各FPN层编码bv_max_cnt的一个实例。如果由于张量组合器162而减少了层数,则以减少的格式为每个所得层(即,为每个编码层)编码一个bv_max_cnt值。步骤2110仅需作为初始化步骤进行,因此一个或多于一个bv_max_cnt值仅需在位流121中的初始化或随机访问点处进行编码。此外,在步骤2110,针对各层的基向量的宽度和高度在SEI消息1313中被编码为bv_width和bv_height。使用在步骤2110解码的最大基向量数量有效地确定编码数据的帧大小,该帧大小适用于编码数据的至少一个帧。处理器205中的控制从步骤2110进行到确定基向量打包格式步骤2120。

[0261] 在步骤2120,子图片编码器556的打包器610确定子图片1216的打包格式和子图片1216的大小。例如,如参考图12B所示,打包格式和大小被确定为使得针对张量115中的各张量的指定大小和计数bv_max_cnt的基向量能够被打包到子图片1216中。即使对于给定图片使用较少的基向量,包含打包基向量的后续图片(即,1200)也具有与第一次调用方法2100时确定的相同大小。因此,子图片编码器(即,514、532、556和572)使用相同大小的帧对连续

的压缩张量集115进行编码,并且不需要发送SPS1310的新实例来用信号通知图片分辨率的变化。如使用帧间预测方法所必需的,改变图片分辨率需要重置帧缓冲器772,这防止了后续图片参考帧缓冲器772中的较早图片。基于最大基向量计数确定图片1200的大小和打包格式允许在所利用的基向量数量变化时进行帧间预测,因为帧缓冲器772不需要重置来容纳不兼容分辨率的图片。因此,即使在要打包的基向量数量发生变化时也保持图片1200的固定大小提高了压缩效率,因为能够跨所得到的打包图片使用帧间预测。处理器205中的控制从步骤2120进行到确定基向量集步骤2130。

[0262] 在步骤2130,分解模块548在处理器205的执行下为张量115中的各张量产生M个基向量的集合,其中M对应于张量115中的相应张量的值bv_max_cnt(“M”)。对于包含256个通道的层,诸如P3和P4等,值M可以设置为固定值25。在两个层(诸如P2和P3等)在组合器162处级联的情况下,值M可以被设置为50,这对应于针对张量115a的每个层(即,在来自张量组合器162的任何层减少之前)保持25个基向量。张量的M个基向量是有序的主分量集合,集合中的各分量捕获张量中的尚未被M个基向量的先前主分量捕获的最大方差量。方差跨张量的通道维度被“捕获”或解释。因此,M个基向量定义如下的空间,在该空间中,张量的各通道可以用比张量的给定通道中的样本数(即,宽度×高度)更少的值来表达。各基向量所解释的方差量是特征值。M个基向量具有相应的一组M个特征值,这些特征值形成严格减少的数字阵列。处理器205中的控制从步骤2130进行到累积特征值步骤2140。

[0263] 在步骤2140,分解模块548在处理器205的执行下,为各值n产生累积的“解释方差”值,其中n的范围从1到M。累积的解释方差值等于从第一个特征值到第n个特征值的特征值的总和。累积的解释方差值指示在步骤2130产生的总特征向量集或主分量集的子集解释了多少方差。处理器205中的控制从步骤2140进行到确定基向量子集步骤2150。

[0264] 在步骤2150,分解模块548在处理器205的执行下选择与“解释方差阈值”相对应的基向量M的子集。解释方差阈值确定要使用的基向量的数量,并且可以根据准确度、面积和编码效率之间的期望折衷来选择以充分解构张量。解释方差阈值可以基于基向量本身中的数据来设置,或者基于期望面积或编码效率或最小期望准确度来设置。选择基向量的子集导致选择M个基向量中的前N个基向量,即在可用(最大)基向量中选择基向量。可以基于每个帧(即,针对要编码的各数据帧)确定解释方差阈值。子集的选择使得由于对累积解释方差的较小贡献而对重建张量作出较小贡献的基向量被省略。子集的选择提高压缩效率,因为子图片1216具有较少的填充区域。子图片1216的未填充区域可以由恒定样本值占据。由于VVC的灵活块结构,恒定样本值具有高度可压缩性。处理器205中的控制从步骤2150进行到编码基向量子集步骤2160。

[0265] 在步骤2160,熵编码器738在处理器205的执行下将张量115中的各张量的值bv_cnt编码到位流121中。值bv_cnt指示执行步骤2150所得到的选择子集。值bv_cnt可以指示对于给定张量,基向量0到N(即,所使用的基向量的数量)可以从1到M不等。步骤2160对用于编码张量的基向量的数量进行编码。指定要使用的基向量的数量的信息可被认为包括如步骤2110中确定的要使用的最大基向量数量和如步骤2150中确定的要使用的实际基向量数量。方法2100终止,并且处理器205中的控制从步骤2160进行回到步骤1490,其中基向量的子集在步骤1490中被编码在位流中。

[0266] 图22示出用于对为了对来自CNN的第一部分的张量进行编码而使用的基向量进行

解码的方法2200。方法2200用于解码来自PCA编码器160所编码的数据的张量。方法2200可以由诸如配置的FPGA、ASIC或ASSP等的设备来实现。方法2100是参考附录A中呈现的基向量打包格式信息SEI消息句法来描述的。可替代地,如下所述,方法2200可以由目的地装置140在处理器205的执行下实现为应用程序233的一个或多于一个软件代码模块。方法2200针对位流143中编码的各视频数据帧重复。实现方法2200的应用程序233的软件代码模块可以存储在例如硬盘驱动器210上和/或存储器206中。方法2200从解码最大基向量计数步骤2210开始。

[0267] 在步骤2210,熵解码器920在处理器205的执行下从SEI消息1313解码bv_max_cnt句法元素。bv_max_cnt句法元素提供用于指定或指示能够用于打包图片1200的最大基向量数量的信息。可以针对从张量组合器162得到的组合表示中的各张量来解码单独的最大值。在步骤2210,还从SEI消息1313解码各组合表示的基向量的宽度和高度。针对图片序列的第一图片解码bv_max_cnt句法元素,并保留值以供后续图片使用,直到指示从位流143解码bv_max_cnt的另一实例为止。例如,可以经由basis_vector_dimensions_update标志指示bv_max_cnt的另一实例。处理器205中的控制从步骤2210进行到确定基向量打包格式步骤2220。

[0268] 在步骤2220,解包器822在处理器205的执行下,根据如步骤2210确定的基向量宽度、高度和最大数量,确定各基向量在子图片1216中打包时的空间位置。基向量通常首先以从左到右的方式、然后以从上到下的方式打包,前进通过有序的基向量列表。基向量通常相邻地且以不重叠的方式打包。在(例如,由于FPN的使用)存在多个层的情况下,针对子图片1216确定各个层的基向量的空间位置。每当bv_max_cnt从位流143解码时(即,当由basis_vector_dimensions_update标志指示时),需要进行步骤2220。当bv_max_cnt未从位流143解码时,可以重用在方法2200的较早调用中进行步骤2220所确定的打包布置。处理器205中的控制从步骤2220进行到解码基向量子集步骤2230。

[0269] 在步骤2230,熵解码器920在处理器205的执行下从SEI消息1313解码bv_cnt句法元素。当basis_vector_cnt_update(其也是从SEI消息1313解码的,指示所使用的基向量计数的更新)需要被传输到目的地装置140时,解码bv_cnt句法元素。步骤2230进行操作以对指示解码中要使用的基向量数量的信息进行解码。用于指定要使用的基向量数量的信息可以被认为包括如2210导出的最大向量数量和步骤2230解码的要使用的实际量这两者。张量115中的各层可以具有单独的基向量计数。用信号通知的基向量计数指示值N,其中N定义了作为M个基向量中的根据定义的阈值选择以充分重建张量的前N个基向量的子集。使用阈值来确立要使用的基向量的数量允许适配张量115的统计数据。bv_cnt句法元素可以基于每个帧(即,针对编码数据的各帧)确定或解码。

[0270] 在需要较少的基向量来达到指定质量水平(即,给定的解释方差阈值)的情况下,可以省略附加基向量,从而提高压缩效率。为了保留大多数基向量,可以使用诸如98%或99%等的阈值。在大多数情况下,保留所有M个基向量,即N等于M,但值N可以以数据相关的方式减小。在可接受较低的质量的情况下,通过相比于进一步增加量化步长大小(即增加量化参数)而言优先降低阈值来实现更好的位率或性能折衷。例如,85%的阈值降低位率,同时与将QP增加到更高水平(例如QP大于45)相比,降低任务性能相对不那么严重,这在期望非常位压缩率的情况下是合适的。95%的阈值降低位率,同时适度降低可实现的任务性能,

这适用于需要在近无损压缩和非常低位率操作点的压缩之间的性能的用例。阈值和QP可以联合优化以产生“Pareto前沿”,即由最佳阈值和QP值组成的位率或任务性能前沿曲线。详尽地测试各种阈值和QP值以导出该Pareto前沿对于源装置110来说通常成本过高,并且可能期望用以指示合适值的启发式规则。实验表明,对于较低的每像素位数(BPP),优选降低阈值而不是增加QP。此外,阈值可能针对(例如,FPN的)不同编码层或从张量组合器162得到的FPN的组合表示而不同。实验表明,可以为分解更多的层给予略高的阈值,从而在位率相对有限的增加下实现任务性能的改善。例如,张量115的P2/P3级联层、P4和P5层的阈值90%、94%和98%比跨所有层应用的固定阈值(诸如90%或95%)提供更好的性能。与分解较少的层(例如P2或P3)相比,分解更多的层(例如P4和P5)具有较小的面积。因此,子图片1216中的打包面积的增量增加小于分解较少的层的类似附加基向量计数。最大基向量计数可以类似地基于每个层而变化。处理器205中的控制从步骤2230进行到解码基向量子图片步骤2240。

[0271] 在步骤2240,图片解码器804在处理器205的执行下对包含打包基向量的子图片1216进行解码。步骤2240根据至少步骤2230解码的信息有效地从编码数据中导出基向量。处理器205中的控制从步骤2240进行到解包基向量步骤2250。

[0272] 在步骤2250,解包器822在处理器205的执行下,根据如步骤2230中在图片1200中使用的用信号通知的基向量数量,从子图片1216中解包数量为bv_cnt的基向量。如图片1200中使用的用信号通知的基向量数量可能小于在步骤2220中确定的打包布置所提供的最大基向量数量。步骤2250至少根据用于指定要使用的基向量数量的信息从编码数据中导出基向量。在多个层被打包到子图片1216中的情况下,根据从SEI消息1313获得的相应基向量计数bv_cnt[layer_idx]值对各层进行解包。方法2200终止,并且处理器205中的控制返回到方法1500,从而恢复步骤1550的处理。

[0273] 图23是示出网络中的拆分点处的不同层的张量之间的关系以及所得到的用于从来自诸如张量组合器162等的张量组合器的编码张量进行再现的选项的示意框图2300。组合张量2310是由src_layer_idx标识的张量149a中的张量。从组合张量2310中,输出一个或两个输出张量,即张量2320和2330,作为张量149的一部分。张量2320保留在张量149中的位置src_layer_idx,而张量2330(如果存在)插入在张量149中的位置dst_layer_idx。图23示出基于从SEI消息解码并参考附录A描述的标志2315(即,layer_extraction_flag、layer_upsample_flag、layer_downsample_flag、resample_vertical_only_flag、resample_horizontal_only_flag)、张量2330(如果存在)与张量2320的宽度和高度相比的空间关系。标志2315在步骤1910确定,并且在步骤1920中被编码为层映射1390。在步骤2010解码标志2315以生成张量2320和张量2330(如果存在)。组合张量2310具有k1+k2个通道,其中k1对应于张量2320的通道,并且k2对应于张量2330的通道,其中边界使用src_channel_offset用信号通知。标志2315提供信号通知以允许对提取的张量进行各种重采样选项。由于张量149需要具有与张量115a相同的维度,因此可使用的标志2315的可用组合依赖于张量115a的维度,张量115a的维度取决于网络和拆分点。

[0274] 在源装置110和目的地装置140的布置中,要打包或解包的基向量数量(即bv_cnt)是相对于打包布置所提供的最大基向量数量、即相对于bv_max_cnt表达的。通常,可以使用大多数最大可用基向量(即,bv_max_cnt),因此bv_max_cnt和bv_cnt之间的差异可能相对

小。当bv_max_cnt和bv_cnt之间的差异小时,可以通过使用指数Golomb句法元素对bv_max_cnt和bv_cnt之间的差异进行编码来提高压缩效率。

[0275] 在PCA编码器160的布置中,通过QP控制器790来配置分别用于均值系数和适用于基向量的系数的子图片编码器532和572,以应用比子图片编码器514和556分别应用于编码均值特征或基向量的量化参数更低的量化参数。子图片编码器532和572中的QP控制器模块790的相应实例可以对基QP应用偏移,该基QP应用于其他地方,诸如应用于编码基向量的CU。QP偏移可以具有诸如负6或负12、负18或负24或其他值等的值,以减小在对均值系数或适用于基向量的系数(诸如在子图片1212和1214中)编码残差时所使用的量化步长大小。应用QP偏移来减小子图片编码系数的量化步长大小减小了PCA解码器170中所产生的解码系数与PCA编码器160中正向变换(例如,524和564)所生成的系数之间的差异。系数对特征图的重建具有乘法效应,这是因为系数在调节各个均值特征或基向量对重建特征图的贡献方面起作用。因此,有损编码对系数的影响导致重建特征图中保真度的大幅损失,从而对CNN头150中的任务性能产生不利影响。因此,系数需要以比基向量更高的质量级别进行编码。还可以将固定的、相对低的QP应用于负责表示系数的子图片或切片。由于系数占据了整个图片1200的相对小区域,因此由于低QP的使用而使用的位率的增加相对小,并且在任何情况下都会提高重建特征图中的保真度。可以通过指定称为“量化组”(QG)的区域以比切片级别更细的粒度来更改QP,该粒度对应于各CTU内可使用“增量QP”句法元素更改QP的最大区域粒度。因此,如果系数和基向量在同一切片中被编码,则增量QP提供了用于以比基向量所需的更高精度对系数进行编码的另一机制。

[0276] 在图片1200的布置中,使用向图块而不是子图片的分割。在使用图块的布置中,均值特征图1220、均值系数1222、基向量1226和基向量系数1224包含在单独的图块中。图块可以独立地在PCA编码器160中编码并在PCA解码器170中解码。

[0277] 在PCA编码器160和PCA解码器170的另一布置中,均值特征被均匀地应用于张量115的所有特征图。换句话说,所有均值系数被指派值“1”,而不是由点积模块524基于每个特征图导出。如果均值系数被指派“1”,则在PCA编码器160中,重建的均值特征图522不需要按均值系数缩放,因此不需要模块524、528、532、536、540,并且522作为542直接传递到模块544。在PCA解码器170中,不需要模块832、836和850,并且解码的均值特征图848作为852被传递到求和模块856。当均值系数被指派值“1”时,子图片1212可从图片1200中省略,因为无需对恒定均值系数值进行编码。当子图片1212从图1200中省略时,包含基向量系数的子图片1214的宽度与打包系数所需的最小宽度相比可能增加。子图片1214的宽度可被设置为打包系数的宽度(即,打包张量115的层中的最大通道数)的1.5倍。增加子图片1214的宽度会导致图片1200具有更宽的纵横比,这更适合于通常遇到的各个基向量的纵横比,因此导致图片1200中未使用的区域更少。可以调整子图片1210、1214和1215的宽度以保持图片结构,使得整个图片1200被子图片占据。使用恒定均值系数还消除了对相关联的量化和逆量化的需求。使用恒定均值系数仍然用于在对导出的基向量进行正向变换(点积)之前对张量115进行“零中心化”的目的,这对于保持正向和逆变换(投影)的可逆操作是必要的。

[0278] 本文呈现的方法能够以适合使用当代基于块的压缩标准(诸如VVC或HEVC等)进行压缩的格式高效地表示张量。基于块的压缩虽然不直观地适用于诸如用于投影基向量以重建特征图的系数等的数据,但诸如通过使用各种变换(包括经训练的二次变换)等,发现块

中的附加非预期冗余。尽管本文呈现的方法是参考“Faster RCNN”和“YOLOv3”网络架构以及这些网络向“主干”部和“头”部的具体分割来描述的,但这些方法适用于对多维张量数据进行操作的任何神经网络,并且适用于这些网络向“主干”部和“头”部的不同分割。

[0279] 产业上的可利用性

[0280] 所描述的布置适用于计算机和数据处理工业,尤其适用于对诸如视频和图像信号等的信号进行编码和解码的数字信号处理,从而实现高压缩效率。

[0281] 还公开了用于在通道或特征图的组中量化浮点张量数据,并使用对数量化域将所得整数值打包到平面帧中的布置。采用对数量化域的量化和逆量化方法能够由于不存在编码大幅度张量值的精确值所消耗的位而实现更大的压缩效率,其中这种精度不会导致所用网络的任务性能的附加改进。

[0282] 前面仅描述了本发明的一些实施例,并且在不脱离本发明的范围和精神的情况下,可以对其进行修改和/或改变,这些实施例是例示性的而不是限制性的。

[0283] 附录A

[0284] 位流中用于表示与基向量打包相关联的元数据(即,1391)的示例SEI消息格式和相关联的语义如下:

[0285]

basis_vector_packing_info(payloadSize) {	**描述符**
basis_vector_dimensions_update	u(1)
if(basis_vector_dimensions_update != 0) {	

[0286]

layer_cnt	u(1)
for(layer_idx = 0; layer_idx < layer_cnt; layer_idx++) {	
bv_max_cnt[layer_idx]	ue(v)
bv_width[layer_idx]	ue(v)
bv_height[layer_idx]	ue(v)
}	
}	
basis_vector_cnt_update	u(1)
if(basis_vector_cnt_update)	
{	
for(layer_idx = 0; layer_idx < layer_cnt; layer_idx++) {	
bv_cnt[layer_idx]	ue(v)
}	
}	
layer_update_mapping_info	u(1)
if(layer_update_mapping_info)	
{	
layer_extraction_flag	u(1)
if(layer_extraction_flag)	
{	
layer_upsample_flag	u(1)
layer_downsampling_flag	u(1)
if(layer_upsample_flag || layer_downsampling_flag)	
{	
resampling_filter_idx	ue(v)
resample_vertical_only_flag	u(1)

[0287]

resample_horizontal_only_flag	u(1)
}	
src_layer_idx	ue(v)
src_channel_offset	ue(v)
dst_layer_idx	ue(v)
}	
}	
}	

[0288] 基向量打包信息语义

[0289] 以上句法结构指定解包基向量平面帧并转换为用于投影到解码张量的基向量以进行推断任务所需的信息。

[0290] 具有描述符u(n)的句法元素指示使用n位对句法元素进行编码,并且将句法元素解释为无符号整数值。具有描述符ue(v)的句法元素指示句法元素被编码为指数Golomb值并被解释为无符号整数值。

[0291] 基向量打包信息SEI消息的持久性从相关联的AU开始,直到基向量打包信息SEI消息的下一次出现或CLVS的结束为止。在下一次出现基向量打包信息SEI消息时,参数基于更新标志进行分组。未更新(即,包含在上述句法表的分支中,其中该分支的更新标志被设置为零)的参数保留其以前的值(如果有)。

[0292] basis_vector_dimensions_update等于1指定了对最大基向量计数和基向量维度的更新,适用于当前帧和后续帧,即,直到下一次出现basis_vector_dimensions_update等于1的基向量打包信息SEI消息为止。位流一致性的要求是,CLVS中的基向量打包信息SEI消息的第一实例的basis_vector_dimensions_update等于1。

[0293] 注意:通常,basis_vector_dimensions_update等于1的基向量打包信息SEI消息与CLVS的第一图片一起发送。

[0294] basis_vector_cnt_update等于1指定了对已使用基向量计数的更新,适用于CLVS中的当前帧和后续帧。已使用基向量计数指示使用了最大基向量中的多少个基向量。

[0295] layer_cnt指定了帧中存在的层数。

[0296] bv_max_cnt[layer_idx]指定了针对layer_idx存在的最大基向量数量。

[0297] bv_width[layer_idx]指定了针对layer_idx的基向量的宽度。

[0298] bv_height[layer_idx]指定了针对layer_idx的基向量图的高度。

[0299] bv_cnt[layer_idx]指定了针对layer_idx存在的基向量数量。位流一致性的要求是,对于范围在0到layer_cnt-1中的各值layer_idx,bv_cnt[layer_idx]小于或等于bv_max_cnt[layer_idx]。当bv_cnt[layer_idx]尚未从位流中解码时,对于范围在0到layer_cnt-1中的layer_idx,将值bv_cnt[layer_idx]设置为等于bv_max_cnt[layer_idx]。

[0300] layer_update_mapping_info等于1指定了对如由张量组合器162确定并由张量分离器172进行的张量重映射的配置的更新。当layer_update_mapping_info等于零时,不会

对张量重映射的配置进行更新,并且任何先前的张量重映射的配置继续应用。

[0301] layer_extraction_flag等于1指定了与相对于张量中的最后一个通道的通道src_channel_offset相对应的张量的一部分,来自层src_layer_idx的张量将被提取以产生新层,以插入到dst_layer_idx处的张量列表中。当layer_extraction_flag等于零时,不进行张量的提取,导致张量149被设置为等于张量149a。当先前未从位流中解析时,layer_extraction_flag被推断为零。

[0302] layer_upsample_flag等于1指定了要提取的张量的一部分被上采样以产生具有更大宽度和高度(例如,输入张量的两倍宽度和两倍高度)的新张量。

[0303] layer_downsample_flag等于1指定了要提取的张量的一部分被下采样以产生具有较小宽度和高度(例如,输入张量的一半宽度和一半高度的)的新张量。

[0304] 位流一致性的要求是,在基向量打包信息SEI消息的一个实例中,layer_upsample_flag和layer_downsample_flag不能两者都等于1。

[0305] resampling_filter_idx指定了在产生新张量时对张量的一部分进行重采样时要应用的滤波的类型。根据resample_vertical_only_flag和resample_horizontal_only_flag,水平地、垂直地、或水平地和垂直地应用重采样。下表列举了可用的滤波:

[0306]

resampling_filter_idx	在下采样时	在上采样时
0	抽取	双线性滤波
1	抗混叠,抽取	双三次滤波
2	保留以供将来使用	最近邻
3-	保留以供将来使用	保留以供将来使用

[0307] resample_vertical_only_flag等于1指定了重采样操作仅在垂直方向上应用。

[0308] resample_horizontal_only_flag等于1指定了重采样操作仅在水平方向上应用。

[0309] 位流一致性的要求是,在基向量映射信息SEI消息的任何实例中,resample_vertical_only_flag和resample_horizontal_only_flag不能两者都等于1。

[0310] src_layer_idx指定了要从张量149a的哪个张量中提取一部分进行重采样,以产生要包括在张量149中的新张量。

[0311] src_channel_offset指定了要从哪个通道偏移量中分割src_layer_idx所标识的张量。src_layer_idx所指定的张量的通道0至src_channel_offset-1被保留并输出为张量149中的相应张量。相对于src_layer_idx所指定的张量的最后一个通道的通道src_channel_offset经受重采样,其结果作为张量149中的附加张量输出。

[0312] dst_layer_idx指定了重采样张量被插入张量149a中以产生张量149的索引。索引值零指定了开头的插入,即创建新的第一个位置,并且索引值等于layer_cnt指定了最后一个位置之后的插入,即附加新张量以创建新的最后一个位置。从1到layer_cnt-1的索引值指示将重采样张量插入到位置dst_layer_idx和dst_layer_idx+1处的张量149a之间的位置以创建张量149。

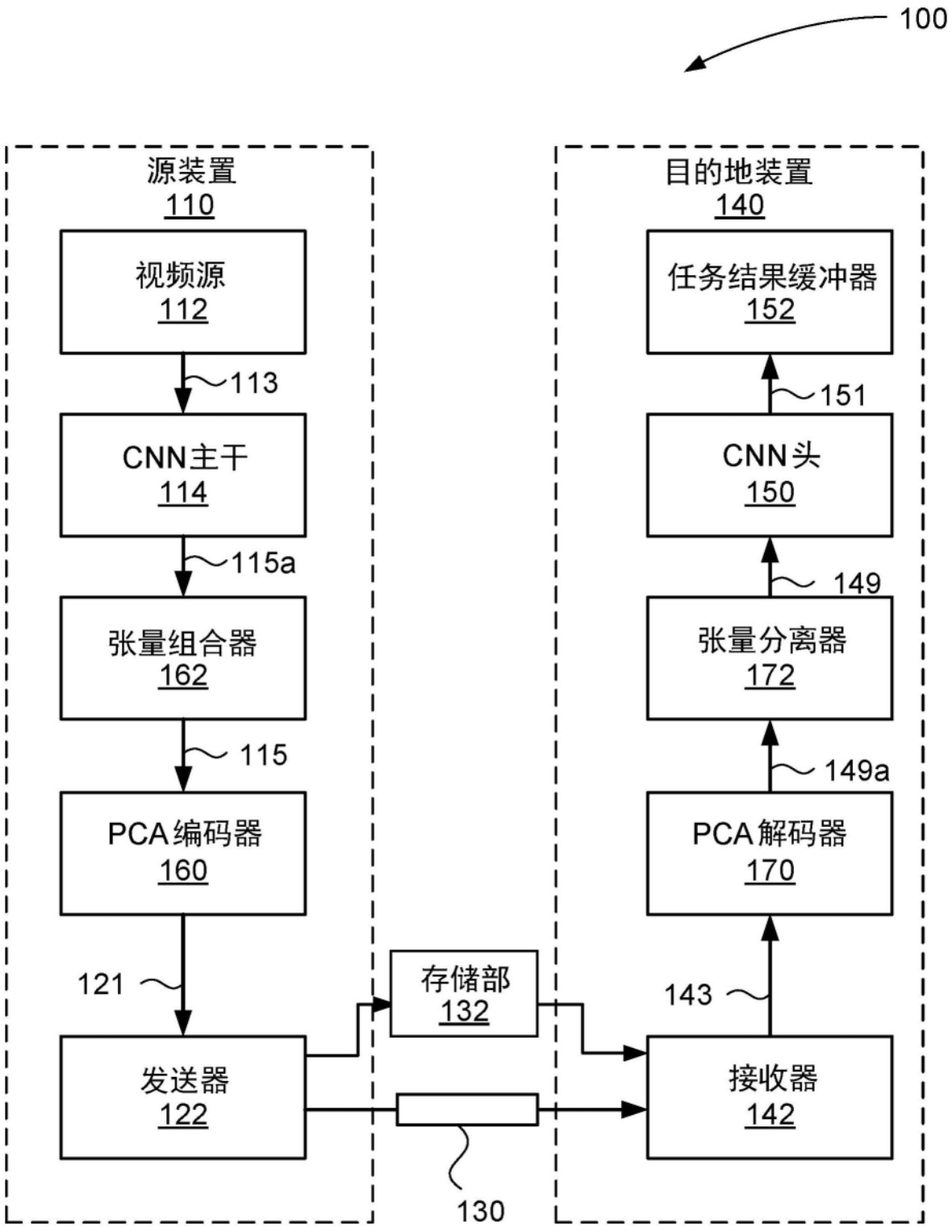

100
源装置
110
视频源
112
113
CNN主干
114
115a
张量组合器
162
115
PCA编码器
160
121
发送器
122
存储部
132
130
目的地装置
140
任务结果缓冲器
152
151
CNN头
150
149
张量分离器
172
149a
PCA解码器
170
143
接收器
142

图1

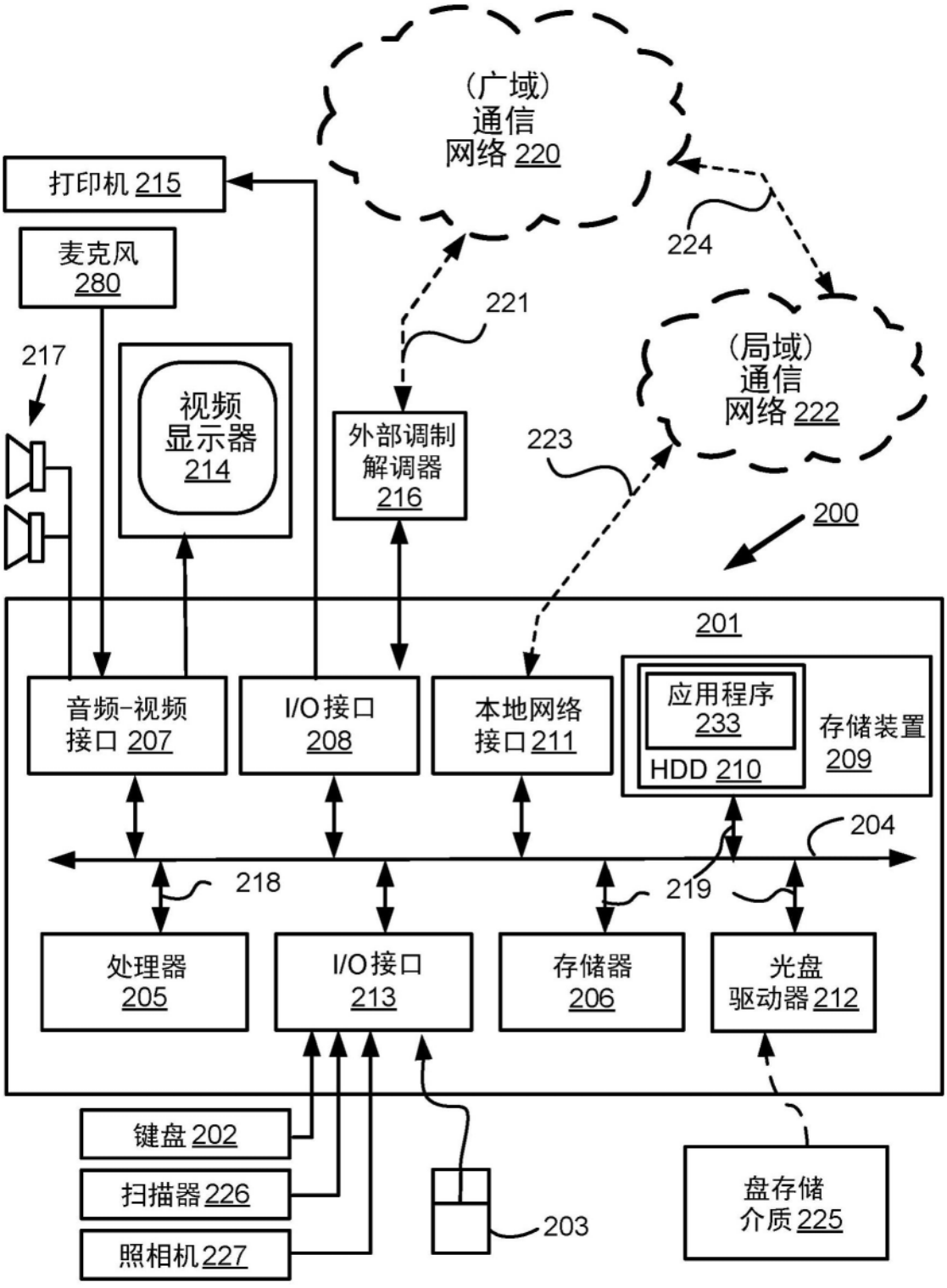

(广域)
通信
网络 220
打印机 215
麦克风
280
224
221
217
视频
显示器
214
外部调制
解调器
216
(局域)
通信
网络 222
223
200
201
音频-视频
接口 207
I/O接口
208
本地网络
接口 211
应用程序
233
HDD 210
存储装置
209
204
218
219
处理器
205
I/O接口
213
存储器
206
光盘
驱动器 212
键盘 202
扫描器 226
照相机 227
203
盘存储
介质 225

图2A

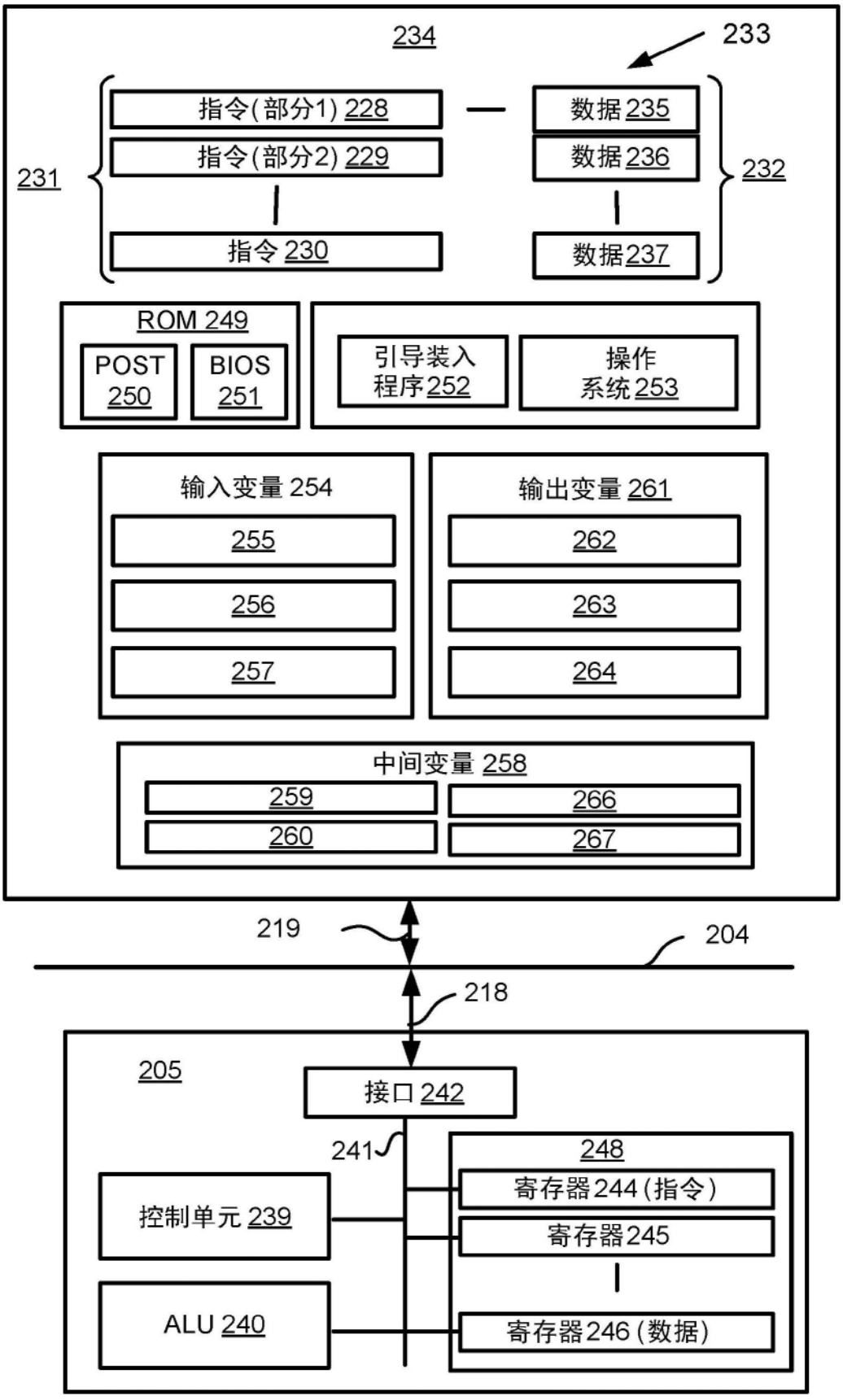

234
233
指令(部分1) 228
指令(部分2) 229
指令 230
231
数据 235
数据 236
数据 237
232
ROM 249
POST 250
BIOS 251
引导装入程序 252
操作系统 253
输入变量 254
255
256
257
输出变量 261
262
263
264
中间变量 258
259
260
266
267
219
204
218
205
接口 242
241
248
寄存器 244 (指令)
寄存器 245
寄存器 246 (数据)
控制单元 239
ALU 240

图2B

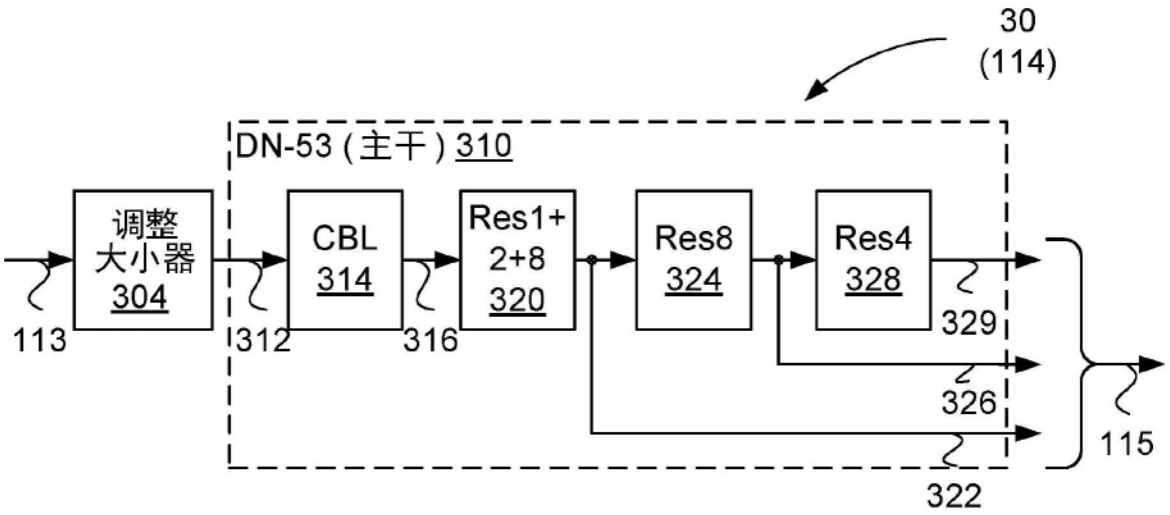

30
(114)
DN-53 (主干) 310
调整
大小器
304
CBL
314
Res1+
2+8
320
Res8
324
Res4
328
113
312
316
329
326
322
115

图3A

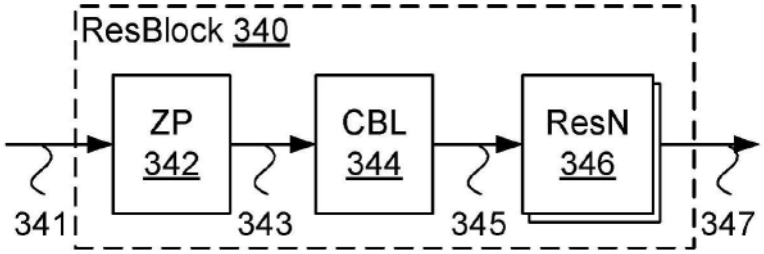

ResBlock 340
ZP
342
CBL
344
ResN
346
341
343
345
347

图3B

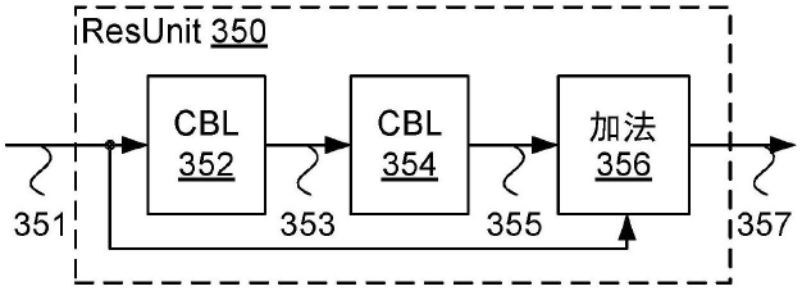

ResUnit 350
CBL
352
CBL
354
加法
356
351
353
355
357

图3C

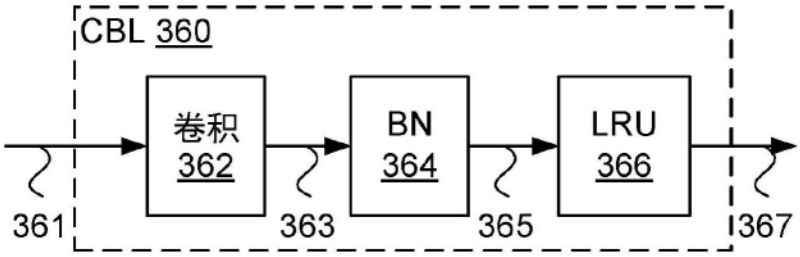


图3D

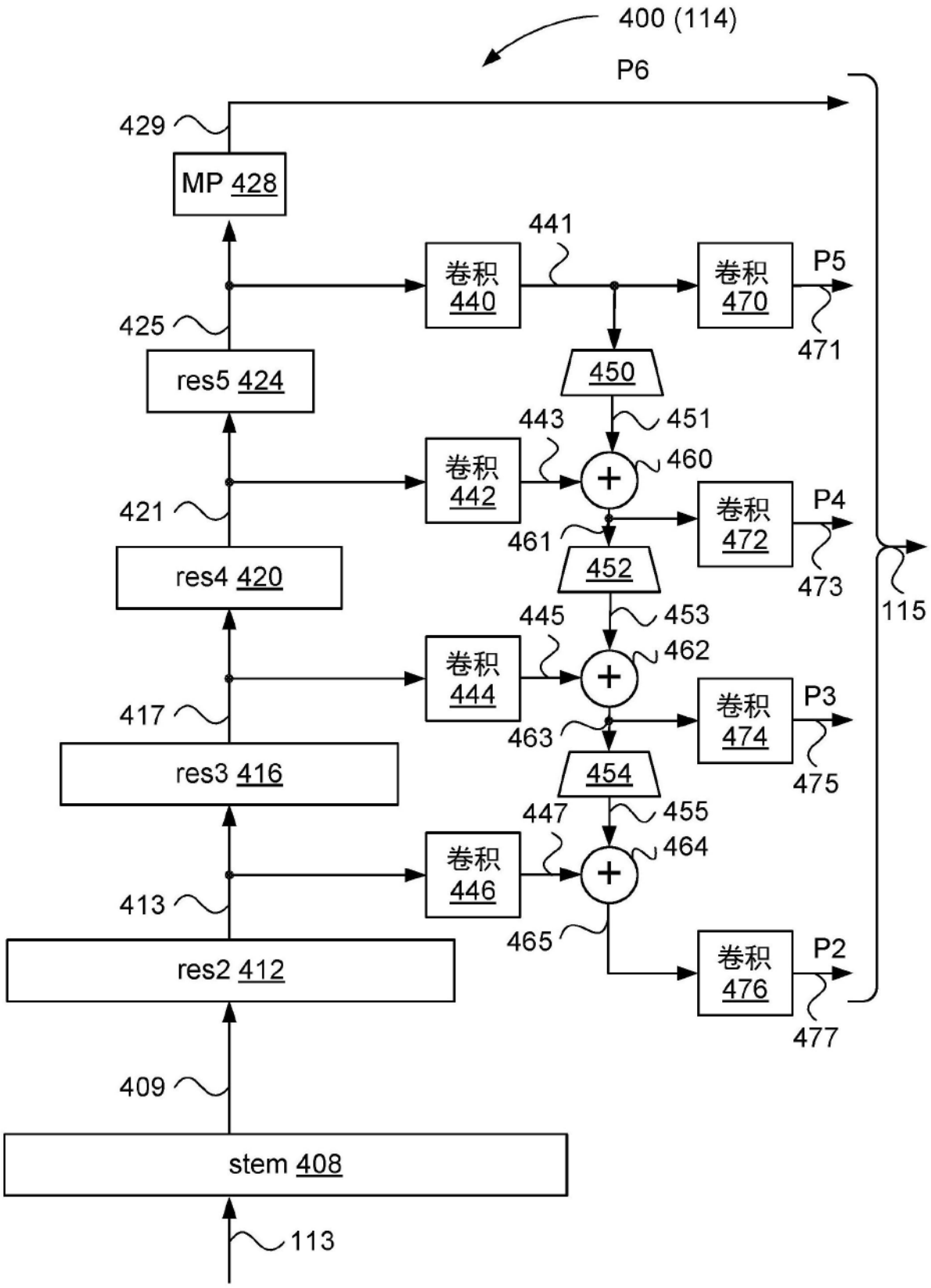

400 (114)
P6
429
MP 428
441
卷积 440
卷积 470
P5
471
425
res5 424
450
443
451
卷积 442
460
卷积 472
P4
473
421
461
res4 420
452
445
453
462
卷积 444
卷积 474
P3
475
417
463
res3 416
454
447
455
464
卷积 446
413
465
res2 412
卷积 476
P2
477
115
409
stem 408
113

图4

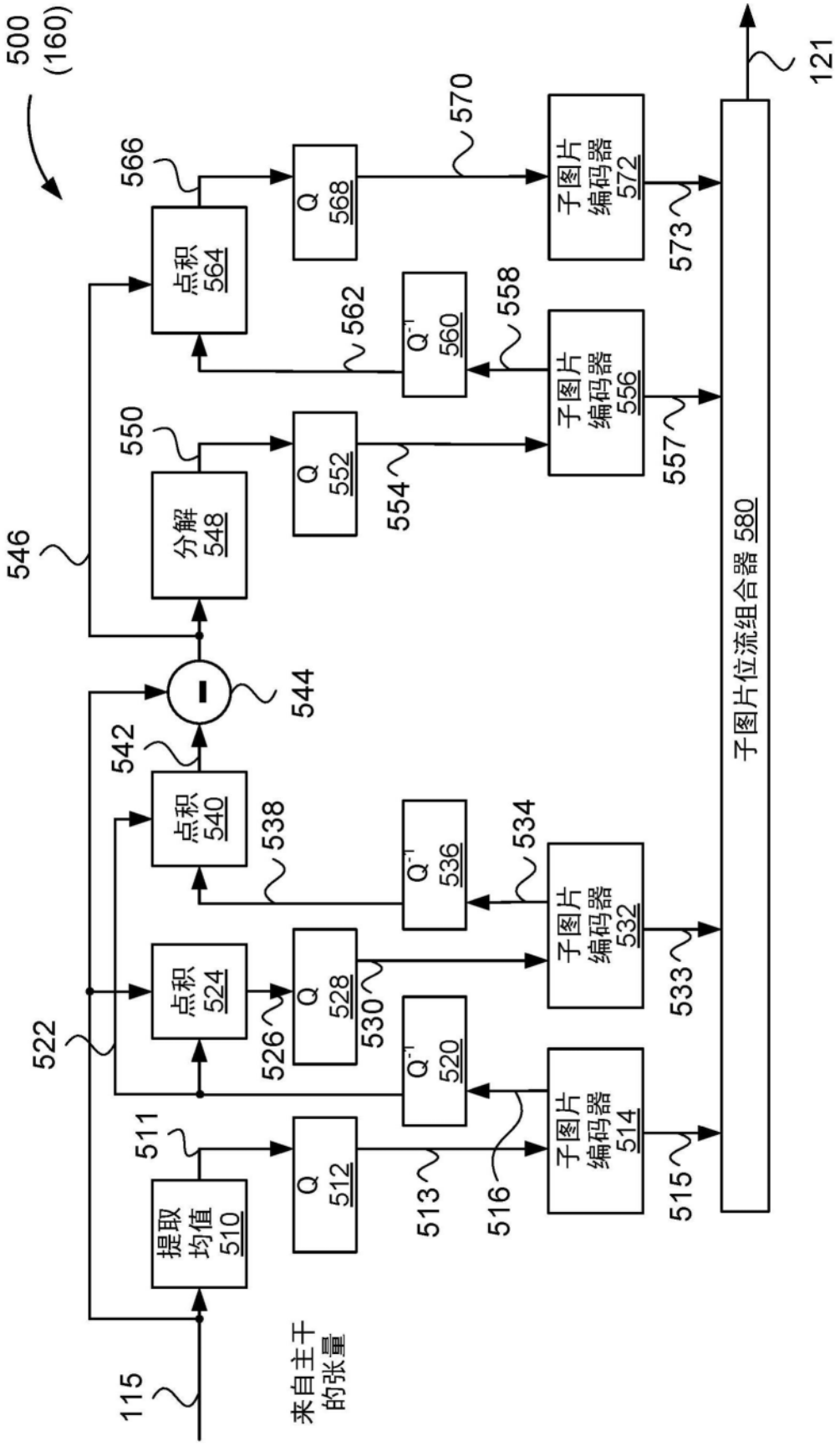

500
(160)
115
来自主干
的张量
提取
均值
510
511
Q
512
513
516
子图片
编码器
514
515
Q⁻¹
520
522
点积
524
Q
528
526
530
子图片
编码器
532
533
Q⁻¹
536
538
534
点积
540
542
544
546
分解
548
550
Q
552
554
子图片
编码器
556
557
Q⁻¹
560
562
558
点积
564
566
Q
568
570
子图片
编码器
572
573
子图片位流组合器 580
121

图5

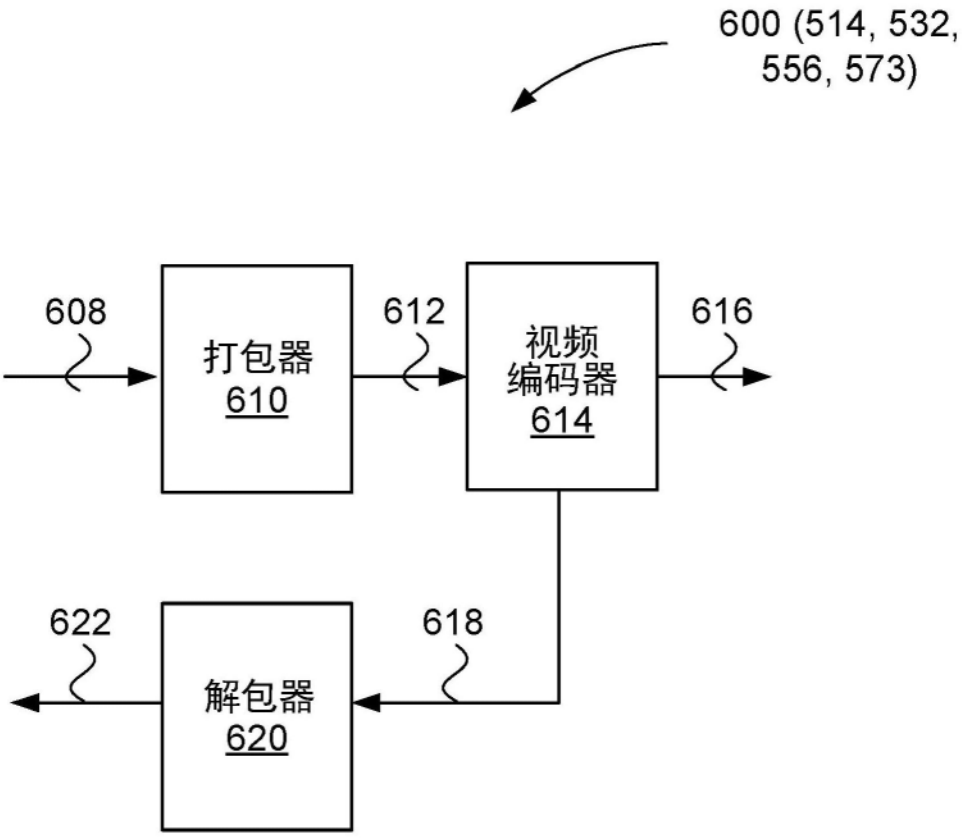

600 (514, 532, 556, 573)
608
打包器 610
612
视频编码器 614
616
622
解包器 620
618

图6

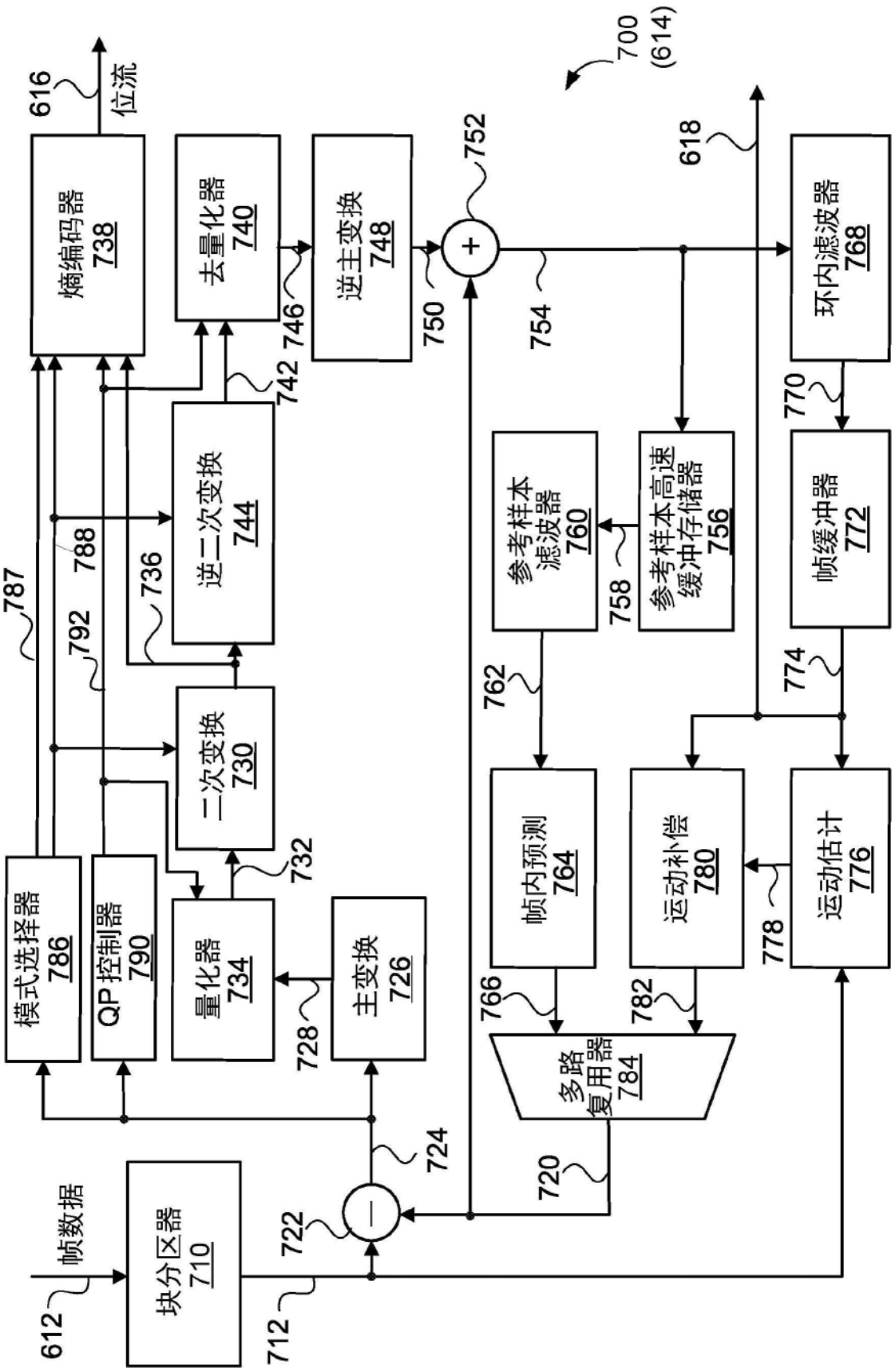

700
(614)
616
位流
熵编码器
738
去量化器
740
逆主变换
748
752
618
环内滤波器
768
746
750
754
742
逆二次变换
744
参考样本
滤波器
760
参考样本高速
缓冲存储器
756
770
帧缓冲器
772
736
788
787
792
758
二次变换
730
762
774
732
模式选择器
786
QP控制器
790
量化器
734
主变换
726
帧内预测
764
运动补偿
780
运动估计
776
778
728
766
782
多路
复用器
784
724
720
722
帧数据
612
块分区器
710
712

图7

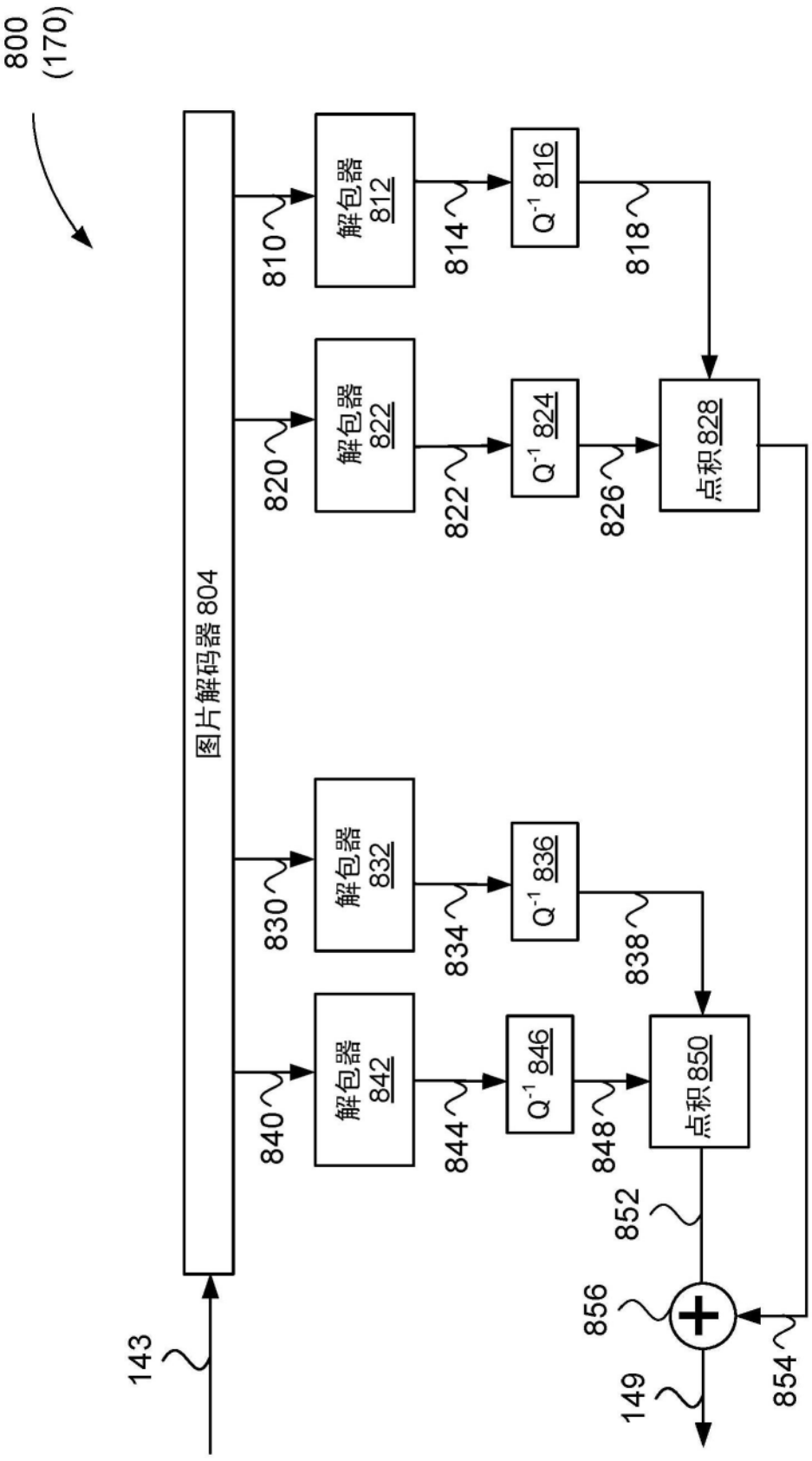

800
(170)
图片解码器 804
143
解包器 812
810
814
Q-1 816
818
解包器 822
820
822
Q-1 824
826
点积 828
解包器 832
830
834
Q-1 836
838
解包器 842
840
844
Q-1 846
848
点积 850
852
856
854
149

图8

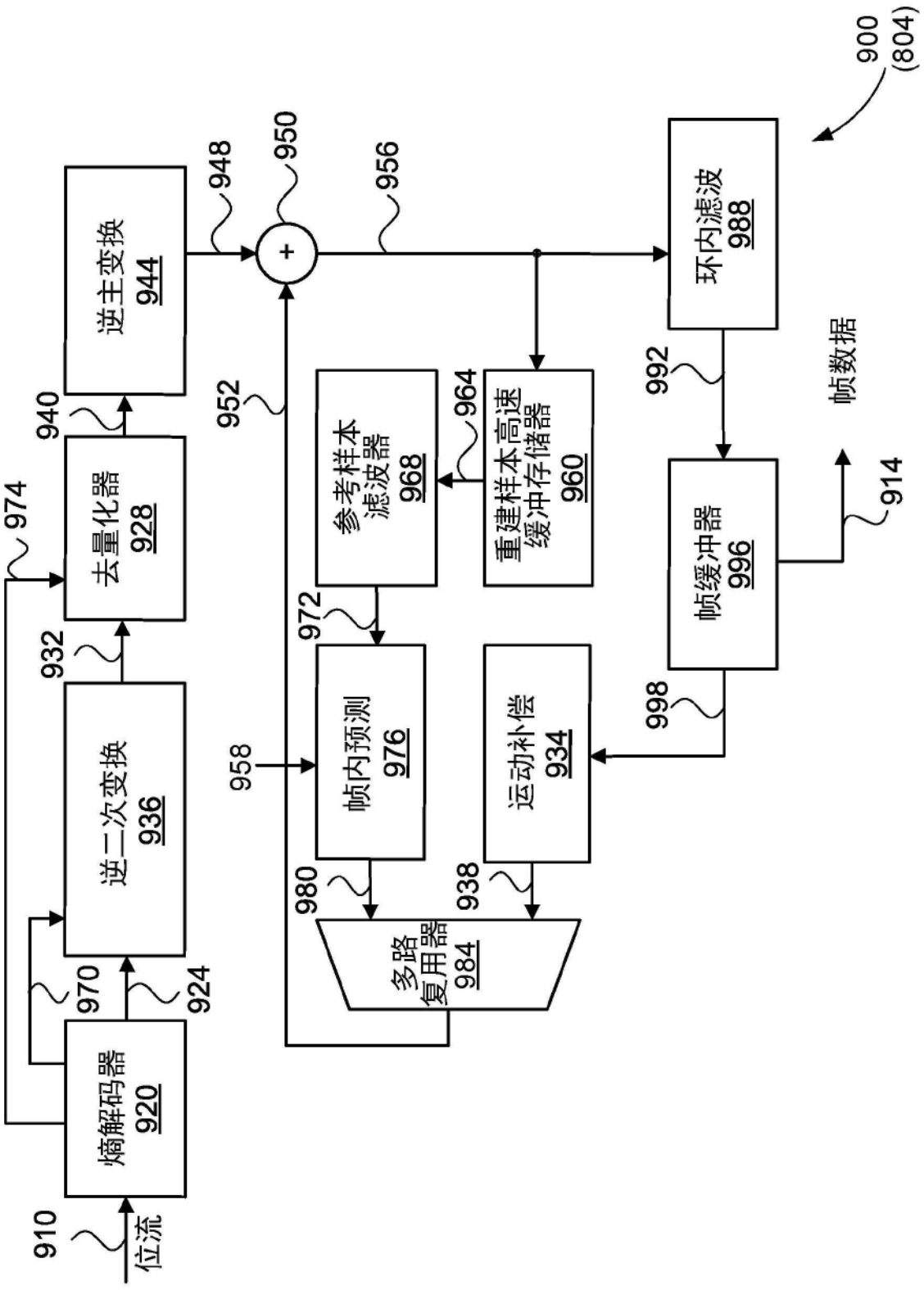

900
(804)
逆主变换
944
948
950
956
环内滤波
988
940
952
964
参考样本
滤波器
968
重建样本高速
缓冲存储器
960
992
帧数据
去量化器
928
974
帧缓冲器
996
914
932
972
逆二次变换
936
帧内预测
976
运动补偿
934
998
958
980
938
多路
复用器
984
970
924
熵解码器
920
910
位流

图9

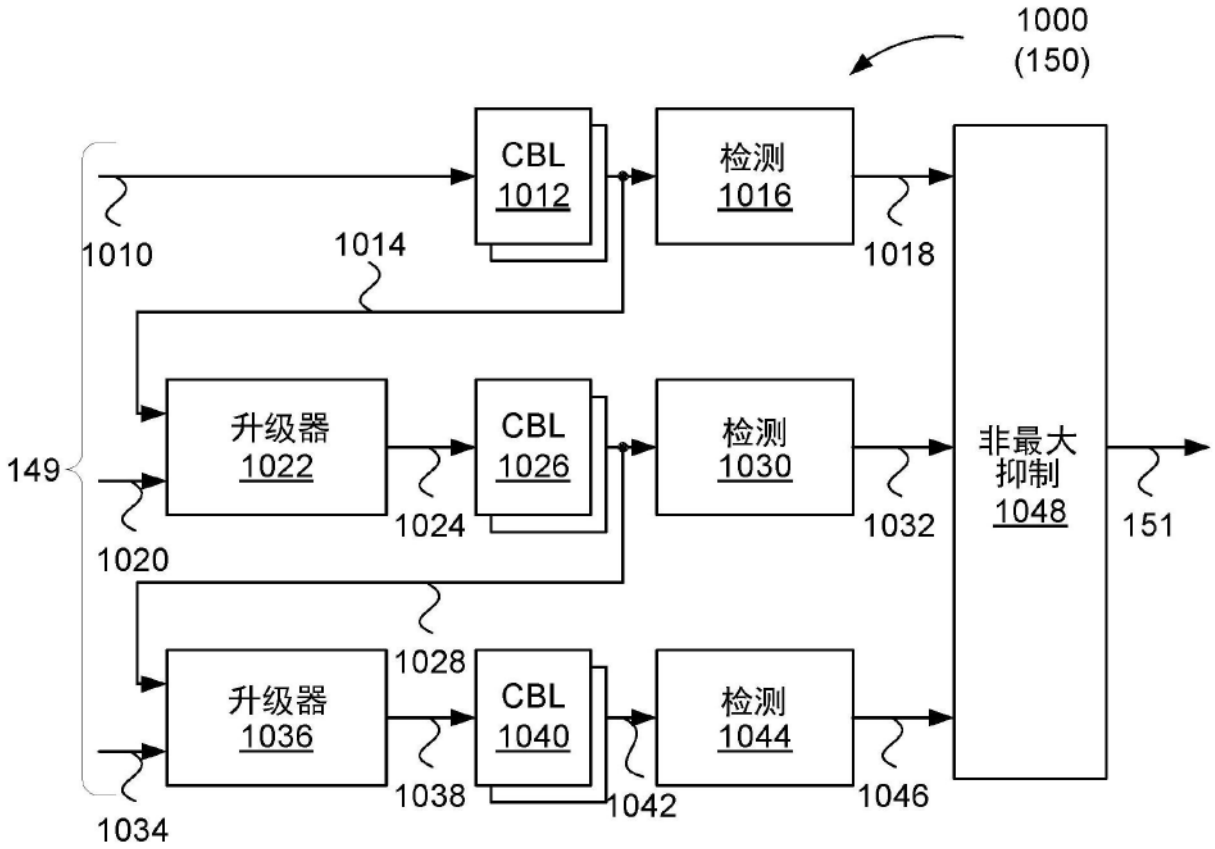

1000
(150)
CBL
1012
检测
1016
1010
1014
1018
升级器
1022
CBL
1026
检测
1030
非最大
抑制
1048
149
1024
1032
151
1020
1028
升级器
1036
CBL
1040
检测
1044
1038
1042
1046
1034

图10A

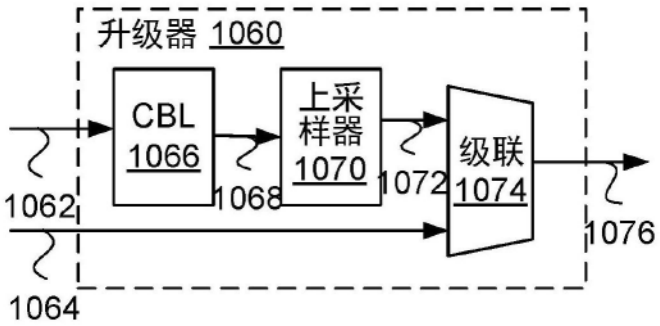

升级器 1060
CBL
1066
上采
样器
1070
级联
1074
1062
1068
1072
1076
1064

图10B

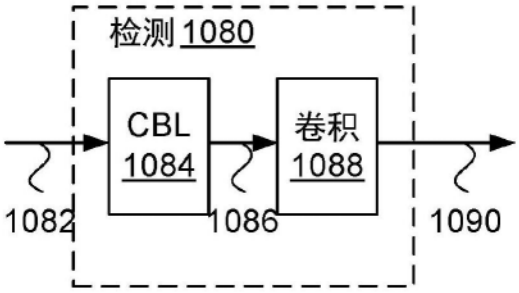


图10C

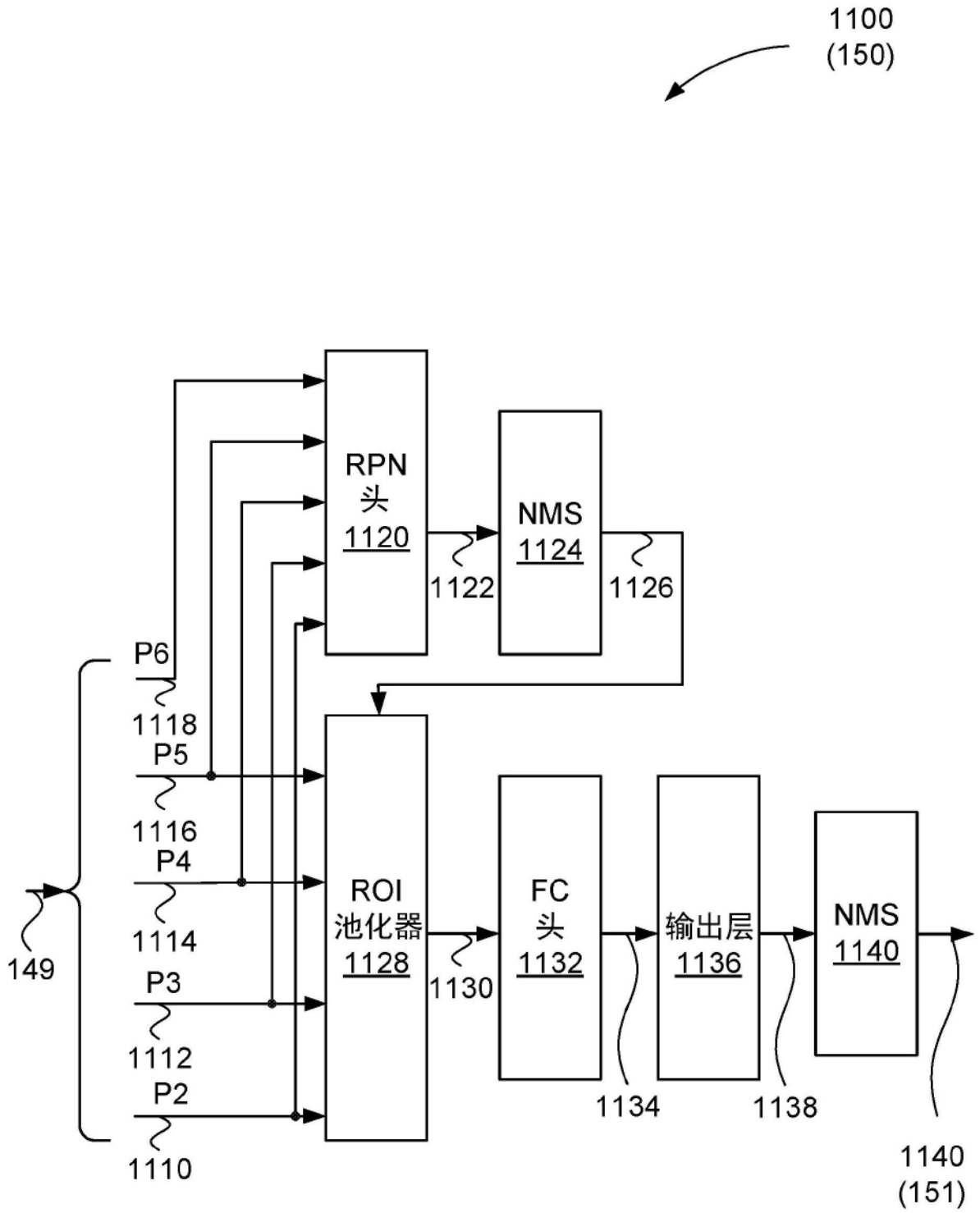

1100
(150)
RPN
头
1120
NMS
1124
1122
1126
P6
1118
P5
1116
P4
1114
P3
1112
P2
1110
149
ROI
池化器
1128
1130
FC
头
1132
1134
输出层
1136
1138
NMS
1140
1140
(151)

图11

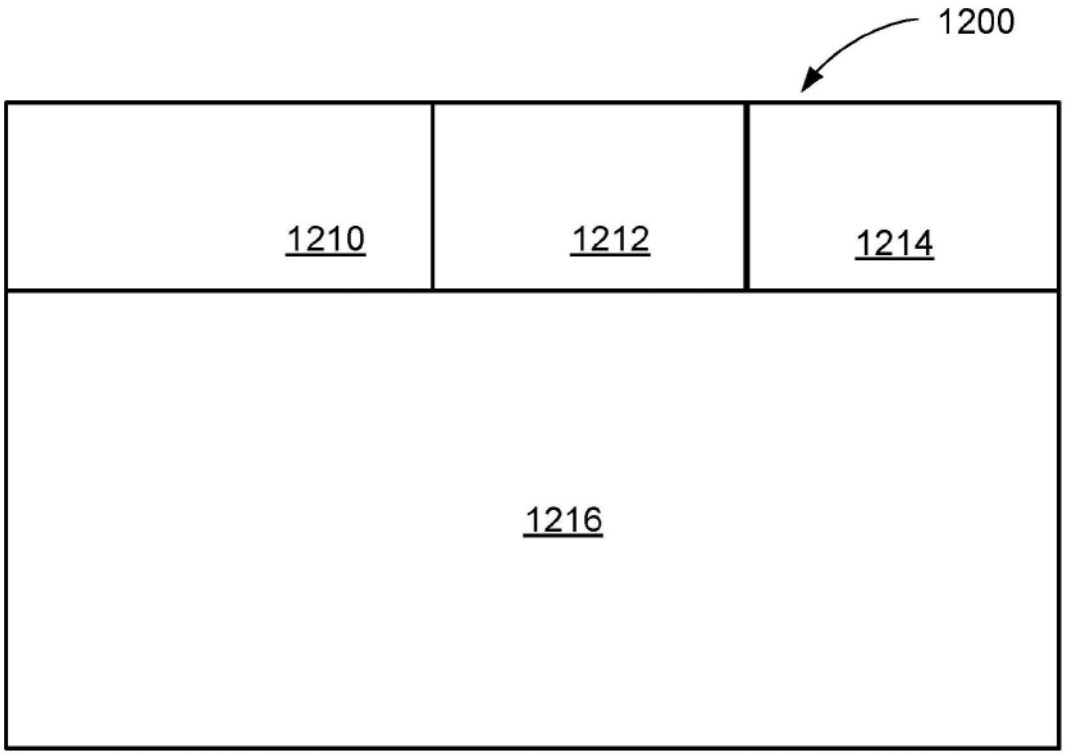

1200
1210
1212
1214
1216

图12A

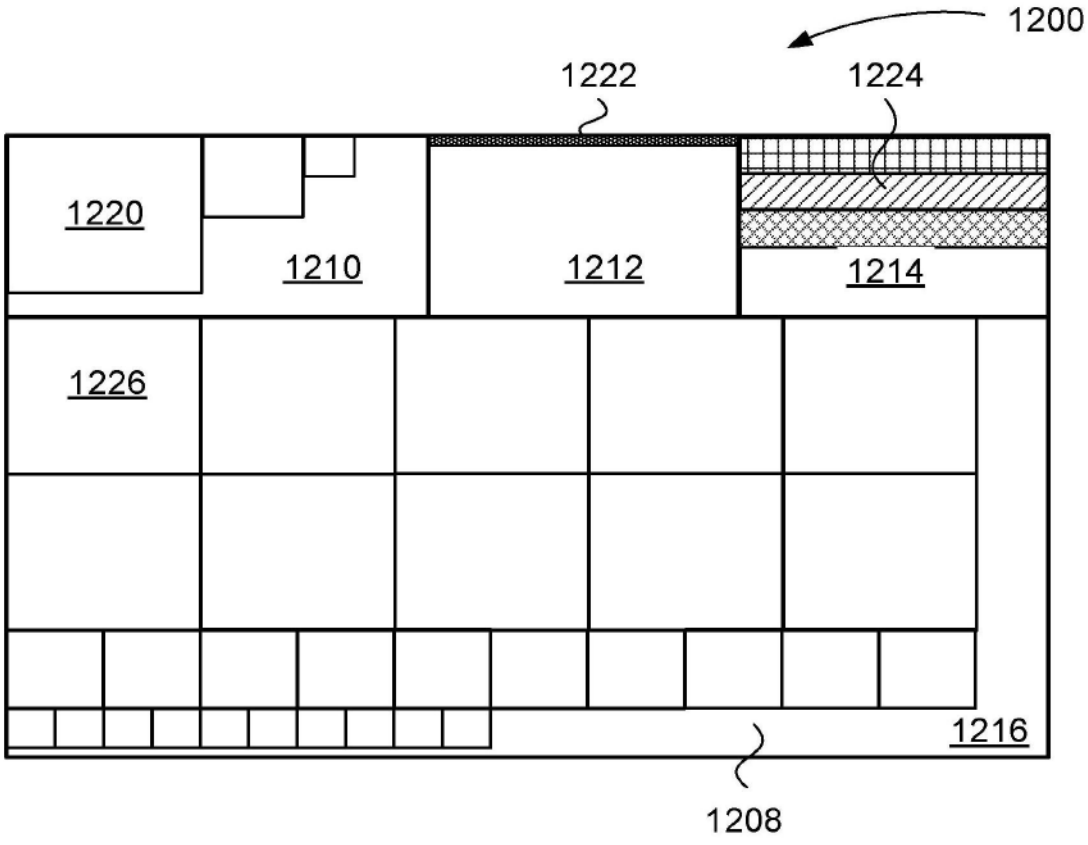

1200
1222
1224
1220
1210
1212
1214
1226
1216
1208

图12B

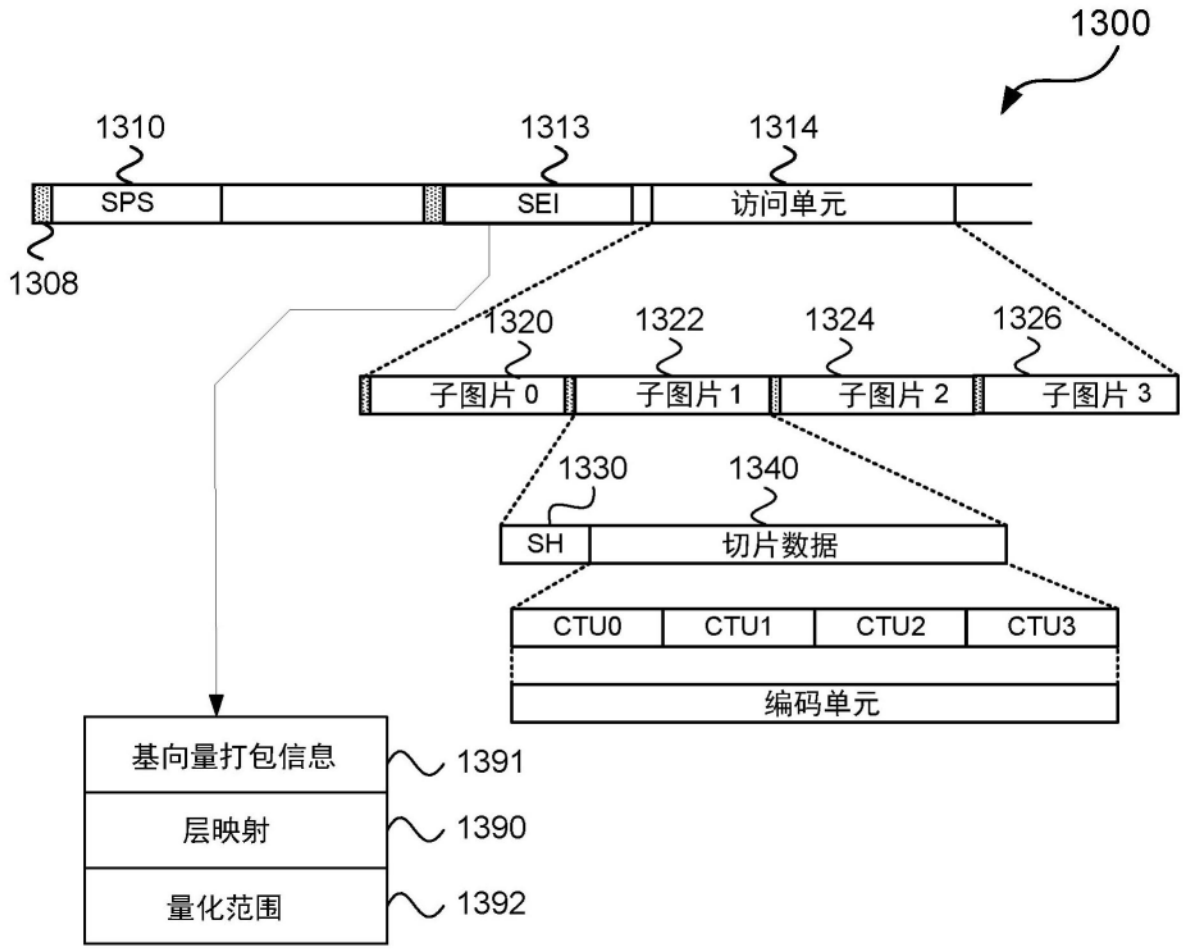

1300
1310
1313
1314
SPS
SEI
访问单元
1308
1320
1322
1324
1326
子图片 0
子图片 1
子图片 2
子图片 3
1330
1340
SH
切片数据
CTU0
CTU1
CTU2
CTU3
编码单元
基向量打包信息
1391
层映射
1390
量化范围
1392

图13

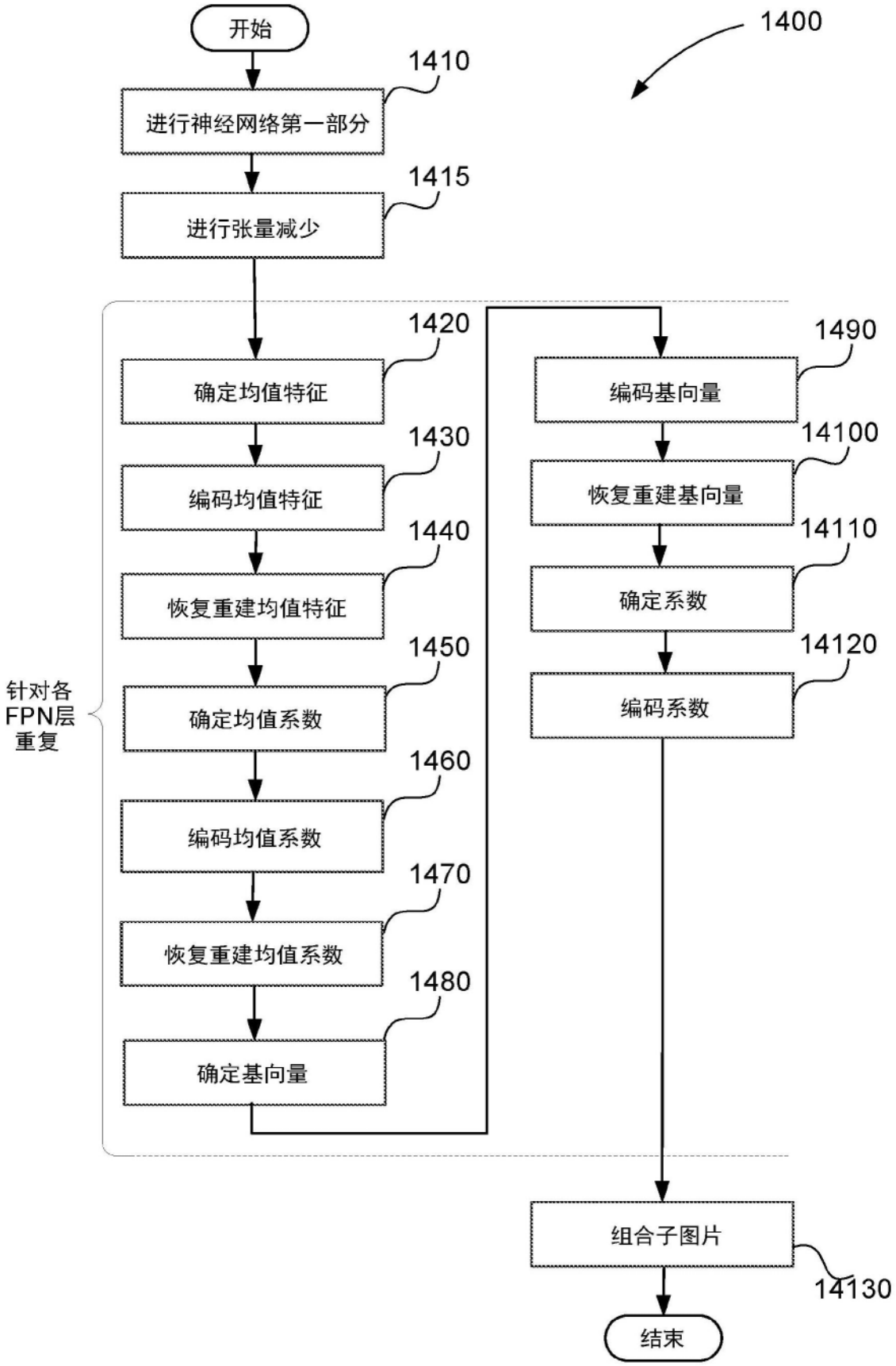

开始
1400
1410
进行神经网络第一部分
1415
进行张量减少
1420
确定均值特征
1430
编码均值特征
1440
恢复重建均值特征
1450
确定均值系数
1460
编码均值系数
1470
恢复重建均值系数
1480
确定基向量
针对各FPN层重复
1490
编码基向量
14100
恢复重建基向量
14110
确定系数
14120
编码系数
组合子图片
14130
结束

图14

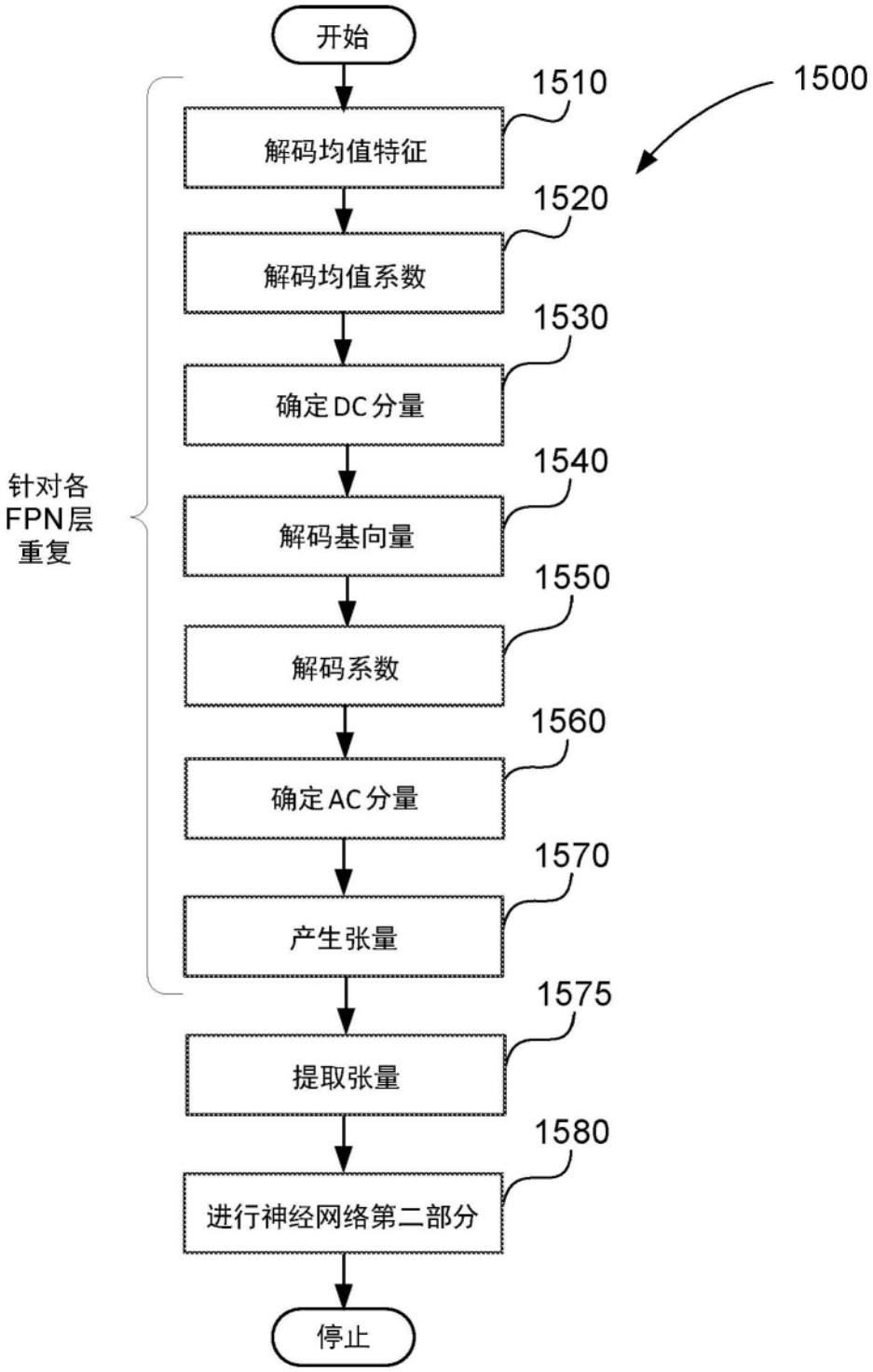

开始
1500
1510
解码均值特征
1520
解码均值系数
1530
确定DC分量
1540
解码基向量
1550
解码系数
1560
确定AC分量
1570
产生张量
1575
提取张量
1580
进行神经网络第二部分
停止
针对各
FPN层
重复

图15

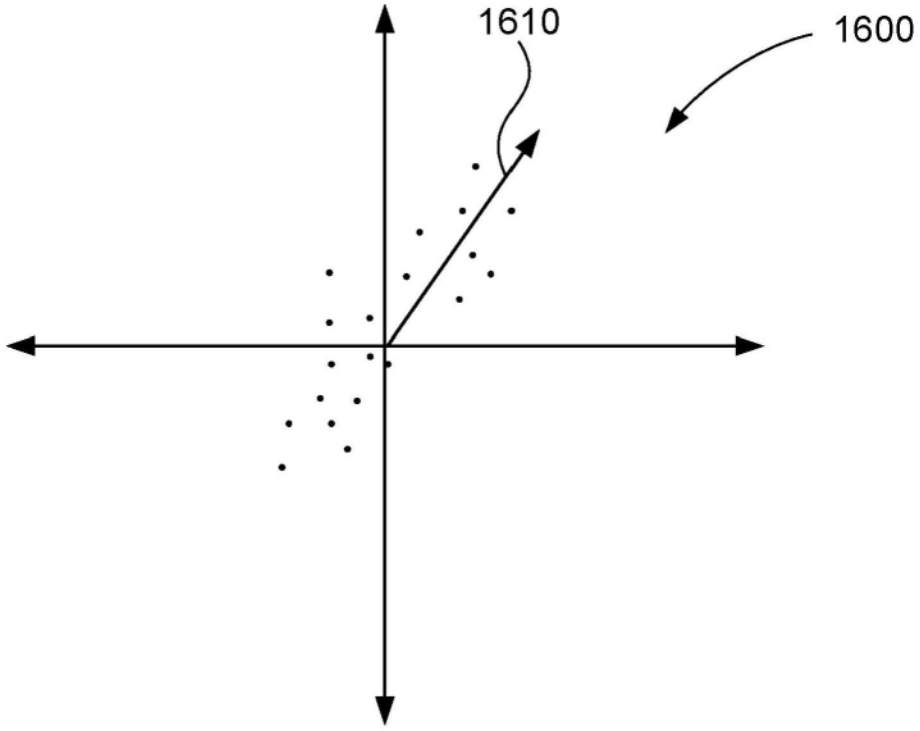


图16A

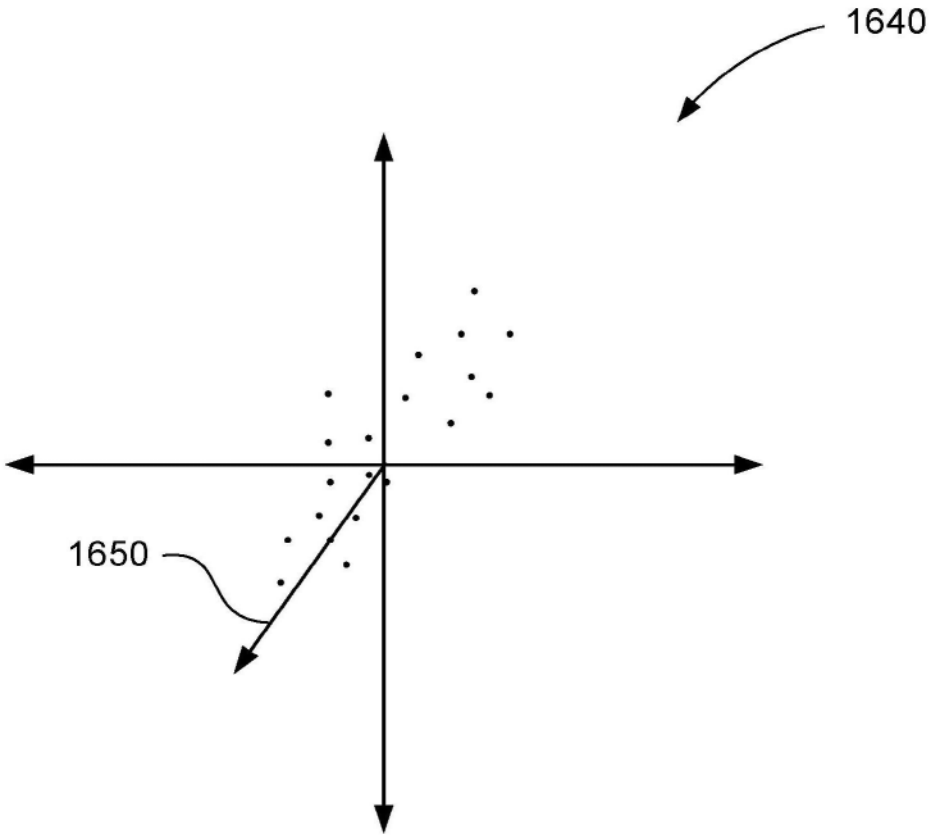


图16B

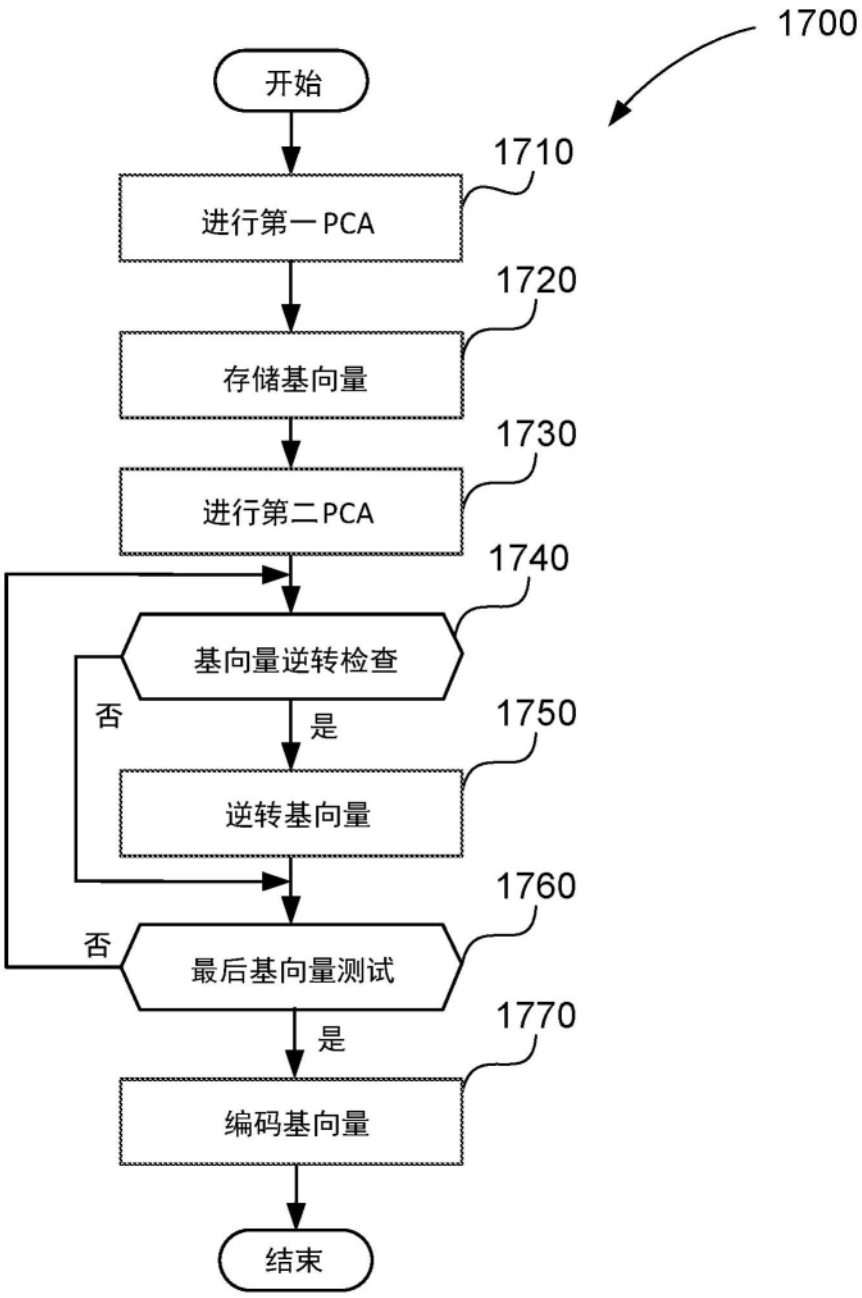

1700
开始
1710
进行第一PCA
1720
存储基向量
1730
进行第二PCA
1740
基向量逆转检查
否
是
1750
逆转基向量
1760
否
最后基向量测试
是
1770
编码基向量
结束

图17

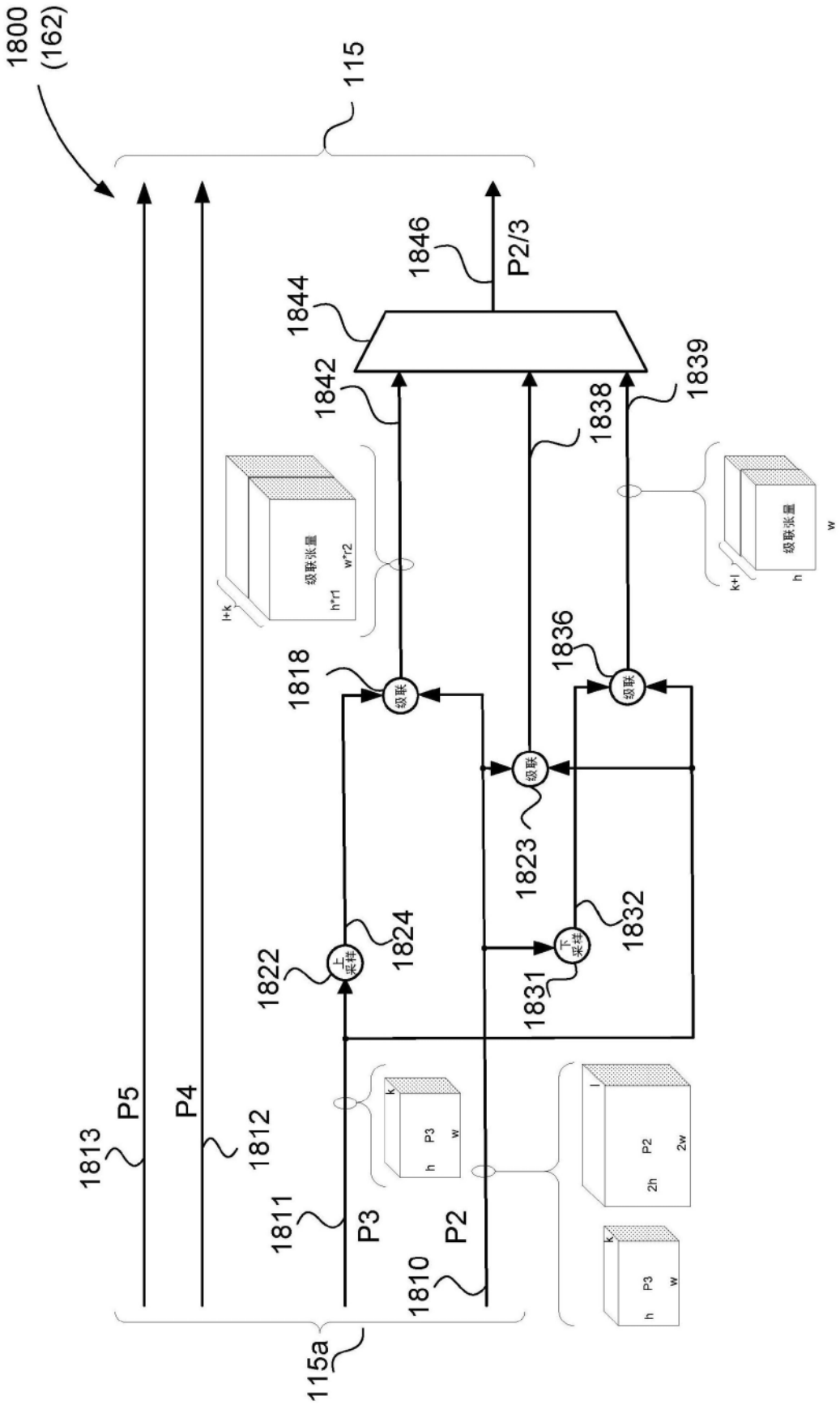

1800
(162)
115
1846
P2/3
1844
1842
1839
1838
1836
1818
1823
1832
1824
1822
1831
1813
P5
P4
1812
1811
P3
P2
1810
115a
级联
上采样
下采样
级联张量
h*r1
w*r2
l+k
k+l
h
w
k
P3
2h
2w
l
P2

图18A

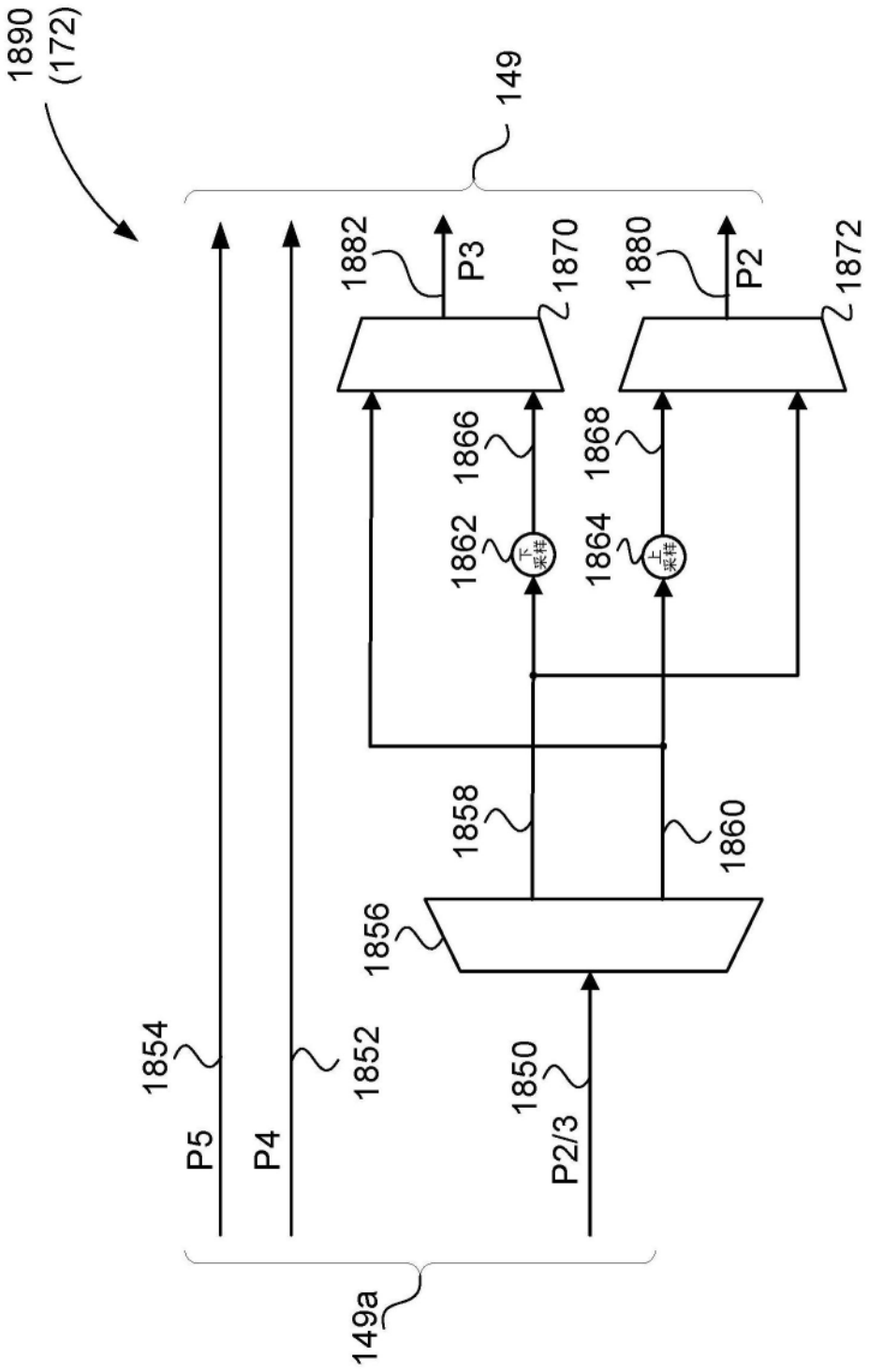

1890
(172)
149
1882
P3
1870
1880
P2
1872
1866
1868
1862
下采样
1864
上采样
1858
1860
1856
1854
1852
1850
P5
P4
P2/3
149a

图18B

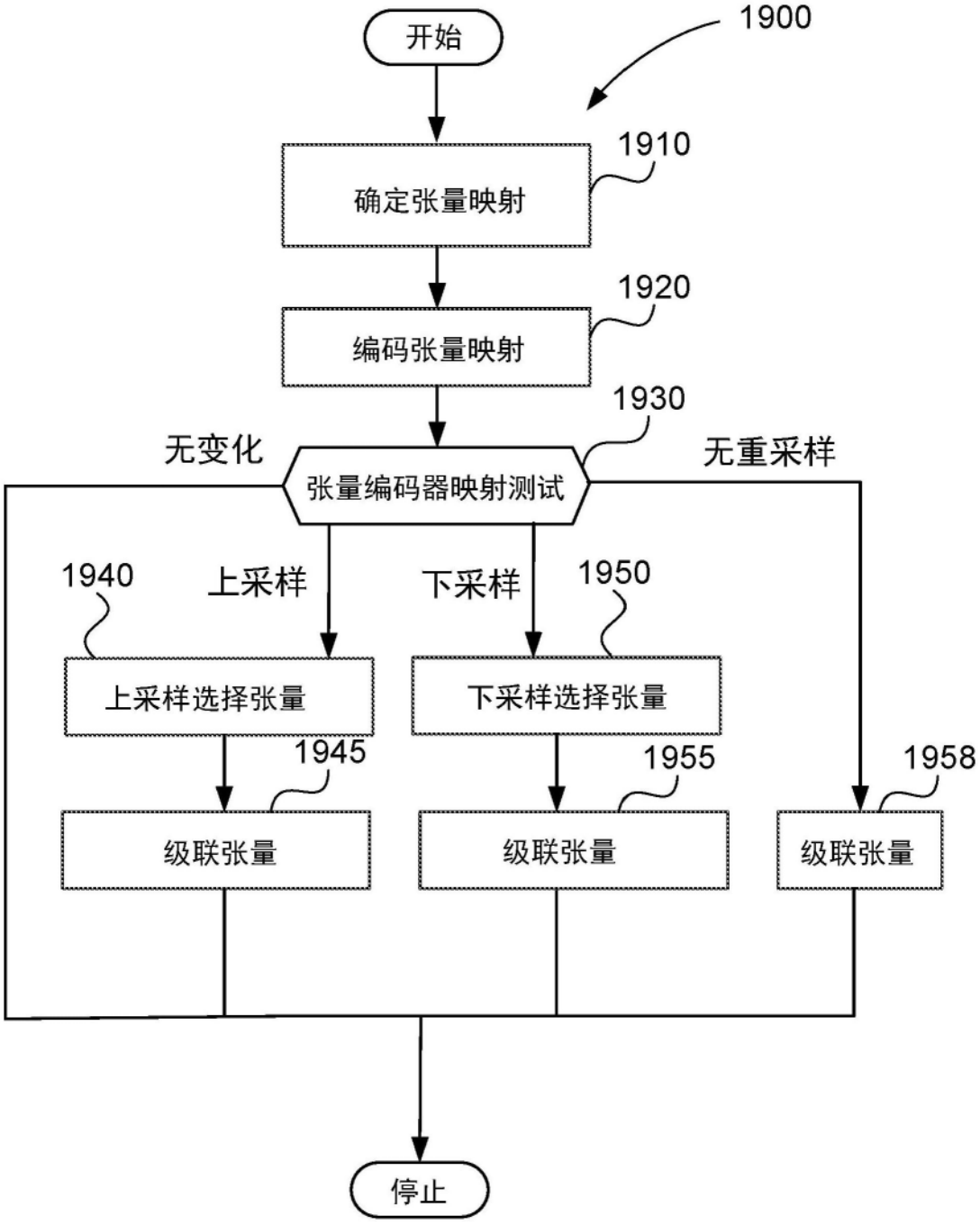

开始
1900
确定张量映射
1910
编码张量映射
1920
张量编码器映射测试
1930
无变化
无重采样
1940
上采样
下采样
1950
上采样选择张量
下采样选择张量
1945
1955
1958
级联张量
级联张量
级联张量
停止

图19

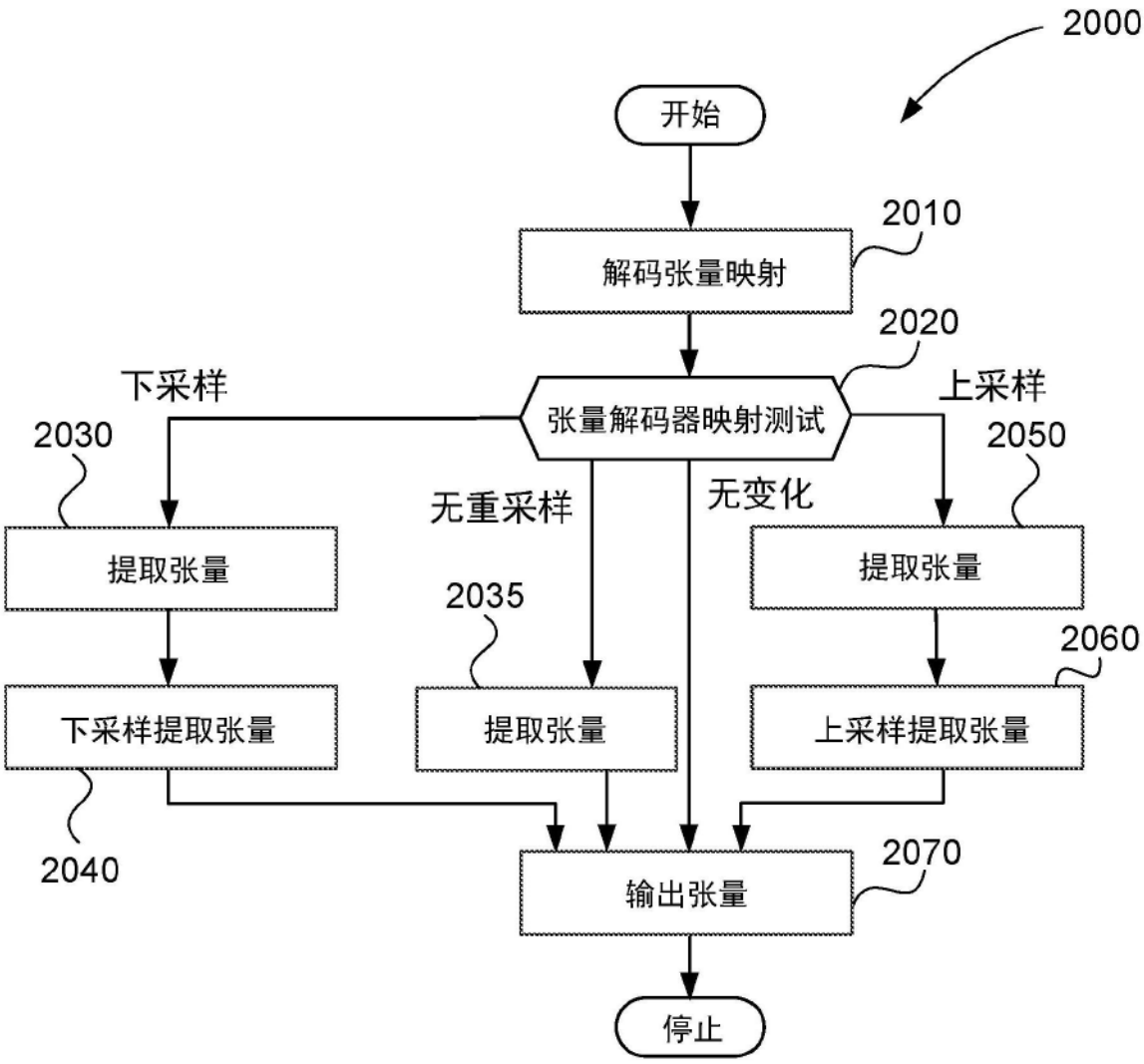

2000
开始
2010
解码张量映射
2020
张量解码器映射测试
下采样
上采样
2030
2050
无重采样
无变化
提取张量
提取张量
2035
2060
下采样提取张量
提取张量
上采样提取张量
2040
2070
输出张量
停止

图20

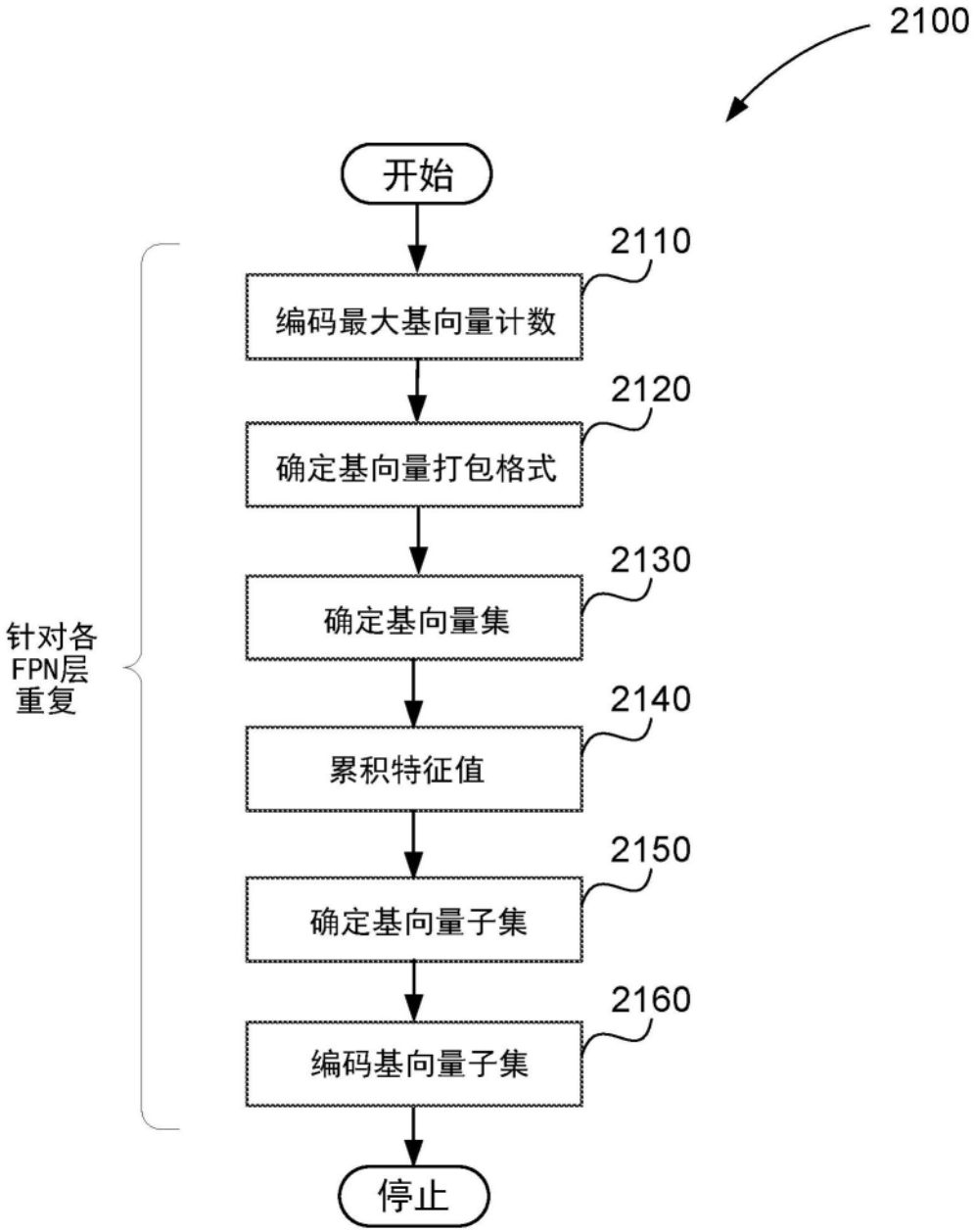

2100
开始
2110
编码最大基向量计数
2120
确定基向量打包格式
2130
确定基向量集
2140
累积特征值
2150
确定基向量子集
2160
编码基向量子集
停止
针对各FPN层重复

图21

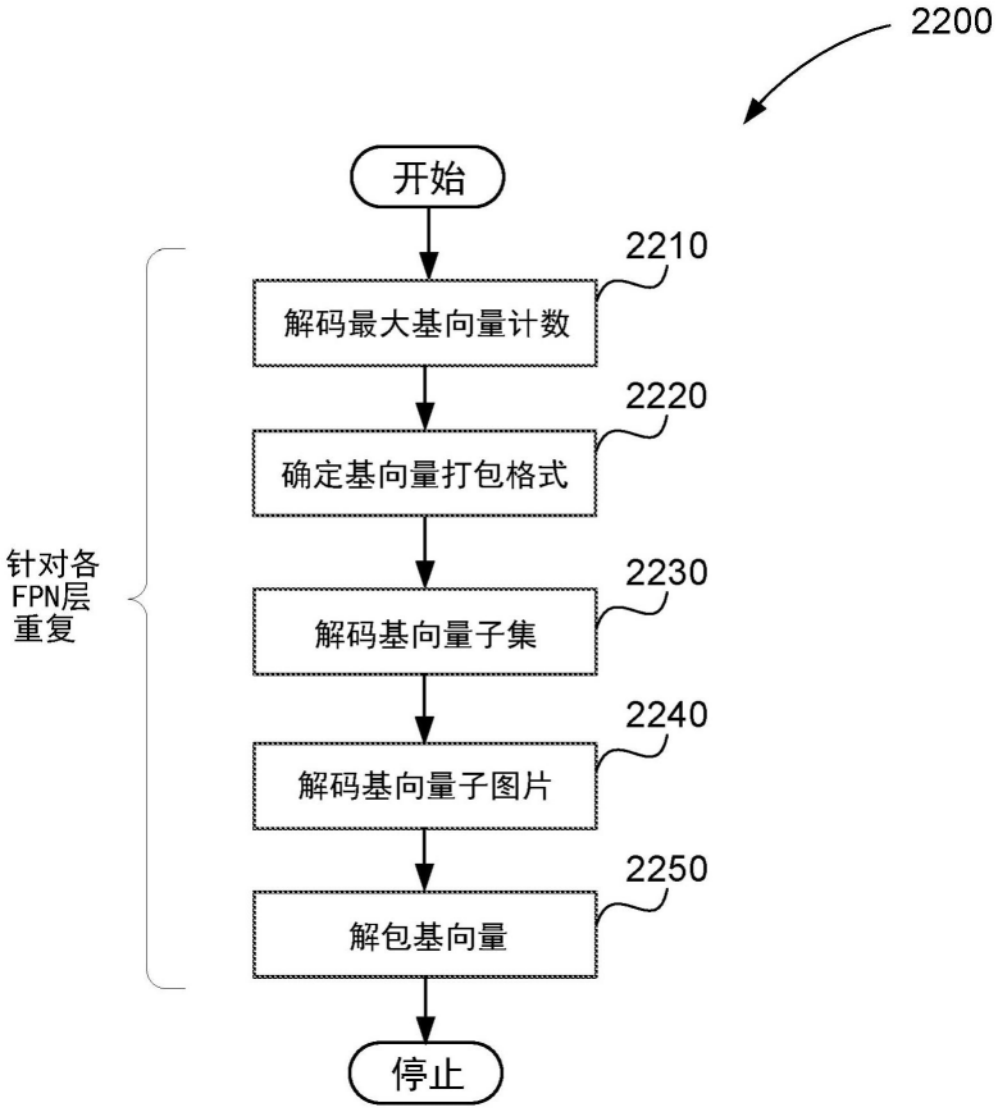

2200
开始
2210
解码最大基向量计数
2220
确定基向量打包格式
2230
解码基向量子集
2240
解码基向量子图片
2250
解包基向量
停止
针对各FPN层重复

图22

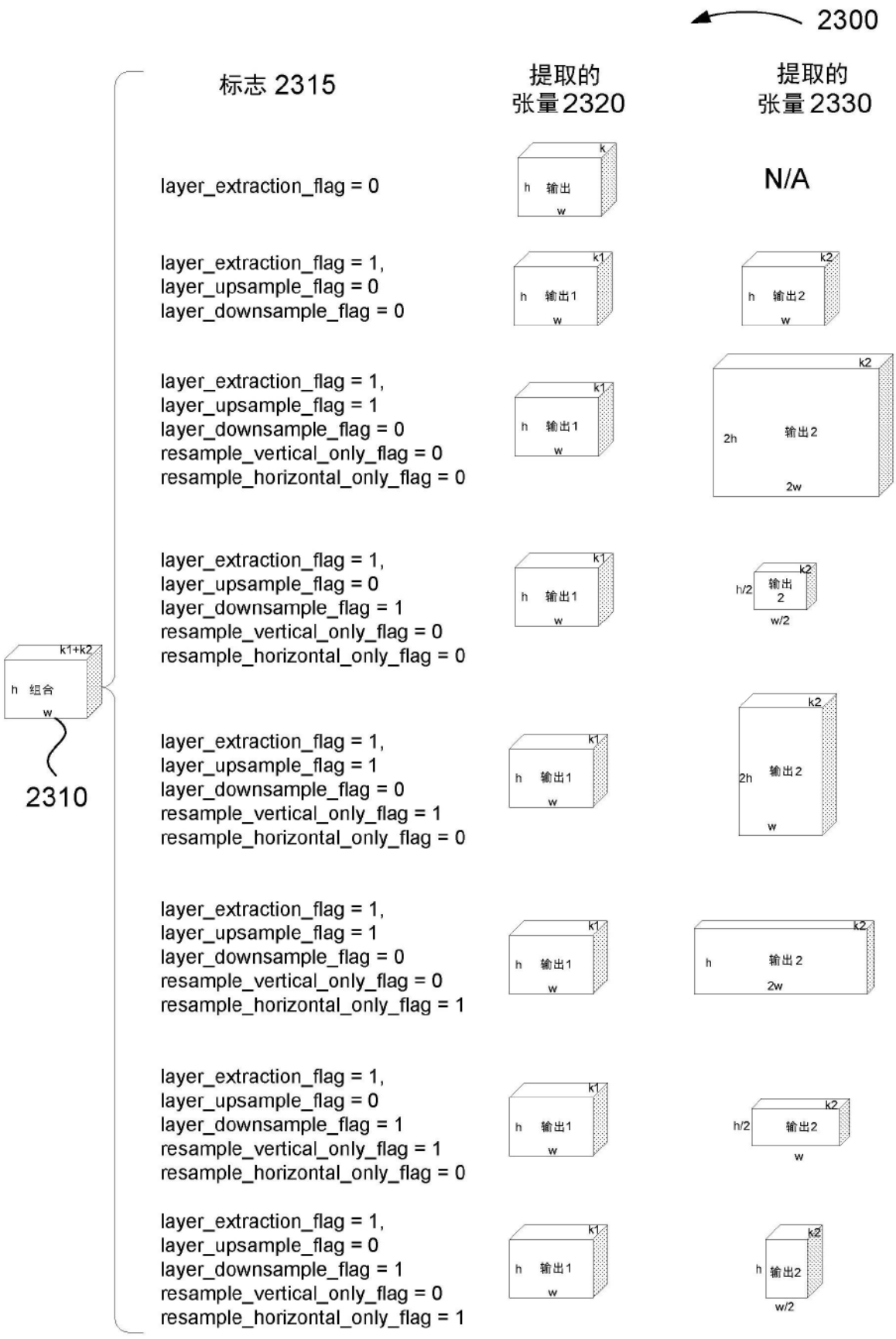

2300
标志 2315
提取的
张量2320
提取的
张量2330
layer_extraction_flag = 0
N/A
layer_extraction_flag = 1,
layer_upsample_flag = 0
layer_downsample_flag = 0
layer_extraction_flag = 1,
layer_upsample_flag = 1
layer_downsample_flag = 0
resample_vertical_only_flag = 0
resample_horizontal_only_flag = 0
layer_extraction_flag = 1,
layer_upsample_flag = 0
layer_downsample_flag = 1
resample_vertical_only_flag = 0
resample_horizontal_only_flag = 0
layer_extraction_flag = 1,
layer_upsample_flag = 1
layer_downsample_flag = 0
resample_vertical_only_flag = 1
resample_horizontal_only_flag = 0
layer_extraction_flag = 1,
layer_upsample_flag = 1
layer_downsample_flag = 0
resample_vertical_only_flag = 0
resample_horizontal_only_flag = 1
layer_extraction_flag = 1,
layer_upsample_flag = 0
layer_downsample_flag = 1
resample_vertical_only_flag = 1
resample_horizontal_only_flag = 0
layer_extraction_flag = 1,
layer_upsample_flag = 0
layer_downsample_flag = 1
resample_vertical_only_flag = 0
resample_horizontal_only_flag = 1
k1+k2
h 组合
w
2310
k
h 输出
w
k1
h 输出1
w
k2
h 输出2
w
2h
2w
h/2
w/2

图23