(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 120019662 A
(43)申请公布日 2025.05.16

(21)申请号 202380071969.5

(22)申请日 2023.07.28

(30)优先权数据
2022252784 2022.10.13 AU

(85)PCT国际申请进入国家阶段日
2025.04.09

(86)PCT国际申请的申请数据
PCT/AU2023/050704 2023.07.28

(87)PCT国际申请的公布数据
WO2024/077325 EN 2024.04.18

(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 19/70(2006.01)
G06N 3/0464(2006.01)
H04N 19/159(2006.01)
H04N 19/169(2006.01)
H04N 19/18(2006.01)
H04N 19/46(2006.01)
H04N 19/50(2006.01)
H04N 19/51(2006.01)
H04N 19/59(2006.01)
H04N 19/61(2006.01)
H04N 19/82(2006.01)
H04N 19/85(2006.01)
H04N 19/90(2006.01)
H04N 19/91(2006.01)
H04N 19/96(2006.01)

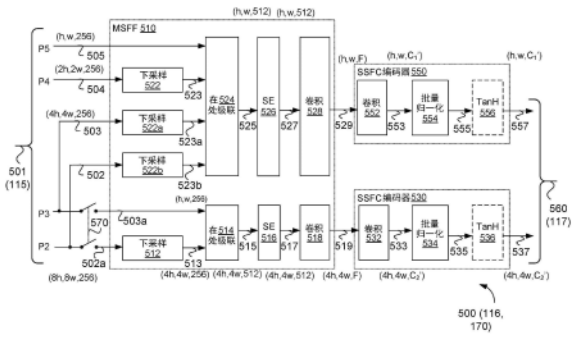
权利要求书3页 说明书38页 附图17页

(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

用于将至少多个张量编码到位流中的系统和方法,所述多个张量形成针对单个帧的分层表示。该方法包括:从形成分层表示的多个张量导出第一信息单元,所述多个张量至少包括第一张量和第二张量,第一张量的特征图与第二张量的特征图相比具有更大的空间分辨率;在第一模式中,将至少第一信息单元编码到位流中。在第二模式中,该方法还包括:从至少第一张量导出第二信息单元;以及将第二信息单元和第一信息单元编码到位流中。



CN 120019662 A

1.一种用于将至少多个张量编码到位流中的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:

从形成所述分层表示的多个张量导出第一信息单元,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;

在第一模式中,将至少所述第一信息单元编码到所述位流中;

在第二模式中,从至少所述第一张量导出第二信息单元;以及

在所述第二模式中,将所述第二信息单元和所述第一信息单元编码到所述位流中。

2.根据权利要求1所述的方法,还包括:基于用于编码的质量配置和要完成的机器任务其中至少之一来确定是在所述第一模式中还是在所述第二模式中进行操作。

3.根据权利要求2所述的方法,还包括:在所述要完成的机器任务是实例分段的情况下,确定在所述第二模式中进行操作。

4.根据权利要求1所述的方法,其中,导出所述第一信息单元包括:至少将所述第一张量和所述第二张量组合到第一组合张量中,并且向所述第一组合张量应用卷积层,接在所述卷积层之后是批量归一化层。

5.根据权利要求4所述的方法,其中,导出所述第一信息单元还包括:将所述批量归一化层的输出提供到TanH层。

6.根据权利要求1所述的方法,其中,

导出所述第一信息单元包括:至少将所述第一张量和所述第二张量组合到第一组合张量中,并且向所述第一组合张量应用第一卷积层和第一批量归一化层,以及

导出所述第二信息单元包括:至少将所述第一张量和另一张量组合到第二组合张量中,并且向所述第二组合张量应用第二卷积层和第二批量归一化层。

7.根据权利要求6所述的方法,其中,

导出所述第一信息单元还包括:将所述第一批量归一化层的输出提供到TanH层,以及

导出所述第二信息单元还包括:将所述第二批量归一化层的输出提供到TanH层。

8.一种用于从位流中解码至少多个张量的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:

从位流中解码至少包括第一信息单元的位流;

在第一模式中,从所述第一信息单元导出形成所述分层表示的多个张量,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;

在第二模式中,从所述位流中解码第二信息单元;以及

在所述第二模式中,从所述第一信息单元的至少一部分和所述第二信息单元导出形成所述分层表示的至少一部分的多个张量,所述第二信息单元与所述第一张量相对应。

9.根据权利要求8所述的方法,还包括:从所述位流中解码是否使用所述第二模式的指示。

10.根据权利要求8所述的方法,其中,在所述第二模式中,使用多路复用器来选择来自所述第二信息单元的多个张量。

11.根据权利要求8所述的方法,其中,使用卷积层来选择与至少所述第一张量相对应

的张量。

12.根据权利要求11所述的方法,其中,在所述第二模式中,所述卷积层接收从所述第一信息单元和所述第二信息单元各自导出的用于至少所述第一张量的张量。

13.根据权利要求11所述的方法,其中,在所述第一模式中,所述卷积层接收(i)来自所述第一信息单元的至少所述第一张量、以及(ii)表示从所述第二信息单元导出的张量的单位矩阵。

14.一种非暂态计算机可读存储介质,其存储程序,所述程序用于执行用于将至少多个张量编码到位流中的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:

从形成所述分层表示的多个张量导出第一信息单元,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;

在第一模式中,将至少所述第一信息单元编码到所述位流中;

在第二模式中,从所述第一张量导出第二信息单元;以及

在所述第二模式中,将所述第二信息单元和所述第一信息单元编码到所述位流中。

15.一种编码器,其被配置为通过以下操作来将至少多个张量编码到位流中,所述多个张量形成针对单个帧的分层表示:

从形成所述分层表示的多个张量导出第一信息单元,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;

在第一模式中,将至少所述第一信息单元编码到所述位流中;

在第二模式中,从至少所述第一张量导出第二信息单元;以及

在所述第二模式中,将所述第二信息单元和所述第一信息单元编码到所述位流中。

16.一种系统,包括:

存储器;以及

处理器,其中所述处理器被配置为执行所述存储器上所存储的代码,所述代码用于实现用于将至少多个张量编码到位流中的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:

从形成所述分层表示的多个张量导出第一信息单元,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;

在第一模式中,将至少所述第一信息单元编码到所述位流中;

在第二模式中,从所述第一张量导出第二信息单元;以及

在所述第二模式中,将所述第二信息单元和所述第一信息单元编码到所述位流中。

17.一种非暂态计算机可读存储介质,其存储程序,所述程序用于执行用于从位流中解码至少多个张量的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:

从位流中解码至少包括第一信息单元的位流;

在第一模式中,从所述第一信息单元导出形成所述分层表示的多个张量,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;

在第二模式中,从所述位流中解码第二信息单元;以及

在所述第二模式中,从所述第一信息单元的至少一部分和所述第二信息单元导出形成所述分层表示的至少一部分的多个张量,所述第二信息单元与至少所述第一张量相对应。

18.一种解码器,其被配置为通过以下操作来从位流中解码至少多个张量,所述多个张量形成针对单个帧的分层表示:

从位流中解码至少包括第一信息单元的位流;

在第一模式中,从所述第一信息单元导出形成所述分层表示的多个张量,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;

在第二模式中,从所述位流中解码第二信息单元;以及

在所述第二模式中,从所述第一信息单元的至少一部分和所述第二信息单元导出形成所述分层表示的至少一部分的多个张量,所述第二信息单元与至少所述第一张量相对应。

19.一种系统,包括:

存储器;以及

处理器,其中所述处理器被配置为执行所述存储器上所存储的代码,所述代码用于实现用于从位流中解码至少多个张量的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:

从位流中解码至少包括第一信息单元的位流;

在第一模式中,从所述第一信息单元导出形成所述分层表示的多个张量,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;

在第二模式中,从所述位流中解码第二信息单元;以及

在所述第二模式中,从所述第一信息单元的至少一部分和所述第二信息单元导出形成所述分层表示的至少一部分的多个张量,所述第二信息单元与至少所述第一张量相对应。

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35U.S.C.§119要求2022年10月13日提交的澳大利亚专利申请2022252784的申请日的权益,该申请通过引用而被全部并入于此,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,并且特别地,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] 卷积神经网络(CNN)是用于应对涉及机器视觉(诸如对象检测、实例分段(instance segmentation)、对象跟踪、人体姿势估计和动作识别等)的用例等的新兴技术。CNN的应用可以涉及作为“云”的一部分耦接到应用服务器的、具有传感器和一些处理能力的“边缘装置”的使用。CNN可能需要相对高的计算复杂度,这超过在利用边缘装置的计算容量或功耗方面典型地可以提供的计算复杂度。以分布式方式执行CNN已成为使用有限能力边缘装置运行前沿网络的一个解决方案。换句话说,分布式处理使得传统边缘装置通过在边缘装置和诸如云服务器等的外部处理部件之间分配处理来仍提供前沿CNN的能力。这种分布式网络架构可以称为“协作智能”,并且提供了诸如将来自网络的第一部分的部分结果针对若干不同的第二部分(可能各部分针对不同的任务进行了优化)进行重用等的益处。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等的网络进行传输。

[0005] CNN典型地包括诸如卷积层和全连接层等的许多层,其中数据以“张量”的形式从以一个层传递到下一层。跨不同装置拆分网络引入了对在CNN内从一个层传递到下一层的中间张量数据进行压缩的需求,这样的压缩可以被称为“特征压缩”,因为中间张量数据经常被称为“特征”或“特征图”并且表示诸如图像帧或视频帧等的输入的部分处理形式。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2-8(ISO/IEC JTC1/SC29/WG2-8)(也称为“运动图片专家组”(MPEG))被指派研究各种上下文中的并且经常与视频相关的压缩技术的任务。WG2“MPEG技术要求”已建立了“机器视频压缩”(VCM)特设组(ad-hoc group),其被委托研究机器消耗所用的压缩、以及特征压缩。特征压缩委托处于发布了“证据征集”(Call for Evidence,CfE)的探索阶段,其征求可以显著优于使用最先进的标准化技术所实现的特征压缩结果的技术。

[0006] CNN典型地需要在训练阶段中预先确定针对各个层的权重,其中在训练阶段中,非常大量的训练数据被传递通过CNN,并且将由经过训练的网络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。获得的结果和期望的结果之间的差异被表示为“损失”并且利用“损失函数”来测量。使用所确定的损失,进行用于更新网络权重的处理(诸

如随机梯度下降(SGD)等)。网络权重更新典型地涉及“梯度”的反向传播,该“梯度”指示要应用于网络权重的差量(delta),从网络的输出层开始并在向网络的输入层时终止,并且覆盖网络的所有中间或“隐藏”层。权重更新的速率由“学习速率”超参数进行缩放,该“学习速率”超参数典型地被设置为便于训练处理在损失方面找到全局最小值(即,针对网络架构和训练数据的最高可能任务性能),同时避免训练处理“困住(stuck)”于局部最小值。困住于局部最小值对应于获得针对网络架构的次优任务性能并且不能找到可导致更高任务性能的新权重值。通过供给被组织成“批(batch)”的输入数据和地面真值数据来重复地更新网络权重,以迭代地细化网络性能,直到不再能实现进一步提高准确度为止。整个训练数据集的迭代形成训练的“纪元(epoch)”,并且训练典型地需要多个纪元以实现针对任务的高水平性能。然后,经训练的网络可用于部署,从而以权重固定且省略针对权重更新的梯度的模式操作。利用输入执行经预训练的CNN并根据该CNN的拓扑来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、通道、高度和宽度。第一维度“批”在对视频数据进行推断时典型地是大小1,并且指示一个帧作为一个批被传递通过CNN。当训练网络时,批维度的值可以增加,使得根据预定的“批大小”在更新网络权重之前在各批中多个帧被传递通过网络。多帧视频可以作为具有大小根据给定视频的帧的数量而增加的批维度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,对视频数据的推断典型地逐个帧来进行。“通道”维度指示给定张量的并发“特征图”的数量,并且高度和宽度维度指示CNN的特定阶段的特征图的大小。通过CNN的通道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] 在进行相对大量的乘法累加(MAC)运算并且相对于存储器写入和读取大量的中间张量、以及读取针对CNN的各层的性能的权重的情况下,CNN的整体复杂度往往相对高。因此,将神经网络分割为部分使得即使在能力较差的边缘装置中也这样实现更复杂的网络。

[0009] 特征压缩可以受益于现有的视频压缩标准,诸如由联合视频专家组(JVET)开发的通用视频编码(VVC)等。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。其他视频压缩标准(诸如高效率视频编码(HEVC)和AV-1等)也可以用于特征压缩应用。

[0010] 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。在要在经打包帧中表示特征图数据的情况下,一般仅具有亮度而不具有颜色通道的单色帧是足够的。在仅存在亮度样本时,所得的单色帧被认为使用“4:0:0色度格式”。

[0011] VVC标准规定“基于块的”架构,其中帧首先被分割为称为“编码树单元”(CTU)的正方形区阵列。在VVC中,CTU一般占据128×128个亮度样本。在使用VVC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在面积上可能较小,其中发生隐式拆分以确保编码块保持在帧中。与各CTU相关联的是定义CTU的区域向块集合的分解的“编码树”(也称为“编码单元”(CU))。适用于仅亮度通道或仅色度通道的块被称为“编

码块”(CB)。编码块的内容的预测被保持在“预测块”(PB)或“预测单元”(PU)中,并且定义要与PB或PU相加组合的样本值的阵列的残差块被称为“变换块”(TB)或“变换单元”(TU),这是由于在TB或TU的生成中典型地使用变换处理。

[0012] 尽管在“单元”和“块”之间存在上述区别,但术语“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。

[0013] 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(一次水平地且一次垂直地)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。

[0014] 可以使用帧内预测或帧间预测处理来生成VVC中的PB或PU。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邻近样本的块的总体(“角度帧内预测”),或者(iv)使用邻近样本和所选择的矩阵系数的矩阵乘法的结果。

[0015] VVC可以用于对来自分离成两个部分的神经网络的第一部分(“主干”)的中间特征图进行压缩。在压缩中,来自主干的特征图被布置到帧中并且从浮点域量化到适合作为视频数据而压缩的样本域。为了减小特征图的空间面积,可以在VVC编码器和解码器与CNN中的发生拆分的中间点之间的接口处实现附加的神经网络层。针对可能不适合于变化的和不可预测的遇到的特征图数据的这种附加网络层进行训练。训练可能不会导致CNN在任务性能方面具有对各种质量的操作点的适应性。编码器和解码器的操作点也可以在操作期间变化,其中需要支持要在解码器侧供给到网络的其余部分的重建张量的变化质量水平。

发明内容

[0016] 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。

[0017] 本公开的一方面提供一种用于将至少多个张量编码到位流中的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:从形成所述分层表示的多个张量导出第一信息单元,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;在第一模式中,将至少所述第一信息单元编码到所述位流中;在第二模式中,从至少所述第一张量导出第二信息单元;以及在所述第二模式中,将所述第二信息单元和所述第一信息单元编码到所述位流中。

[0018] 本公开的另一方面提供一种用于从位流中解码至少多个张量的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:从位流中解码至少包括第一信息单元的位

流;在第一模式中,从所述第一信息单元导出形成所述分层表示的多个张量,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;在第二模式中,从所述位流中解码第二信息单元;以及在所述第二模式中,从所述第一信息单元的至少一部分和所述第二信息单元导出形成所述分层表示的至少一部分的多个张量,所述第二信息单元与所述第一张量相对应。

[0019] 本公开的另一方面提供一种非暂态计算机可读存储介质,其存储程序,所述程序用于执行用于将至少多个张量编码到位流中的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:从形成所述分层表示的多个张量导出第一信息单元,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;在第一模式中,将至少所述第一信息单元编码到所述位流中;在第二模式中,从所述第一张量导出第二信息单元;以及在所述第二模式中,将所述第二信息单元和所述第一信息单元编码到所述位流中。

[0020] 本公开的另一方面提供一种编码器,其被配置为通过以下操作来将至少多个张量编码到位流中,所述多个张量形成针对单个帧的分层表示:从形成所述分层表示的多个张量导出第一信息单元,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;在第一模式中,将至少所述第一信息单元编码到所述位流中;在第二模式中,从至少所述第一张量导出第二信息单元;以及在所述第二模式中,将所述第二信息单元和所述第一信息单元编码到所述位流中。

[0021] 本公开的另一方面提供一种系统,包括:存储器;以及处理器,其中所述处理器被配置为执行所述存储器上所存储的代码,所述代码用于实现用于将至少多个张量编码到位流中的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:从形成所述分层表示的多个张量导出第一信息单元,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;在第一模式中,将至少所述第一信息单元编码到所述位流中;在第二模式中,从所述第一张量导出第二信息单元;以及在所述第二模式中,将所述第二信息单元和所述第一信息单元编码到所述位流中。

[0022] 本公开的另一方面提供一种非暂态计算机可读存储介质,其存储程序,所述程序用于执行用于从位流中解码至少多个张量的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:从位流中解码至少包括第一信息单元的位流;在第一模式中,从所述第一信息单元导出形成所述分层表示的多个张量,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;在第二模式中,从所述位流中解码第二信息单元;以及在所述第二模式中,从所述第一信息单元的至少一部分和所述第二信息单元导出形成所述分层表示的至少一部分的多个张量,所述第二信息单元与至少所述第一张量相对应。

[0023] 本公开的另一方面提供一种解码器,其被配置为通过以下操作来从位流中解码至少多个张量,所述多个张量形成针对单个帧的分层表示:从位流中解码至少包括第一信息单元的位流;在第一模式中,从所述第一信息单元导出形成所述分层表示的多个张量,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;在第二模式中,从所述位流中解码第二信息单元;以及在所述第二模式中,从所述第一信息单元的至少一部分和所述第二信息单元导出形成所述分层表

示的至少一部分的多个张量,所述第二信息单元与至少所述第一张量相对应。

[0024] 本公开的另一方面提供一种系统,包括:存储器;以及处理器,其中所述处理器被配置为执行所述存储器上所存储的代码,所述代码用于实现用于从位流中解码至少多个张量的方法,所述多个张量形成针对单个帧的分层表示,所述方法包括:从位流中解码至少包括第一信息单元的位流;在第一模式中,从所述第一信息单元导出形成所述分层表示的多个张量,所述多个张量至少包括第一张量和第二张量,所述第一张量的特征图与所述第二张量的特征图相比具有更大的空间分辨率;在第二模式中,从所述位流中解码第二信息单元;以及在所述第二模式中,从所述第一信息单元的至少一部分和所述第二信息单元导出形成所述分层表示的至少一部分的多个张量,所述第二信息单元与至少所述第一张量相对应。

[0025] 还公开了其他方面。

附图说明

[0026] 现在将参考以下的附图来说明本发明的至少一个实施例,其中:

[0027] 图1是示出分布式机器任务系统的示意框图;

[0028] 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;

[0029] 图3A是示出CNN的主干部的功能模块的示意框图;

[0030] 图3B是示出图3A的残差块的示意框图;

[0031] 图3C是示出图3A的残差单元的示意框图;

[0032] 图3D是示出图3A的CBL模块的示意框图;

[0033] 图4是示出CNN的可替代主干部的功能模块的示意框图;

[0034] 图5是示出用于在压缩之前降低张量维度的跨层张量瓶颈编码器的示意框图;

[0035] 图6示出用于进行CNN的第一部分、使用瓶颈编码器进行收缩、并且对所得到的收缩特征图进行编码的方法;

[0036] 图7A是示出用于多个压缩张量的打包布置的示意框图;

[0037] 图7B是示出用于多个压缩张量的位流的示意框图;

[0038] 图8是示出视频编码器的功能模块的示意框图;

[0039] 图9是示出视频解码器的功能模块的示意框图;

[0040] 图10是示出用于在压缩之后恢复张量维度的跨层张量瓶颈解码器的示意框图;

[0041] 图11示出用于对位流进行解码、重建去相关特征图、并且执行CNN的第二部分的方法;

[0042] 图12A是示出CNN的头部的示意框图;

[0043] 图12B是示出图12A的升级器模块的示意框图;

[0044] 图12C是示出图12A的检测模块的示意框图;

[0045] 图13是示出CNN的可替代头部的示意框图;

[0046] 图14是用于将张量编码成降维形式的方法;以及

[0047] 图15是示出用于将张量从降维形式解码成原始维度的张量的方法。

具体实施方式

[0048] 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。

[0049] 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某任务结果的基于服务器场(“云”)的应用。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稍后处理(根据需要可能用于多个不同的任务)。机器任务的示例包括对象检测和实例分段，两者都产生超过交并比(IoU)的阈值(诸如0.5等)的检测所用的被测量为“平均精度均值”(mAP)的任务结果。另一示例机器任务是对象跟踪，其中平均对象跟踪准确度(MOTA)得分作为典型的任务结果。

[0050] 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。视频压缩标准典型地对以平面阵列布置的一些给定位深度(诸如每样本8位或10位等)的整数样本进行操作。

[0051] 张量典型地具有以下维度：批大小、通道数、高度和宽度。例如，维度[1,256,76,136]的张量将被认为包含一个张量的批，该张量包含二百五十六(256)个特征图，各特征图的大小为136×76。对于视频数据，典型地一次对一个帧进行推断，而不是使用包含多个帧的张量，这样得到批大小为1。

[0052] 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，该分布式机器任务系统100实现被分割成两个部分(例如，一个部分可以在边缘装置中且另一部分可以在云服务器中)的神经网络。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经编码数据中解码特征图的方法。可以实现这些方法，使得压缩数据被编码以降低位率，同时适应在输入数据中遇到的变化的统计信息。这样，系统100提供对输入张量数据进行“实时”训练(或“细化训练”)以在编码器和解码器中生成针对活动使用的网络的权重更新的能力。尽管对任务网络进行训练需要针对输入数据的地面真值，但细化训练仅应用于网络的形成瓶颈编码器和解码器的一部分。细化训练的目标是以最小的退化保存通过的数据。因此，在向瓶颈编码器的输入处的张量形成针对瓶颈解码器的输出的地面真值。细化训练减轻了对充分训练预定网络权重以预期所有可想到的输入数据的需求。

[0053] 参考图6(其示出用于进行CNN的第一部分的方法600)和图11(其示出用于进行CNN的第二部分的方法1100)来说明图1。还参考图7A(其示出特征图从压缩张量到单色视频帧中的打包布置)和图7B(其示出在对张量进行编码和解码时使用的位流格式)。

[0054] 系统100实现“FasterRCNN”网络，该“FasterRCNN”网络用于对象检测并且在所描述的示例中在典型地被描述为“P层”的中间点处被拆分成主干部和头部。可以在系统100中实现诸如“MaskRCNN”等的其他网络。值得注意的是，FasterRCNN和MaskRCNN的主干具有卷积、批量归一化和激活函数等的相同拓扑和维度。FasterRCNN的头是MaskRCNN的头的子集，其中MaskRCNN包括“掩码头”，用于除了生成FasterRCNN和MaskRCNN两者中存在的边界框输出之外，还生成实例分段图。掩码头包括两个卷积层，并针对从(将参考图13来说明的)RoIAlign阶段产生的各“感兴趣区”产生分段图。因此，MaskRCNN可以用于进行对象检测和

实例分段这两者,其中由于使用掩码头而在网络头中具有附加复杂度。

[0055] 系统100包括用于以经编码视频位流123的形式从视频源112生成经编码张量数据的源装置110。系统100还包括用于对经编码视频位流123形式的张量数据进行解码以产生任务结果153的目的地装置140。通信通道130用于将经编码视频位流123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中的一者或两者可以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(Ethernet)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)上捕获经编码视频数据的应用。

[0056] 源装置140根据存储在存储器206中且在处理器205(参见图2A)的执行下进行的方法600来操作。方法600可以使用诸如配置的FPGA、ASIC或ASSP等的设备来实现。可替代地,如以下所述,方法600可以由源装置110在处理器205的执行下实现为应用程序233(参见图2A)的一个或多于一个软件代码模块。实现方法600的应用程序233的软件代码模块例如可以驻留在(将关于图2A来说明的)硬盘驱动器210和/或存储器206中。方法600对针对视频数据的一帧的张量进行编码,并且包括更新对张量进行编码所使用的权重的功能性。经更新权重用于基于当前使用的权重与具有在源装置110接收图像帧时被更新的权重的内部模型相比的性能来对视频数据的后续帧进行编码。

[0057] 视频源112提供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。

[0058] 源装置140通过进行CNN的第一部分(被称为主干网络114)来开始机器任务,以产生中间张量。中间张量被示出为115。为了便于位流123的位率降低,源装置140使用被称为“瓶颈编码器”(被示出为编码器116)的压缩网络来降低张量115的维度。编码器116操作以降低张量115的维度。在目的地装置140中,“瓶颈解码器”(被示出为150)恢复张量维度,使得输出张量151与张量115的张量维度相对应。由于瓶颈编码器116和瓶颈解码器150包括可训练的网络层,因此存在指定权重的需要。

[0059] 对于任务网络(包括主干网络114和头网络152),离线训练(或“预训练”)是指定权重的一个可能机制。使用离线训练的权重的一个缺点是训练处理需要预期各种输入数据的非常宽的范围。典型的视频压缩标准通过提供一定程度的数据适应性(诸如通过使用“上下文自适应二进制算术编码”(CABAC)来对变化的输入数据统计信息进行建模等)来适应广泛变化的输入。神经网络典型地使用预定的(固定的)权重,因此不适应输入数据。张量115可以形成由主干114中的特征金字塔网络(FPN)产生的多尺度表示,该多尺度表示例如在编码器116处使用称为“多尺度特征压缩”(MSFC)的途径被一起“融合”到单个张量中。MSFC通常用于将所有FPN层合并到单个张量中,并将单个张量的空间维度降低到FPN层张量中的最小空间分辨率张量的空间维度。将所有FPN层合并到单个张量中是以FPN的分解较少的(较大的)层的空间细节为代价实现的。

[0060] 参考图6的方法600来说明源装置110的操作。方法600以进行神经网络第一部分步

骤610来开始。

[0061] 在步骤610,CNN主干114从视频源112接收视频帧数据113的一帧,并且进行整体CNN的特定早期层,诸如与CNN的“主干”相对应的层等。步骤610输出张量115。CNN 114的主干层可以针对各帧产生多个张量作为输出,这多个张量例如与将包括FPN的主干应用于输入图像的不同空间尺度相对应。由具有FPN的主干114产生的张量115形成包括特征图的数据的帧数据113的分层表示(hierarchical representation)。分层表示的各连续层具有在前层的宽度的一半和高度的一半。在系统100执行“YOLOv3”网络时,FPN可以得到张量115中的与从主干114输出的三个层相对应的三个张量,其中张量115具有不同的空间分辨率和通道数。在系统100正在进行诸如“Faster RCNN X101-FPN”或“Mask RCNN X101-FPN”等的网络时,张量115包括四层的张量P2至P5。尽管在主干模块114中进行的第一部分的层可以被称为整体任务网络的“主干”,但源装置110和目的地装置140之间的层的特定分割不需要与来自在任务网络中典型地被定义为“主干”的层的边界结果相对应。术语“主干”和“头”在本文中用于指代网络向第一部分和第二部分的任何分割,这样的分割可基于除机器任务网络架构以外的考虑(诸如源装置110或目的地装置140中的可用计算资源等)来选择。

[0062] 方法600在处理器205的控制下从步骤610继续到进行瓶颈编码步骤615。在步骤615,源装置110通过进行与使用在目的地装置140中也关联的权重集合的压缩操作相对应的多个网络层,来使用瓶颈编码器116减少张量115的数量和维度。步骤615接收至少一个张量(在该至少一个张量中,已部分进行了用于机器任务的神经网络(也就是说,已实现了主干网络))作为输入,并且产生张量117。对于帧数据113的给定帧,张量115借助于使用主干114中的FPN来包括在FPN层之间具有不同空间分辨率的多个张量。该“多尺度”表示在瓶颈编码器116中被转换成更少的张量,例如,被转换成针对帧数据113的各帧的减少张量117内的一个(“基础层”)或两个(“基础层”和“增强层”)张量(其也具有降低的空间维度)。在编码器116处实现的跨层融合和空间减少是用于融合神经网络的特征的深度特征压缩的一个方法的示例。在所描述的示例中,使用被称为“多尺度特征压缩”(MSFC)的技术来实现跨层融合。以下参考图5来说明MFSC的操作。如参考图5所述的MSFC的变形是可能的,这些变形包括针对各FPN层应用单尺度特征融合(SSFC),而没有跨层融合。针对各FPN层应用单独的SSFC阶段使得能够针对各层进行通道数减少,但不提供用于具有FPN层的分辨率中的较大分辨率的层的空间减少的任何手段,也不利用任何跨层冗余。步骤615操作以针对与图像帧相关的数据进行数据压缩(在所描述的由主干模块114处理的示例中)。数据压缩由编码器的神经网络使用当前关联或应用的权重集合来进行。参考图14的方法1400来说明在步骤615处实现的跨层融合和空间减少的方法。处理器205中的控制从步骤615进入编码压缩张量步骤620。

[0063] 在步骤620,量化和打包模块184将来自浮点域的张量117量化为整数样本(诸如10位样本等)。模块184还使用参考图7A所述的打包格式将张量117的各通道的特征图打包到帧中以产生经打包帧185。视频编码器120将经打包帧185编码到位流123中作为压缩张量121。处理器205中的控制从步骤620进入进行瓶颈解码步骤625。

[0064] 在步骤625,张量117被供给到瓶颈解码器118。瓶颈解码器118利用目的地装置140已知的权重集合进行操作,以从张量117产生恢复张量119。恢复张量119具有与张量115相同的张量维度。张量119表示张量115的退化版本,其中由于张量117的收缩维度以及模块

116和118的权重对于输入张量统计信息的最优度而导致损失。利用与瓶颈解码器150中所使用的相同的权重来初始化瓶颈解码器118。参考图15所述的方法1500示出用于从张量117产生恢复张量119的一个途径。处理器205中的控制从步骤625进入测量瓶颈性能步骤630。

[0065] 在步骤630,确定或获取到用于使用步骤640处所确定的权重来评估数据压缩的值。例如,均方误差(MSE)模块178将张量115与张量119进行比较,从而产生MSE值179,该MSE值179指示由于向张量117的降低维度的转换和从张量117的降低维度的转换而导致的损失。随着时间的经过,张量115的统计信息可以改变,导致被视为MSE 179的降低的、瓶颈编码器116和瓶颈解码器118的性能降低,尽管也可以预期MSE的短期变化。MSE 179提供目的地装置140中的张量151的预期信号质量的指示,除了视频编码器120的有损视频编码处理不包括在该指示中以外。因此,MSE 179给出瓶颈编码器和解码器(即,116、118、150)在保存从主干114传递到头152的张量数据时的性能的指示。值179是使用输入到编码器116的张量115以及在步骤615和630处使用与神经网络相关联的权重的数据压缩之后的张量119来获得或获取到的。在其他实现中,除MSE以外的机制可以用于评估数据压缩。处理器205中的控制从步骤630进入训练状态步骤635。

[0066] 在步骤635,处理器205基于MSE值179来更新训练状态变量。训练状态变量存储在存储器206中,并且指示源装置110当前是否正在对模块集合进行训练。如果MSE值179降至低于阈值,则训练状态变量被设置为“TRAINING(训练)”状态。可以以多个方式(例如,使用先前生成的MSE值179的移动平均、基于与MSE值的移动平均相比或与预定或可配置阈值相比等的期望MSE值)来确定阈值。如果在一段时间的训练之后没有进行权重更新,则训练状态变量可以被重置以停止进一步的训练。如果训练状态变量指示训练正在进行中,则处理器205中的控制从步骤635进入进行可训练瓶颈编码/解码步骤640(图6中的“训练(TRAIN)”)。独立于诸如MSE值179等的度量,步骤635可以操作以周期性地(例如,每天一次或每周一次)进入TRAINING状态,从而提供持续的手段以确定是否即使在不存在退化性能的明确指示(诸如测量的MSE值179的下降等)的情况下也可以生成改进的模型。如果训练状态变量没有指示训练正在进行中,则处理器205中的控制从步骤635进入编码权重更新标志步骤660(图6中的“未训练(NOT TRAIN)”)。

[0067] 在步骤640,可训练瓶颈编码器170和可训练瓶颈解码器174操作以对张量115进行压缩和解压缩以产生张量175。可训练瓶颈编码器170具有与编码器116相同的结构。可训练瓶颈解码器174具有与解码器150相同的结构。模块170和模块116之间以及模块174和模块150之间的结构兼容性使得在模块170和174中生成的权重能够分别被传送到模块116和150以用于后续处理。压缩张量171从可训练瓶颈编码器170输出并被传递到可训练瓶颈解码器174。使用当前存在于模块170和174中的权重来进行前向通过模块170和174。模块170和174的网络拓扑和层维度分别与模块116和118的网络拓扑和层维度相对应。模块170和174的操作分别符合参考图14和图15所述的方法1400和1500。处理器205中的控制从步骤640进入测量可训练瓶颈性能步骤645。

[0068] 在步骤645,确定或获取用于使用步骤640处所确定的权重来评估数据压缩的值。步骤645使用与步骤630相同的机制来评估数据压缩性能。例如,均方误差(MSE)模块180通过对张量115和175进行均方误差计算来产生测量损失181。测量损失181值提供了模块170和174在借助于使用降维张量171在维度上收缩之后准确地恢复张量的能力的度量。处理器

205中的控制从步骤645进入反向传播步骤650。

[0069] 在步骤650,使用“反向传播”的处理来在模块170和174中进行权重更新,由此基于试图最小化测量损失181的处理(诸如随机梯度下降(SGD)等)来更新权重。可训练瓶颈编码器170和可训练瓶颈解码器174借助于由于反向传播而引起的正在进行的权重更新,能够适应张量115的统计信息的这种变化,这导致与MSE 179相比实现更低MSE 181的潜力。权重更新的速率由“学习速率”缩放。较高的学习速率一般导致用较少的反向传播操作(即,较快的训练处理)使测量损失最小化,但遭受由于过度调整权重而引起的训练中的不稳定性的风险,这阻止了损失函数的局部最小值的发现。较小的学习速率可能需要较长的时间来训练网络,但较小的学习速率值也不太可能过度调整速率。在使用诸如一个图像等的小批大小时,期望较小的学习速率以减少统计上可能是异常值的个体帧的影响。典型的学习速率可以是诸如0.01或0.001等的值,并且可以通过批大小的倒数或者通过批大小的平方根的倒数来缩放,以达到给定批的最终学习速率。学习速率可以随时间变化,其中在网络权重远离权重的最终值时,最初使用较大的值,并且在网络在MSE方面接近最佳或可接受状态时,稍后使用较小的值。在瓶颈编码器和解码器的细化训练的上下文中,较小的学习速率(诸如在0.001至0.0001之间等)是合适的。尽管模块170和174一次接收一个张量,但输入张量可以被分组为大小大于1的批。增加批大小可以改善训练,因为各权重更新步骤受到各种输入的影响。增加批大小会增加针对模块170和174的存储器需求。此外,对于视频数据,来自主干114的连续帧或连续张量中的统计变化不太明显,这降低了使用较大批大小的益处。还可以使用更新权重的其他处理,诸如利用动量和缩放并将权重衰减与梯度更新解耦的“AdamW”优化器。降低张量171形成瓶颈编码器和解码器(即,170和174)在遇到实际输入数据(即,转换为张量115的帧数据113)时对该实际输入数据进行训练或适配的结果。处理器205中的控制从步骤650进入权重更新确定步骤655。

[0070] 在步骤655,触发模块182将MSE 179与MSE 181进行比较。如果观察到MSE 179在一段时间中低于MSE 181并且差超过阈值,则发起权重更新处理。该一段时间可以与多个帧、移动平均或其组合相对应,其指示与如由模块170和174指示的可实现性能相比模块116和118的低于标准的性能。用于发起权重更新的试探法适于捕获当前使用中的用于张量降低和恢复(即,张量115向张量117并最终向张量119或151的转换)的权重不再很好地适合于接收到的输入数据的统计信息的点。在模块170和174中使用的权重对于编码在位流123中是昂贵的,因此在每次更新操作时不被送出到目的地装置140。例如基于在使用当前活跃权重的同时检测到性能退化而从源装置110到目的地装置140的从步骤655的确定产生的经更新权重的较不频繁传输对于充足的系统性能是足够的。步骤655的结果是是否以权重更新标志750的形式进行权重更新的决定。步骤625至655进行操作以确定是否要改变在下一轮数据压缩中应用的数据压缩所用的权重集合。基于MSE值181和179的比较来确定是否要实现改变。如果要进行权重更新或改变,则训练状态变量也被重置以停止对方法600的后续调用的进一步训练。处理器205中的控制从步骤655进入编码权重更新标志步骤660。

[0071] 在步骤660,要参考图8说明的熵编码器838将权重更新标志编码到位流123中作为权重更新标志750。更新权重的指示可以包括在与当前经打包帧185相关联的补充增强信息(SEI)消息中。SEI消息744包含权重并且可以与当前经打包帧185相关联,该当前经打包帧185包括权重更新标志750、或者具有包含权重更新标志750的单独SEI消息。可替代地,包含

权重的SEI消息的存在可以是从源装置140到目的地装置140的要进行权重更新的指示。处理器205中的控制从步骤660进入权重更新标志测试步骤665。

[0072] 在步骤665,如果权重更新标志750指示触发模块182中的确定是要改变或更新权重(如图6所示,“更新(UPDATE)”),则处理器205中的控制进入加载经更新权重步骤667。否则,如果权重更新标志不指示触发模块182中的进行权重的更新的确定,则处理器205中的控制从步骤665进入在进行神经网络第一部分步骤675处处理视频数据的下一帧(如图6所示,“无更新(NO UPDATE)”)。

[0073] 在步骤667,瓶颈解码器权重176从瓶颈解码器174被传递到瓶颈解码器118并被传递到权重编码器186。来自可训练瓶颈编码器170的瓶颈编码器权重172被加载到或应用到瓶颈编码器116。权重的变化可以指向特定存储的值集合或替换先前存储的值。在步骤667,编码器116的关联从步骤615处所使用的权重集合改变到权重172。将权重172与编码器116相关联可以涉及将权重172加载到由编码器116引用的存储器206的区中、或者更改指针以选择权重172以用于编码器116中的后续使用。处理器205中的控制从步骤667进入编码经更新权重步骤670。

[0074] 在步骤670,权重编码器186对瓶颈解码器权重176进行编码以产生经编码权重187。使用多路复用器122将经编码权重187存储在位流123中(作为图7B的权重752)。对权重187进行编码可以涉及使用可变长度码字来直接对权重187进行编码以对各权重值进行压缩。可以采用诸如上下文自适应二进制算术编码(CABAC)等的算术编码方案来对权重187进行编码。可替代地,经编码信息可以表示权重之间的差量,例如相对于由瓶颈解码器118先前使用并且也由瓶颈解码器146借助于先前权重更新和同步初始状态已知的权重的差量。可用于表示权重的一个示例句法是标准ISO/IEC 15938-17(有时称为“神经网络编码”或“神经网络表示”或“MPEG-NNR”),尽管可以使用用于将神经网络权重高效地编码到位流中的其他手段,诸如“开放神经网络交换中间表示”等。在步骤670完成时,针对来自视频源112的一帧完成对部分任务结果(即,张量115)的编码。发生前进到后续帧(诸如下一帧等)。方法600中的其余步骤与来自视频源112的后续帧相关,这些其余步骤示出如步骤655处确定的经更新权重的使用或不使用。处理器205中的控制从步骤670进入进行第二神经网络第一部分步骤675。

[0075] 在步骤675,与步骤610类似,对来自视频源112的帧数据的后续帧(例如帧113a)进行神经网络第一部分,从而产生用于帧113a的经更新张量,诸如张量115a等。处理器205中的控制从步骤675进入进行第二瓶颈编码步骤680。

[0076] 在步骤680,与步骤615类似,瓶颈编码器116进行经更新张量115a的编码以产生经更新张量117a。如在步骤655处确定的,如果在步骤667处进行权重更新,则步骤680使用由瓶颈编码器116从可训练瓶颈编码器172接收到的瓶颈编码器权重172来进行编码。换句话说,使用关联的经更新权重集合176来进行数据压缩。处理器205中的控制从步骤680进入编码第二压缩张量步骤685。

[0077] 在步骤685,与步骤620类似,利用视频编码器120将根据帧700的格式表示为经打包帧的经更新张量117a编码到视频位流121中作为压缩帧N+1(参见图7B的746)。利用多路复用器122将视频位流121多路复用到位流123中。源装置140在处理器205的执行下,继续对来自主干114的视频数据113的连续帧的特征图进行编码,其中如由触发模块182所确定的,

时不时地更新诸如模块116和118的可训练层等的可训练层。作为方法600的操作的结果,位流123包含来自神经网络的第一部分的经编码张量,并且在一些实例中包含用于降低张量维度的权重的动态更新。更新取决于瓶颈编码器116相对于利用瓶颈编码器170和解码器174的可实现性能的性能。在模块170和174中对接收到的数据进行训练利用如下的性质:对于压缩任务,目标是以最小损失恢复输入数据。换句话说,对于压缩任务,地面真值是输入数据。由模块170和174进行的训练与存在于模块116、118和150中的“部署”网络权重的使用同时发生,并且使用相同的输入数据。因此,由模块170和174进行的训练可以被认为是对当前输入数据的“过拟合”。然而,由于网络能够进行动态权重更新,因此该过拟合行为可以被认为是数据驱动自适应。对细化训练的支持减轻了增加瓶颈编码器和解码器的训练的复杂度和/或瓶颈编码器和解码器自身的复杂度的需求,以适应更广泛的统计多样性的输入数据。考虑到系统100中提供的从输入帧113到任务结果153的数据路径(即,模块114、116、150和152),该路径的可训练部分与模块116和150相对应,从而形成数据路径中的总CNN层的子集。

[0078] 参考图11的方法1100来说明系统100的目的地装置140部分的操作。方法1100涉及在源装置处编码的数据,其中已部分进行了用于机器任务的神经网络处理(即,利用主干网络114)。方法1100可以由存储在存储器206上并由处理器205的执行控制的应用程序233的软件代码模块来实现。方法1100从解码经打包帧步骤1110开始。

[0079] 在步骤1110,解多路复用器142接收到位流123。解多路复用器142从位流123提取与位流121相对应的视频位流143。视频位流143被供给到视频解码器146以产生经解码经打包帧162。经解码经打包帧162被传递到解包和去量化模块160。模块160将来自经打包帧162的各特征图从整数样本域提取和逆量化到浮点域,从而将特征图布置成张量,诸如经解码张量147等。处理器205中的控制从步骤1110进入进行瓶颈解码步骤1120。

[0080] 在步骤1120,包含针对诸如卷积层等的层的权重的瓶颈解码器150将张量147转换为与张量147相比具有增加的维度的张量151。瓶颈解码器150操作,以解码与已经在源装置110处进行压缩的图像(包括视频的图像帧的情况)相关的数据。由瓶颈解码器150的神经网络使用当前关联的权重集合来进行解码。瓶颈解码器150使用与由瓶颈编码器116实现的编码或压缩相关联的解码方法(例如MFSC)。瓶颈解码器150的操作符合参考图15所述的方法1500。处理器205中的控制从步骤1120进入进行神经网络第二部分步骤1130。

[0081] 在步骤1130,头模块152进行能够在系统100中实现的整体神经网络的第二部分以产生任务结果153。任务结果153存储在一般在存储器206中实现的任务结果缓冲器154中。方法1100相应地在步骤1130处实现神经网络机器任务的其余部分。处理器205中的控制从步骤1130进入解码权重更新标志步骤1140。

[0082] 在步骤1140,要参考图9所述的熵解码器920从位流123解码权重更新标志750。如由触发模块182所确定的,权重更新标志750提供用于指示是否要改变用于瓶颈解码的权重集合(也就是说,是否要更新瓶颈解码器150中的权重)的信息。处理器205中的控制从步骤1140进入权重更新指示测试步骤1150。

[0083] 在步骤1150,应用233基于经解码权重更新标志来确定是否要更新权重。如果由权重更新标志750指示权重的更新或改变(步骤1150处的“更新”),则处理器205中的控制进入解码经更新权重步骤1160。如关于步骤670所述,信息可以与权重值或权重值之间的差量相

关。否则,处理器205中的控制进入解码经打包第二帧步骤1180(步骤1150处的“无更新”)。

[0084] 在步骤1160,使用解多路复用器142从位流123提取用于指示经编码权重145的信息。权重解码器148将经编码权重145转换或解码为经解码权重149。处理器205中的控制从步骤1160进入应用经更新权重步骤1170。

[0085] 在步骤1170,更新瓶颈解码器150以使用或应用经解码权重149,从而维持权重与源装置140中的瓶颈解码器118中存在的权重的同步。换句话说,更新瓶颈解码器150的神经网络关联以使用来自步骤1120中所使用的权重的权重149。更新可以指向特定存储的值集合或替换先前存储的值。在步骤1170完成时,在主干114和头152中实现的CNN任务已针对一个帧进行,并且发生前进到后续帧(诸如下一帧等)。处理器205中的控制从步骤1170进入步骤1180。

[0086] 在步骤1180处,与步骤1110类似,视频解码器146对来自位流123的第二经编码帧746进行解码以产生第二经打包帧(例如,帧147a)。处理器205中的控制从步骤1180进入进行第二瓶颈解码步骤1190。

[0087] 在步骤1190,瓶颈解码器150使用权重149(在由权重更新标志750指示的情况下)对帧147a进行解码以产生第二经解码张量151a。目的地装置140继续使用张量151a运行CNN头152以产生另一任务结果(诸如结果153a等),并且继续利用时不时更新的瓶颈解码器中所使用的权重(如由经解码权重更新标志750指示)来对后续帧进行解码。

[0088] 任务结果缓冲器154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各自的功能性可以在一些实现中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。

[0089] 尽管以上说明了示例装置,但源装置110和目的地装置140各自典型地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(其可被配置为呈现任务结果151的显示装置)、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。

[0090] 计算机模块201典型地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许可计算机系统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典型地将包括所谓的"防火墙"装置或具有类似功能性的装置。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。

[0091] 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,其中前者典型地根据通用串行总线(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典型地包括硬盘驱动器(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)。典型地设置有光盘驱动器212以用作数据的非易失性源。可以使用例如光盘(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。典型地,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经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。

[0092] 计算机模块201的组件205至213典型地经由互连总线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。

[0093] 在适当或期望的情况下,可以使用计算机系统200来实现源装置110和目的地装置140以及以下所述的方法。特别地,可以将源装置110、目的地装置140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。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)来实现源装置110、目的地装置140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。

[0094] 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源装置110和目的地装置140以及所述方法的有利设备。

[0095] 典型地将软件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。

[0096] 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/或视频数据或经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、以及包括电子邮件传输和网站上所记录的信息等的因特网或内联网。

[0097] 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。

[0098] 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。

[0099] 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。典型地将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,并且典型地检查处理器205、存储器234(209、206)和典型地还存储在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,BIOS251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(bootstrap loader program)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。

[0100] 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。

[0101] 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在寄存器区段中典型地包括多个存储用寄存器244至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典型地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。

[0102] 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至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。

[0103] 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。

[0104] 瓶颈编码器116、瓶颈解码器148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。瓶颈编码器116、瓶颈解码器148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。

[0105] 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:

[0106] 提取操作,用于从存储器位置228、229、230提取或读取指令231;

[0107] 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

[0108] 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。

[0109] 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。

[0110] 图4至图15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,并且典型地通过处理器205中的寄存器区段244、245、247、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。

[0111] 图3A是示出可以用作CNN主干114的CNN的主干部310的功能模块的示意框图。主干部114有时被称为“DarkNet-53”,并且形成“YOLOv3”对象检测网络的主干。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。

[0112] 如图3A中所示,视频数据113被传递到大小调整器模块304。该大小调整器模块304将帧调整到适合于CNN主干310处理的分辨率,从而产生调整大小后的帧数据312。如果帧数据113的分辨率已经适合于CNN主干310,则不需要大小调整器模块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。

[0113] 参考图3D,CBL模块360将张量361作为输入。该张量361被传递到卷积层362以产生张量363。在卷积层362具有1的步幅并且填充被设置为k个样本(其中卷积核的大小是2k+1)

时,张量363具有与张量361相同的空间维度。在卷积层362具有较大的步幅(诸如2等)时,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的大小被减半。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,例如降低到先前值的0.1倍。

[0114] 返回到图3A,张量316从CBL块314被传递到残差块11(Res11)模块320。模块320包括在内部分别包含1个、2个和8个残差单元的三个残差块的顺次级联。

[0115] 参考如图3B所示的ResBlock 340描述诸如存在于模块320中等的残差块。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,其中残差块340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。

[0116] 参考如图3C中所示的ResUnit 350来描述诸如单元346等的残差单元。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,ResUnit350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。

[0117] 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328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。

[0118] 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个网络层、第90个网络层和第105个网络层处分离。各张量可以具有与下一张量不同的分辨率。各张量的分辨率可以在相应张量之间在高度和宽度上加倍。在形成输出张量115时,层张量322、326和329提供包括用于编码到位流的特征图的数据的帧数据的分层表示。分离点取决于CNN 310。

[0119] 图4是示出可以用作CNN主干114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的替代。帧数据113被输入并经由张量409、413、417、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。

[0120] 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(LeakyReLU激活)。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被分别传递到1×1横向卷积模块446、444、442和440。模块440、442、444和446分别产生张量441、443、445、447。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P5 471。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。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度。张量429、471、473、475和477形成CNN主干400的输出张量115。在形成输出张量115时,张量429、471、473、475和477的FPN提供包括用于编码到位流的特征图的数据的帧数据的分层表示。

[0121] 图5是示出一个类型的瓶颈编码器500的示意框图,该瓶颈编码器500可以用作瓶颈编码器116,或者当在支持具有权重更新的梯度的反向传播的情况下实现时,用作可训练瓶颈编码器170。图7A示出从压缩张量到单色视频帧中的特征图的打包布置。

[0122] 图14示出用于使用图5的瓶颈编码器500来降低张量维度的方法1400。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,如以下所述,方法1400可以由源装置110在处理器205的执行下作为应用程序233的一个或多于一个软件代码模块来实现。实现方法1400的应用程序233的软件代码模块例如可以驻留在硬盘驱动器210和/或存储器206中。针对位流123中的压缩数据的各帧重复方法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。

[0123] 方法1400从选择第一FPN张量步骤1410开始。瓶颈编码器500接收与张量115相对应的FPN张量501,并且根据方法1400进行操作以将接收到的张量的维度限制到更少的层并且具有减小的空间大小。在分离的神经网络的第一部分(主干114)和第二部分(头152)之间应用瓶颈编码器和解码器使得能够减小经打包张量数据帧中的空间区域。通过使用瓶颈编码器和瓶颈解码器之间的接口作为神经网络的第一部分和第二部分(114和152)的拆分点来实现空间面积的减小。瓶颈编码器116充当附加到神经网络的第一部分的附加层,并且瓶颈解码器150充当预置到神经网络的第二部分的附加层。张量115可以被认为包括第一张量集合(例如P2、P3、P4和P5)和第二张量集合(例如P2和P3),其中第二张量集合的特征图是第一张量集合的子集。第二张量集合的特征图包括具有第一张量集合的特征图的空间分辨率中的较大空间分辨率的张量。属于第一张量集合和第二张量集合这两者的第一张量(例如P2或P3)在基础层(第一信息单元)和增强层(第二信息单元)中表示。属于第一张量集合但不属于第二张量集合的第二张量(例如P4或P5)在基础层(第一信息单元)中表示,但不在增强层(第二信息单元)中表示。第一信息单元典型地对张量115内的所有张量(即P2至P5)进行编码,并且第二信息单元典型地对由第一信息单元编码的张量子集(例如P2和P3)进行编码,该子集包含具有P2至P5张量的特征图分辨率中的较大特征图分辨率的张量。张量集合的其他组合也是可能的。第二信息单元可以例如仅对张量P2进行编码或者可以对张量P2至P4进行编码。第一信息单元可以仅包括张量P2至P4,其中张量P5被单独打包到帧700中以用

于编码到位流123中。在第一信息单元包括张量P2至P4时,则第二信息单元可以包括张量P2或张量P2和P3。

[0124] 任务结果对瓶颈的灵敏度取决于任务的性质。对于对象检测,存在较小的空间灵敏度,并且因此FPN的较大层的空间下采样对所得到的mAP的危害较小。特别是对于FPN的空间较大张量,实例分段和所得到的分段图对空间细节的损失更灵敏,因此受益于不太严重的空间下采样。在所描述的布置中,瓶颈编码器500在两个尺度处进行操作。第一尺度覆盖所有的FPN层,并且第二尺度覆盖FPN层的子集。第二尺度可以包括较大层(例如P2和P3(502和503)),由此提供针对这些较高分辨率特征图的附加保真度。输入FPN张量501包括层P2 502、P3 503、P4 504和P5 505。层P2至P4(502、503、504)的空间分辨率是P5 505的空间分辨率的2的幂倍数。在P5 505具有宽度和高度(w,h)的情况下,P2至P4(502,503,504)分别具有尺寸(8w,8h)、(4w,4h)、(2w,2h)。换句话说,各个张量具有形成在连续张量之间在宽度和高度方面加倍的指数序列的分辨率。层P2至P5各自具有256个通道。尽管基础层被描述为强制性的并且增强层被描述为可选的(取决于诸如保真度等的情形),但包括第二增强层的布置是可能的。第二增强层仅在包括增强层时才能够被包括,并且第二增强层的张量是第一增强层的张量的子集(如上所述)。例如,仅在(第一)增强层与P2和P3相关的情况下,第二子集的张量才可以用于P2。换句话说,第二增强层针对已经由第一增强层增强的特定张量提供进一步的保真度。级联增强层通过将各渐进增强层的张量包括在经打包帧中来在提供增量质量改进时提供了进一步的灵活性。

[0125] 在上述示例中,输入P2至P5与由图4的CNN主干400生成的分层特征金字塔网络输出(P2 477、P3 475、P4 473和P5 471)相对应。如果CNN主干基于图3A实现并且输出张量329、326和322,则张量329、326和322被导出成两个张量集合:第一组具有一个FPN层329,并且第二组具有两个FPN层322和326。具有一个FPN层329的第一组具有最小的空间分辨率,并且不需要瓶颈编码器116中的MSFF模块510的操作,其中张量329作为张量529被直接传递到SSFC编码器550。第二组由瓶颈编码器116中的MSFF 510处理,其中张量326作为张量503被传递进入(输入),并且张量322作为张量502被传递进入。在步骤1410处,瓶颈编码器500在处理器205的执行下,选择张量501中相邻的多个张量作为第一多个张量(诸如四个张量P2 502、P3 503、P4 504和P5 505等)。张量501形成帧数据113的分层表示,该分层表示由将FPN应用于帧数据113产生。在FPN中(即,在模块412、416、420和424处)使用等于两个卷积阶段的步幅导致在根据例如从P2到P5的分解水平进行排序时,张量501中的张量的空间维度随着各相应张量在宽度和高度上减半。尽管层具有不同的空间分辨率,但在张量501的层P2至P5之间存在一定程度的层间相关性。只要张量首先在空间上缩放到相同的分辨率(例如,要组合的张量中的最小分辨率),利用层间相关性就许可相对于跨层的张量的级联的通道数减少。由于下采样操作的更高比率,对空间分辨率大大不同的张量进行组合会典型地导致更高分辨率张量中的细节的相对高损失。例如,将P2 502缩放到P5 505需要将宽度和高度减小到其先前值的八分之一,以将面积减小到P2 502面积的六十四分之一。对于诸如实例分段等的依赖于空间细节的任务,mAP劣化。对于网络头依赖更高分辨率层的小对象的检测,发生由于较大层的这种高下采样而引起的mAP的减小。处理器205中的控制从步骤1410进入生成第一瓶颈张量步骤1420。

[0126] 在步骤1420,MSFF模块510(参见图5)在处理器205的执行下组合第一张量集合的

各张量(即,502、503、504、505),以产生组合张量529。组合张量529被编码为压缩张量557。组合张量529形成FPN层张量的“基础层”表示。下采样模块522、522a、522b分别对具有较大空间尺度的张量(即,在2h、2w、256处的P4 504和在4h、4w、256处的P3 503和在8h、8w、256处的P2 502)进行操作。模块522、522a和522b进行下采样以匹配最小张量(即,在h、w、256处的P5 505)的空间尺度,从而分别产生下缩放P5张量523、523a、523b。级联模块524进行张量505、523、523a和523b的逐个通道的级联,以产生维度h、w、512的级联张量525。级联张量525被传递到挤压和激发(SE)模块526以产生张量527。SE模块526顺次进行全局池化、通道数减少的全连接层、修正线性单元激活、恢复通道数的第二全连接层、以及sigmoid激活函数,以产生缩放张量。张量525根据缩放张量而缩放以产生作为张量527的输出。SE块526能够被训练以基于第一全连接层输出自适应地更改通过的张量中的不同通道的加权。第一全连接层输出将针对各通道的各特征图减少到单个值。各单个值被传递通过非线性激活单元(ReLU)以创建适合于其他通道的加权的该单元的条件表示,其中由第二全连接层进行向全通道数的恢复。因而,SE块526能够在从张量525产生张量527时提取非线性通道间相关性,其程度比纯粹利用卷积(线性)层可能的程度更大。由于张量525和527包含512个通道(两个FPN层的级联的结果),因此由SE块526实现的去相关跨越两个FPN层P5和P4。

[0127] 张量527被传递到卷积层528。卷积层528实现一个或多于一个卷积层以产生第一组合张量529,其中通道数被减少到F个通道(典型地是256个通道)(即,F=256)。

[0128] SSFC编码器550的操作降低组合张量529的维度以产生压缩张量557。组合张量529被传递到编码器550的卷积层552。编码器550产生张量553。张量553具有从256减小到较小值C’(诸如64等)的通道数。值96也可以用于C’,这导致要参考图7A说明的针对经打包帧的更大面积要求。张量553被传递到批量归一化模块554以产生张量555。批量归一化张量555具有与张量553相同的维度。张量555被传递到TanH层556。TanH层556按照层536实现双曲正切(TanH)层以产生压缩张量557。压缩张量557具有与张量553相同的维度。步骤1420进行操作以从张量501导出或解码第一信息单元。处理器205中的控制从步骤1420进入确定第二瓶颈张量存在步骤1430。

[0129] 在步骤1430,处理器205确定是否生成和编码第二张量集合,第二经编码张量集合被指示为第二瓶颈张量537。步骤1430处的确定可以取决于系统100的配置和要在头网络152处完成的机器任务中的至少一个。例如,如果系统100被配置为进行需要高度空间敏锐度的任务(诸如实例分段等),则可以做出包括增强层的确定,从而许可目的地装置140实现更高的mAP。在系统100被配置为进行需要较低程度的空间敏锐度(或“常规质量”)的任务(诸如对象检测等)时,可以做出省略增强层的确定,从而节省附加层的位率开销。如果系统100被配置用于“常规质量”操作,则不生成第二张量集合,并且系统100被认为以“第一模式”操作。如果系统100被配置用于“高质量”操作,则生成第二张量集合并且系统100被设置为以“第二模式”操作。如果由目的地装置140正在进行的机器任务需要空间细节(例如)实例分段的相对高保留,则做出包括第二张量集合的确定。处理器205中的控制从步骤1420进入编码第二瓶颈张量存在指示步骤1440。在系统100的布置中,任务结果153的消费者(诸如人类操作员或聚合来自许多网络的任务的算法等)可以确定对更高质量的需求,并且经由带外通信通道向源装置110用信号通知包括(或省略)增强层的指示。一个示例布置将涉及目的地装置140的神经网络头152进行通用人物检测,并且在检测到人物时向源装置110用

信号通知以包括增强层,然后进行更有能力的可替代网络头作为网络头152。更有能力的可替代网络头可以进行具有更大特异性的对象检测任务,例如识别感兴趣的人物。

[0130] 在步骤1440,熵编码器838在处理器205的执行下,将指示在步骤1420处做出的以第一模式(仅基础层)或第二模式(基础层和增强层)操作的决定的标志编码到位流123中。该标志可以作为标志751被包括在SEI消息744中。处理器205中的控制从步骤1430进入第二瓶颈张量存在测试步骤1450。

[0131] 在步骤1450,软件233确定是否生成第二瓶颈张量537。如果存在指示要生成第二瓶颈张量的标志(步骤150处为“存在(PRESENT)”),则处理器205中的控制从步骤1450进入选择第二FPN张量步骤1460。否则,如果标志不存在(步骤1450处为“不存在(ABSENT)”),则方法1400终止。在实现步骤1450时终止方法400(在步骤1450处“不存在”)可以被认为是编码器500的第一操作模式。从步骤1450进入步骤1460以及以下的步骤可以被认为编码器500的第二操作模式。因此,步骤1450基于质量配置(quality configuration)和要完成的机器任务中的至少一个来确定是以第一模式还是第二模式进行操作。按照上述示例,如果要完成的机器任务是实例分段,则确定以第二模式的操作。

[0132] 在步骤1460,选择第二FPN张量集合。第二FPN张量集合是作为步骤1410的一部分所选择的张量的子集,并且一般包括具有较大空间分辨率和相邻空间尺度的张量(例如P2 502和P3 503)。激活开关570,使得将第二FPN张量集合指派给后续处理阶段。特别地,在开关270闭合时,张量502被提供为502a,并且张量503被提供为503a。处理器205中的控制从步骤1460进入生成第二瓶颈张量步骤1470。

[0133] 在步骤1470,MSFF模块510在处理器205的执行下,组合第二张量中的各张量(即502a、503a)以产生如参考图5所述的组合张量519。组合张量519以FPN层张量的子集的形式提供“增强层”表示。下采样模块512对具有较大空间尺度的张量(即,在8h、8w、256处的P2 502)进行操作,从而下采样以匹配较小张量(即,在4h、4w、256处的P3 503)的空间尺度,从而产生下缩放P2张量513。级联模块514进行张量503a和513的逐个通道的级联,以产生级联张量515。张量515具有维度4h、4w、512。级联张量515被传递到挤压和激发(SE)模块516以产生张量517。SE模块516以与参考SE模块526所述相同的方式操作。张量517被传递到卷积层518。卷积层518以与卷积层528类似的方式操作以产生第二组合张量519。第二组合张量519具有减少到F个通道(典型地是256个通道)(即,F=256)的通道数。作为模块512、514、516和518的结果,两个FPN层的张量被减小到单个张量,该单个张量具有与输入FPN层张量相同的通道数和两个FPN层张量中较小者的空间分辨率。降维是利用若干网络层实现的,并且依赖于训练层(例如516和518的层)而不是要利用的相关性的即时确定。

[0134] SSFC编码器530进行操作以进一步降低组合张量519的维度以产生压缩张量537。组合张量519被传递到卷积层532以产生张量533。张量533具有从256减小到较小值C’(诸如64等)的通道数。张量533被传递到批量归一化模块534以产生张量535。批量归一化张量535具有与张量533相同的维度。张量535被传递到TanH层536以产生压缩张量537。压缩张量537具有与张量533相同的维度。双曲正切(TanH)层的使用将张量537内的值的动态范围压缩到[-1,1],从而去除异常值。层532、534和536分别以与层552、554和556类似的方式操作。

[0135] 压缩张量557和537提供使用如下的卷积运算所获得的帧数据113的特征图的信息单元:(i)对张量505、504、503和502的MSFF 510和SSFC编码器550,以及(如果存在)(ii)对

张量503和502的MSFF 510和SSFC编码器530。压缩张量557和537提供与瓶颈经编码张量117相对应的张量集合560。如果编码器500正以第一模式操作(在1450处“不存在”并且开关570闭合),则张量557提供张量560。张量560被作为张量117提供到量化和打包模块184,以用于由视频编码器120编码到位流中。

[0136] 在瓶颈编码器500的布置中,TanH模块536和556被省略,从而导致张量535和555分别作为张量537和557传递。换句话说,省略536的布置中的输出537涉及将组合张量519应用于卷积层532和批量归一化层。省略556的布置中的输出557涉及将组合张量529应用于卷积层532和批量归一化层。省略TanH模块导致保留异常值或大幅度值,其中实验发现这些异常值或大幅度值对头152中的最终任务性能做出不成比例的贡献。

[0137] 在图7A中示出示例单个单色视频帧(帧700)。帧700与经打包和经量化的特征图数据185相对应。TanH在去除异常值方面的性质得到适合对帧700的位深度的线性量化的分布。压缩张量557的通道被打包为特定大小的特征图,诸如帧700中的区714中的特征图712等。压缩张量537的通道(如果(如在步骤1430处所确定的)存在)被打包为帧700的区域715中的不同大小的特征图,诸如特征图710等。压缩张量537的一个通道对应于由一个矩形区域(诸如区域710等)指示的一个特征图。压缩张量557的一个通道对应于由一个矩形区域(诸如区域712等)指示的一个特征图。区714和区716(如果存在)形成张量557和537的经打包表示,其中张量557和537一旦被视频编码器120压缩就分别形成第一信息单元和第二信息单元。第一信息单元和第二信息单元可以诸如通过使用帧700的单独切片、图块或子图片或完全单独的图片等以许可独立编码和解码的方式存储。视频解码器148可以仅对区714(即,第一信息单元)进行解码,并丢弃区716(即,第二信息单元),并且仍向CNN头152提供张量151以产生任务结果,尽管保真度低于区716或第二信息单元已被解码时可实现的保真度。

[0138] 图7B是示出对张量数据进行编码的位流723(其可以是位流123的一部分)的示意框图。压缩帧n 742包含如参考图7A所述布置的压缩张量。SEI消息744包括权重更新标志750,并且如果由权重更新标志750指示,则包括神经网络权重752。压缩帧N+1 746包含使用如从神经网络权重752导出的权重的压缩张量。第二瓶颈张量(诸如537等)的存在可以被编码在SEI消息744中作为存在标志751。

[0139] 图8是示出视频编码器120(也称为特征图编码器)的功能模块的示意框图。视频编码器120对经打包特征图帧185(在图7A的示例中示出为帧700)进行编码以产生视频位流121。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120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120,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120。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120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120包括模块810至890,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。

[0140] 尽管图8的视频编码器120是通用视频编码(VVC)视频编码流水线的示例,但其他视频编码标准或实现也可采用模块810至890的处理阶段。还可以从存储器206、硬盘驱动器210、CD-ROM、蓝光盘(Blu-ray Disc™)或其他计算机可读存储介质读取(或者向存储器206、硬盘驱动器210、CD-ROM、蓝光盘或其他计算机可读存储介质写入)帧数据185(和位流121)。另外,帧数据185(和位流121)可以从外部源(诸如连接到通信网络220的服务器或射频接收器等)接收(或被发送到外部源)。通信网络220可提供有限的带宽,从而需要在视频编码器120中使用速率控制以避免在帧数据185难以压缩时使网络饱和。帧数据185可以是所使用的配置文件支持的任何色度格式和位深度,例如,VVC标准的“主10”配置文件的4:0:0、4:2:0,样本精度为八(8)至十(10)位。

[0141] 块分区器810首先将帧数据185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个亮度样本,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810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总是具有作为二的幂的边长。因此,根据CTU的亮度编码树和色度编码树,从块分区器810输出表示为812的当前CB(根据CTU的一个或多于一个块上的迭代来前进)。

[0142] 可以按光栅扫描顺序扫描由帧数据185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。

[0143] 视频编码器120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输出该图片,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经解码图片缓冲器中,从而导致多帧时延。

[0144] 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。

[0145] 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。

[0146] 对于各CTU,视频编码器120在两个阶段中操作。在第一阶段(称为“搜索”阶段)中,块分区器810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。测试阶段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185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121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。

[0147] 视频编码器120针对各CB(例如,CB 812)产生由箭头820指示的预测块(PB)。PB 820是关联的CB 812的内容的预测。减法器模块822产生PB 820和CB 812之间的表示为824的差(或“残差”,指差在空间域中)。差824是PB 820和CB 812中的相应样本之间的块大小差。差824被变换、量化和表示为由箭头836指示的变换块(TB)。PB 820以及关联的TB 836典型地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。

[0148] 候选编码块(CB)是从视频编码器120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120中的预测PB组合时,TB 836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 812之间的差。

[0149] 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的失真典型地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886使用差824可以确定从各候选PB得到的估计,以确定预测模式887。预测模式887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。

[0150] 根据率失真确定最佳模式典型地是使用拉格朗日优化的变型来实现的。

[0151] 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810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886中应用候选模式的拉格朗日优化处理,选择具有最低成本测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888,该索引也由熵编码器838编码到位流121中。

[0152] 在视频编码器120的操作的第二阶段(称为“编码”阶段),在视频编码器120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)。

[0153] 熵编码器838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。对于给定切片,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。

[0154] 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121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121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编码技术使得能够对使“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。

[0155] 句法元素的值向一个或多于一个bin的序列的分解被称为句法元素的“二值化”。二值化可以包括较早bin的值上的较晚bin的有条件存在,从而使可变bin长度二值化成为可能。另外,各bin可以与多于一个上下文相关联。针对bin的上下文的选择被称为“上下文建模”。上下文建模可以取决于句法元素中的较早bin、邻近句法元素(即,来自邻近块的邻近句法元素)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。

[0156] 熵编码器838还支持缺少上下文的bin(被称为“旁路bin”)。利用“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121中具有一位的编码成本,并且一般在bin值的概率分布中不存在(不容易利用)统计偏移的情况下使用。不存在上下文节省了存储器并降低了复杂度,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。

[0157] 熵编码器838使用上下文编码的和旁路编码的bin的组合来对量化参数892进行编码,并且如果用于当前CB,则对二次变换索引888进行编码。使用QP控制器模块890所生成的“差量QP”来对量化参数892进行编码。在已知为“量化组”的各区域中,最多用信号通知差量QP一次。量化参数892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892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888。

[0158] 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。

[0159] 多路复用器模块884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864输出PB 820。候选预测模式不需要包括视频编码器120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邻近样本导出的垂直和水平梯度用样本填入PB。附近重建样本典型地包括在当前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滤波和传播的重建邻近样本填入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。

[0160] 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,各模式使用从邻近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。

[0161] 模块864还可以通过使用“块内复制”(IBC)方法从当前帧中的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。对于128×128CTU,发生向64×64个象限(有时称为“虚拟流水线数据单元”(VPDU))的分割。可参考区域包括当前CTU中的所有CU已被解码的VPDU、以及先前CTU中的VPDU(除在当前CTU是切片、图块或子图片中的第一个时以外),直到128×128个亮度样本的总区域。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854(即在环路滤波之前),因此需要与帧缓冲器872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,访问邻近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大小得到更加对齐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。

[0162] 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。这样自然视频典型地由摄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,这相比于各种变换的显著低频系数,更适合变换跳过编码。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,并且这在使用块内复制来产生特征图数据的预测块时也适用。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。

[0163] 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)途径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。

[0164] 在先前重建的邻近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)。

[0165] 对于帧间预测,由运动补偿模块880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882,并由多路复用器模块884输出该预测块882作为PB 820。此外,对于帧间预测,单个编码树典型地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。

[0166] 典型地使用“图片组”(GOP)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。从邻近块(“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邻近块共享相同的运动向量。

[0167] 根据运动向量878和参考图片索引来选择样本。运动向量878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。运动参数典型地包括参考帧索引,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。

[0168] 已经确定和选择了PB 820,并且在减法器822处从原始样本块中减去PB 820,获得具有最低编码成本的残差(表示为824),并且对该残差进行有损压缩。有损压缩是从通过向残差系数的正向变换产生的系数的量化处理产生的,其中这些系数准备好被熵编码到位流中。正向主变换模块826对差824应用正向变换,将差824从空间域转换到频域,并产生由箭头828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826来变换差824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。

[0169] 变换826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824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828被设置为零(即,被丢弃)。剩余的主变换系数828被传递到量化器模块834。根据与CB相关联的量化参数892对主变换系数828进行量化,以产生主变换系数832。除了量化参数892之外,量化器模块834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892可以不同。主变换系数832被传递到正向二次变换模块830,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836表示的变换系数。正向主变换典型地是可分离的,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826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)。

[0170] 模块830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828的左上4×4子块)或四十八(48)个样本(布置为主变换系数828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(诸如水平地和垂直地应用的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。

[0171] 量化参数892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892可以随着用信号通知的“差量量化参数”而定期地变化。对于包含在给定区域内的CU(称为“量化组”),用信号通知差量量化参数(差量QP)一次。如果CU大于量化组大小,则利用CU的TB之一,用信号通知差量QP一次。即,通过熵编码器838针对CU的第一量化组用信号通知差量QP一次,而不针对CU的任何后续量化组用信号通知差量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892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,最近邻途径用于从大小小于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836被供给至熵编码器838以在位流121中编码。典型地,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的残差系数集合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典型地遵循反向对角线扫描模式。另外,使用差量QP句法元素将量化参数892编码到位流121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888被编码到位流121中。

[0172] 如上所述,视频编码器120需要访问与视频解码器中看到的经解码帧表示相对应的帧表示。因此,残差系数836被传递通过逆二次变换模块844,该逆二次变换模块844根据二次变换索引888操作以产生由箭头842表示的中间逆变换系数。由去量化器模块840根据量化参数892对中间逆变换系数842进行逆量化,以产生由箭头846表示的逆变换系数。去量化器模块840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834中进行的正向缩放。逆变换系数846被传递到逆主变换模块848,以产生TU的残差样本(由箭头850表示)。逆主变换模块848水平和垂直地应用DCT-2变换,这受参考正向主变换模块826描述的最大可用变换大小的约束。由逆二次变换模块844进行的逆变换的类型与由正向二次变换模块830进行的正向变换的类型相对应。由逆主变换模块848进行的逆变换的类型与由主变换模块826进行的主变换的类型相对应。求和模块852将残差样本850和PU 820相加以产生CU的重建样本(由箭头854指示)。

[0173] 重建样本854被传递到参考样本高速缓冲存储器856和环内滤波器模块868。典型地使用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856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。最小依赖关系典型地包括沿着一行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856将参考样本(由箭头858表示)供给至参考样本滤波器860。样本滤波器860应用平滑操作以产生经滤波参考样本(由箭头862指示)。经滤波参考样本862由帧内预测模块864使用以产生由箭头866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864产生样本块(即866)。通过模块864使用诸如DC、平面或角度帧内预测等的技术来生成样本块866。也可以使用矩阵乘法途径来产生样本块866,其中邻近参考样本作为输入,并且由视频编码器120从矩阵集合中选择矩阵,使用索引在位流121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144使用。

[0174] 环内滤波器模块868对重建样本854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边界对齐的平滑化,以减少由不连续而产生的伪影。环内滤波器模块868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868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。

[0175] 从环内滤波器模块868输出由箭头870表示的滤波样本。滤波样本870被存储在帧缓冲器872中。帧缓冲器872典型地具有存储数个(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,典型地不使用片上存储器来存储帧缓冲器872。如此,对帧缓冲器872的访问在存储器带宽方面是昂贵的。帧缓冲器872将参考帧(由箭

头874表示)提供至运动估计模块876和运动补偿模块880。

[0176] 运动估计模块876估计多个“运动向量”(表示为878),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872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882)。滤波参考样本882形成可供模式选择器886的潜在选择用的进一步候选模式。此外,对于给定CU,PB 820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880根据支持运动向量中的子像素精度的滤波处理来产生PB 820。如此,运动估计模块876(其对许多候选运动向量进行操作)与运动补偿模块880(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度。当视频编码器120针对CU选择帧间预测时,运动向量878被编码到位流121中。在图9中示出视频解码器146(也称为特征图解码器)。尽管图9的视频解码器146是通用视频编码(VVC)视频解码流水线的示例,但其他视频编解码器也可用于进行本文所述的处理阶段。如图9所示,位流143被输入到视频解码器146。位流143可以从存储器206、硬盘驱动器210、CD-ROM、蓝光盘或其他非暂态计算机可读存储介质中读取。可替代地,位流143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经编码句法元素。

[0177] 位流143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143中提取句法元素,并将句法元素的值传递给视频解码器146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。

[0178] 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143中解码句法元素。经解码的句法元素被用于重建视频解码器146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。典型地与来自先前解码的CB的样本数据相结合,参数被用于生成PB。

[0179] 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840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144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。

[0180] 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826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。

[0181] 在求和模块950处,残差样本948被加到经解码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。

[0182] 重建样本高速缓冲存储器960以类似于视频编码器120的重建样本高速缓冲存储器856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,通过作为代替使用典型地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143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,产生被表示为980的帧内预测样本的块。帧内预测模块976支持模块864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。

[0183] 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形成经解码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“邻近样本”导出的。邻近样本是与当前块相邻的样本,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。

[0184] 当CB的预测模式在位流143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以产生用于经解码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120的环内滤波模块868一样,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。来自帧缓冲器996的帧被输出为经解码帧162。

[0185] 图10是示出跨层张量逆瓶颈解码器1000的示意框图,该跨层张量逆瓶颈解码器1000与用于在压缩之后恢复张量维度的解码器150(以及类似地,解码器118和174)相对应。图15示出用于使用图10的瓶颈解码器150恢复张量维度的方法1500。方法1500可以使用诸如配置的FPGA、ASIC或ASSP的设备来实现。可替代地,如以下所述,方法1500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。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。针对位流123中的各压缩数据帧重复方法1500。方法1500可以存储在计算机可读存储介质上和/或存储器206中。方法1500提供“可切换”部件以解码包含整个FPN的压缩表示并且可选地包含FPN的一部分的附加压缩表示的位流,并且将FPN上的两个表示(整个和附加部分)组合成最终的经解码FPN张量以供提供给CNN头152。

[0186] 解码器150接收如通过视频解码器146和模块160对位流143的操作所生成的张量147。张量147包括张量1011和1021作为第一信息单元和第二信息单元。张量1021与张量537的经解码版本相对应。类似地,张量1011与张量557的经解码版本相对应。如果通过编码器

500的第一模式的操作来生成位流143,则由视频解码器146从位流中解码至少张量1011。还可以解码与模式操作相关的标志。如果通过编码器500的第一模式的操作生成位流143,则也由视频解码器146解码张量1021。方法1500从解码第一瓶颈张量步骤1510开始。

[0187] 在步骤1510,在处理器205的执行下实现SSFC解码器1010。SSFC解码器进行神经网络层以对张量147的第一经解码压缩张量1011进行解压缩以产生第一经解码组合张量1017。卷积层1012接收具有C'＝64个通道的张量1011并且输出具有F＝256个通道的张量1013。张量1013被传递到批量归一化层1014。批量归一化层输出张量1015。张量1015被传递到参数化泄漏修正线性(PReLU)层1016。PReLU层1016输出张量1017。张量1017被传递到MSFR模块1030,该MSFR模块1030生成张量1051、1053、1055和1037,从而形成经解码FPN层的“基础层”。基础层提供了与在包括附加的“增强层”经解码FPN层时存在的保真度相比更低程度的保真度。上采样模块1032、1034和1036接收张量1017并以2倍、4倍和8倍比例执行插值以分别产生张量1033、1035和1037。例如,张量1033的宽度和高度是张量1017的宽度和高度的两倍。

[0188] 张量1037形成来自MSFR模块1030的一个输出并且被传递到下采样模块1042。下采样模块1042利用2的因子、水平地和垂直地对张量1037进行下采样,以产生具有与张量1035相同维度的张量1043。张量1043被提供给卷积层1048,该卷积层1048输出张量1049。求和模块1054将张量1035和1049相加以产生张量1055作为MSFR模块1030的输出。下采样模块1040利用2的因子、水平地和垂直地对张量1035进行下采样,以产生具有与张量1033相同维度的张量1041。张量1041被提供给卷积层1046,该卷积层1046输出张量1047。求和模块1052将张量1033和1047相加以产生张量1053作为MSFR模块1030的输出。下采样模块1038利用2的因子、水平地和垂直地对张量1033进行下采样,以产生具有与张量1017相同维度的张量1039。张量1039被提供给卷积层1044,该卷积层1044输出张量1045。求和模块1050将张量1017和1045相加以产生张量1051作为MSFR模块1030的输出。

[0189] 步骤1510生成张量1051、1053、1055和1037。张量1051、1053、1055和1037形成图像帧的分层表示,并且可以被认为包括第一张量(例如,1037P'2或1055P'3)和第二张量(例如,1053P'4或1051P'5),第一张量中的特征图与第二张量的特征图相比具有更大的空间分辨率。处理器205中的控制从步骤1160进入解码第二瓶颈张量存在指示步骤1520。

[0190] 在步骤1520,熵解码器920在处理器205的执行下解码来自位流123的指示,该指示用于指示位流123是否包括第二瓶颈张量。可以根据从SEI消息744获得的经解码存在标志751来确定第二瓶颈张量的存在。处理器205中的控制从步骤1520进入第二瓶颈张量存在测试步骤1530。

[0191] 在步骤1530,应用233执行以确定步骤1520处的位流123的解码是否指示包括第二瓶颈张量。如果存在标志751指示了第二瓶颈张量被编码(步骤1530处为“存在”),则处理器205中的控制从步骤1530进入解码第二瓶颈张量步骤1540。否则,如果未确定第二瓶颈张量的存在(步骤1530处为“不存在”),则处理器205中的控制进入组合第一张量和第二张量步骤1550。在第二瓶颈张量未被确定为存在(1530处为“不存在”)时,解码器1000可以被认为以第一(基础)操作模式操作。在第二瓶颈张量被确定为存在(1530处为“存在”)时,解码器1000可以被认为以第二(增强)操作模式操作。

[0192] 在步骤1540,在处理器205的执行下实现SSFC解码器1020。SSFC解码器1020进行神

经网络层以对张量147的第二经解码压缩张量1021进行解压缩以产生第二经解码组合张量1027。

[0193] 卷积层1022接收具有C'＝64个通道的张量1021并且输出具有F＝256个通道的张量1023。张量1023被传递到批量归一化层1024。批量归一化层1024输出张量1025。张量1025被传递到PReLU层1026。PReLU层1026输出张量1027。

[0194] MSFR模块1030使用上采样模块1060、下采样模块1062、卷积层1064和求和模块1066来生成经解码张量1061和1067。张量1027被传递到上采样模块1060。上采样模块1060进行插值以产生宽度和高度是张量1027的宽度和高度的两倍的张量1061。张量1061从MSFR模块1030输出并被传递到下采样模块1062。下采样模块1062对张量1061进行下采样以产生具有与张量1027相同维度的张量1063。张量1063被提供给具有1的步幅的卷积层1064,该卷积层1064输出张量1065。求和模块1066将张量1065和1027相加以产生张量1067,该张量1067从MSFR模块1030输出。处理器205中的控制从步骤1540进入步骤1550。

[0195] 在步骤1550,产生经解码FPN张量P'2至P'5,即1051、1053、1073和1077。来自MSFR模块1030的基础层部分的张量1051和1053准备好被传递到CNN头150。在步骤1550处根据张量1055和1037以及可选地根据张量1071和1078确定张量1073和1077。如果确定为不使用增强层(即,省略第二瓶颈张量),则多路复用器1074和1076分别输出张量1055和1037作为张量1073和1077。在省略第二瓶颈张量时,向CNN头提供允许进行任务的所有FPN层的张量,然而由于较高分辨率层(例如P2和P3)中的较低空间保真度而具有降低的任务性能。由于仅使用基础层而导致的降低的任务性能往往限制针对实例分段的最大可实现的映射,其中在该实例分段中,在视频编码器120中达到近无损压缩并且瓶颈编码器和解码器内的损失被最小化。(除基础层之外还)存在增强层可以将最大可实现的mAP增加到几乎是神经网络作为单个操作(即,在不分离成两个部分的情况下)运行时所实现的mAP。

[0196] 如果确定包括增强层(步骤1030处为"存在"),则进行卷积1070和1072。卷积1070采用沿着通道维度级联的张量1055和1067作为输入并产生输出张量1071。卷积1072采用沿着通道维度级联的张量1037和1061作为输入并产生输出张量1078。多路复用器1074和1076传递张量1071和1078作为张量1073和1077。在包括增强层时,针对P'2和P'3的输出张量1071和1073已增加了空间保真度,这有利于诸如实例分段等的任务。

[0197] 因此,在第二模式中,导出形成图像数据的分层表示的至少一部分的多个张量。这些张量是从张量1021以及在步骤1510处解码的与至少第一张量相关的张量的至少一部分(例如,1051和1053)导出的。

[0198] 在方法1500的布置中,多路复用器1074和1076被省略,并且根据是否使用增强层的确定来初始化卷积1072和1074。在使用增强层时,预训练的权重用于初始化卷积1072和1074。因此,在第二(增强)模式中,卷积层接收针对从第一信息单元和第二信息单元中的每一个导出的P'2和P'3张量(每一个是第一张量的示例)的张量。当增强层不在使用中(在第一模式中)时,初始化卷积1072和1074,使得与输入张量1055和1037相对应的卷积权重形成单位矩阵(identity matrix),并且与输入张量1061和1067相对应的权重被归零。利用卷积模块1072和1074将单位矩阵应用于输入张量1055和1037导致分别输出张量1055和1037作为1071和1078,其中输入张量1061和1067由于其对应的权重被归零而对输出没有贡献。

[0199] 方法1500在实现步骤1550时终止,已产生了准备用于利用CNN头150的处理的经解

码FPN。针对编码在位流123中的视频数据的各帧重新调用方法1500。

[0200] 瓶颈编码器116和瓶颈解码器150的利用FPN层张量的基本层和增强层表示的操作提供了质量可扩展性的形式。为了确保增强层作为“差量”的预期操作以提高经解码FPN张量119或151相对于从主干114发出的FPN张量115的保真度,必须在增强层不活动的情况下初始训练与基础层相关联的瓶颈编码器和解码器中的可训练层。SE模块526、卷积528、SSFC编码器550、SSFC解码器1010、卷积1044、1046和1048被训练以提供瓶颈编码器和解码器的基础层能力。为了训练增强层,瓶颈编码器和解码器中的与基础层相关联的模块是固定的,并且与增强层相关联的模块被设置为可训练。然后,增强层模块(SE模块516、卷积518、SSFC编码器530、SSFC解码器1020、卷积1064)以及用于将基础层和增强层张量合并在一起的两个卷积1070和1072学习以在仅使用基础层所实现的性能之上提供性能的“差量”改进。

[0201] 图12A是示出用于对象检测的CNN的头部152的示意框图。取决于要在目的地装置140中进行的任务,可以用不同的网络替代CNN头152。输入张量151被分离成各层的张量(即,张量1210、1220和1234)。张量1210被传递到CBL模块1212以产生张量1214。张量1214被传递到检测模块1216和升级器模块1222。检测模块1216进行操作以检测边界框1218。边界框1218采用检测张量的形式。边界框1218被传递到非最大抑制(NMS)模块1248。NMS模块1248选择由检测模块生成的多个输入之一以产生检测结果153。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,进行按原始视频宽度和高度的缩放。升级器模块1222产生按原始视频宽度和高度缩放的升级张量1224。升级张量1224被传递到CBL模块1226。CBL模块1226产生张量1228作为输出。张量1228被传递到检测模块1230和升级器模块1236。检测模块1230以与检测模块1216类似的方式操作,并且产生检测张量1232。检测张量1232被供给到NMS模块1248。

[0202] 升级器模块1236以与模块1260相同的方式进行操作,并且输出升级张量1238。升级张量1238被传递到CBL模块1240。CBL模块1240以与模块1212和1226相同的方式进行操作以将张量1242输出到检测模块1244。检测模块1244以与检测模块1216类似的方式进行操作,并且产生检测张量1246。检测张量1246被供给到NMS模块1248。

[0203] CBL模块1212、1226和1240各自包含五个CBL模块的级联,各CBL模块如参考图3D所述。升级器模块1222和1236是如图12B所示的升级器模块1260的各实例。

[0204] 升级器模块1260接受张量1262和张量1264作为输入。张量1262被传递到CBL模块1266以产生张量1268。使用最近邻插值或其他各种方法,张量1268被传递到上采样器1270以产生上采样张量1272。级联模块1274通过将上采样张量1272与输入张量1264级联来产生张量1276。

[0205] 检测模块1216、1230和1244是如图12C所示的检测模块1280的实例。检测模块1260接收张量1282,该张量1282被传递到CBL模块1284以产生张量1286。张量1286被传递到卷积模块1288,该卷积模块1288实现检测核。检测核应用1×1核以在三个层处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,典型地是三(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。模块1288输出张量1290。常数“5”表示四个边界框属性(框中心x,y和大小尺度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,典型地是三个层,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的处理以丢弃冗余框,诸如相似尺度的重叠预测等,从而得到边界框的最终集合作为对象检测的输出。

[0206] 图13是示出如可以针对模块152实现的CNN的可替代头部1300的示意框图。头部1300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,主干部分400)、区建议网络和检测网络。向头部1300的输入是张量151。张量151分别包括P2至P6层张量1310、1312、1314、1316和1318。P2至P6层张量1310、1312、1314、1316和1318被输入到区建议网络(RPN)头模块1320。RPN头模块1320对输入张量进行卷积,从而产生中间张量。中间张量被馈送到模块1320中的两个后续的同级层中,一个层用于分类,以及一个层用于边界框或“感兴趣区”(ROI)回归,从而生成分类和边界框1322的输出。分类和边界框1322被传递到NMS模块1324。NMS模块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1326。

[0207] 边界框1326被传递到感兴趣区(ROI)对齐(align)模块1328(或“RoIAlign”阶段)。ROI对齐模块1328还接收张量P2至P6,并使用双线性插值操作从各种输入大小图中产生固定大小的特征图。在由ROI对齐模块1328进行的操作中,子采样是从接收感兴趣区中的作为3×3个子区的多个子区的双线性插值产生的,以产生感兴趣输出区作为输出张量中的一个输出值。

[0208] 在CNN主干400和CNN头1300的布置中,从输出张量115中省略“P6”层张量429,并且在CNN头1300中,通过对P5张量1316进行步幅等于2的“Maxpool”操作来产生P6输入张量1318。由于P6层可以从P5层重建,因此无需单独编码和解码第一张量集合或第二张量集合中的作为显式FPN层的P6层。

[0209] 向ROI对齐模块1328的输入是P2至P5特征图1310、1312、1314和1316(分别对应于图10的1077、1073、1053和1051)以及感兴趣区建议1326。来自1326的各建议(ROI)与特征图(1310～1316)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1310～1316的基础部分。例如,根据以下规则,选择特征图1310～1316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是典型的框大小。ROI对齐模块1328由此根据建议1326裁剪传入的特征图,从而产生张量1330。张量1330被馈送到全连接(FC)神经网络头1332。FC头1332进行两个全连接层以产生类得分和边界框预测子差量张量1334。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。边界框预测差量张量是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1336进行,输出层模块1336接收张量1334并进行滤波操作以产生滤波后的张量1338。最终处理利用与对象分类以及边界框的确对应于对象的置信度(“对象性”值)有关的指示,针对各FPN层的张量中的各位置,编码一个或多于一个边界框。低得分(低分类)对象不再被进一步考虑。非最大抑制模块1340接收滤波后的张量1334,并通过去除具有较低分类得分的重叠框来去除被编码在所接收的张量中的重叠的边界框,从而得到推断输出张量151。

[0210] 在源装置110和目的地装置140的布置中,主干114和头152被省略,并且瓶颈编码器和解码器(即,170、174、116、118和150)可作为端到端经学习图像压缩和解压缩神经网络进行操作,从而采用图像帧作为向编码阶段的输入并从解码阶段输出经解码图像帧。根据需要在操作期间训练端到端经学习图像压缩网络以适应改变的输入帧数据。这使得能够使

用动态更新以匹配被压缩的特定视频或图像的潜在较小网络,而不是依赖于需要已经在各种源材料上训练以实现一致性能的预训练网络。可能需要适配的不同类型的源材料的示例包括屏幕内容、照相机捕获内容(在各种照明和其他条件下)、以及渲染内容等。用于测量瓶颈编码器和解码器的性能的度量可以不同于MSE,例如,可以在步骤630和645处产生MS-SSIM。在系统100可操作以提供可训练的端到端经学习图像压缩的情况下,MS-SSIM的性能度量可能是有用的。

[0211] 瓶颈编码器和解码器(即,170、174、116、118和150)的初始权重可以通过使用适合于原始任务网络(即,适合于由主干114和头152形成的网络)的数据集和地面真值进行训练来导出。可以在主干114和头152的网络权重固定并且允许更新插入的瓶颈编码器和解码器的网络权重的情况下进行使用这种数据集的训练。在任何细化训练之前,可以在源装置110和目的地装置140中使用这样的初始权重。这样的初始权重可以表现出如下的趋势:在训练期间,当基于各通道测量时,MSE在通道之间变化。在与最终任务性能相对应的损失函数下通道之间的MSE方差指示各通道对最终任务结果的相对贡献。在源装置110的布置中,模块178和180产生每通道权重MSE,使得对最终任务性能做出较小贡献的通道在其对最终MSE的贡献方面被“降额”或按比例缩小。基于预定加权的MSE的每通道(或“逐通道”)缩放使得能够在不会过度分配重要性以保留对最终任务得分做出相对较小贡献的通道的情况下适应细化训练。

[0212] 在瓶颈编码器和解码器的布置中,模块116、120、170、174和150操作以将所有FPN层的张量合并成单个张量,该单个张量具有根据用于一个图像的FPN张量中的最小分辨率张量所设置的维度。换句话说,参考图5和图10,一个MSFF模块510将用于所有FPN层张量(例如,502、503、504和505)的张量融合在一起成单个张量,并且MSFR模块1030重建用于所有FPN层的张量(例如,1077、1073、1053和1051)。另外,如果需要,开关570可以闭合,使得用于P2和P3张量的附加数据可以如关于图5所述被编码,并且如针对图10关于与MSFR 1030相关联的SSFC解码器1020所述被解码。

[0213] 在源装置110的布置中,多个经训练权重集合可用,例如,针对屏幕内容、照相机捕获内容、渲染内容而优化的权重。源装置110可操作以在可用的预定权重中选择最优权重集合,并用信号通知目的地装置140所选择的权重。源装置110可以“测试”模块170和174中的各权重集合,以确定应使用哪个集合。帧数据113的内容类型的变化可能导致如由模块182测量到的性能的降低,该降低提示应该使用预定权重的哪个集合的重新评估。

[0214] 在源装置110的布置中,模块170和174可操作以训练与增强层(第二信息单元)以及卷积1070和1072相关联的权重,但不训练与基础层(第一信息单元)相关联的权重。与位流123中的权重更新相关联的信令支持指示在做出更新权重的确定时仅更新增强层。

[0215] 在源装置110的另一布置中,模块170和174可操作以将基础层和增强层训练为单独的训练阶段。在对基础层进行训练时,增强层被禁用,从而允许导出最佳的基础层权重。一旦在源装置110中确定了针对基本层的更新权重并将这些更新权重通信到目的地装置140,则在可以启用增强层之前,增强层需要重新训练。由于增强层与已被重新训练的基础层组合地操作,因此需要重新训练。一旦增强层权重已被训练用于新基础层上的操作,则在增强层可以被重新启用以用于张量115的压缩之前,增强层权重必须被通信到目的地装置140。

[0216] 在基础层和增强层被单独训练的布置中,使用附加的权重更新标志(即,除权重更新标志750之外的标志)来指示要对哪个层进行权重更新。权重752包括针对所指示的层的权重。

[0217] 在系统100的布置中,增强层537的各特征图被表示为可应用于基向量集合的系数集合。使用主成分分析(PCA)方法(诸如奇异值分解(SVD)等)从增强层导出基向量。当PCA方法在使用中时,帧700的区716包括基向量和系数,其中增强层537的每特征图每基向量一个系数。在源装置110中,使用与基向量的点积来进行从增强层537向系数的变换。在目的地装置140中,利用系数和基向量的点积进行从系数回到重建增强层1021的变换。PCA方法可以应用于基础层和增强层这两者,从而产生两个基向量集合和两个系数集合。在文献“[VCM Track 1]Tensor compression using VVC”(ISO/IEC JTC 1/SC 29/WG 2文档m59591)中详细描述了PCA编码器和PCA解码器的操作。PCA方法可以应用于基础层和增强层这两者,其中针对各层产生单独的基向量和系数。PCA方法可以仅应用于增强层,其中仅针对该层产生基向量和系数,同时将基础层的所有特征图直接打包到帧700中。

[0218] 在将单独的瓶颈编码器应用于FPN层的不同非重叠张量集合的情况下,PCA方法可以独立地应用于所有所得压缩张量中的任何压缩张量。在PCA编码器在使用中的情况下,从SSFC编码器550和530的输出(即,从张量557和537(其是TanH模块556和536(如果存在)的输出或批量归一化模块554和534的输出))接收到向PCA编码器的输入张量。输出基向量、系数和平均特征图被转发以用于量化和打包到帧700中。在PCA解码器在使用中的情况下,从经打包帧700获得输入基向量、系数和平均特征图,并且输入基向量、系数和平均特征图被供给到各PCA解码器,各PCA解码器输出张量1011和1021以分别由SSFC解码器1010和1020使用。在系统100的布置中,在使用PCA编码器和解码器的情况下,批量归一化554和534被推迟直到PCA解码器之后为止,即,模块554和534从SSFC编码器550和530中省略,并且在SSFC解码器1010和1020中的PCA解码器之后进行。

[0219] 在系统100的布置中,不论要进行的任务是对象检测还是实例分段,源装置140都实现MaskRCNN主干,并且利用针对MaskRCNN网络的预训练权重来初始化。在进行对象检测时,目的地装置140在152处实现FasterRCNN头,并且利用针对FasterRCNN网络的预训练权重来初始化。如果实现了FasterRCNN头,则需要瓶颈解码器150充当从MaskRCNN主干生成但被供给到FasterRCNN头的特征图之间的接口。瓶颈编码器116在针对MaskRCNN主干和头的训练方面保持优化,因为在进行编码时可能不知道头网络将进行什么。为了准备针对瓶颈解码器150的初始权重,可以进行“混合”训练处理。混合训练处理包括实例化FasterRCNN网络,其中利用MaskRCNN权重初始化主干,而利用FasterRCNN权重初始化头。利用MSFC(瓶颈编码器和解码器)的MaskRCNN训练相对应的权重来初始化MSFC编码器116。然后,仅瓶颈解码器150被设置为可训练,并且所有其他网络层被设置为固定。进行训练操作,该训练操作训练瓶颈解码器,以不仅对压缩FPN张量进行解码,而且使所得特征图适应以匹配向FasterRCNN头的预期输入,从而导致最小化损失。作为训练的结果,瓶颈解码器150提供MaskRCNN主干和FasterRCNN头之间的适配,其中瓶颈编码器16保持针对能力更强的网络(即,MaskRCNN)进行优化。当利用针对“混合”操作系统(具有FasterRCNN头的MaskRCNN主干)训练的权重而初始化瓶颈解码器150时,来自源装置110的所得位流在其生成时适合于对象检测和实例分段这两者,也就是说,相同位流可以稍后在无需任何转码或其他操作的

情况下用于这两个任务。在由源装置110生成的单个位流可以用于向不同的神经网络头提供张量(假设与各神经网络头相对应的不同神经网络共享相同的主干拓扑和维度)时,源装置110被称为支持“共享主干”操作模式。

[0220] 在系统100的另一布置中,与分别针对SSFC编码器550和SSFC解码器1010(即,基本层)的C’通道数相比,针对SSFC编码器530和SSFC解码器1020(即,增强层)的C’通道数减小。增强层可以使用32的C’值。这样的布置将较少的较大大小的特征图(即,710)打包到帧700中。重新训练应用于增强层的瓶颈编码器和瓶颈解码器的能力许可使用较少的通道来对P2和P3层中存在的增强细节进行编码,从而适应跨适用张量的全通道数(即,跨P2和P3FPN层的256个通道)遇到的变化的统计信息。

[0221] 产业上的可利用性

[0222] 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。

[0223] 关于图1中的权重编码描述的布置提供了能够通过时不时地经过细化训练处理(如通过使用中的瓶颈编码器和解码器的性能的持续监测被认为有必要)来适应输入视频数据的动态变化的统计信息的系统。在确定提供改进的性能的细化权重时,更新编码器和解码器中主动使用的权重以使用细化权重,从而维持训练处理期间的操作。使用性能的持续监测允许基于数据输入的变化或基于推断操作期间的特定数据类型输入来实现MFSC单元116和MFSC解码单元150的训练。因此,可以在不需要针对特定图像类型(诸如自然或计算机生成的图像)或情境的单独的系统外训练的情况下调整或更新特征压缩操作。

[0224] 如关于图14和图15所述,关于图5和图10所述的布置允许在高性能和低性能要求之间的灵活操作。针对各系统不需要不同的架构,而是提供了不需要加载不同网络的灵活性。

[0225] 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。

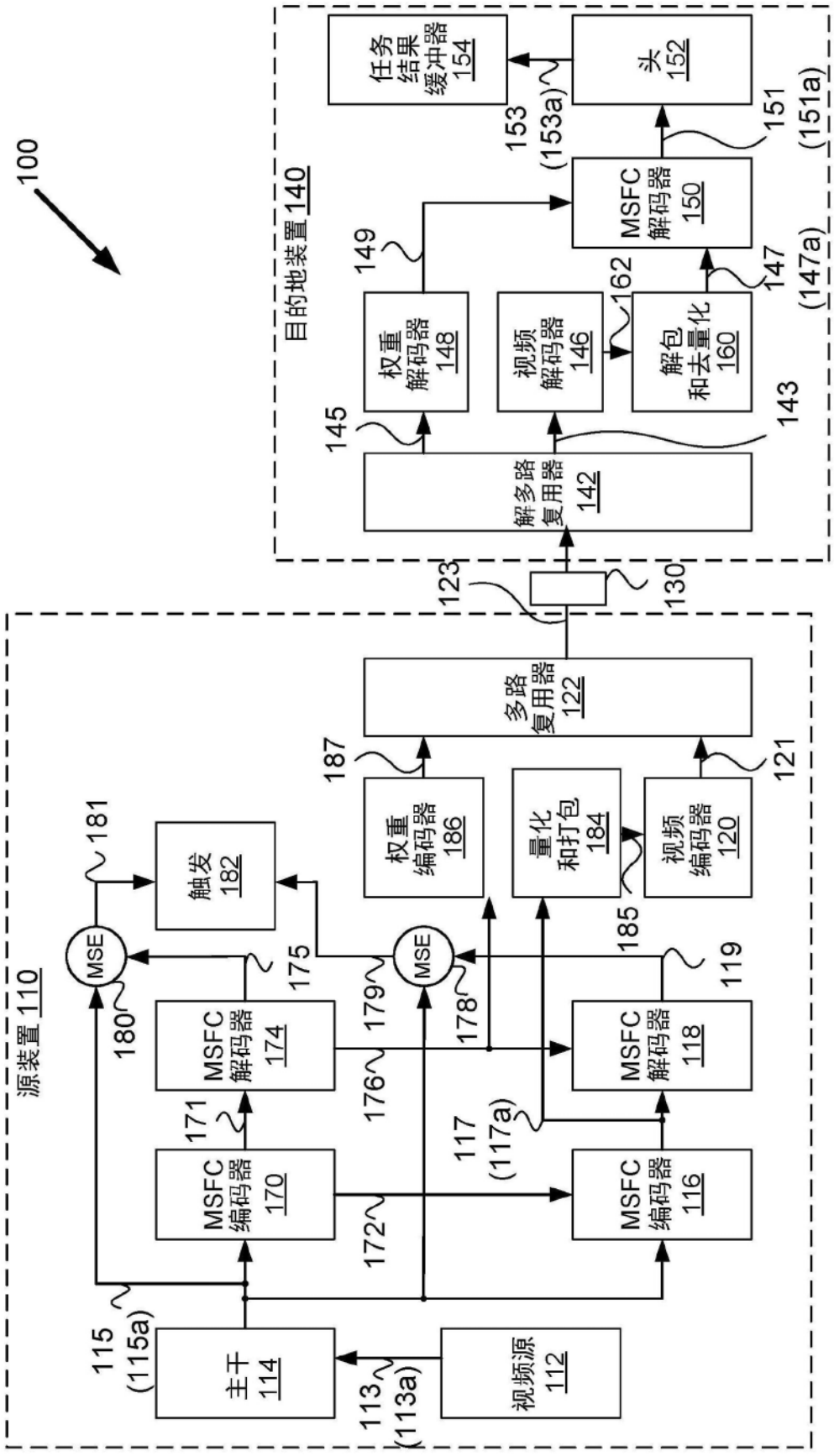

100
目的地装置 140
任务结果缓冲器 154
头 152
153 (153a)
151 (151a)
MSFC 解码器 150
149
147 (147a)
162
权重解码器 148
视频解码器 146
解包和去量化 160
145
143
解多路复用器 142
123
130
多路复用器 122
187
121
权重编码器 186
量化和打包 184
视频编码器 120
185
181
触发 182
MSE
180
175
179
178
119
MSFC 解码器 174
MSFC 解码器 118
176
171
117 (117a)
MSFC 编码器 170
MSFC 编码器 116
172
源装置 110
115 (115a)
主干 114
113 (113a)
视频源 112

图1

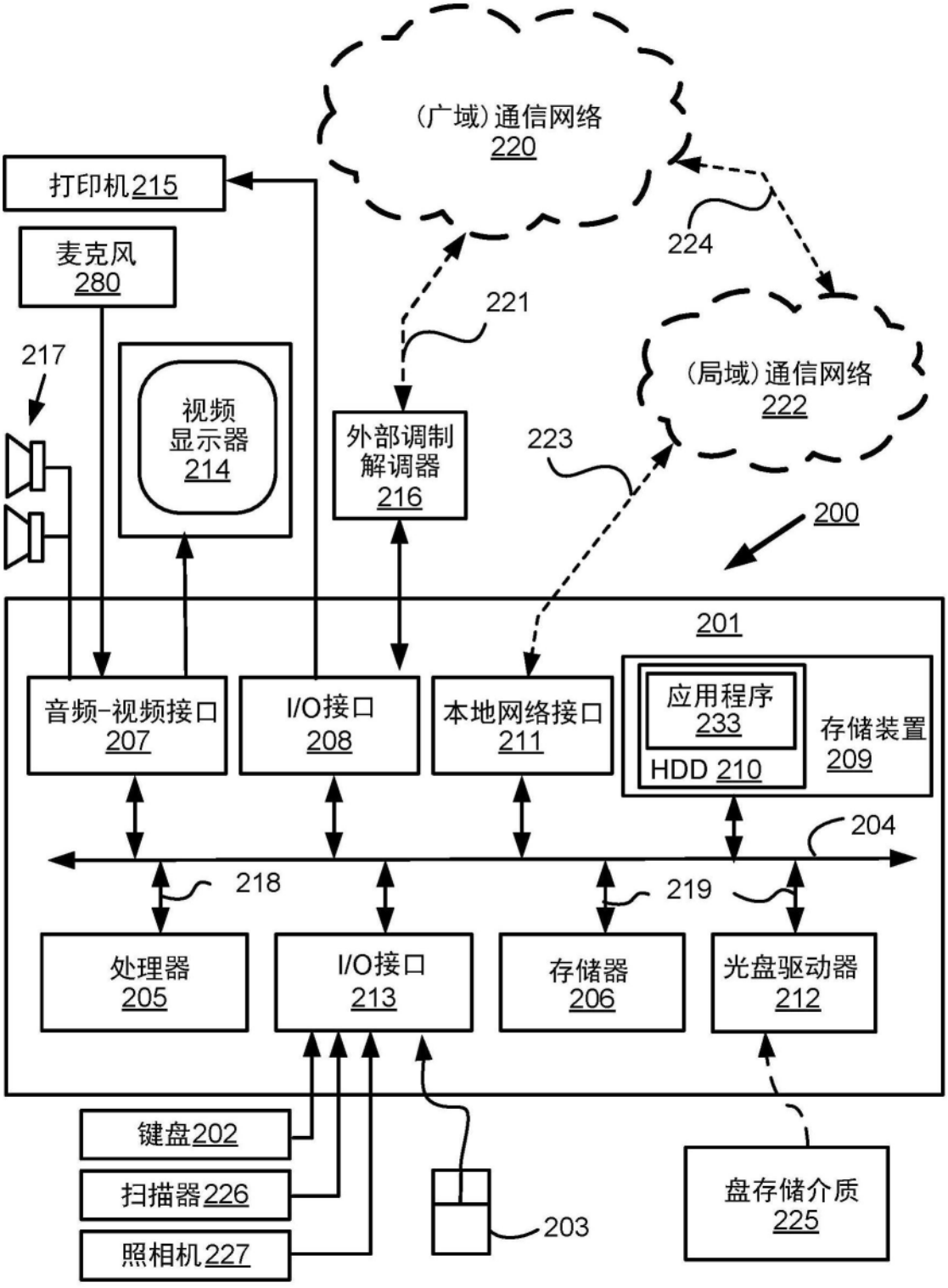

(广域)通信网络
220
打印机215
麦克风
280
224
221
217
视频
显示器
214
外部调制
解调器
216
(局域)通信网络
222
223
200
201
音频-视频接口
207
I/O接口
208
本地网络接口
211
应用程序
233
HDD 210
存储装置
209
204
218
219
处理器
205
I/O接口
213
存储器
206
光盘驱动器
212
键盘202
扫描器226
照相机227
203
盘存储介质
225

图2A

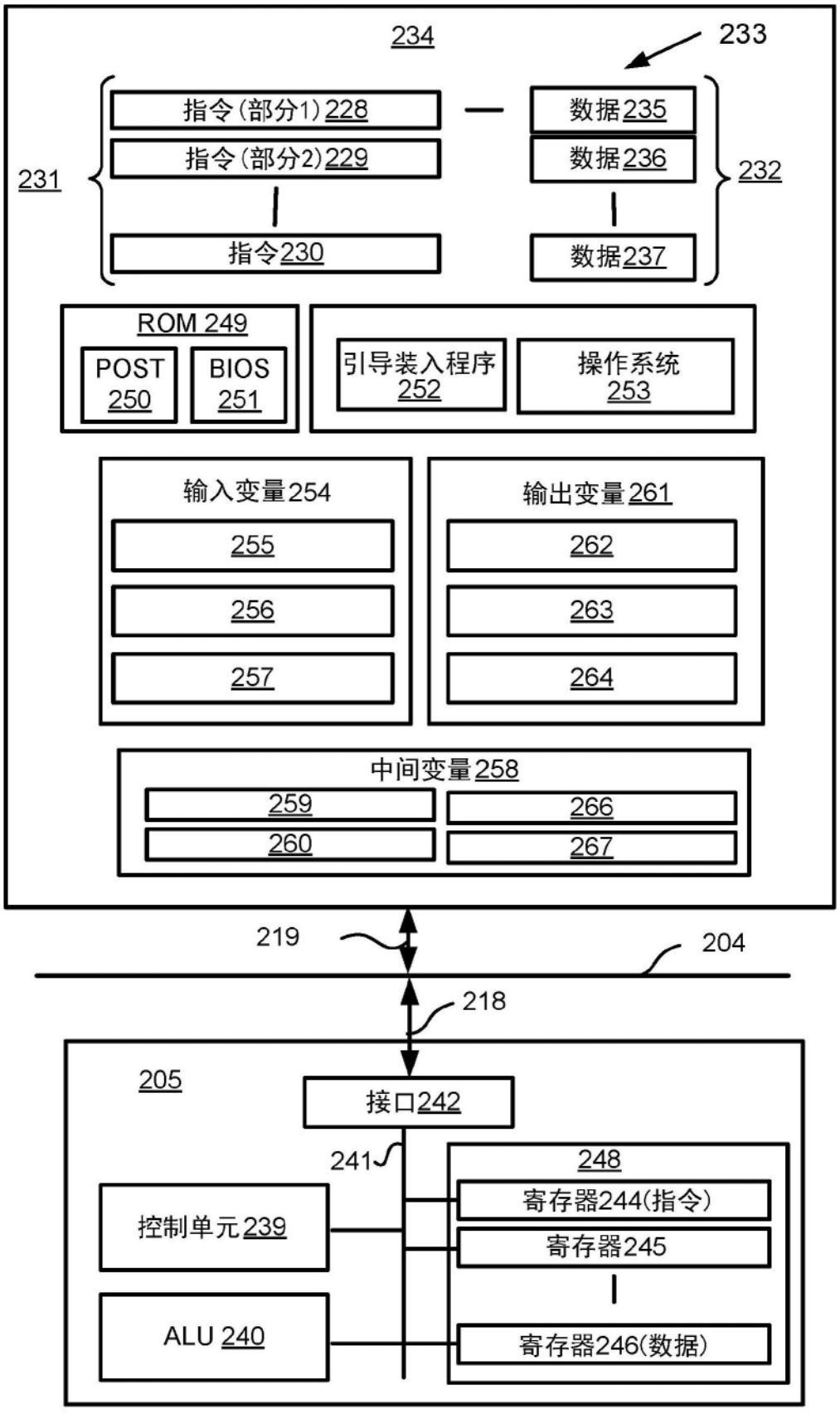

234
233
指令(部分1)228
指令(部分2)229
指令230
数据235
数据236
数据237
231
232
ROM 249
POST 250
BIOS 251
引导装入程序 252
操作系统 253
输入变量254
255
256
257
输出变量261
262
263
264
中间变量258
259
260
266
267
219
204
218
205
接口242
241
248
寄存器244(指令)
寄存器245
寄存器246(数据)
控制单元239
ALU 240

图2B

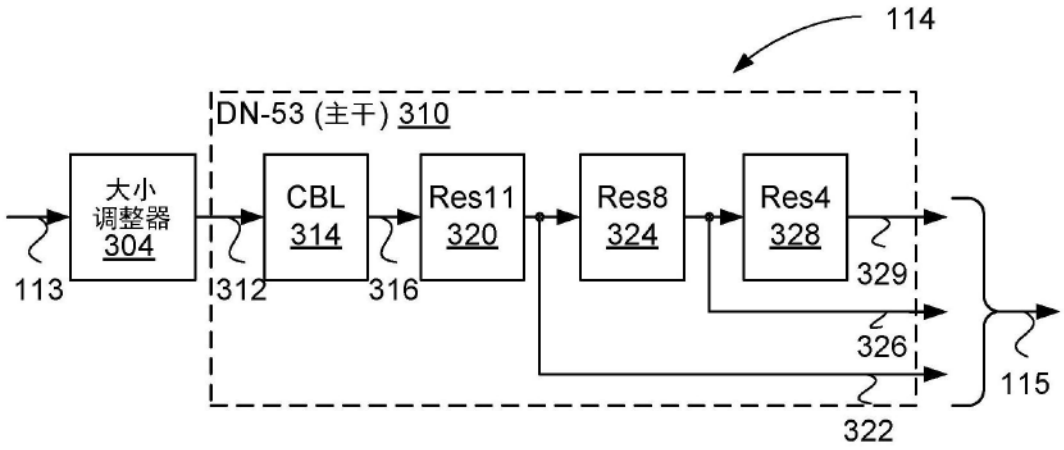


图3A

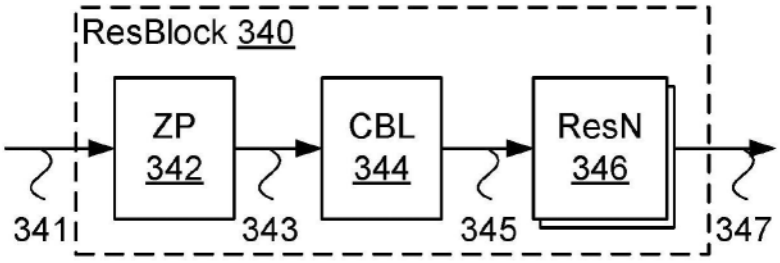


图3B

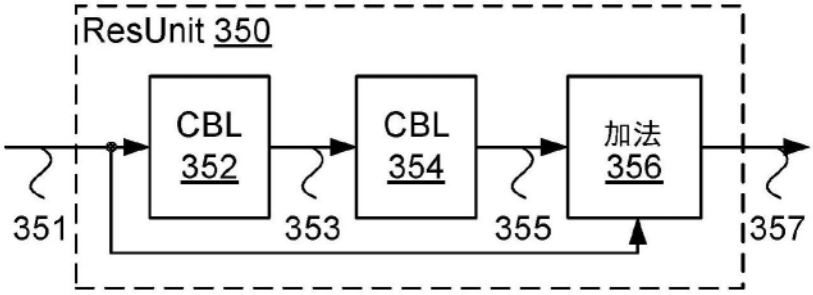


图3C

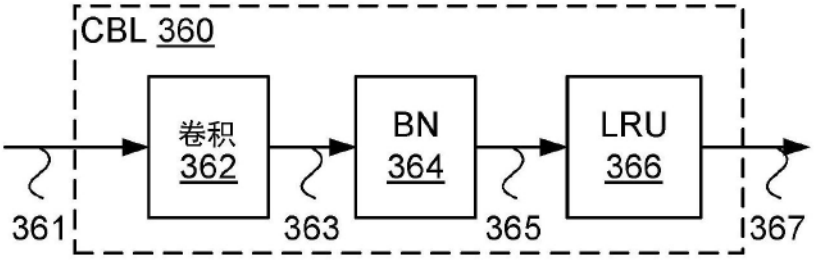


图3D

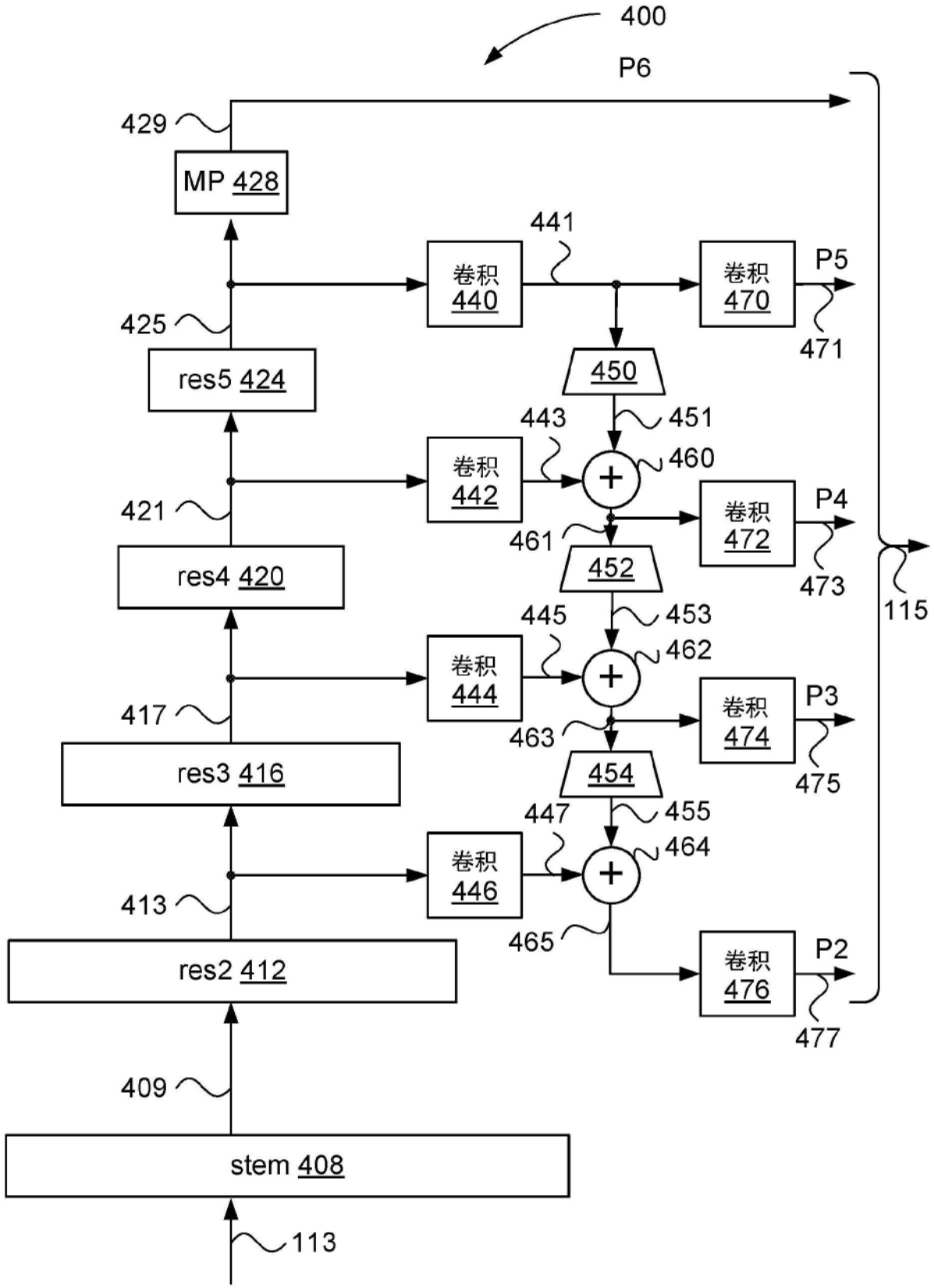

400
P6
429
MP 428
441
卷积 440
卷积 470
P5
471
425
res5 424
450
443
451
卷积 442
460
卷积 472
P4
473
421
461
res4 420
452
445
453
462
卷积 444
卷积 474
P3
475
417
463
res3 416
454
447
455
464
卷积 446
413
465
卷积 476
P2
477
115
res2 412
409
stem 408
113

图4

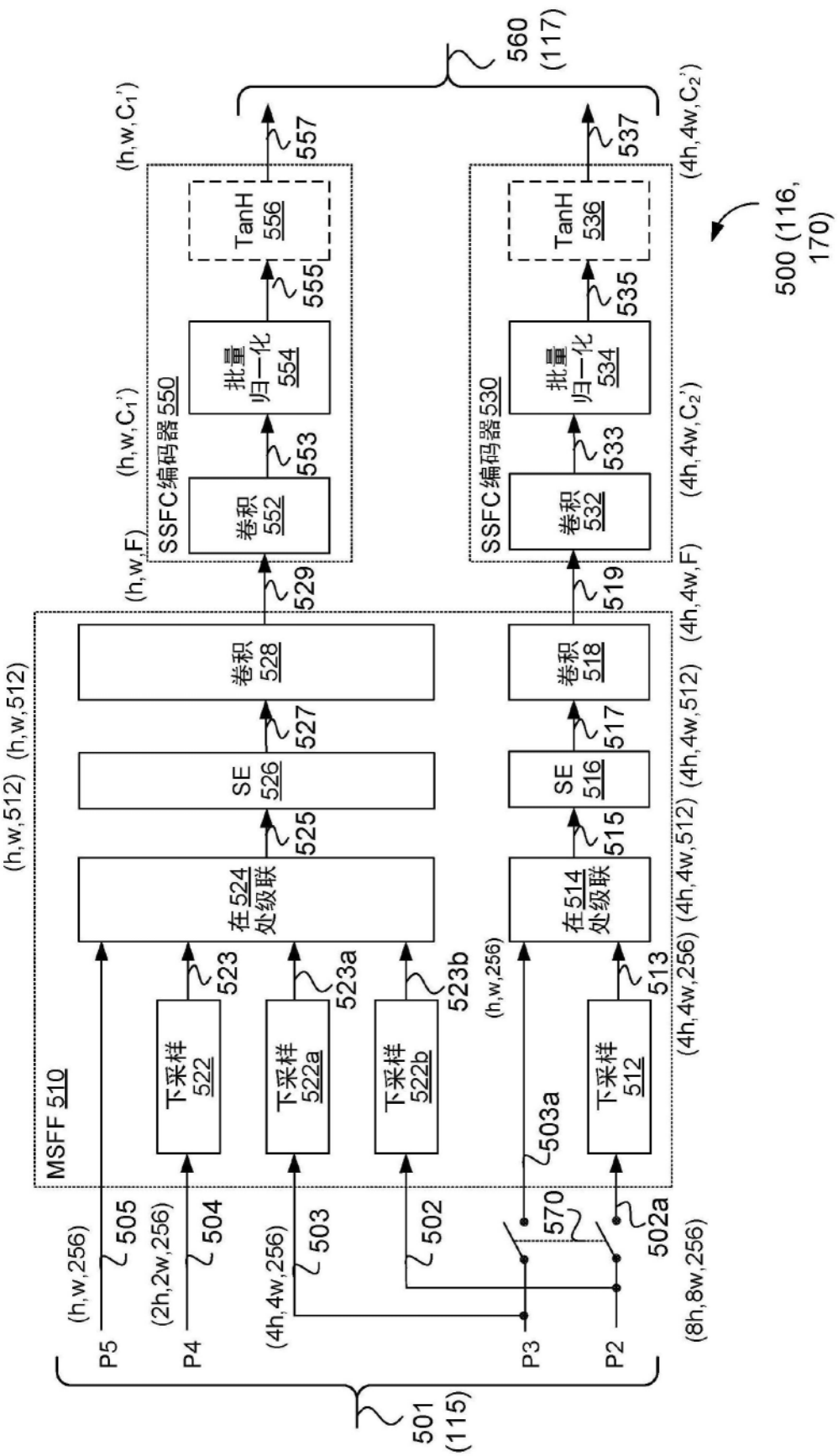

560 (117)
500 (116, 170)
MSFF 510
SSFC编码器 550
SSFC编码器 530
卷积 552
批量归一化 554
TanH 556
卷积 532
批量归一化 534
TanH 536
卷积 528
SE 526
在 524 处级联
卷积 518
SE 516
在 514 处级联
下采样 522
下采样 522a
下采样 522b
下采样 512
501 (115)
P5
P4
P3
P2
505
504
503
502
503a
502a
570
523
523a
523b
513
515
517
519
527
529
553
555
557
535
533
537
(h,w,256)
(2h,2w,256)
(4h,4w,256)
(8h,8w,256)
(h,w,512)
(4h,4w,512)
(h,w,F)
(4h,4w,F)
(h,w,C1')
(4h,4w,C2')

图5

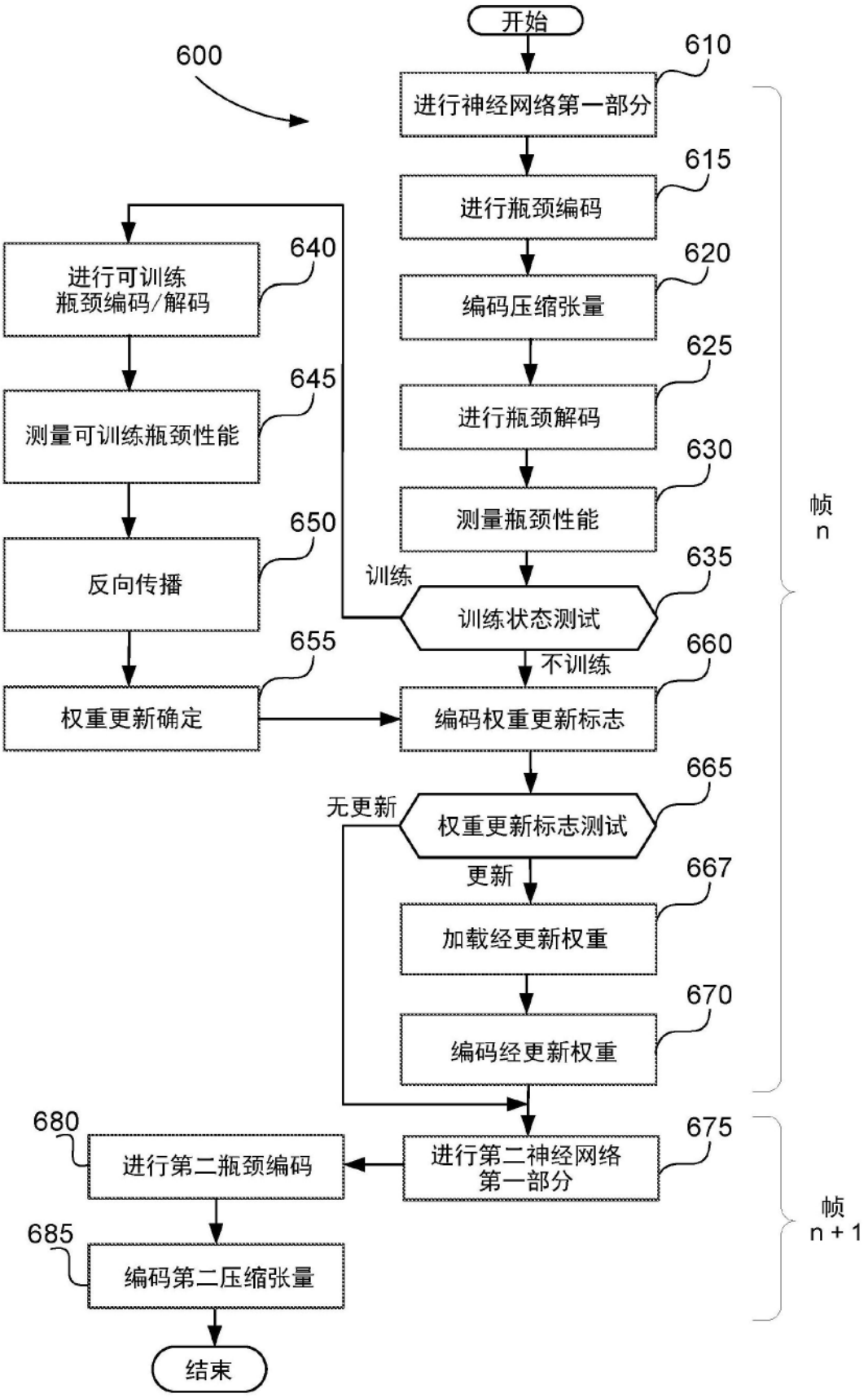

开始
600
610
进行神经网络第一部分
615
进行瓶颈编码
620
编码压缩张量
625
进行瓶颈解码
630
测量瓶颈性能
635
训练状态测试
训练
不训练
640
进行可训练瓶颈编码/解码
645
测量可训练瓶颈性能
650
反向传播
655
权重更新确定
660
编码权重更新标志
665
权重更新标志测试
无更新
更新
667
加载经更新权重
670
编码经更新权重
帧 n
675
进行第二神经网络第一部分
680
进行第二瓶颈编码
685
编码第二压缩张量
帧 n + 1
结束

图6

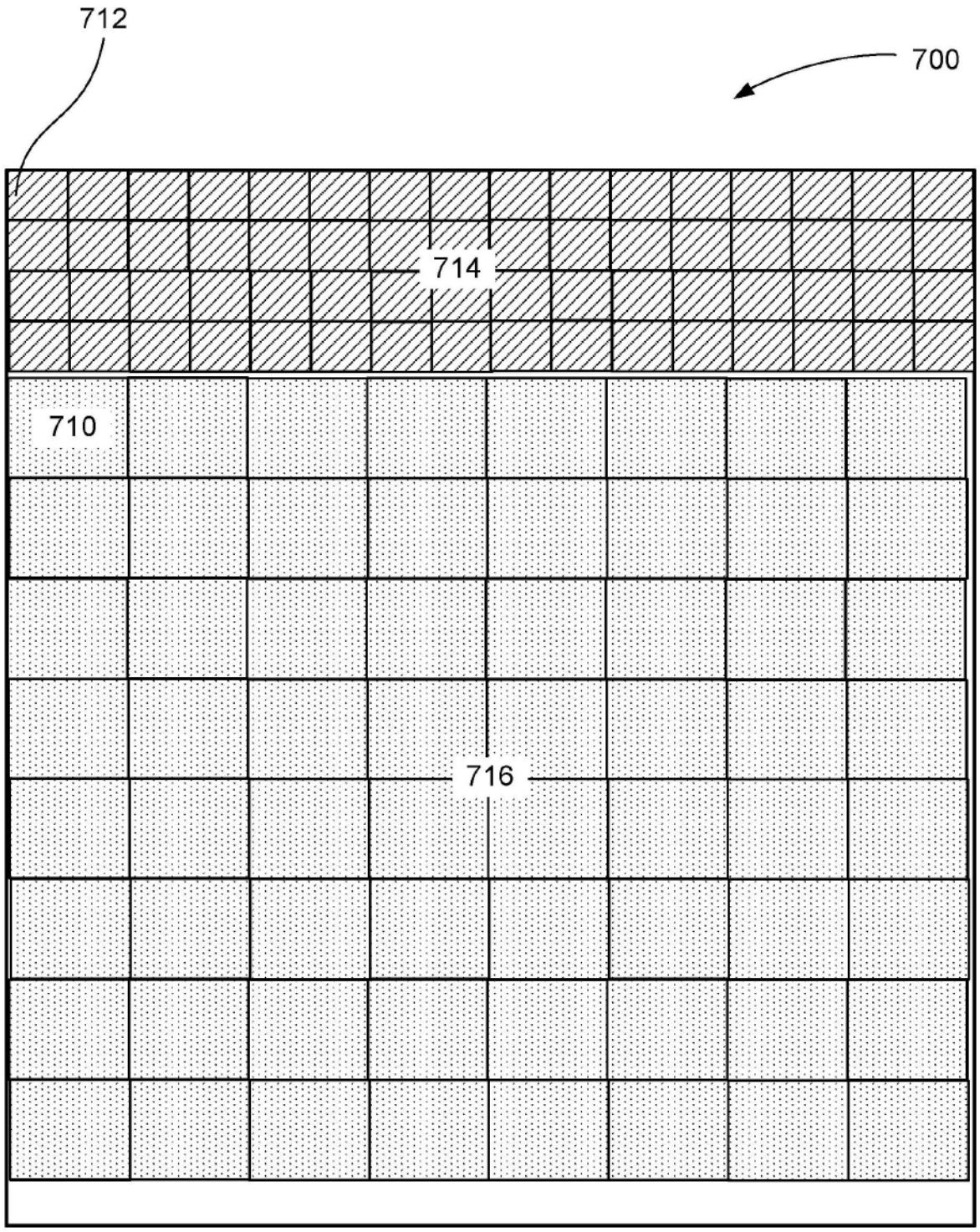


图7A

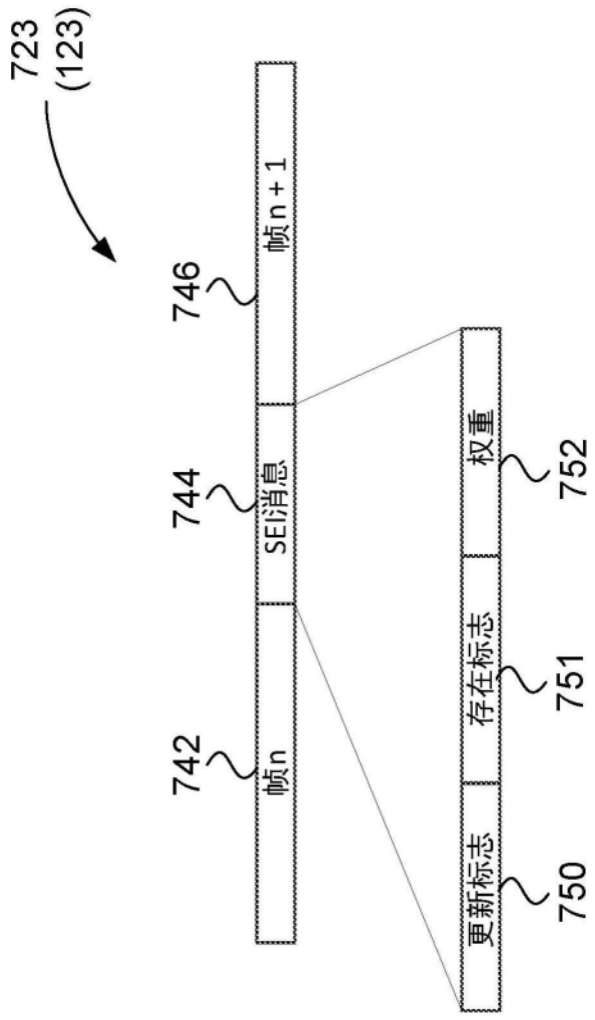


图7B

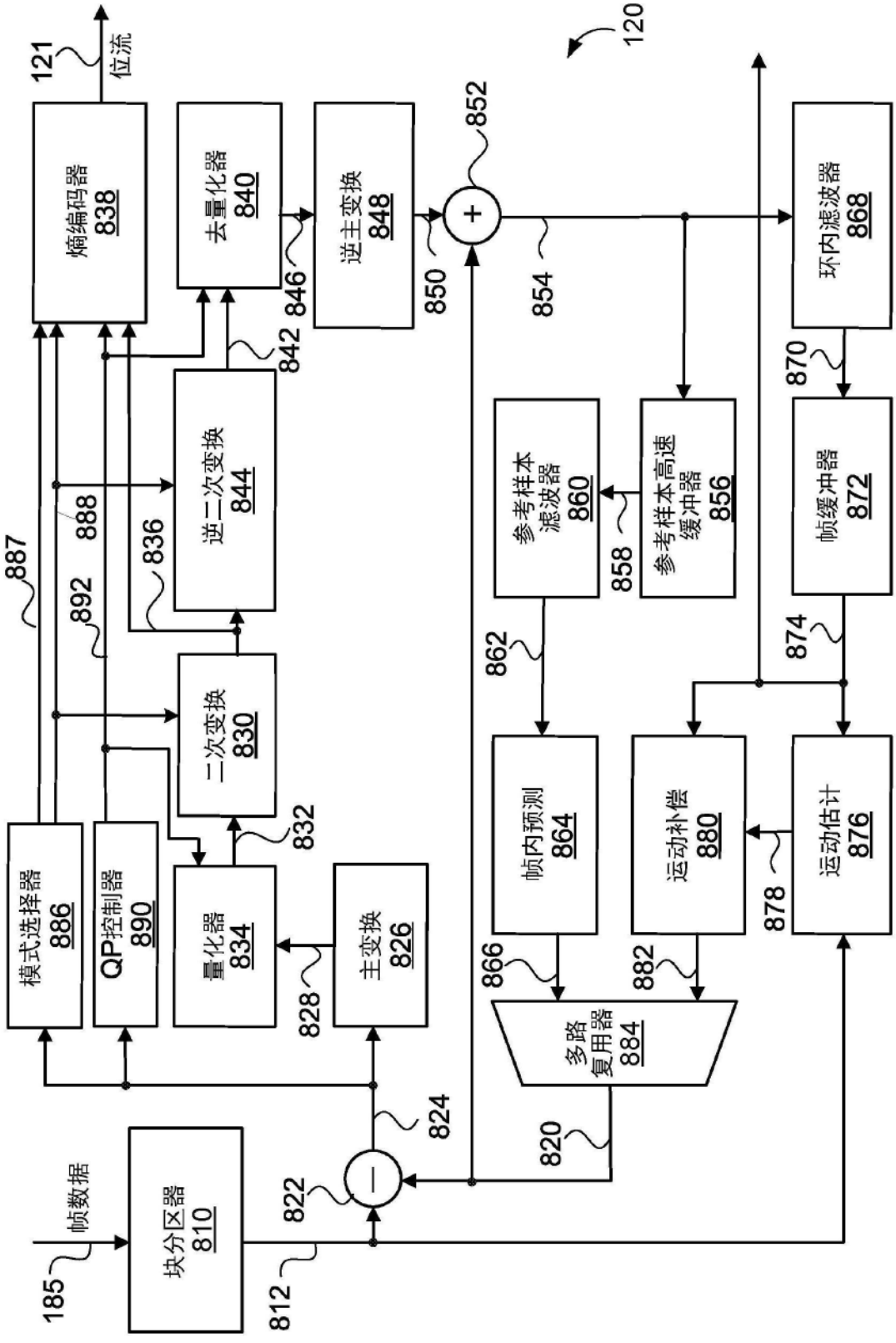

120
121
位流
熵编码器 838
去量化器 840
逆主变换 848
852
850
854
环内滤波器 868
870
842
846
逆二次变换 844
836
888
887
892
参考样本滤波器 860
858
参考样本高速缓冲器 856
帧缓冲器 872
862
874
二次变换 830
832
模式选择器 886
QP控制器 890
量化器 834
主变换 826
828
帧内预测 864
运动补偿 880
运动估计 876
878
866
882
多路复用器 884
824
820
822
帧数据
185
块分区器 810
812

图8

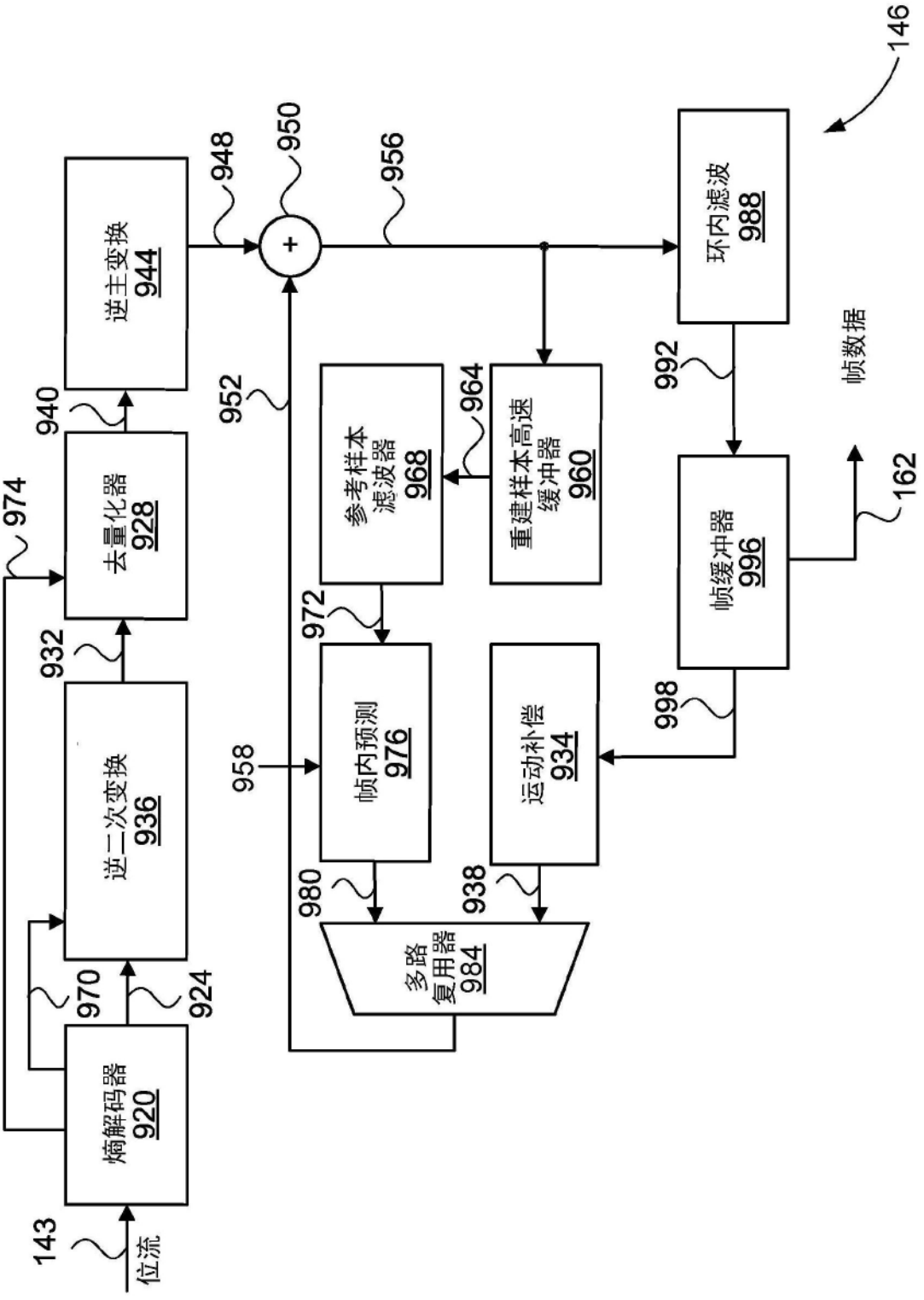

146
逆主变换
944
948
950
956
环内滤波
988
952
参考样本
滤波器
968
964
重建样本高速
缓冲器
960
992
帧数据
940
去量化器
928
974
帧缓冲器
996
162
932
逆二次变换
936
972
帧内预测
976
运动补偿
934
998
958
980
938
多路
复用器
984
970
924
熵解码器
920
143
位流

图9

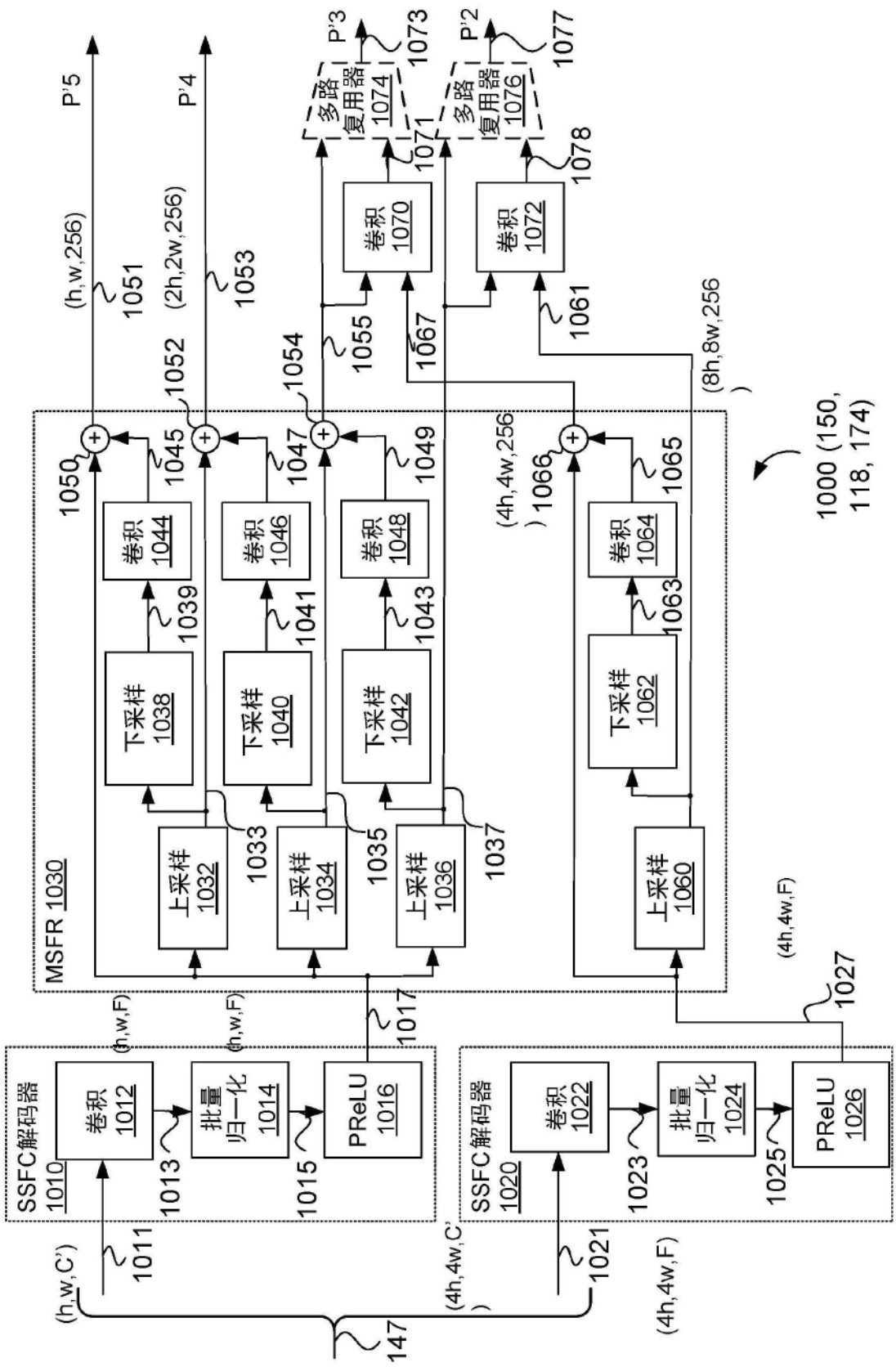

SSFC解码器 1010
卷积 1012
批量归一化 1014
PReLU 1016
SSFC解码器 1020
卷积 1022
批量归一化 1024
PReLU 1026
MSFR 1030
上采样 1032
上采样 1034
上采样 1036
下采样 1038
下采样 1040
下采样 1042
卷积 1044
卷积 1046
卷积 1048
上采样 1060
下采样 1062
卷积 1064
卷积 1070
卷积 1072
多路复用器 1074
多路复用器 1076
1000 (150, 118, 174)
(h,w,C')
(h,w,F)
(4h,4w,C')
(4h,4w,F)
(h,w,256)
(2h,2w,256)
(4h,4w,256)
(8h,8w,256)
P'5
P'4
P'3
P'2

图10

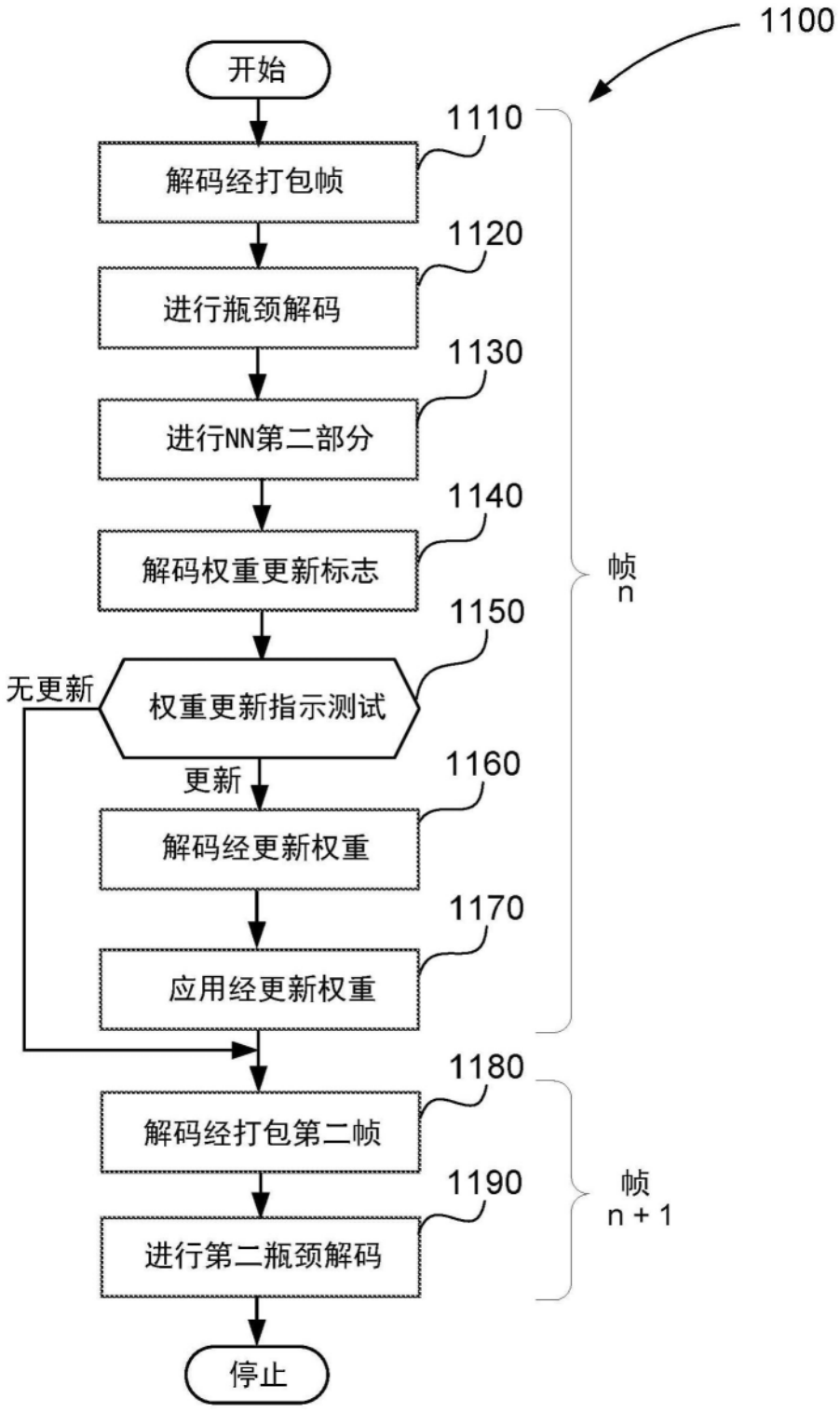

1100
开始
1110
解码经打包帧
1120
进行瓶颈解码
1130
进行NN第二部分
1140
解码权重更新标志
1150
权重更新指示测试
无更新
更新
1160
解码经更新权重
1170
应用经更新权重
帧 n
1180
解码经打包第二帧
1190
进行第二瓶颈解码
帧 n + 1
停止

图11

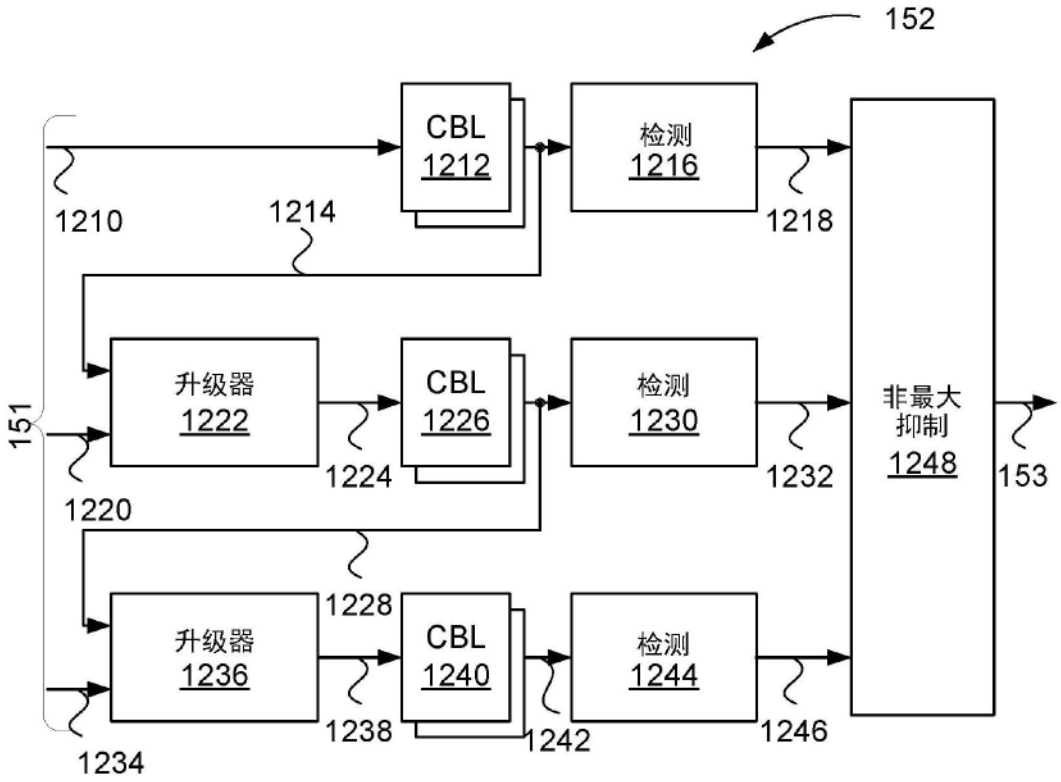


图12A

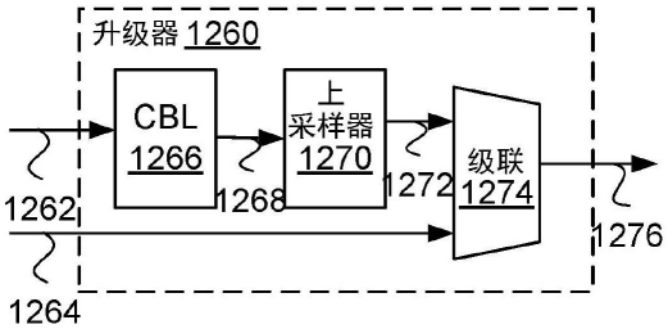


图12B

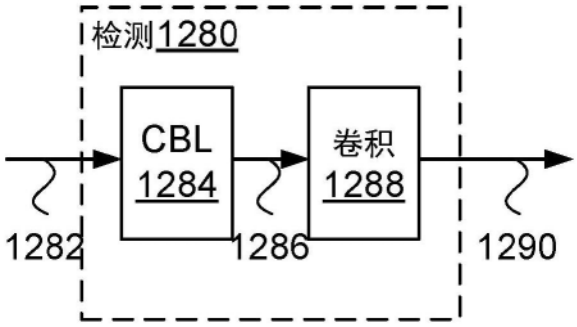


图12C

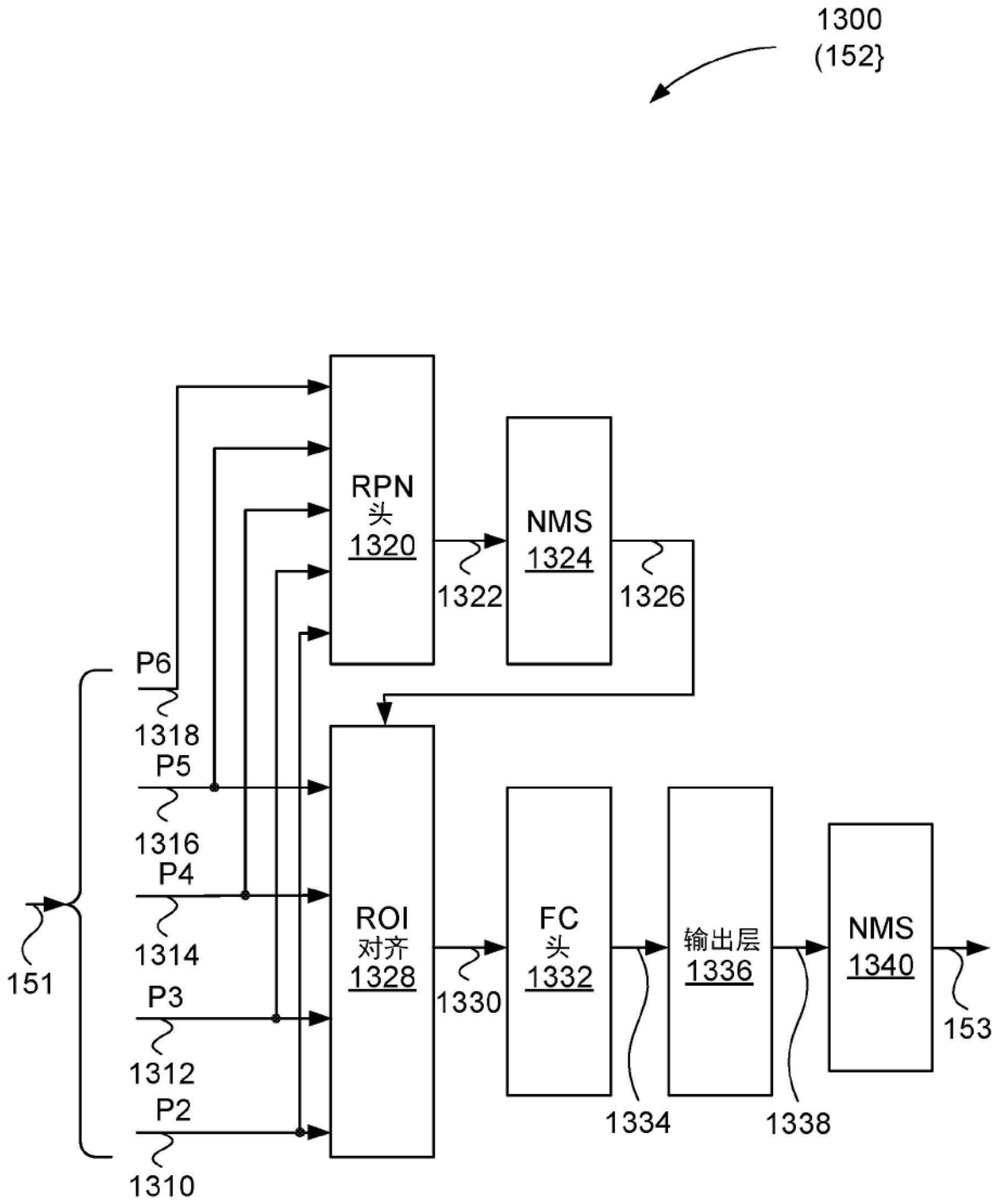

1300
{152}
RPN
头
1320
NMS
1324
1322
1326
P6
1318
P5
1316
P4
1314
P3
1312
P2
1310
151
ROI
对齐
1328
1330
FC
头
1332
1334
输出层
1336
1338
NMS
1340
153

图13

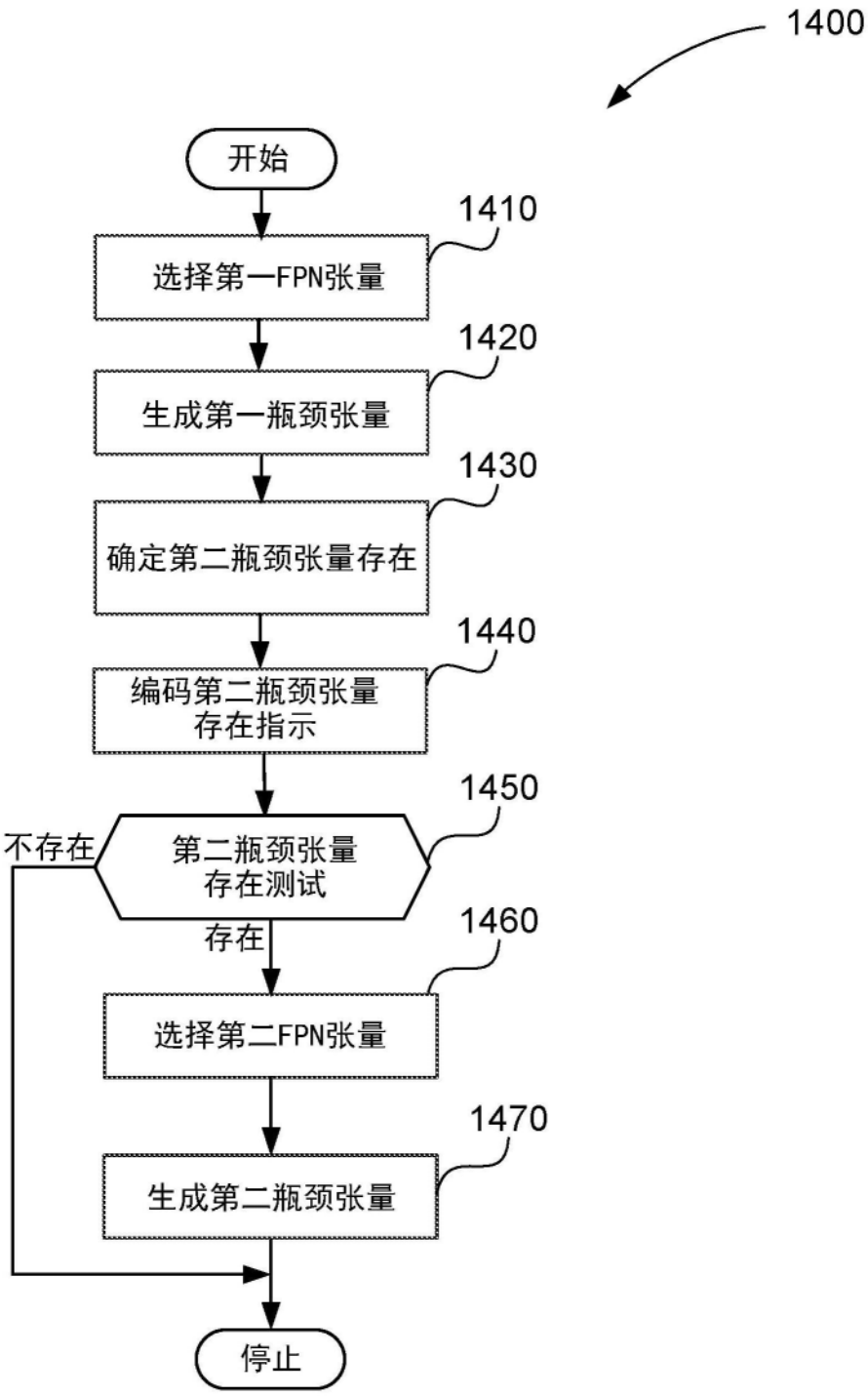

1400
开始
1410
选择第一FPN张量
1420
生成第一瓶颈张量
1430
确定第二瓶颈张量存在
1440
编码第二瓶颈张量
存在指示
1450
第二瓶颈张量
存在测试
不存在
存在
1460
选择第二FPN张量
1470
生成第二瓶颈张量
停止

图14

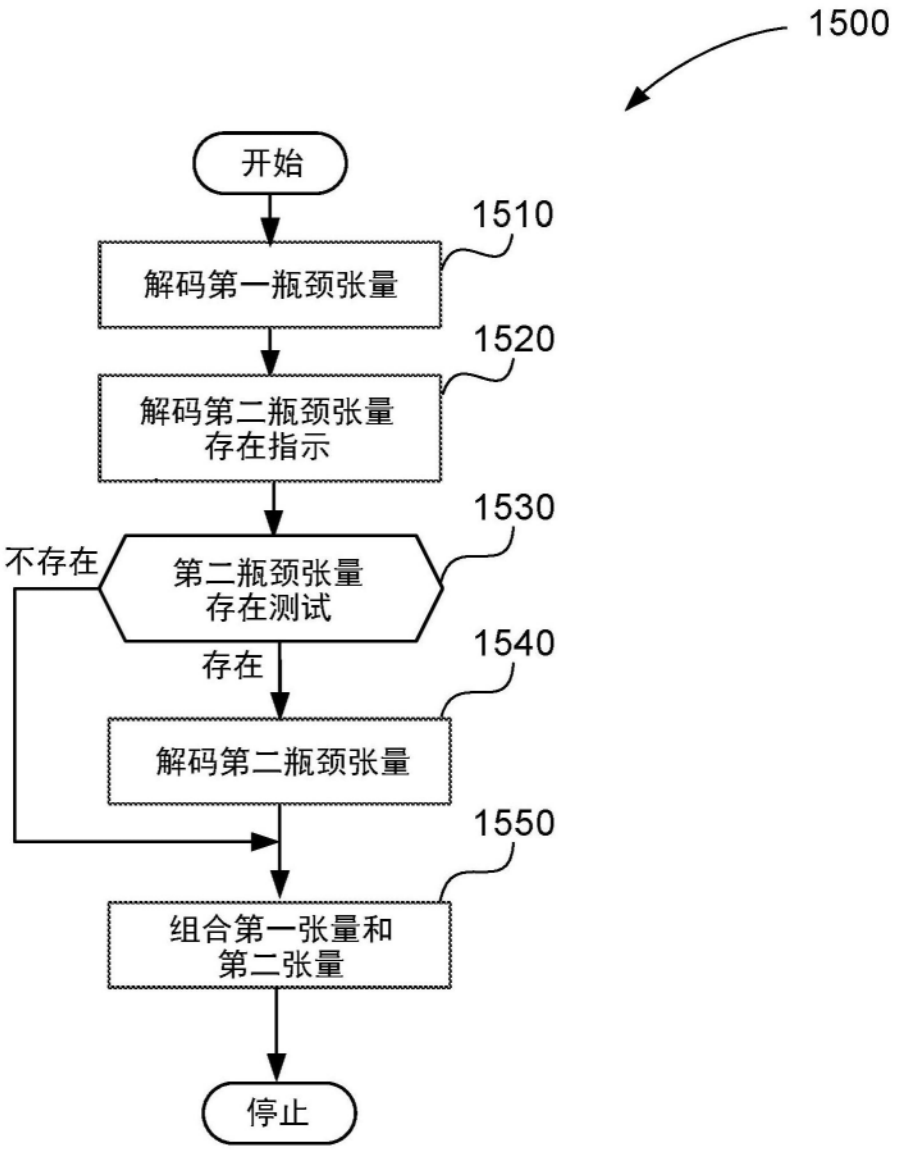

1500
开始
1510
解码第一瓶颈张量
1520
解码第二瓶颈张量
存在指示
1530
不存在
第二瓶颈张量
存在测试
存在
1540
解码第二瓶颈张量
1550
组合第一张量和
第二张量
停止

图15