



(12) 发明专利

(10) 授权公告号 CN 107533669 B

(45) 授权公告日 2021.07.20

(21) 申请号 201680024438.0

(22) 申请日 2016.03.10

(65) 同一申请的已公布的文献号
申请公布号 CN 107533669 A

(43) 申请公布日 2018.01.02

(30) 优先权数据
62/154,089 2015.04.28 US
14/848,283 2015.09.08 US

(85) PCT国际申请进入国家阶段日
2017.10.26

(86) PCT国际申请的申请数据
PCT/US2016/021886 2016.03.10

(87) PCT国际申请的公布数据
W02016/175923 EN 2016.11.03

(73) 专利权人 高通股份有限公司
地址 美国加利福尼亚州

(72) 发明人 R·B·托瓦

(74) 专利代理机构 上海专利商标事务所有限公司
31100
代理人 周敏 陈炜

(51) Int.Cl.
G06N 3/08 (2006.01)
G06N 3/04 (2006.01)

(56) 对比文件
US 2003/0018595 A1,2003.01.23
CN 1529495 A,2004.09.15
Roberto Rigamonti等.《Learning
Separable Filters》.《CVPR 2013》.2013,

审查员 张李一

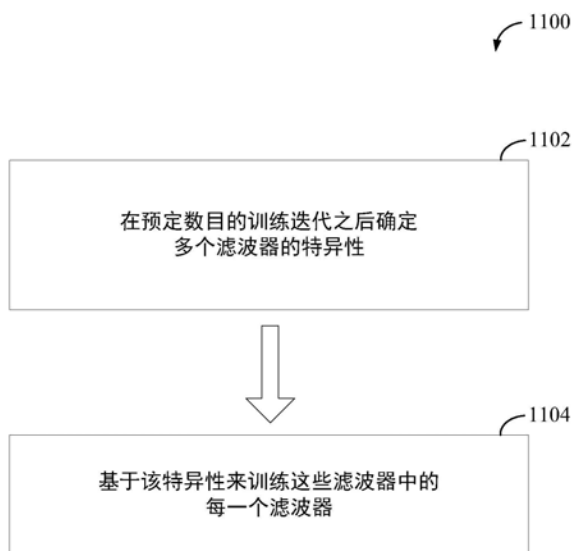
权利要求书3页 说明书13页 附图13页

(54) 发明名称

滤波器特异性作为用于神经网络的训练准则

(57) 摘要

一种训练神经网络模型的方法包括在预定数目的训练迭代之后确定多个滤波器的特异性。该方法还包括基于该特异性来训练这些滤波器中的每一个滤波器。



1. 一种训练神经网络模型的计算机实现的方法,包括:

在预定数目的训练迭代之后基于熵、滤波器的权重从原始的滤波器的权重的改变、方差权重值、每个滤波器的权重与其它滤波器的权重中的每个滤波器的权重的差异、与其它滤波器的互相关、或者其组合中的至少一者来确定多个滤波器中的每一个滤波器的特异性;

至少部分地基于所述特异性来训练所述多个滤波器中的每一个滤波器;以及

基于存储器规范、功率规范或其组合中的至少一者来确定所述神经网络实现的目标复杂度,

其中所确定的目标复杂度被用于以下至少一者:选择性地训练滤波器、对要训练的滤波器进行优先级排序、或者其组合。

2. 如权利要求1所述的计算机实现的方法,其特征在于,进一步包括,至少部分地基于所述特异性确定是否要继续对每一个滤波器的训练。

3. 如权利要求2所述的计算机实现的方法,其特征在于,进一步包括,在所述多个滤波器中的特定滤波器的特异性大于阈值时,停止对所述特定滤波器的训练。

4. 如权利要求2所述的计算机实现的方法,其特征在于,进一步包括,当所述预定数目的训练迭代之后特定滤波器的特异性的改变小于阈值时,停止对所述特定滤波器的训练。

5. 如权利要求2所述的计算机实现的方法,其特征在于,进一步包括,当所述预定数目的训练迭代之后特定滤波器的特异性小于阈值时,从所述神经网络模型中消除所述特定滤波器。

6. 如权利要求5所述的计算机实现的方法,其特征在于,进一步包括,在消除所述特定滤波器之后继续对所述神经网络模型的训练。

7. 如权利要求1所述的计算机实现的方法,其特征在于,在误差函数用所述多个滤波器的特异性的池化测量扩增时训练所述神经网络模型。

8. 如权利要求1所述的计算机实现的方法,其特征在于,进一步包括:

至少部分地基于所述多个滤波器中的每一个滤波器的特异性来对要应用于输入的滤波器进行优先级排序;以及

至少部分地基于所述神经网络模型的目标复杂度来选择数个经优先级排序的滤波器。

9. 一种用于训练神经网络模型的装置,包括:

存储器单元;以及

耦合至所述存储器单元的至少一个处理器,所述至少一个处理器被配置成:

在预定数目的训练迭代之后基于熵、滤波器的权重从原始的滤波器的权重的改变、方差权重值、每个滤波器的权重与其它滤波器的权重中的每个滤波器的权重的差异、与其它滤波器的互相关、或者其组合中的至少一者来确定多个滤波器中的每一个滤波器的特异性;

至少部分地基于所述特异性来训练所述多个滤波器中的每一个滤波器;以及

基于存储器规范、功率规范或其组合中的至少一者来确定所述神经网络实现的目标复杂度,

其中所确定的目标复杂度被用于以下至少一者:选择性地训练滤波器、对要训练的滤波器进行优先级排序、或者其组合。

10. 如权利要求9所述的装置,其特征在於,所述至少一个处理器被进一步配置成至少部分地基于所述特异性确定是否要继续对每一个滤波器的训练。

11. 如权利要求10所述的装置,其特征在於,所述至少一个处理器被进一步配置成在所述多个滤波器中的特定滤波器的特异性大于阈值时,停止对所述特定滤波器的训练。

12. 如权利要求10所述的装置,其特征在於,所述至少一个处理器被进一步配置成当所述预定数目的训练迭代之后特定滤波器的特异性的改变小于阈值时,停止对所述特定滤波器的训练。

13. 如权利要求10所述的装置,其特征在於,所述至少一个处理器被进一步配置成当所述预定数目的训练迭代之后特定滤波器的特异性小于阈值时,从所述神经网络模型中消除所述特定滤波器。

14. 如权利要求13所述的装置,其特征在於,所述至少一个处理器被进一步配置成在消除所述特定滤波器之后继续对所述神经网络模型的训练。

15. 如权利要求9所述的装置,其特征在於,所述至少一个处理器被进一步配置成在误差函数用所述多个滤波器的特异性的池化测量扩增时训练所述神经网络模型。

16. 如权利要求9所述的装置,其特征在於,所述至少一个处理器被进一步配置成:
至少部分地基于所述多个滤波器中的每一个滤波器的特异性来对要应用于输入的滤波器进行优先级排序;以及

至少部分地基于所述神经网络模型的目标复杂度来选择数个经优先级排序的滤波器。

17. 一种训练神经网络模型的设备,包括:

用于在预定数目的训练迭代之后基于熵、滤波器的权重从原始的滤波器的权重的改变、方差权重值、每个滤波器的权重与其它滤波器的权重中的每个滤波器的权重的差异、与其它滤波器的互相关、或者其组合中的至少一者来确定多个滤波器中的每一个滤波器的特异性的装置;

用于至少部分地基于所述特异性来训练所述多个滤波器中的每一个滤波器的装置;以及

用于基于存储器规范、功率规范或其组合中的至少一者来确定所述神经网络实现的目标复杂度的装置,

其中所确定的目标复杂度被用于以下至少一者:选择性地训练滤波器、对要训练的滤波器进行优先级排序、或者其组合。

18. 如权利要求17所述的设备,其特征在於,进一步包括,用于至少部分地基于所述特异性确定是否要继续对每一个滤波器的训练的装置。

19. 如权利要求18所述的设备,其特征在於,进一步包括,用于在所述多个滤波器中的特定滤波器的特异性大于阈值时,停止对所述特定滤波器的训练的装置。

20. 如权利要求18所述的设备,其特征在於,进一步包括,用于当所述预定数目的训练迭代之后特定滤波器的特异性的改变小于阈值时,停止对所述特定滤波器的训练的装置。

21. 如权利要求18所述的设备,其特征在於,进一步包括,用于当所述预定数目的训练迭代之后特定滤波器的特异性小于阈值时,从所述神经网络模型中消除所述特定滤波器的装置。

22. 如权利要求21所述的设备,其特征在於,进一步包括,用于在消除所述特定滤波器

之后继续对所述神经网络模型的训练的装置。

23. 一种用于训练神经网络模型的非瞬态计算机可读介质,所述计算机可读介质具有记录于其上的程序代码,所述程序代码由处理器执行并且包括:

用于在预定数目的训练迭代之后基于熵、滤波器的权重从原始的滤波器的权重的改变、方差权重值、每个滤波器的权重与其它滤波器的权重中的每个滤波器的权重的差异、与其它滤波器的互相关、或者其组合中的至少一者来确定多个滤波器中的每一个滤波器的特异性的程序代码;

用于至少部分地基于所述特异性来训练所述多个滤波器中的每一个滤波器的程序代码;以及

用于基于存储器规范、功率规范或其组合中的至少一者来确定所述神经网络实现的目标复杂度的程序代码,

其中所确定的目标复杂度被用于以下至少一者:选择性地训练滤波器、对要训练的滤波器进行优先级排序、或者其组合。

滤波器特异性作为用于神经网络的训练准则

[0001] 相关申请的交叉引用

[0002] 本申请根据35U.S.C.§119(e)要求于2015年4月28日提交的题为“FILTER SPECIFICITY AS TRAINING CRITERION FOR NEURAL NETWORKS (滤波器特异性作为用于神经网络的训练准则)”的美国临时专利申请No.62/154,089的权益,该临时申请的公开通过援引被明确地整体纳入于此。

技术背景

[0003] 领域

[0004] 本公开的某些方面一般涉及神经系统工程,并且尤其涉及用于使用滤波器特异性作为用于神经网络的训练准则的系统和方法。

[0005] 技术背景

[0006] 可包括一群互连的人工神经元(例如,神经元模型)的人工神经网络是一种计算设备或者表示将由计算设备执行的方法。

[0007] 卷积神经网络是一种前馈人工神经网络。卷积神经网络可包括神经元集合,其中每个神经元具有感受野并且共同地拼出一输入空间。卷积神经网络(CNN)具有众多应用。具体地,CNN已被广泛使用于模式识别和分类领域。

[0008] 深度学习架构(诸如,深度置信网络和深度卷积网络)是分层神经网络架构,其中神经元的第一层的输出变成神经元的第二层的输入,神经元的第二层的输出变成神经元的第三层的输入,以此类推。深度神经网络可被训练以识别特征阶层并因此它们被越来越多地用于对象识别应用。类似于卷积神经网络,这些深度学习架构中的计算可在处理节点群体上分发,其可被配置在一个或多个计算链中。这些多层架构可每次训练一层并可使用反向传播微调。

[0009] 其他模型也可用于对象识别。例如,支持向量机(SVM)是可被应用于分类的学习工具。支持向量机包括对数据进行分类的单独的超平面(例如,决策边界)。该超平面由监督式学习定义。期望的超平面增加训练数据的余裕。换言之,超平面应该具有到训练示例的最大的最小距离。

[0010] 尽管这些解决方案在数个分类基准上取得了优异的结果,但它们的计算复杂度可能极其高。另外,模型的训练是有挑战性的。

[0011] 概述

[0012] 在本公开的一个方面,公开了一种用于训练神经网络模型的方法。该方法包括在预定数目的训练迭代之后确定多个滤波器的特异性。该方法还包括基于该特异性来训练这些滤波器中的每一个滤波器。

[0013] 本公开的另一方面涉及一种设备,其包括用于在预定数目的训练迭代之后确定多个滤波器的特异性的装置。该设备还包括用于基于该特异性来训练这些滤波器中的每一个滤波器的装置。

[0014] 在本公开的另一方面,一种非瞬态计算机可读介质具有记录于其上的非瞬态程序

代码。该程序代码被处理器执行并且包括用于在预定数目的训练迭代之后确定多个滤波器的特异性的程序代码。该程序代码还包括用于基于该特异性来训练这些滤波器中的每一个滤波器的程序代码。

[0015] 本公开的另一方面涉及一种用于训练神经网络模型的装置,该装置具有存储器以及耦合至该存储器的一个或多个处理器。该处理器被配置成在预定数目的训练迭代之后确定多个滤波器的特异性。该处理器还被配置成基于该特异性来训练这些滤波器中的每一个滤波器。

[0016] 本公开的附加特征和优点将在下文描述。本领域技术人员应该领会,本公开可容易被用作修改或设计用于实施与本公开相同的目的的其他结构的基础。本领域技术人员还应认识到,这样的等效构造并不脱离所附权利要求中所阐述的本公开的教导。被认为是本公开的特性的新颖特征在其组织和操作方法两方面连同进一步的目的和优点在结合附图来考虑以下描述时将被更好地理解。然而,要清楚理解的是,提供每一幅附图均仅用于解说和描述目的,且无意作为对本公开的限定的定义。

[0017] 附图简述

[0018] 在结合附图理解下面阐述的详细描述时,本公开的特征、本质和优点将变得更加明显,在附图中,相同附图标记始终作相应标识。

[0019] 图1解说根据本公开的一些方面的使用片上系统(SOC)(包括通用处理器)来设计神经网络的示例实现。

[0020] 图2解说根据本公开的各方面的系统的示例实现。

[0021] 图3A是解说根据本公开的各方面的神经网络的示意图。

[0022] 图3B是解说根据本公开的各方面的示例性深度卷积网络(DCN)的框图。

[0023] 图4是解说根据本公开的各方面的可将人工智能(AI)功能模块化的示例性软件架构的框图。

[0024] 图5是解说根据本公开的各方面的智能手机上AI应用的运行时操作的框图。

[0025] 图6是解说用于后向传播的常规过程的流程图。

[0026] 图7-10解说根据本公开的各方面的各种训练迭代之后的滤波器的示例。

[0027] 图11和12是解说根据本公开的各方面的用于训练滤波器的方法的流程图。

[0028] 详细描述

[0029] 以下结合附图阐述的详细描述旨在作为各种配置的描述,而无意表示可实践本文中所描述的概念的仅有的配置。本详细描述包括具体细节以便提供对各种概念的透彻理解。然而,对于本领域技术人员将显而易见的是,没有这些具体细节也可实践这些概念。在一些实例中,以框图形式示出众所周知的结构和组件以避免湮没此类概念。

[0030] 基于本教导,本领域技术人员应领会,本公开的范围旨在覆盖本公开的任何方面,不论其是与本公开的任何其他方面相独立地还是组合地实现的。例如,可以使用所阐述的任何数目的方面来实现装置或实践方法。另外,本公开的范围旨在覆盖使用作为所阐述的本公开的各个方面的补充或者与之不同的其他结构、功能性、或者结构及功能性来实践的此类装置或方法。应当理解,所披露的本公开的任何方面可由权利要求的一个或多个元素来实施。

[0031] 措辞“示例性”在本文中用于表示“用作示例、实例或解说”。本文中描述为“示例

性”的任何方面不必被解释为优于或胜过其他方面。

[0032] 尽管本文描述了特定方面,但这些方面的众多变体和置换落在本公开的范围之内。虽然提到了优选方面的一些益处和优点,但本公开的范围并非旨在被限定于特定益处、用途或目标。相反,本公开的各方面旨在能宽泛地应用于不同的技术、系统配置、网络和协议,其中一些作为示例在附图以及以下对优选方面的描述中解说。详细描述和附图仅仅解说本公开而非限定本公开,本公开的范围由所附权利要求及其等效技术方案来定义。

[0033] 对于神经网络和/或其它系统而言,滤波器可被用来确定输入的内容。例如,可向音频输入应用滤波器以确定是否存在特定语音。在另一示例中,可向图像应用滤波器以确定该图像中是否存在特定物体,诸如汽车。在一些情形中,当训练集较小、模型过大、和/或训练集包括误差时,滤波器的质量可被降低。具有降低质量的滤波器可被称为弱滤波器。弱滤波器或者部分学习的滤波器可能不提供被指定以确定输入的内容的区别性基础。

[0034] 因此,本公开的各方面涉及改进滤波器的质量。具体而言,在一种配置中,在训练神经网络模型时,在预定数目的训练迭代之后确定一个或多个滤波器的特异性。此外,在该配置中,网络基于该特异性确定是否要继续训练每个滤波器。

[0035] 图1解说了根据本公开的某些方面使用片上系统(SOC) 100进行前述的网络训练的示例实现,SOC 100可包括通用处理器(CPU)或多核通用处理器(CPU) 102。与计算设备(例如,具有权重的神经网络)相关联的变量(例如,神经信号和突触权重)、系统参数,延迟,频率槽信息和任务信息可被存储在与神经处理单元(NPU) 108相关联的存储器块中或专用存储器块118中。在通用处理器102处执行的指令可从与CPU 102相关联的程序存储器加载或可从专用存储器块118加载。

[0036] SOC 100还可包括为具体功能定制的附加处理块(诸如图形处理单元(GPU) 104、数字信号处理器(DSP) 106、连通性块110(其可包括第四代长期演进(4G LTE)连通性、无执照Wi-Fi连通性、USB连通性、蓝牙连通性等))以及例如可检测和识别姿势的多媒体处理器112。SOC 100还可包括传感器处理器114、图像信号处理器(ISP)、和/或导航120(其可包括全球定位系统)。SOC可基于ARM指令集。

[0037] SOC 100还可包括为具体功能定制的附加处理块(诸如GPU 104、DSP 106、连通性块110(其可包括第四代长期演进(4G LTE)连通性、无执照Wi-Fi连通性、USB连通性、蓝牙连通性等))以及例如可检测和识别姿势的多媒体处理器112。在一种实现中,NPU实现在CPU、DSP、和/或GPU中。SOC 100还可包括传感器处理器114、图像信号处理器(ISP)、和/或导航120(其可包括全球定位系统)。

[0038] SOC 100可基于ARM指令集。在本公开的一方面,被加载到通用处理器102中的指令可包括用于在预定数目的训练迭代之后确定多个滤波器的特异性的代码。被加载到通用处理器102中的指令还可包括用于至少部分地基于该特异性来训练该多个滤波器中的每一个滤波器的代码。

[0039] 图2解说根据本公开的某些方面的系统200的示例实现。如图2中所解说的,系统200可具有多个可执行本文所描述的方法的各种操作的局部处理单元202。每个局部处理单元202可包括局部状态存储器204和可存储神经网络的参数的局部参数存储器206。另外,局部处理单元202可具有用于存储局部模型程序的局部(神经元)模型程序(LMP)存储器208、用于存储局部学习程序的局部学习程序(LLP)存储器210、以及局部连接存储器212。此外,

如图2中所解说的,每个局部处理单元202可与用于为该局部处理单元的各局部存储器提供配置的配置处理器单元214对接,并且与提供各局部处理单元202之间的路由的路由连接处理单元216对接。

[0040] 深度学习架构可通过学习在每一层中以逐次更高的抽象程度来表示输入、藉此构建输入数据的有用特征表示来执行对象识别任务。以此方式,深度学习解决了传统机器学习的主要瓶颈。在深度学习出现之前,用于对象识别问题的机器学习办法可能严重依赖人类工程设计的特征,或许与浅分类器相结合。浅分类器可以是两类线性分类器,例如,其中可将特征向量分量的加权和与阈值作比较以预测输入属于哪一类。人类工程设计的特征可以是由拥有领域专业知识的工程师针对具体问题领域定制的模式或内核。相反,深度学习架构可学习以表示与人类工程师可能会设计的相似的特征,但它是通过训练来学习的。另外,深度网络可以学习以表示和识别人类可能还没有考虑过的新类型的特征。

[0041] 深度学习架构可以学习特征阶层。例如,如果向第一层呈递视觉数据,则第一层可学习以识别输入流中的相对简单的特征(诸如边)。在另一示例中,如果向第一层呈递听觉数据,则第一层可学习以识别特定频率中的频谱功率。取第一层的输出作为输入的第二层可以学习以识别特征组合,诸如对于视觉数据识别简单形状或对于听觉数据识别声音组合。例如,更高层可学习以表示视觉数据中的复杂形状或听觉数据中的词语。再高层可学习以识别常见视觉对象或口语短语。

[0042] 深度学习架构在被应用于具有自然阶层结构的问题时可能表现特别好。例如,机动车辆的分类可受益于首先学习以识别轮子、挡风玻璃、以及其他特征。这些特征可在更高层以不同方式被组合以识别轿车、卡车和飞机。

[0043] 神经网络可被设计成具有各种连通性模式。在前馈网络中,信息从较低层被传递到较高层,其中给定层中的每个神经元向更高层中的神经元进行传达。如上所述,可在前馈网络的相继层中构建阶层式表示。神经网络还可具有回流或反馈(也被称为自顶向下(top-down))连接。在回流连接中,来自给定层中的神经元的输出可被传达给相同层中的另一神经元。回流架构可有助于识别跨越不止一个按顺序递送给该神经网络的输入数据组块的模式。从给定层中的神经元到较低层中的神经元的连接被称为反馈(或自顶向下)连接。当高层级概念的识别可辅助辨别输入的特定低层级特征时,具有许多反馈连接的网络可能是有助益的。

[0044] 参照图3A,神经网络的各层之间的连接可以是全连接的(302)或局部连接的(304)。在全连接网络302中,第一层中的神经元可将它的输出传达给第二层中的每个神经元,从而第二层中的每个神经元将从第一层中的每个神经元接收输入。替换地,在局部连接网络304中,第一层中的神经元可连接至第二层中有限数目的神经元。卷积网络306可以是局部连接的,并且被进一步配置成使得与针对第二层中每个神经元的输入相关联的连接强度被共享(例如,308)。更一般化地,网络的局部连接层可被配置成使得一层中的每个神经元将具有相同或相似的连通性模式,但其连接强度可具有不同的值(例如,310、312、314和316)。局部连接的连通性模式可能在更高层中产生空间上相异的感受野,这是由于给定区域中的更高层神经元可接收到通过训练被调谐为到网络的总输入的受限部分的性质的输入。

[0045] 局部连接的神经网络可能非常适合于其中输入的空间位置有意义的问题。例如,

被设计成识别来自车载相机的视觉特征的网络300可发展具有不同性质的高层神经元,这取决于它们与图像下部关联还是与图像上部关联。例如,与图像下部相关联的神经元可学习以识别车道标记,而与图像上部相关联的神经元可学习以识别交通信号灯、交通标志等。

[0046] DCN可以用受监督式学习来训练。在训练期间,DCN可被呈递图像(诸如限速标志的经裁剪图像326),并且可随后计算“前向传递(forward pass)”以产生输出322。输出322可以是对应于特征(诸如“标志”、“60”和“100”)的值向量。网络设计者可能希望DCN在输出特征向量中针对其中一些神经元输出高得分,例如与经训练的网络300的输出322中所示的“标志”和“60”对应的那些神经元。在训练之前,DCN产生的输出很可能是不正确的,并且由此可计算实际输出与目标输出之间的误差。DCN的权重可随后被调整以使得DCN的输出得分与目标更紧密地对准。

[0047] 为了调整权重,学习算法可为权重计算梯度向量。该梯度可指示在权重被略微调整情况下误差将增加或减少的量。在顶层,该梯度可直接对应于连接倒数第二层中的活化神经元与输出层中的神经元的权重的值。在较低层中,该梯度可取决于权重的值以及所计算出的较高层的误差梯度。权重可随后被调整以减小误差。这种调整权重的方式可被称为“反向传播”,因为其涉及在神经网络中的“反向传递(backward pass)”。

[0048] 在实践中,权重的误差梯度可能是在少量示例上计算的,从而计算出的梯度近似于真实误差梯度。这种近似方法可被称为随机梯度下降法。随机梯度下降法可被重复,直到整个系统可达成的误差率已停止下降或直到误差率已达到目标水平。

[0049] 在学习之后,DCN可被呈递新图像326并且在网络中的前向传递可产生输出322,其可被认为是该DCN的推断或预测。

[0050] 深度置信网络(DBN)是包括多层隐藏节点的概率性模型。DBN可被用于提取训练数据集的阶层式表示。DBN可通过堆叠多层受限波尔兹曼机(RBM)来获得。RBM是一类可在输入集上学习概率分布的人工神经网络。由于RBM可在没有关于每个输入应该被分类到哪个类的信息的情况下学习概率分布,因此RBM经常被用于无监督式学习。使用混合无监督式和受监督式范式,DBN的底部RBM可按无监督方式被训练并且可以用作特征提取器,而顶部RBM可按受监督方式(在来自先前层的输入和目标类的联合分布上)被训练并且可用作分类器。

[0051] 深度卷积网络(DCN)是卷积网络的网络,其配置有附加的池化和归一化层。DCN已在许多任务上达成现有最先进的性能。DCN可使用受监督式学习来训练,其中输入和输出目标两者对于许多典范是已知的并被用于通过使用梯度下降法来修改网络的权重。

[0052] DCN可以是前馈网络。另外,如上所述,从DCN的第一层中的神经元到下一更高层中的神经元群连接跨第一层中的神经元被共享。DCN的前馈和共享连接可被利用于进行快速处理。DCN的计算负担可比例如类似大小的包括回流或反馈连接的神经网络小得多。

[0053] 卷积网络的每一层的处理可被认为是空间不变模版或基础投影。如果输入首先被分解成多个通道,诸如彩色图像红色、绿色和蓝色通道,那么在该输入上训练的卷积网络可被认为是三维的,其具有沿着该图像的轴的两个空间维度以及捕捉颜色信息的第三维度。卷积连接的输出可被认为在后续层318和320中形成特征图,该特征图(例如,320)中的每个元素从先前层(例如,318)中一定范围的神经元以及从该多个通道中的每一个通道接收输入。特征图中的值可以用非线性(诸如矫正) $\max(0, x)$ 进一步处理。来自毗邻神经元的值可被进一步池化(这对应于降采样)并可提供附加的局部不变性以及维度缩减。还可通过

特征图中神经元之间的侧向抑制来应用归一化,其对应于白化。

[0054] 深度学习架构的性能可随着有更多被标记的数据点变为可用或随着计算能力提高而提高。现代深度神经网络用比仅仅十五年前可供典型研究者使用的计算资源多数千倍的计算资源来例行地训练。新的架构和训练范式可进一步推升深度学习的性能。经矫正的线性单元可减少被称为梯度消失的训练问题。新的训练技术可减少过度拟合 (over-fitting) 并因此使更大的模型能够达成更好的普遍化。封装技术可抽象出给定的感受野中的数据并进一步提升总体性能。

[0055] 图3B是解说示例性深度卷积网络350的框图。深度卷积网络350可包括多个基于连通性和权重共享的不同类型的层。如图3B所示,该示例性深度卷积网络350包括多个卷积块(例如,C1和C2)。每个卷积块可配置有卷积层、归一化层(LNorm)、和池化层。卷积层可包括一个或多个卷积滤波器,其可被应用于输入数据以生成特征图。尽管仅示出了两个卷积块,但本公开不限于此,而是,根据设计偏好,任何数目的卷积块可被包括在深度卷积网络350中。归一化层可被用于对卷积滤波器的输出进行归一化。例如,归一化层可提供白化或侧向抑制。池化层可提供在空间上的降采样聚集以实现局部不变性和维度缩减。

[0056] 例如,深度卷积网络的平行滤波器组可任选地基于ARM指令集被加载到SOC 100的CPU 102或GPU 104上以达成高性能和低功耗。在替换实施例中,平行滤波器组可被加载到SOC 100的DSP 106或ISP 116上。另外,DCN可访问其他可存在于SOC上的处理块,诸如专用于传感器114和导航120的处理块。

[0057] 深度卷积网络350还可包括一个或多个全连接层(例如,FC1和FC2)。深度卷积网络350可进一步包括逻辑回归(LR)层。深度卷积网络350的每一层之间是要被更新的权重(未示出)。每一层的输出可以用作深度卷积网络350中后续层的输入以从第一卷积块C1处提供的输入数据(例如,图像、音频、视频、传感器数据和/或其他输入数据)学习阶层式特征表示。

[0058] 图4是解说可使人工智能(AI)功能模块化的示例性软件架构400的框图。使用该架构,应用402可被设计成可使得SOC 420的各种处理块(例如CPU 422、DSP 424、GPU 426和/或NPU 428)在该应用402的运行时操作期间执行支持计算。

[0059] AI应用402可配置成调用在用户空间404中定义的功能,例如,这些功能可提供对指示该设备当前操作位置的场景的检测和识别。例如,AI应用402可取决于识别出的场景是办公室、报告厅、餐馆、还是室外环境(诸如湖泊)而以不同方式配置话筒和相机。AI应用402可向与在场景检测应用编程接口(API)406中定义的库相关联的经编译程序代码作出请求以提供对当前场景的估计。该请求可最终依赖于配置成基于例如视频和定位数据来提供场景估计的深度神经网络的输出。

[0060] 运行时引擎408(其可以是运行时框架的经编译代码)可进一步可由AI应用402访问。例如,AI应用402可使得运行时引擎请求特定时间间隔的场景估计或由应用的用户接口检测到的事件触发的场景估计。在使得运行时引擎估计场景时,运行时引擎可进而发送信号给在SOC 420上运行的操作系统410(诸如Linux内核412)。操作系统410进而可使得在CPU 422、DSP 424、GPU 426、NPU 428、或其某种组合上执行计算。CPU 422可被操作系统直接访问,而其他处理块可通过驱动器(诸如用于DSP 424、GPU 426、或NPU 428的驱动器414-418)被访问。在示例性示例中,深度神经网络可被配置成在处理块的组合(诸如CPU 422和GPU

426) 上运行,或可在NPU 428(如果存在的话)上运行。

[0061] 图5是解说智能手机502上的AI应用的运行时操作500的框图。AI应用可包括预处理模块504,该预处理模块504可被配置(例如,使用JAVA编程语言被配置)成转换图像506的格式并随后对该图像进行剪裁和/或调整大小(508)。经预处理的图像可接着被传达给分类应用510,该分类应用510包含场景检测后端引擎512,该场景检测后端引擎512可被配置(例如,使用C编程语言被配置)成基于视觉输入来检测和分类场景。场景检测后端引擎512可被配置成进一步通过缩放(516)和剪裁(518)来预处理(514)该图像。例如,该图像可被缩放和剪裁以使所得到的图像是224像素×224像素。这些维度可映射到神经网络的输入维度。神经网络可由深度神经网络块520配置以使得SOC 100的各种处理块进一步借助深度神经网络来处理图像像素。深度神经网络的结果可随后被取阈(522)并被传递通过分类应用510中的指数平滑块524。经平滑的结果可接着使得智能手机502的设置和/或显示改变。

[0062] 在一种配置中,机器学习模型被配置成用于在预定数目的训练迭代之后确定多个滤波器的特异性;以及至少部分地基于该特异性来训练该多个滤波器中的每一个滤波器。该模型包括确定装置和/或训练装置。在一个方面,确定装置和/或训练装置可以是配置成执行所叙述功能的通用处理器102、与通用处理器102相关联的程序存储器、存储器块118、局部处理单元202、和/或路由连接处理单元216。在另一种配置中,前述装置可以是配置成执行由前述装置所叙述的功能的任何模块或任何设备。

[0063] 根据本公开的某些方面,每个局部处理单元202可被配置成基于模型的一个或多个期望功能特征来确定模型的参数,以及随着所确定的参数被进一步适配、调谐和更新来使这一个或多个功能特征朝着期望的功能特征发展。

[0064] 滤波器特异性作为用于神经网络的训练准则

[0065] 在常规系统中,可指定滤波器来修改和/或增强输入,诸如图像。另外,滤波器可被用来确定输入中是否存在特定元素。例如,滤波器可确定在图像的3x3像素部分中是否存在水平线。因此,系统可通过应用各种类型的滤波器来确定输入中是否存在特定内容。相应地,各种滤波器可被用来对输入进行分类。本公开的各方面关于作为输入的图像来描述。尽管如此,本公开的各方面不限于图像并且构想了其它类型的输入,诸如音频、视频、传感器数据和/或其它输入数据。

[0066] 在常规系统中,卷积可被指定用于图像的线性滤波。具体而言,卷积输出是输入像素的加权和。权重矩阵可被称为卷积内核或滤波器。卷积可通过线性化图像和线性化滤波器的矩阵相乘来获得。

[0067] 另外,常规神经网络将后向传播用于训练。后向传播对于每个输入值使用已知输出以计算损耗函数梯度(即,误差)。也就是说,指定后向传播以从网络的可修改权重计算网络的误差的梯度。在大多数情形中,梯度被用来确定减少误差的滤波器的权重。

[0068] 图6解说用于后向传播的流程图600的示例。在该示例中,图像的已知输出(即,标记)是猫。因此,该标记被设为猫并且预测标记应当为猫。此外,猫的图像602被输入到第一卷积层604。第一卷积层604将卷积的结果输出到第二卷积层606。此外,第二卷积层606将卷积的结果输出到第三卷积层608。最终,从第三卷积层608输出预测标记610。当然,本公开的各方面并不限于三个卷积层并且可按需指定更多或更少的卷积层。

[0069] 图像的标记可被表示为向量。例如,基于图6的示例,用于猫和狗的标记可以是诸

如 (1, 0) 之类的向量, 以使得 1 是用于猫的向量并且 0 是用于狗的向量。此外, 作为示例, 基于网络的训练, 网络可关于输入图像为猫具有 50% 的置信度。该置信度可被称为预测概率。也就是说, 关于猫的预测标记输出可以为 0.5。因此, 关于该猫的误差是 1 与 0.5 之差 (即, (1-0.5)) 的函数。此外, 基于训练, 网络可关于输入图像为狗具有 20% 的置信度。因此, 关于狗的预测标记输出可以为 0.2。因此, 关于狗的误差是 0 与 0.2 之差 (即, (0-0.2)) 的函数。

[0070] 误差函数被用来基于标记与预测标记之差来训练网络。具体而言, 可如下确定误差:

[0071] 误差 = f (标记 - 预测标记) (1)

[0072] 误差函数基于以输入图像为基础的网络置信度输出 (output)。此外, 基于所计算的误差 (诸如, 关于猫的 0.5), 滤波器的权重根据下式来调整:

$$[0073] \quad \frac{\partial \text{error}}{\partial w_{ij}} = \frac{\partial \text{error}}{\partial \text{output}_j} \frac{\partial \text{output}_j}{\partial \text{activation_fcn}_j} \frac{\partial \text{activation_fcn}_j}{\partial w_{ij}} \quad (2)$$

$$[0074] \quad \Delta w_{ij} = -\alpha \frac{\partial \text{error}}{\partial w_{ij}} \quad (3)$$

[0075] 式 2 和 3 是常规后向传播等式。在式 2 和 3 中, w_{ij} 是神经元 i 和 j 之间的权重。此外, 如式 2 中所示, 误差 (error) 的偏导数相对于滤波器的权重 (w_{ij}) 来计算。式 2 的右侧基于常规链规则。为了使用梯度下降来更新权重 (w_{ij}), 选择学习速率 α 。如式 3 中所示, 被添加到旧权重 w_{ij} 的权重改变 Δw_{ij} 等于学习速率和梯度的乘积乘以 -1。指定 -1 以更新误差函数的最小值而非最大值的方向。

[0076] 基于式 3, 如果滤波器的权重较高并且该权重对误差作出了贡献, 则该权重与较低且未对误差作出贡献的权重相比可被减小较大量。另外, 基于式 3, 如果误差的偏导数为非零, 则更新滤波器。

[0077] 在大部分情形中, 由于误差函数是全局函数, 因此网络的个体权重被忽视, 只要该误差被减小使得预测标记是准确的。全局函数可被用来生成各种特定滤波器。如先前所讨论的, 当训练集较小、模型过大、和/或训练集包括差错时, 可获得弱滤波器。

[0078] 图 7 解说了弱滤波器集合 702 与强滤波器集合 704 对照。如图 7 所示, 弱滤波器 702 不具有明确的清晰度。例如, 弱滤波器 702 中的每一个被一般化且不具有良好定义的轮廓。相反, 强滤波器 704 的清晰度大于弱滤波器 702 的清晰度, 从而各个线和角是可见的。强滤波器 704 改进了对输入的特定特征的检测, 诸如图像中是否存在一条或多条水平线。

[0079] 作为示例, 网络可被分配在狗和猫之间进行区分的任务。在该示例中, 可能存在有限数目的训练样本或者训练中的误差。例如, 可能仅针对白猫和黑狗的图像来训练滤波器。因此, 在该示例中, 滤波器可因有限的训练样本而被一般化为各种灰色阴影。因此, 在该示例中, 灰度滤波器可在黑狗和白猫之间进行区分。尽管如此, 灰度滤波器可能不在不同颜色的猫和狗 (诸如褐色狗或橙色猫) 之间进行区分。因此, 减轻对例如因有限数目的训练样本和/或训练中的误差而较弱的滤波器的使用是合乎期望的。尽管如此, 训练样本中的误差可能是未知的。例如, 真实世界训练集的偏差可能是未知或受限的。具体而言, 来自移动设备的训练集可被限制为仅仅用户所捕捉的图像。

[0080] 为了减轻因受限训练集和/或训练中的误差而导致的较弱滤波器, 本公开的各方面涉及基于从后向传播计算出的误差来训练网络以及基于滤波器特异性来训练网络。滤波

器特异性指代滤波器的强度。

[0081] 滤波器的特异性可基于各种因素(诸如权重的分布)来计算。在一种配置中,滤波器特异性基于熵 $\sum_i p(w_i) \log_2 p(w_i)$ 来测量。熵是滤波器中像素差异的测量。在另一配置中,滤波器特异性基于从原始值的变化 $\sum_i w_i(t) - w_i(t_0)$ 来测量。在又一配置中,滤波器特异性基于权重值的方差 $\frac{1}{N} \sum_i (w_i - \bar{w})^2$ 来测量。在又一种配置中,滤波器特异性基于与其它滤波器的差异 $\sum_b \sqrt{\sum_i (w_{a,i} - w_{b,i})^2}$ 来测量。在另一配置中,滤波器特异性基于与其它滤波器的互相关来测量。在前述等式中,w是滤波器权重。尽管已列出若干度量,但本公开旨在涵盖任何相似度量。

[0082] 此外,在一种配置中,为滤波器特异性指定误差项。例如,滤波器特异性误差可通过 $\frac{1}{\text{滤波器特异性}}$ 来计算。在该示例中,如果滤波器特异性较低,则误差较大。替换地,如果滤波器特异性较高,则误差较低。

[0083] 因此,基于本公开的各方面,训练被指定为除获得具有改进特异性的滤波器之外还输出正确标记。也就是说,本公开的各方面涉及获得因各个特征而异的滤波器。

[0084] 图8解说了从第一训练迭代(历元1)训练的滤波器800以及第90次训练迭代(历元90)之后相同滤波器800的示例。训练迭代有时可被称为训练轮次。在该示例中,数据集可具有特定数目的图像,诸如一万。训练使用来自数据集的图像以基于权重更新等式(式3)来调整滤波器的权重。可在对来自数据集的特定数目的图像(诸如一百个图像)进行训练之后调整滤波器的权重。因此,在该示例中,基于一万个图像和在一百个图像之后更新权重,权重将在每个历元期间被调整一百次。

[0085] 如图8所示,在第一训练轮次时,每个滤波器具有特定熵。例如,第一滤波器802具有2.006的熵并且第二滤波器804具有2.018的熵。第一训练轮次中的滤波器从低熵到高熵来排序。此外,如图8所示,每个滤波器的熵在第90个训练轮次(历元90)之后被修改。第90个训练轮次中的滤波器从低熵到高熵来排序。应注意,由于历元1和历元90中的滤波器均从低熵到高熵排序,因此相同滤波器在每幅图中不具有相同的位置。也就是说,历元1的第一滤波器808可能是或者可能不是历元90的第一滤波器808。换言之,历元1的第一滤波器802相比于相邻滤波器可具有熵的较大改变,从而历元1的第一滤波器802例如可以是历元90的第11滤波器814。

[0086] 更具体地,在第一训练轮次时,滤波器是随机化的。此外,在某一数目(诸如90)的训练轮次之后,当滤波器不基于特异性来调整时,一些滤波器具有高特异性,而其它滤波器具有低特异性。在该示例中,特异性是基于滤波器的熵来确定的。

[0087] 如图8所示,在约90个训练轮次之后,历元90的滤波器的熵得到增大。仍然,一些滤波器可能在训练轮次之间具有很小改变。例如,历元90的第二滤波器810具有2.494的熵并且历元90的第三滤波器812具有2.584的熵。因此,第二滤波器810和第三滤波器812的熵自初始训练轮次起具有很小的改变。

[0088] 如先前所讨论的,本公开的各方面不限于仅基于熵来确定滤波器的特异性。特异性还可经由其它计算来确定。

[0089] 另外,如先前所讨论的,使精细的滤波器确定对象的具体特征是合乎期望的。例

如,历元90的第11滤波器814具有由较低熵表示的低精细水平。因此,历元90的第11滤波器814可能仅向后续层添加附加噪声。相反,历元90的第46滤波器816具有较高熵并且包括关于水平线的细节。因此,第46滤波器816可确定图像中是否存在水平线。

[0090] 在一些情形中,在减小来自后向传播的误差的同时改进滤波器特异性是合乎期望的。可为后向传播指定联合成本函数以确定来自后向传播的总误差。在一种配置中,联合成本函数是标记误差和滤波器特异性的函数的求和。例如,联合成本函数为(标记误差+1/(滤波器特异性))。也就是说,误差函数用滤波器的特异性的合计测量来扩增。联合成本函数可基于网络中所有滤波器特异性的总计测量或者可以是基于层的。另外,期望减小训练期间的总误差。因此,在一种配置中,如果基于误差计算出的权重导致低滤波器特异性,则改变滤波器权重以改进滤波器特异性。

[0091] 在另一配置中,具有特定特异性的滤波器的训练被终止以降低计算成本。也就是说,具有大于或等于阈值的特异性的滤波器的学习被停止,从而滤波器的权重不再被更新。

[0092] 图9解说了用于训练的滤波器集合900的示例。第33训练轮次(历元33)的滤波器从低熵到高熵来排序。如图9所示,在历元33的滤波器900中,第43滤波器904具有4.2的熵,并且第64滤波器906具有4.1078的熵。此外,如图9所示,历元90的滤波器900基于第90训练轮次从低熵到高熵排序。在历元90的滤波器900中,第38滤波器908具有4.151的熵,并且第57滤波器910具有4.550的熵。第38滤波器908的熵和第57滤波器910的熵均大于阈值。因此,在该示例中,第38滤波器908和第57滤波器910的训练可被终止以降低计算成本。该阈值可以是预定的和/或基于训练结果来调整。

[0093] 附加或替换地,在一种配置中,网络可随时间推移监视滤波器的特异性。作为示例,如图9所示,来自历元33的第43滤波器904和历元90的第38滤波器908相关。另外,如图9所示,历元33的第64滤波器906和历元90的第57滤波器910相关。

[0094] 如图9所示,历元33的第43滤波器904和第64滤波器906的熵尚未被增大或减小超过第33轮次和第90轮次之间的特定阈值。因此,在该示例中,在滤波器的特异性大于阈值后,随时间推移监视特异性的改变。相应地,在特异性的改变随时间推移小于阈值时,滤波器的训练可被终止以降低计算成本。

[0095] 在又一配置中,基于目标网络复杂度来选择滤波器以进行训练。目标网络复杂度可以基于存储器规范(诸如,被指定用于网络的存储器量)和/或功率规范(诸如,被指定用于网络的计算功率量)。在该配置中,可基于所确定的复杂度来训练特定数目的滤波器。该特定数目的滤波器可以在基于滤波器特异性终止对一个或多个滤波器的训练之前被选择或者在基于滤波器特异性终止对一个或多个滤波器的训练之后被选择。

[0096] 附加或替换地,基于目标网络复杂度,网络可对特定滤波器进行优先级排序以继续训练。该优先级排序可以在基于滤波器特异性终止对一个或多个滤波器的训练之前执行或者在基于滤波器特异性终止对一个或多个滤波器的训练之后执行。

[0097] 在另一配置中,在基于滤波器特异性终止对一个或多个滤波器的训练之前或者在基于滤波器特异性终止对一个或多个滤波器的训练之后,滤波器可基于每个滤波器的特异性而被优先级排序。该优先级排序可被用来确定可被应用于输入的滤波器。此外,根据本公开的一个方面,可基于目标网络复杂度从经优先级排序的滤波器中选择数个滤波器。

[0098] 在又一配置中,在预定数目的训练轮次之后,具有小于阈值的特异性的滤波器可

从训练中消除。图10解说了90个训练轮次之后的滤波器集合1000的示例。如图10所示,滤波器基于熵从低到高来排序。在该示例中,头六个滤波器1002-1012具有小于阈值(诸如,3)的熵。因此,由于滤波器1002-1012的熵小于3,因此头六个滤波器1002-1012被从训练中消除。另外,由于弱滤波器可向输入添加附加噪声,因此从训练中移除的滤波器可不被应用于输入。

[0099] 图11解说了训练神经网络模型的方法1100。在框1102,在预定数目的训练迭代之后确定多个滤波器的特异性。在框1104,基于该特异性来训练这些滤波器中的每一个滤波器。

[0100] 图12解说了用于训练神经网络模型的流程图1200。在框1202处,发起对多个滤波器的训练。如先前所讨论的,滤波器可被用来确定输入中是否存在特定元素。此外,可使用后向传播来训练滤波器。另外,在框1204,神经网络模型在预定数目的训练迭代之后确定该多个滤波器的特异性。此外,在框1206,神经网络模型基于所确定的特异性来确定是否要继续训练滤波器。

[0101] 在一种配置中,如果滤波器的所确定特异性大于阈值,则网络停止训练该滤波器(框1208)。附加或替换地,当预定数目的训练迭代之后特定滤波器的特异性的改变小于阈值时,网络停止训练该滤波器(框1208)。在另一配置中,如框1210中所示,当预定数目的训练迭代之后特定滤波器的特异性小于阈值时,从神经网络模型中消除滤波器(框1210)。

[0102] 在另一配置中,可基于所确定的指定继续对滤波器的训练。例如,如果特异性大于第一阈值并小于第二阈值,则训练可继续。附加或替换地,当预定数目的训练迭代之后特定滤波器的特异性的改变大于第一阈值并且小于第二阈值时,网络继续训练该滤波器(框1212)。

[0103] 以上所描述的方法的各种操作可由能够执行相应功能的任何合适的装置来执行。这些装置可包括各种硬件和/或软件组件和/或模块,包括但不限于电路、专用集成电路(ASIC)、或处理器。一般而言,在附图中有解说的操作的场合,那些操作可具有带相似编号的相应配对装置加功能组件。

[0104] 如本文所使用的,术语“确定”涵盖各种各样的动作。例如,“确定”可包括演算、计算、处理、推导、研究、查找(例如,在表、数据库或其他数据结构中查找)、探知及诸如此类。另外,“确定”可包括接收(例如接收信息)、访问(例如访问存储器中的数据)、及类似动作。而且,“确定”可包括解析、选择、选取、确立及类似动作。

[0105] 如本文所使用的,引述一系列项目中的“至少一个”的短语是指这些项目的任何组合,包括单个成员。作为示例,“a、b或c中的至少一个”旨在涵盖:a、b、c、a-b、a-c、b-c、以及a-b-c。

[0106] 结合本公开所描述的各种解说性逻辑框、模块、以及电路可用设计成执行本文所描述功能的通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列信号(FPGA)或其他可编程逻辑器件(PLD)、分立的门或晶体管逻辑、分立的硬件组件或其任何组合来实现或执行。通用处理器可以是微处理器,但在替换方案中,处理器可以是任何市售的处理器、控制器、微控制器、或状态机。处理器还可以被实现为计算设备的组合,例如DSP与微处理器的组合、多个微处理器、与DSP核心协同的一个或多个微处理器、或任何其它此类配置。

[0107] 结合本公开所描述的方法或算法的步骤可直接在硬件中、在由处理器执行的软件模块中、或在这两者的组合中体现。软件模块可驻留在本领域所知的任何形式的存储介质中。可使用的存储介质的一些示例包括随机存取存储器 (RAM)、只读存储器 (ROM)、闪存、可擦除可编程只读存储器 (EPROM)、电可擦除可编程只读存储器 (EEPROM)、寄存器、硬盘、可移动盘、CD-ROM, 等等。软件模块可包括单条指令、或许多条指令, 且可分布在若干不同的代码段上, 分布在不同的程序间以及跨多个存储介质分布。存储介质可被耦合到处理器以使得该处理器能从/向该存储介质读写信息。在替换方案中, 存储介质可以被整合到处理器。

[0108] 本文所公开的方法包括用于实现所描述的方法的一个或多个步骤或动作。这些方法步骤和/或动作可以彼此互换而不会脱离权利要求的范围。换言之, 除非指定了步骤或动作的特定次序, 否则具体步骤和/或动作的次序和/或使用可以改动而不会脱离权利要求的范围。

[0109] 所描述的功能可在硬件、软件、固件或其任何组合中实现。如果以硬件实现, 则示例硬件配置可包括设备中的处理系统。处理系统可以用总线架构来实现。取决于处理系统的具体应用和整体设计约束, 总线可包括任何数目的互连总线和桥接器。总线可将包括处理器、机器可读介质、以及总线接口的各种电路链接在一起。总线接口可用于尤其将网络适配器等经由总线连接至处理系统。网络适配器可用于实现信号处理功能。对于某些方面, 用户接口 (例如, 按键板、显示器、鼠标、操纵杆, 等等) 也可以被连接到总线。总线还可以链接各种其他电路, 诸如定时源、外围设备、稳压器、功率管理电路以及类似电路, 它们在本领域中是众所周知的, 因此将不再进一步描述。

[0110] 处理器可负责管理总线和一般处理, 包括执行存储在机器可读介质上的软件。处理器可用一个或多个通用和/或专用处理器来实现。示例包括微处理器、微控制器、DSP处理器、以及其他能执行软件的电路系统。软件应当被宽泛地解释成意指指令、数据、或其任何组合, 无论是被称作软件、固件、中间件、微代码、硬件描述语言、或其他。作为示例, 机器可读介质可包括随机存取存储器 (RAM)、闪存、只读存储器 (ROM)、可编程只读存储器 (PROM)、可擦式可编程只读存储器 (EPROM)、电可擦式可编程只读存储器 (EEPROM)、寄存器、磁盘、光盘、硬驱动器、或者任何其他合适的存储介质、或其任何组合。机器可读介质可被实施在计算机程序产品中。该计算机程序产品可以包括包装材料。

[0111] 在硬件实现中, 机器可读介质可以是处理系统中与处理器分开的一部分。然而, 如本领域技术人员将容易领会的, 机器可读介质或其任何部分可在处理系统外部。作为示例, 机器可读介质可包括传输线、由数据调制的载波、和/或与设备分开的计算机产品, 所有这些都可由处理器通过总线接口来访问。替换地或补充地, 机器可读介质或其任何部分可被集成到处理器中, 诸如高速缓存和/或通用寄存器文件可能就是这种情形。虽然所讨论的各种组件可被描述为具有特定位置, 诸如局部组件, 但它们也可按各种方式来配置, 诸如某些组件被配置成分布式计算系统的一部分。

[0112] 处理系统可以被配置为通用处理系统, 该通用处理系统具有一个或多个提供处理器功能性的微处理器、以及提供机器可读介质中的至少一部分的外部存储器, 它们都通过外部总线架构与其他支持电路系统链接在一起。替换地, 该处理系统可以包括一个或多个神经元形态处理器以用于实现本文所述的神经元模型和神经系统模型。作为另一替换方案, 处理系统可以用带有集成在单块芯片中的处理器、总线接口、用户接口、支持电路系统、

和至少一部分机器可读介质的专用集成电路 (ASIC) 来实现,或者用一个或多个现场可编程门阵列 (FPGA)、可编程逻辑器件 (PLD)、控制器、状态机、门控逻辑、分立硬件组件、或者任何其他合适的电路系统、或者能执行本公开通篇所描述的各种功能性的电路的任何组合来实现。取决于具体应用和加诸于整体系统上的总设计约束,本领域技术人员将认识到如何最佳地实现关于处理系统所描述的功能性。

[0113] 机器可读介质可包括数个软件模块。这些软件模块包括当由处理器执行时使处理系统执行各种功能的指令。这些软件模块可包括传送模块和接收模块。每个软件模块可以驻留在单个存储设备中或者跨多个存储设备分布。作为示例,当触发事件发生时,可以从硬驱动器中将软件模块加载到RAM中。在软件模块执行期间,处理器可以将一些指令加载到高速缓存中以提高访问速度。随后可将一个或多个高速缓存行加载到通用寄存器文件中以供处理器执行。在以下述及软件模块的功能性时,将理解此类功能性是在处理器执行来自该软件模块的指令时由该处理器来实现的。

[0114] 如果以软件实现,则各功能可作为一条或多条指令或代码存储在计算机可读介质上或藉其进行传送。计算机可读介质包括计算机存储介质和通信介质两者,这些介质包括促成计算机程序从一地到另一地转移的任何介质。存储介质可以是能被计算机访问的任何可用介质。作为示例而非限定,此类计算机可读介质可包括RAM、ROM、EEPROM、CD-ROM或其他光盘存储、磁盘存储或其他磁存储设备、或能用于携带或存储指令或数据结构形式的期望程序代码且能被计算机访问的任何其他介质。另外,任何连接也被正当地称为计算机可读介质。例如,如果软件是使用同轴电缆、光纤电缆、双绞线、数字订户线 (DSL)、或无线技术 (诸如红外 (IR)、无线电、以及微波) 从web网站、服务器、或其他远程源传送而来,则该同轴电缆、光纤电缆、双绞线、DSL或无线技术 (诸如红外、无线电、以及微波) 就被包括在介质的定义之中。如本文中所使用的盘 (disk) 和碟 (disc) 包括压缩碟 (CD)、激光碟、光碟、数字多用碟 (DVD)、软盘、和蓝光[®]碟,其中盘 (disk) 常常磁性地再现数据,而碟 (disc) 用激光来光学地再现数据。因此,在一些方面,计算机可读介质可包括非瞬态计算机可读介质 (例如,有形介质)。另外,对于其他方面,计算机可读介质可包括瞬态计算机可读介质 (例如,信号)。上述的组合应当也被包括在计算机可读介质的范围内。

[0115] 因此,一些方面可包括用于执行本文中给出的操作的计算机程序产品。例如,此类计算机程序产品可包括其上存储 (和/或编码) 有指令的计算机可读介质,这些指令能由一个或多个处理器执行以执行本文中所描述的操作。对于一些方面,计算机程序产品可包括包装材料。

[0116] 此外,应当领会,用于执行本文中所描述的方法和技术的模块和/或其它合适装置能由用户终端和/或基站在适用的场合下载和/或以其他方式获得。例如,此类设备能被耦合至服务器以促成用于执行本文中所描述的方法的装置的转移。替换地,本文所述的各种方法能经由存储装置 (例如,RAM、ROM、诸如压缩碟 (CD) 或软盘等物理存储介质等) 来提供,以使得一旦将该存储装置耦合至或提供给用户终端和/或基站,该设备就能获得各种方法。此外,可利用适于向设备提供本文所描述的方法和技术的任何其他合适的技术。

[0117] 将理解,权利要求并不被限定于以上所解说的精确配置和组件。可在以上所描述的方法和装置的布局、操作和细节上作出各种改动、更换和变形而不会脱离权利要求的范围。

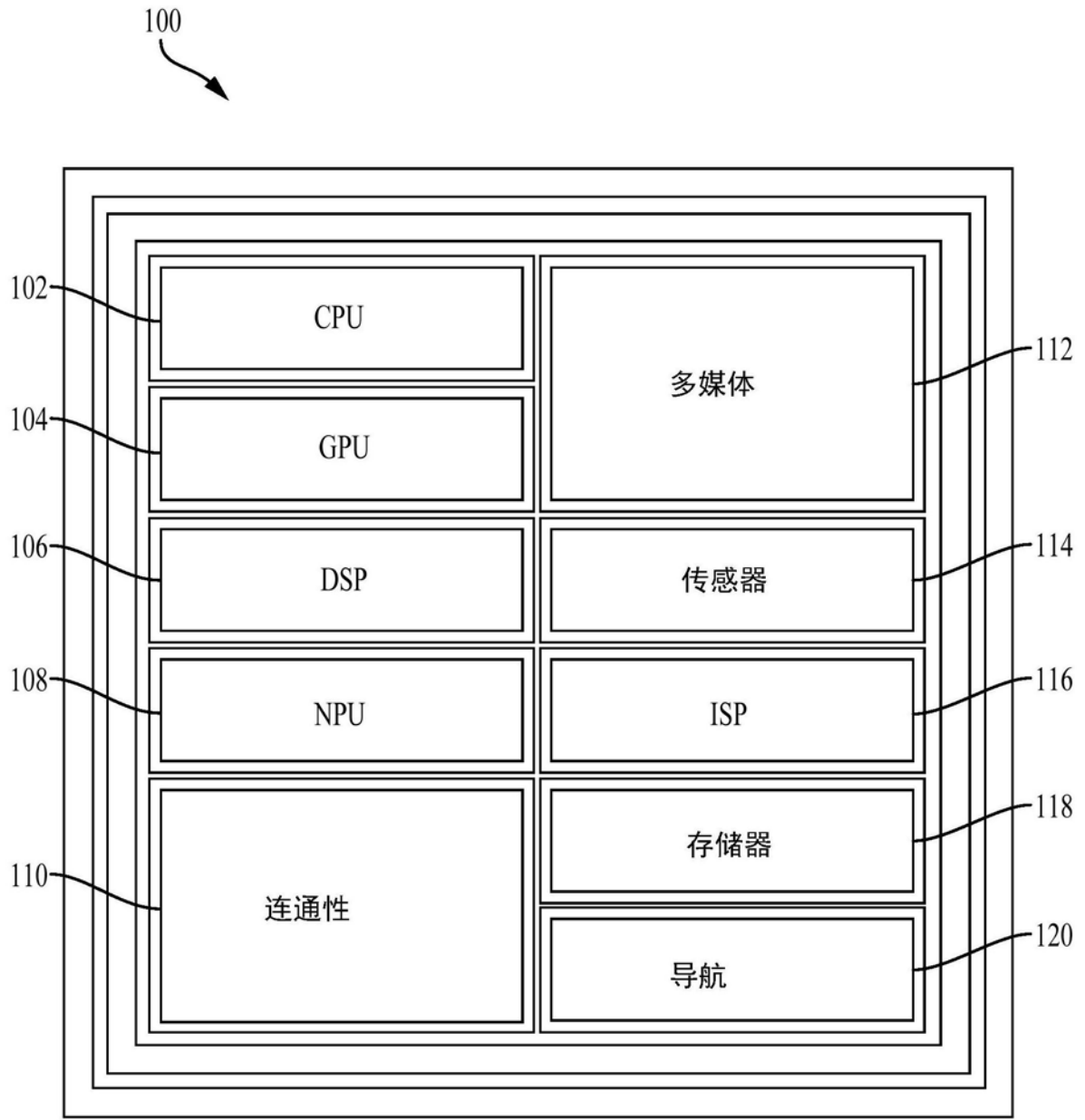


图1

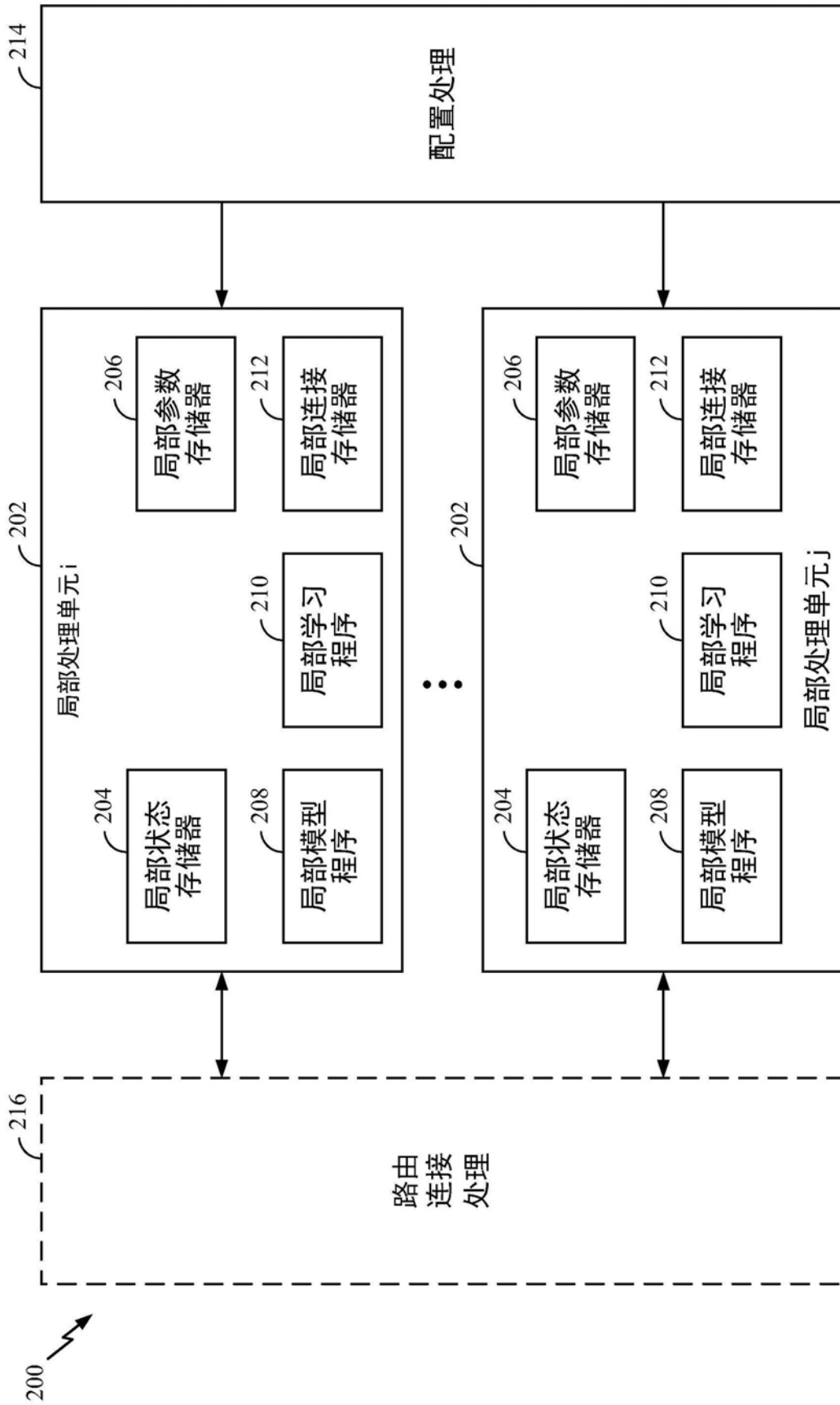


图2

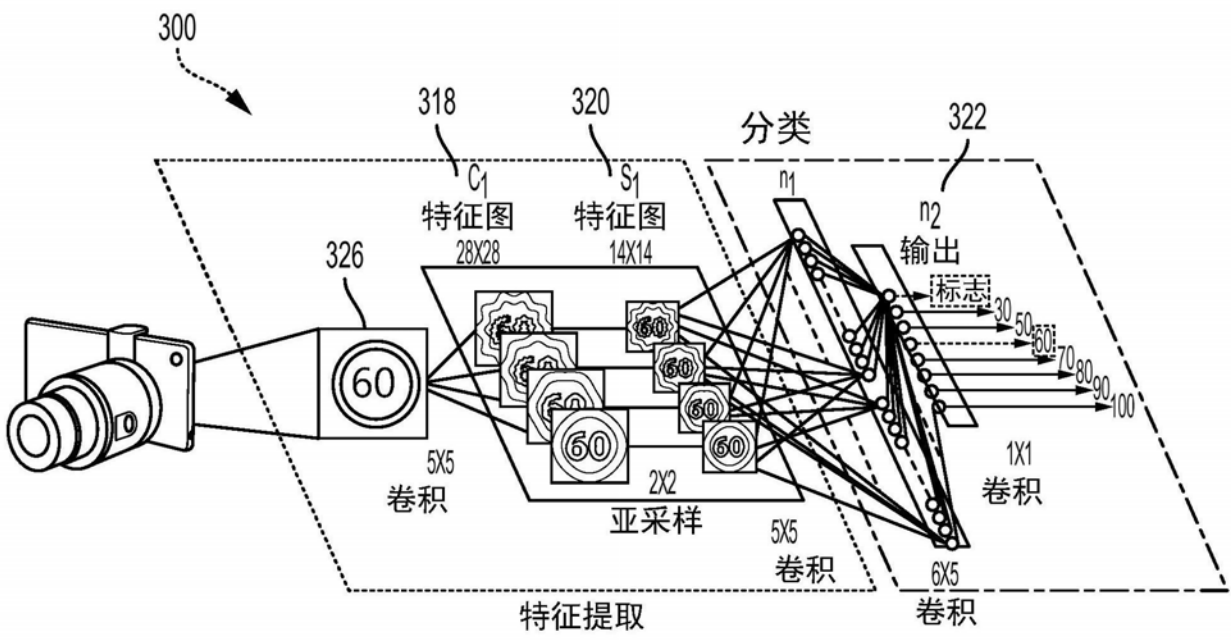
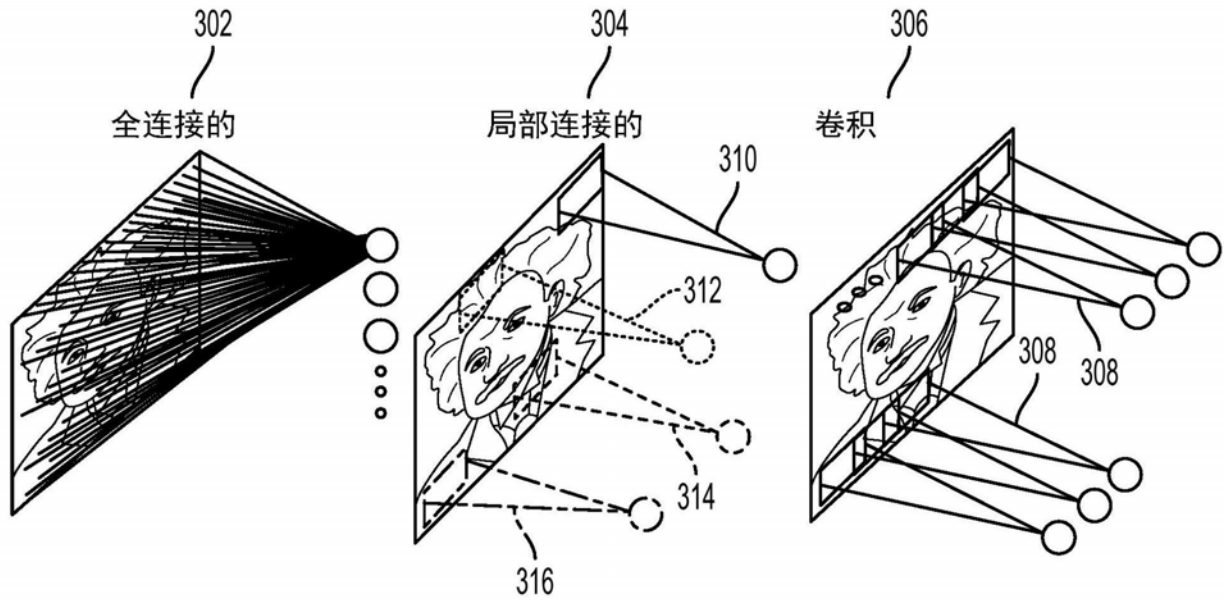


图3A

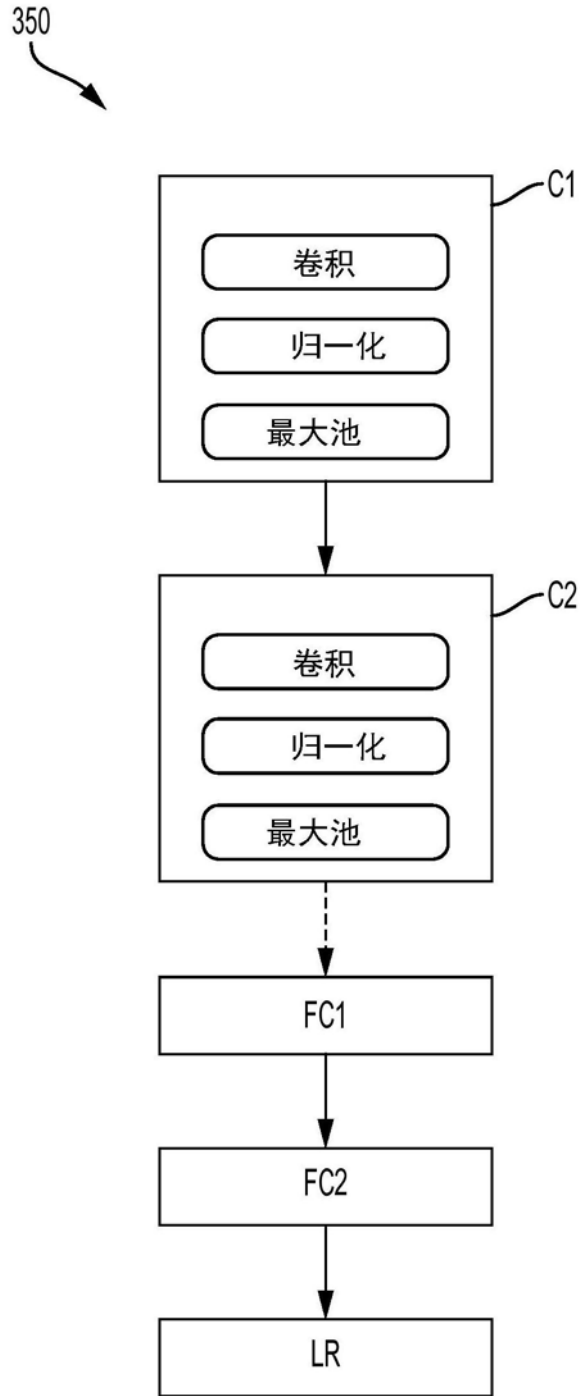
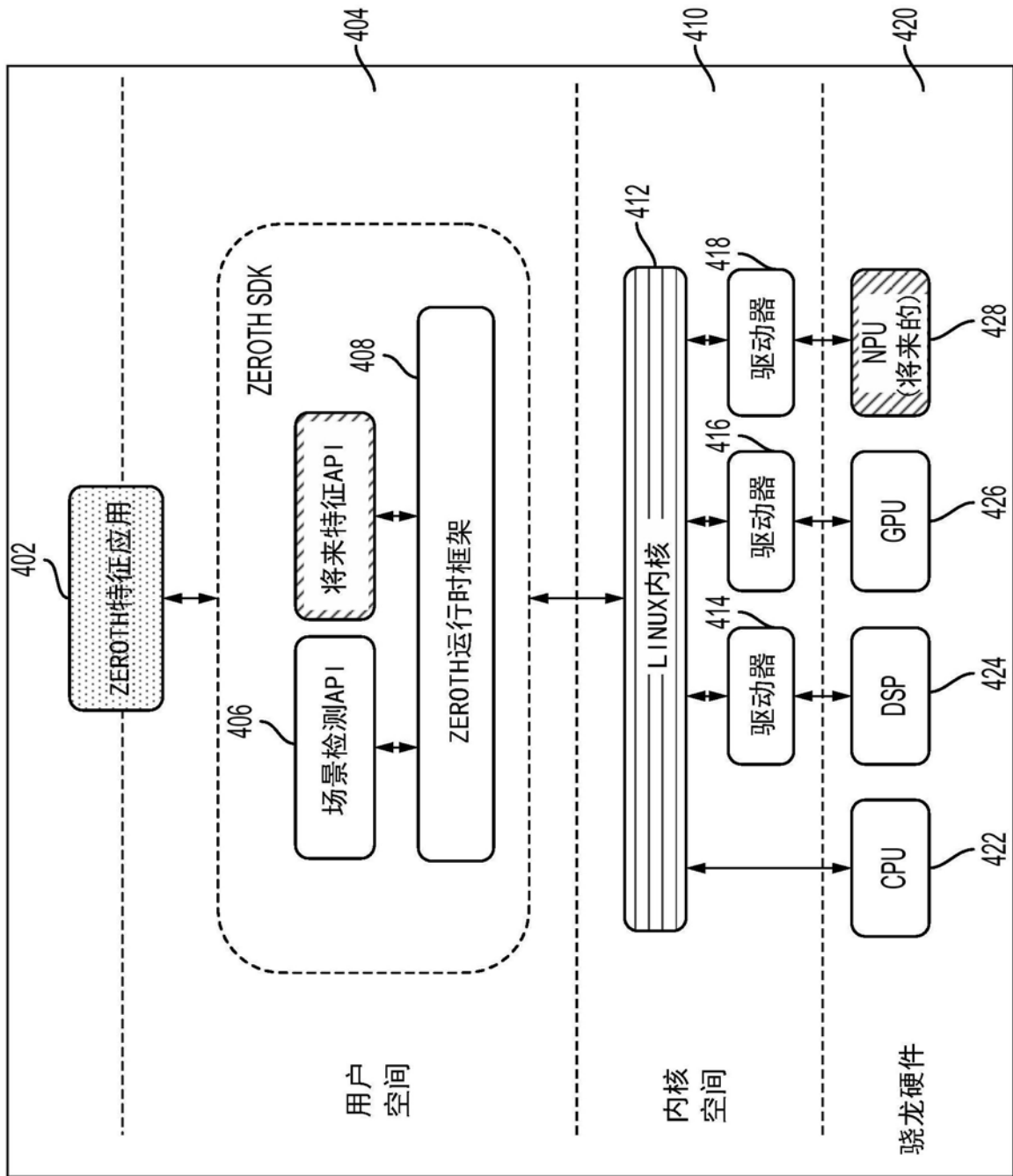


图3B



400 ↗

图4

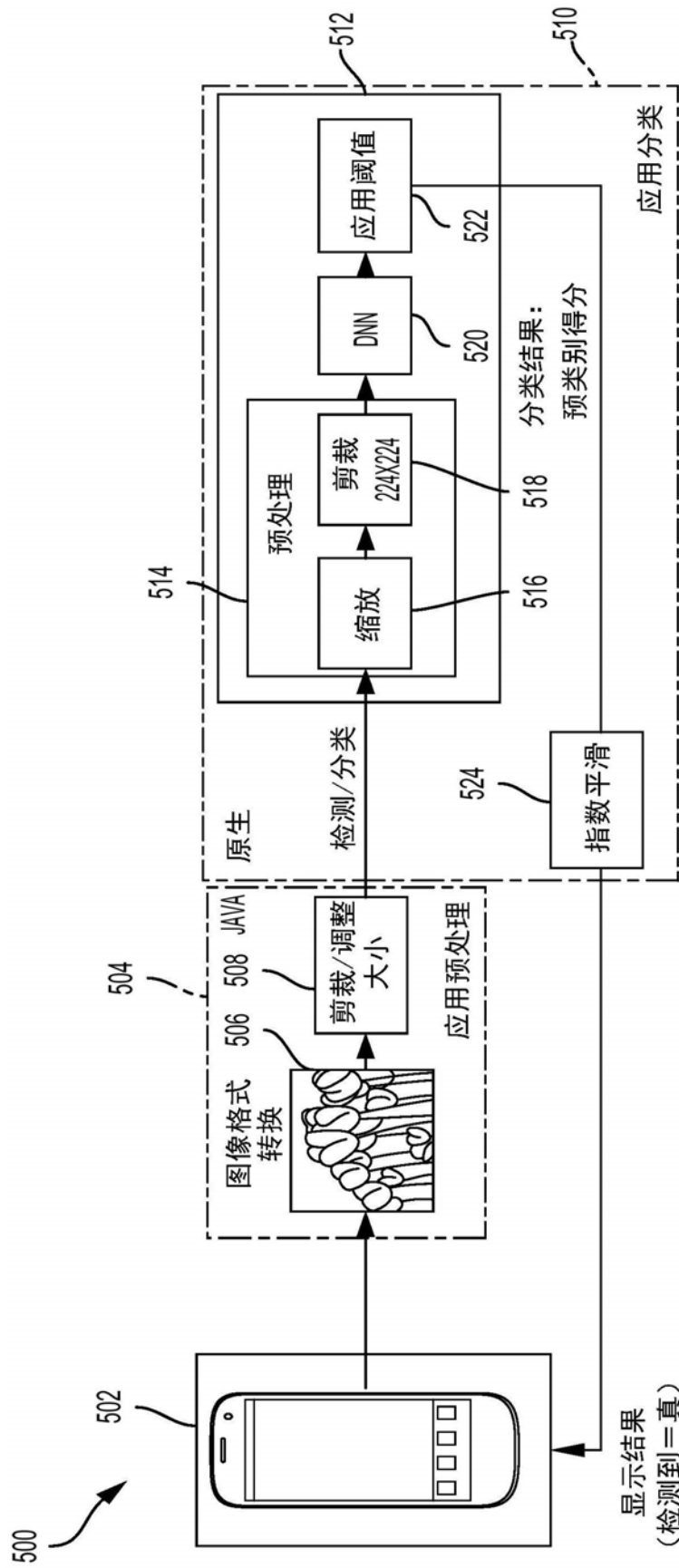


图5

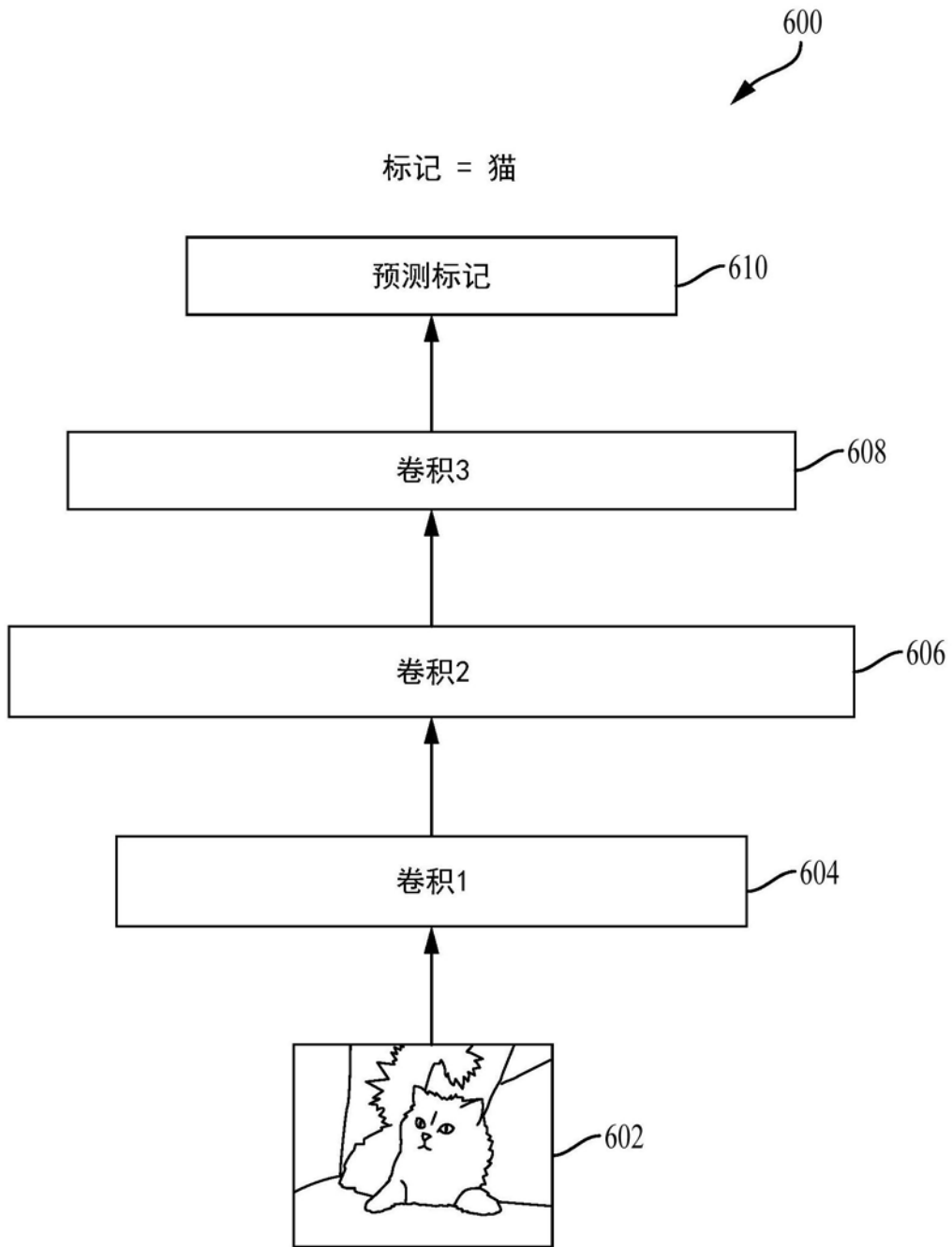


图6



图7

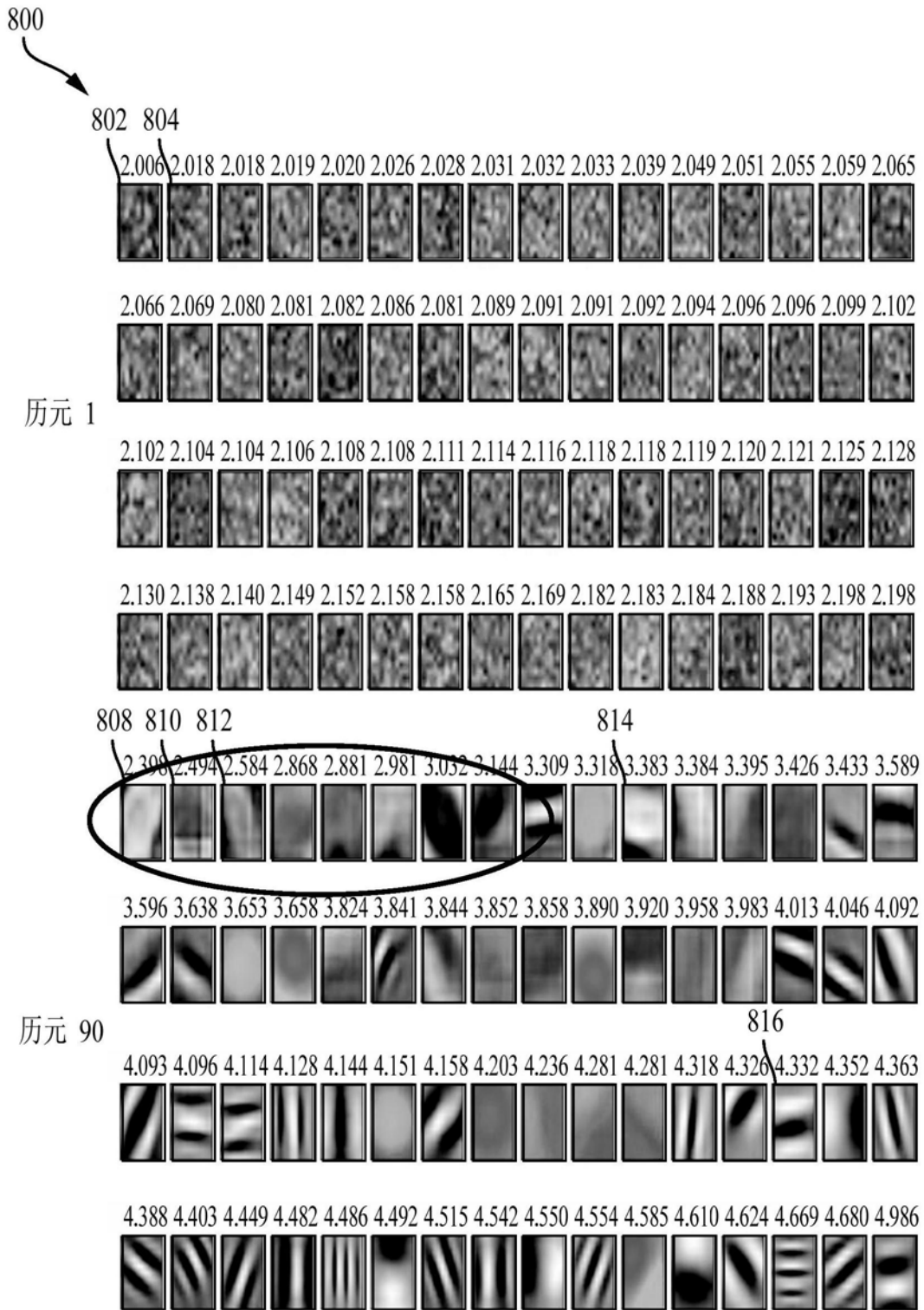


图8

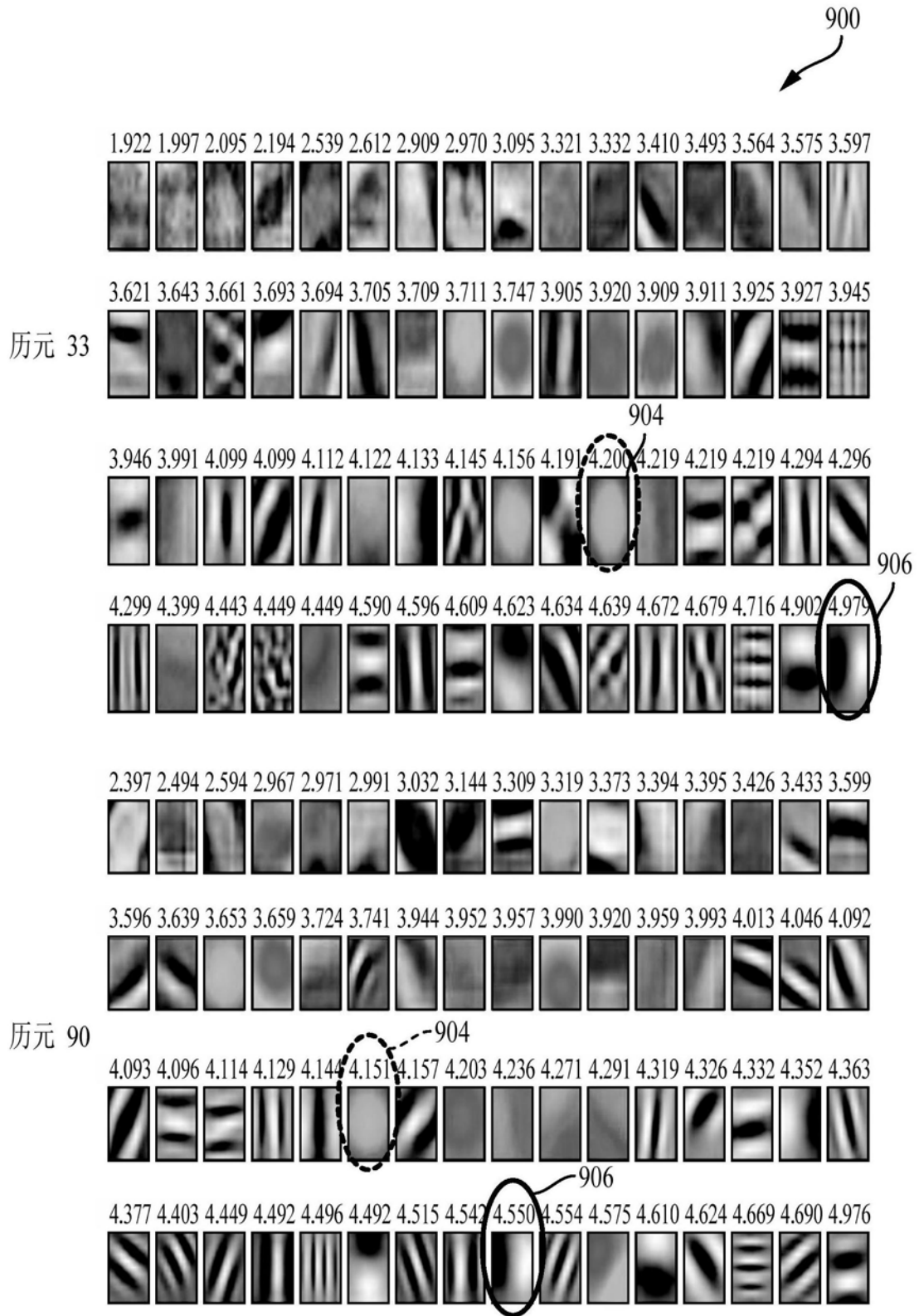


图9

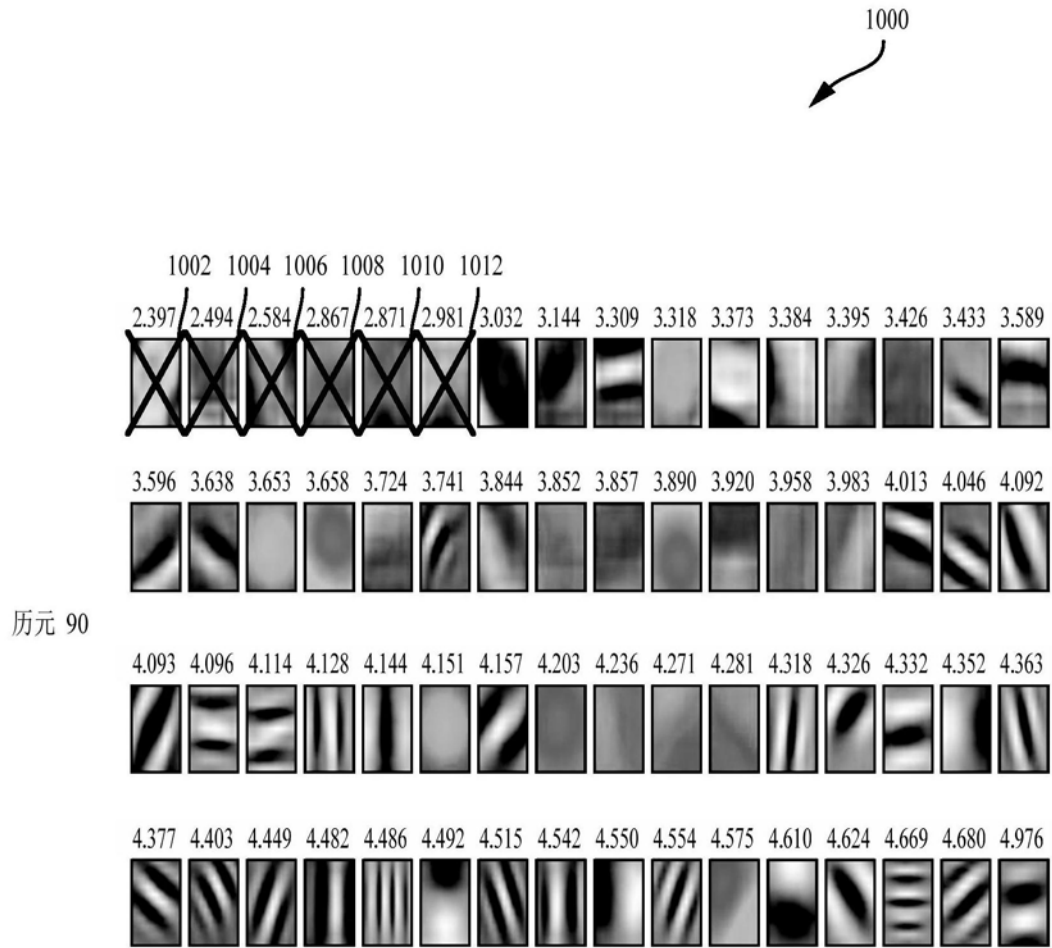


图10

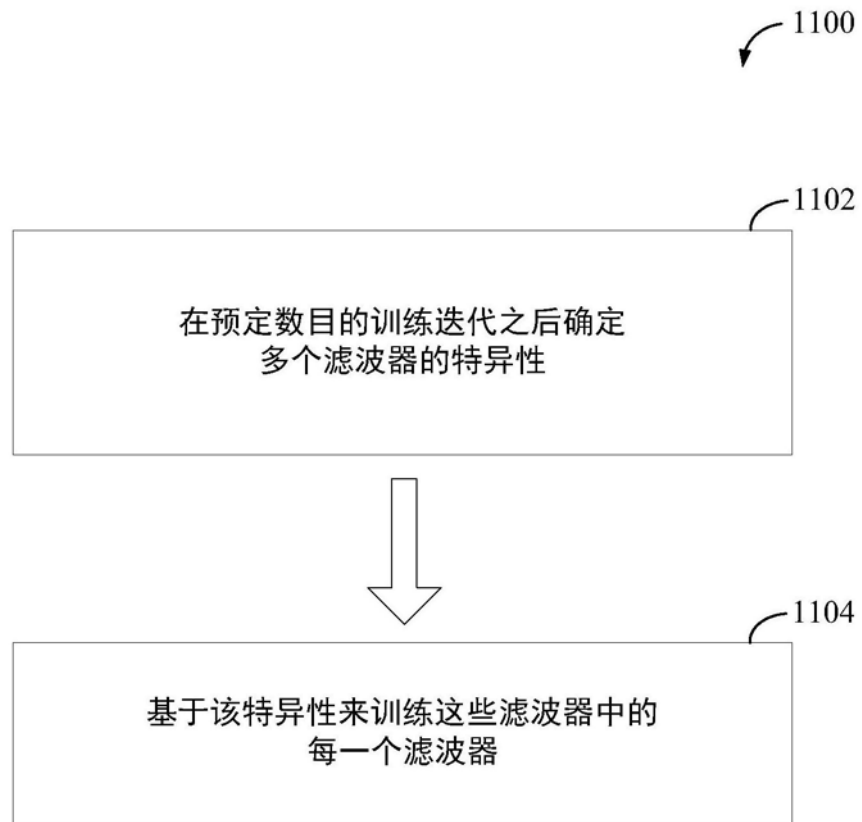


图11

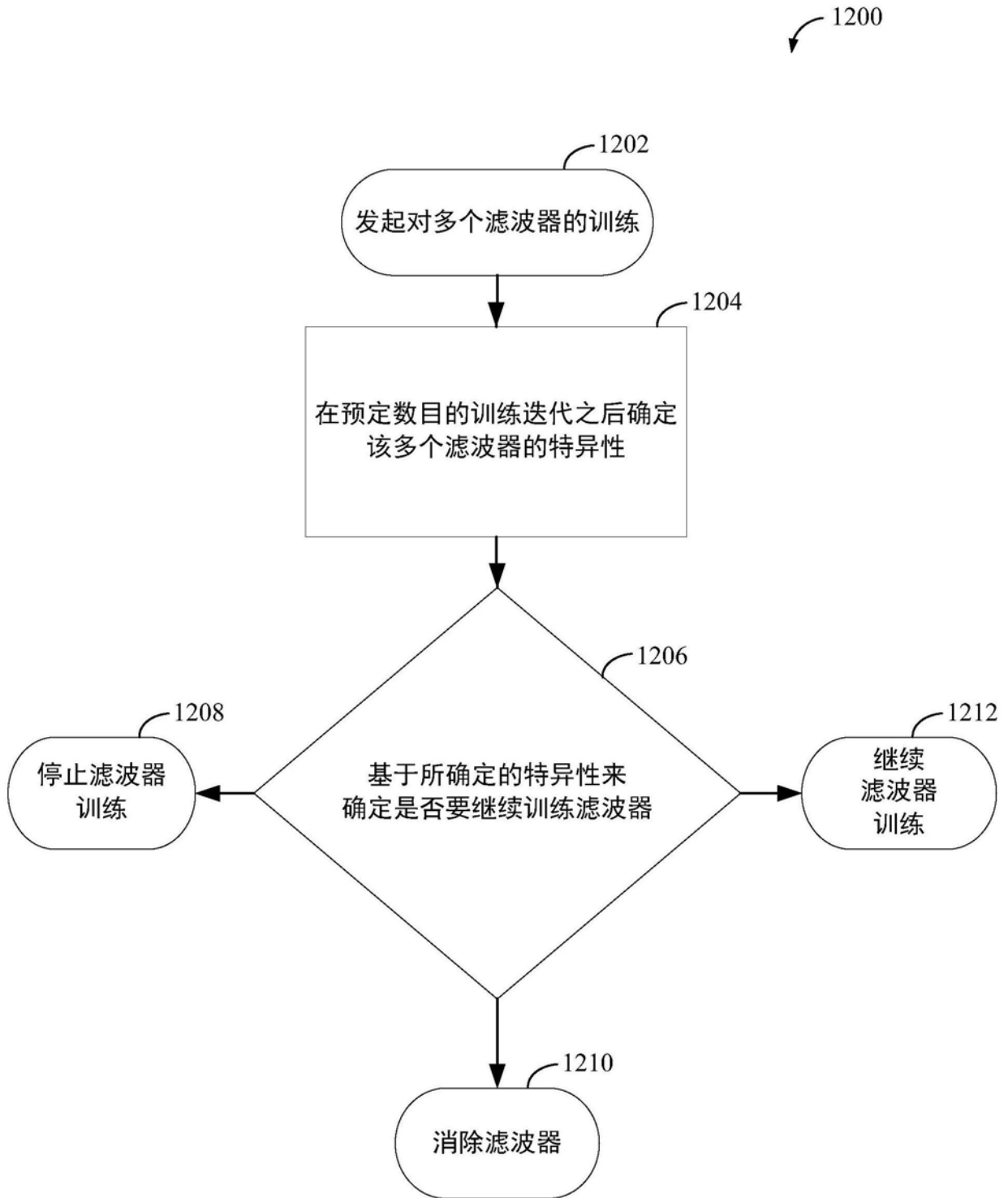


图12