



(12)发明专利申请

(10)申请公布号 CN 110175671 A  
(43)申请公布日 2019.08.27

(21)申请号 201910351894.1

(22)申请日 2019.04.28

(71)申请人 华为技术有限公司

地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 陈鑫 谢凌曦 田奇

(74)专利代理机构 北京龙双利达知识产权代理有限公司 11329

代理人 张振 张欣

(51) Int. Cl.

G06N 3/04(2006.01)

G06N 3/08(2006.01)

G06F 16/55(2019.01)

G06K 9/00(2006.01)

G06K 9/62(2006.01)

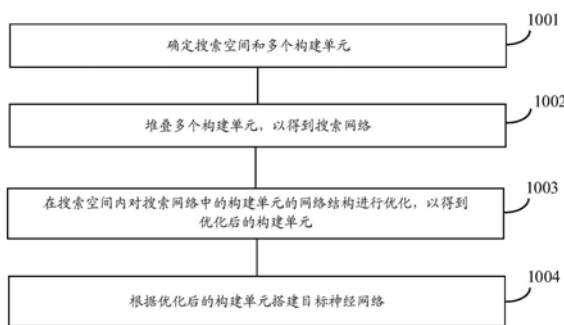
权利要求书3页 说明书27页 附图12页

(54)发明名称

神经网络的构建方法、图像处理方法及装置

(57)摘要

本申请公开了人工智能领域中计算机视觉领域的一种神经网络的构建方法、图像处理方法及装置。其中,该神经网络的构建方法包括:确定搜索空间和多个构建单元;堆叠所述多个构建单元,以得到搜索网络,所述搜索网络是用于搜索神经网络结构的神经网络;在所述搜索空间内对所述搜索网络中的构建单元的网络结构进行优化,以得到优化后的构建单元,其中,在优化过程中搜索空间逐渐减小,构建单元数量逐渐增加,搜索空间的减小和构建单元数量的增加使得所述优化过程中产生的显存消耗在预设范围内;根据所述优化后的构建单元搭建所述目标神经网络。本申请能够在显存资源一定的情况下,构建出较好地满足应用需求的神经网络。



1. 一种神经网络的构建方法,其特征在于,包括:

确定搜索空间和多个构建单元,其中,所述搜索空间是根据待构建的目标神经网络的应用需求确定的,所述多个构建单元是根据所述搜索空间以及构建所述目标神经网络的设备的显存资源的大小确定的,所述构建单元是由多个节点之间通过神经网络的基本操作连接得到的一种网络结构,所述构建单元是用于构建神经网络的基础模块;

堆叠所述多个构建单元,以得到搜索网络,所述搜索网络是用于搜索神经网络结构的神经网络;

在所述搜索空间内对所述搜索网络中的构建单元的网络结构进行优化,以得到优化后的构建单元;

其中,对所述搜索网络中的构建单元的网络结构进行优化的优化过程包括N个阶段,第i个阶段和第j个阶段为所述N个阶段中的任意两个阶段,所述搜索空间在所述第i个阶段的大小大于所述搜索网络在所述第j个阶段的大小,所述搜索网络在所述第i个阶段时包含的构建单元的数量小于所述搜索空间在所述第j个阶段时包含的构建单元的数量,所述搜索网络的搜索空间的减小和所述搜索网络的构建单元数量的增加使得所述优化过程中产生的显存消耗在预设范围内,所述搜索网络在第N个阶段包含的构建单元的数量与所述目标神经网络包含的构建单元的数量之差在预设范围内,所述目标神经网络包含的构建单元的数量是根据所述目标神经网络的应用需求确定的,N为大于1的正整数,i和j均为小于或者等于N的正整数,并且i小于j;

根据所述优化后的构建单元搭建所述目标神经网络。

2. 如权利要求1所述的方法,其特征在于, $j=i+1$ 。

3. 如权利要求1或2所述的方法,其特征在于,所述优化后的构建单元的各个节点之间的连接关系中包含的第一类操作的数量在预设范围内,所述第一类操作是不包含神经网络可训练参数的操作。

4. 如权利要求1-3中任一项所述的方法,其特征在于,所述搜索网络中的构建单元包括第一类构建单元,所述第一类构建单元是输入特征图的数量和大小分别与输出特征图的数量和大小相同的构建单元。

5. 如权利要求1-4中任一项所述的方法,其特征在于,所述搜索网络中的构建单元包括第二类构建单元,所述第二类构建单元的输出特征图的分辨率是输入特征图的 $1/M$ ,所述第二类构建单元的输出特征图的数量是输入特征图的数量M倍,M为大于1的正整数。

6. 一种图像处理方法,其特征在于,包括:

获取待处理图像;

根据目标神经网络对所述待处理图像进行分类,得到所述待处理图像的分类结果。

其中,所述目标神经网络由多个优化后的构建单元搭建而成,所述多个优化后的构建单元是通过对搜索网络中的构建单元的网络结构进行N个阶段的优化得到的,第i个阶段和第j个阶段为所述N个阶段中的任意两个阶段,所述搜索空间在所述第i个阶段的大小大于所述搜索网络在所述第j个阶段的大小,所述搜索网络在所述第i个阶段时包含的构建单元的数量小于所述搜索空间在所述第j个阶段时包含的构建单元的数量,所述搜索网络的搜索空间的减小和所述搜索网络的构建单元数量的增加使得所述优化过程中产生的显存消耗在预设范围内,所述搜索网络在第N个阶段包含的构建单元的数量与所述目标神经网络

包含的构建单元的数量的差异在预设范围内,所述目标神经网络包含的构建单元的数量是根据所述目标神经网络的应用需求确定的, $N$ 为大于1的正整数, $i$ 和 $j$ 均为小于或者等于 $N$ 的正整数,并且 $i$ 小于 $j$ 。

7.如权利要求6所述的方法,其特征在于, $j=i+1$ 。

8.一种神经网络构建装置,其特征在于,包括:

存储器,用于存储程序;

处理器,用于执行所述存储器存储的程序,当所述存储器存储的程序被执行时,所述处理器用于执行以下过程:

确定搜索空间和多个构建单元,其中,所述搜索空间是根据待构建的目标神经网络的应用需求确定的,所述多个构建单元是根据所述搜索空间以及构建所述目标神经网络的设备的显存资源的大小确定的,所述构建单元是由多个节点之间通过神经网络的基本操作连接得到的一种网络结构,所述构建单元是用于构建神经网络的基础模块;

堆叠所述多个构建单元,以得到搜索网络,所述搜索网络是用于搜索神经网络结构的神经网络;

在所述搜索空间内对所述搜索网络中的构建单元的网络结构进行优化,以得到优化后的构建单元;

其中,对所述搜索网络中的构建单元的网络结构进行优化的优化过程包括 $N$ 个阶段,第 $i$ 个阶段和第 $j$ 个阶段为所述 $N$ 个阶段中的任意两个阶段,所述搜索空间在所述第 $i$ 个阶段的大小大于所述搜索网络在所述第 $j$ 个阶段的大小,所述搜索网络在所述第 $i$ 个阶段时包含的构建单元的数量小于所述搜索空间在所述第 $j$ 个阶段时包含的构建单元的数量,所述搜索网络的搜索空间的减小和所述搜索网络的构建单元数量的增加使得所述优化过程中产生的显存消耗在预设范围内,所述搜索网络在第 $N$ 个阶段包含的构建单元的数量与所述目标神经网络包含的构建单元的数量的差异在预设范围内,所述目标神经网络包含的构建单元的数量是根据所述目标神经网络的应用需求确定的, $N$ 为大于1的正整数, $i$ 和 $j$ 均为小于或者等于 $N$ 的正整数,并且 $i$ 小于 $j$ ;

根据所述优化后的构建单元搭建所述目标神经网络。

9.如权利要求8所述的装置,其特征在于, $j=i+1$ 。

10.如权利要求8或9所述的装置,其特征在于,所述优化后的构建单元的各个节点之间的连接关系中包含的第一类操作的数量在预设范围内,所述第一类操作是不包含神经网络可训练参数的操作。

11.如权利要求8-10中任一项所述的装置,其特征在于,所述搜索网络中的构建单元包括第一类构建单元,所述第一类构建单元是输入特征图的数量和大小分别与输出特征图的数量和大小相同的构建单元。

12.如权利要求8-11中任一项所述的装置,其特征在于,所述搜索网络中的构建单元包括第二类构建单元,所述第二类构建单元的输出特征图的分辨率是输入特征图的 $1/M$ ,所述第二类构建单元的输出特征图的数量是输入特征图的数量 $M$ 倍, $M$ 为大于1的正整数。

13.一种图像处理装置,其特征在于,包括:

存储器,用于存储程序;

处理器,用于执行所述存储器存储的程序,当所述存储器存储的程序被执行时,所述处

理器用于执行以下过程：

获取待处理图像；

根据目标神经网络对所述待处理图像进行分类，得到所述待处理图像的分类结果。

其中，所述目标神经网络由多个优化后的构建单元搭建而成，所述多个优化后的构建单元是通过对搜索网络中的构建单元的网络结构进行N个阶段的优化得到的，第i个阶段和第j个阶段为所述N个阶段中的任意两个阶段，所述搜索空间在所述第i个阶段的大小大于所述搜索网络在所述第j个阶段的大小，所述搜索网络在所述第i个阶段时包含的构建单元的数量小于所述搜索空间在所述第j个阶段时包含的构建单元的数量，所述搜索网络的搜索空间的减小和所述搜索网络的构建单元数量的增加使得所述优化过程中产生的显存消耗在预设范围内，所述搜索网络在第N个阶段包含的构建单元的数量与所述目标神经网络包含的构建单元的数量差异在预设范围内，所述目标神经网络包含的构建单元的数量是根据所述目标神经网络的应用需求确定的，N为大于1的正整数，i和j均为小于或者等于N的正整数，并且i小于j。

14. 如权利要求13所述的装置，其特征在于， $j=i+1$ 。

15. 一种计算机可读存储介质，其特征在于，所述计算机可读介质存储用于设备执行的程序代码，该程序代码包括用于执行如权利要求1-5或者6-7中任一项所述的方法。

16. 一种芯片，其特征在于，所述芯片包括处理器与数据接口，所述处理器通过所述数据接口读取存储器上存储的指令，以执行如权利要求1-5或者6-7中任一项所述的方法。

## 神经网络的构建方法、图像处理方法及装置

### 技术领域

[0001] 本申请涉及人工智能领域,并且更具体地,涉及一种神经网络的构建方法、图像处理方法及装置。

### 背景技术

[0002] 人工智能(artificial intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说,人工智能是计算机科学的一个分支,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式作出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法,使机器具有感知、推理与决策的功能。人工智能领域的研究包括机器人,自然语言处理,计算机视觉,决策与推理,人机交互,推荐与搜索, AI基础理论等。

[0003] 随着人工智能技术的快速发展,神经网络(例如,深度神经网络)近年来在图像、视频以及语音等多种媒体信号的处理与分析中取得了很大的成就。一个性能优良的神经网络往往拥有精妙的网络结构,而这需要具有高超技能和丰富经验的人类专家花费大量精力进行构建。为了更好地构建神经网络,人们提出了通过神经网络结构搜索(neural architecture search, NAS)的方法来搭建神经网络,通过自动化地搜索神经网络结构,从而得到性能优异的神经网络结构。

[0004] 传统方案常采用可微分的神经网络结构搜索方法来搭建神经网络,该搜索方法一般是根据一定数量的构建单元搭建成一个搜索网络,然后在搜索空间内对搜索网络中构建单元的各个节点之间的连接关系进行优化,以得到优化的构建单元,最后再根据优化的构建单元搭建目标神经网络。该搜索方法在优化过程中,会将所有可能的操作都放到搜索空间中,这就导致优化过程中需要巨大的显存空间,从而只能堆叠成较浅的搜索网络。而最终要构建的目标神经网络往往层次较深,这就导致搜索网络与目标神经网络之间存在较大的深度差异,并且由于较浅的搜索网络优化得到的构建单元不是完全适用于较深的目标神经网络,从而使得最终搭建的目标神经网络可能无法很好地满足应用需求。

### 发明内容

[0005] 本申请提供一种神经网络的构建方法、图像处理方法、装置、计算机可读存储介质和芯片,以更好地构建满足需求的神经网络。

[0006] 第一方面,提供了一种神经网络的构建方法,该方法包括:确定搜索空间和多个构建单元;堆叠多个构建单元,以得到搜索网络,该搜索网络是用于搜索神经网络结构的神经网络;在搜索空间内对搜索网络中的构建单元的网络结构进行优化,以得到优化后的构建单元;根据优化后的构建单元搭建目标神经网络。

[0007] 其中,上述搜索空间是根据待构建的目标神经网络的应用需求确定的,上述多个构建单元是根据搜索空间以及构建目标神经网络的设备的显存资源的大小确定的,另外,

构建单元是由多个节点之间通过神经网络的基本操作连接得到的一种网络结构,构建单元是用于构建神经网络的基础模块。

[0008] 其中,对搜索网络中的构建单元的网络结构进行优化的优化过程包括N个阶段,第i个阶段和第j个阶段为N个阶段中的任意两个阶段,搜索空间在第i个阶段的大小大于搜索网络在第j个阶段的大小,搜索网络在第i个阶段时包含的构建单元的数量小于搜索空间在第j个阶段时包含的构建单元的数量,搜索网络的搜索空间的减小和搜索网络的构建单元数量的增加使得优化过程中产生的显存消耗在预设范围内,搜索网络在第N个阶段包含的构建单元的数量与目标神经网络包含的构建单元的数量差异在预设范围内,上述目标神经网络包含的构建单元的数量是根据目标神经网络的应用需求确定的,N为大于1的正整数,i和j均为小于或者等于N的正整数,并且i小于j。

[0009] 上述搜索空间是根据待构建的目标神经网络的应用需求确定的,具体包括:上述搜索空间是根据目标神经网络的处理数据的类型确定的。

[0010] 具体地,当上述目标神经网络用于处理图像数据的神经网络时,上述搜索空间包含的操作的种类和数量要与图像数据的处理相适应。

[0011] 例如,当目标神经网络是用于处理图像数据的神经网络时,上述搜索空间可以包含卷积操作,池化操作,跳连接(skip-connect)操作等等。

[0012] 当上述目标神经网络用于处理语音数据时,上述搜索空间包含的操作的种类和数量要与语音数据的处理相适应。

[0013] 例如,当目标神经网络是用于处理语音数据的神经网络时,上述搜索空间可以包含激活函数(如ReLU、Tanh)等等。

[0014] 具体地,上述目标神经网络包含的构建单元的数量是根据目标神经网络的应用需求确定的,包括:上述目标神经网络包含的构建单元的数量是根据目标神经网络要处理的数据类型和/或计算的复杂度确定的。

[0015] 例如,当上述目标神经网络用于处理一些简单的文本数据时,目标神经网络包含较少数量的构建单元即可,当上述目标神经网络用于处理一些比较复杂的图像数据时,目标神经网络需要包含数量较多的构建单元。

[0016] 再如,当目标神经网络需要处理的数据复杂度较高时,目标神经网络需要包含数量较多的构建单元;当目标神经网络需要处理的数据复杂度较低时,目标神经网络需要较少数量的构建单元即可。

[0017] 可选地,上述显存资源可以替换为缓存资源,该缓存资源是用于构建神经网络的设备在优化过程中用于存放运算数据的内存或者存储单元。

[0018] 上述缓存资源具体可以包括显存资源。

[0019] 可选地,上述堆叠多个构建单元,以得到搜索网络,包括:按照预设的堆叠方式将所述多个构建单元依次堆叠起来,以得到搜索网络,其中,在该搜索网络中,位于搜索网络前面的构建单元的输出是位于搜索网络的后面的构建单元的输入。

[0020] 上述预设的堆叠方式可以包括在什么位置堆放什么类型的构建单元以及堆叠的数量等等。

[0021] 本申请中,在对构建单元的网络结构进行优化的过程中,减少搜索空间节省的显存资源可以用来增加构建单元的数量,从而能够在显存资源有限的情况下,尽可能的堆叠

得到构建单元数量与最终要搭建的目标神经网络的构建单元数量比较接近的搜索网络。使得优化后的构建单元能够更好地适用于目标神经网络的搭建,进而使得根据优化后的构建单元搭建成的目标神经网络能够更好地满足应用需求。

[0022] 具体地,本申请在搜索网络的构建单元的网络结构的优化过程中,通过逐渐减少搜索空间的大小,并增加搜索网络的构建单元的数量,能够在构建出能够较好满足应用需求的目标神经网络的情况下,减少优化过程显存资源的依赖,使得在优化过程中仅仅依赖较少的显存资源就能够得到较好的满足应用需求的目标神经网络,也在一定程度上提高了显存资源的利用率。

[0023] 一般地,如果搜索网络的网络深度与待构建的目标神经网络的网络深度比较接近时,搜索网络中优化得到的构建单元比较适合由于搭建目标神经网络。神经网络的深度与包含的构建单元的数量是正相关的关系,因此,当搜索网络的构建单元的数量与目标神经网络的构建单元数量比较接近时,搜索网络的网络深度与目标神经网络的网络深度也比较接近。

[0024] 可选地,上述搜索空间在第 $i$ 个阶段的大小为 $S_i$ ,上述搜索空间在第 $j$ 个阶段的大小为 $S_j$ ,上述搜索网络在第 $i$ 个阶段包含的构建单元数量为 $L_i$ 个,上述搜索网络在第 $j$ 个阶段包含的构建单元数量为 $L_j$ 个,其中,上述 $L_j-L_i$ 大小是根据 $S_i-S_j$ 的大小确定的,或者,上述 $S_i-S_j$ 的大小是根据 $L_j-L_i$ 的大小确定的。

[0025] 具体地,在上述两个阶段中,可以预先设定 $S_i-S_j$ 的大小,然后再根据 $S_i-S_j$ 的大小确定 $L_j-L_i$ 的大小,使得由于搜索空间减小节省的显存资源与构建单元增加导致多消耗的显存资源的差值在一定阈值范围内。

[0026] 在上述两个阶段中,也可以预先设定 $L_j-L_i$ 的大小,然后再根据 $L_j-L_i$ 的大小确定 $S_i-S_j$ 的大小,使得由于构建单元增加导致多消耗的显存资源与搜索空间减小节省的显存资源的差值在一定阈值范围内。

[0027] 可选地,上述 $N$ 的大小是预先设置的。

[0028] 上述 $N$ 的大小可以根据目标神经网络的构建需求来确定。具体地,当目标神经网络需要在较短的时间内构建完成时,可以将 $N$ 设置成一个较小的数值,当目标神经网络可以在较长的时间内构建完成时,可以将 $N$ 设置成一个较大的数值。

[0029] 应理解,在本申请中,只要上述 $N$ 个阶段中存在至少两个阶段满足搜索空间减小,构建单元数量增加即可,而不必使得每两个相邻的阶段都满足搜索空间减小,构建单元数量增加的要求。

[0030] 例如,上述 $N=4$ ,第2个阶段相比于第1阶段以及第4阶段相对于第3阶段均满足:搜索空间减小,搜索网络的构建单元数量增加。而第2个阶段和第3个阶段的搜索空间以及搜索网络包含的构建单元数量均没有发生变化。

[0031] 结合第一方面,在第一方面的某些实现方式中, $j=i+1$ 。

[0032] 当 $j=i+1$ 时,在优化过程中,任意两个相邻阶段之间都会满足搜索空间逐渐减小,搜索网络的构建单元逐渐增加,使得优化过程比较平稳。

[0033] 可选地,上述 $N$ 个阶段中,搜索网络在任意两个相邻阶段的构建单元的数量变化值相同,搜索空间在任意两个相邻阶段的大小变化值也相同。

[0034] 在上述优化过程中,构建单元的数量变化以及搜索空间的大小变化都是均匀的,

优化的过程更加平稳。

[0035] 可选地,上述第 $i+1$ 个阶段相对于第 $i$ 个阶段增加的构建单元的数量可以是根据上述数值 $N$ ,以及搜索网络在优化前包含的构建单元的数量,以及目标神经网络中的构建单元的数量来确定的。

[0036] 例如,搜索网络在第 $i+1$ 个阶段相对于第 $i$ 个阶段增加的构建单元的数量为 $X$ ,优化开始前搜索网络包含的构建单元数量为 $U$ ,目标神经网络中的构建单元的数量为 $V$ ,那么, $X$ 可以根据公式 $X = (U - V) / N$ 计算得到。

[0037] 应理解,在优化过程中,搜索空间的大小降低的幅度以及搜索网络构建单元数量的增加幅度可以根据多种方式来确定,只要能够确保优化过程中搜索网络的搜索空间的减小和所述搜索网络的构建单元数量的增加使得所述优化过程中产生的显存消耗在预设范围内即可。

[0038] 在实际应用中可以先预先设定搜索空间大小降低的幅度,然后再确定搜索网络构建单元数量增加的幅度;也可以预先设定搜索网络的大小,再确定搜索空间大小降低的幅度。本申请对此不做限定,所有确保显存消耗在预设范围内的实现方式都在本申请的保护范围内。

[0039] 结合第一方面,在第一方面的某些实现方式中,优化后的构建单元的各个节点之间的连接关系中包含的第一类操作的数量在预设范围内,第一类操作是不包含神经网络可训练参数的操作。

[0040] 本申请通过将第一类操作的数量限制在一定范围,使得最终搭建的目标神经网络的可训练参数保持在相对稳定的水平,进而使得目标神经网络的性能保持稳定。

[0041] 具体地,上述第一类操作是不包含可训练参数的操作,如果此类操作过多会导致包含可训练参数的其他操作较少,从而神经网络总体的可训练参数较少,神经网络的特征表达能力降低。

[0042] 由于在构建单元数量较多的搜索网络中进行结构搜索稳定性不足,会导致每次搜索得到的构建单元中第一类操作的数量具有一定的差异,搜索得到的神经网络结构(即构建单元)在相应任务上的性能表现波动。限制第一类操作的数量可以使得由搜索得到的神经网络结构搭建的测试网络的可训练参数保持在相对稳定的水平,从而减小在相应任务上的性能波动。

[0043] 结合第一方面,在第一方面的某些实现方式中,搜索网络中的构建单元包括第一类构建单元,第一类构建单元是输入特征图的数量和大小分别与输出特征图的数量和大小相同的构建单元。

[0044] 结合第一方面,在第一方面的某些实现方式中,搜索网络中的构建单元包括第二类构建单元,第二类构建单元的输出特征图的分辨率是输入特征图的 $1/M$ ,第二类构建单元的输出特图的数量是输入特征图的数量 $M$ 倍, $M$ 为大于1的正整数。

[0045] 第二方面,提供了一种图像处理方法,该方法包括:获取待处理图像;根据目标神经网络对所述待处理图像进行分类,得到所述待处理图像的分类结果,其中,该目标神经网络是根据第一方面中的任意一种实现方式构建得到的神经网络。

[0046] 应理解,第二方面中的图像处理方法所采用的目标神经网络在进行图像分类之前,还需要再根据训练图像对该目标神经网络进行训练,训练得到的目标神经网络就可以

对待处理图像进行分类。

[0047] 也就是说,可以采用第一方面中的神经网络结构搜索方法得到目标神经网络,接下来,再根据训练图像对该目标神经网络进行训练,训练完成后就可以用该目标神经网络对待处理图像进行分类了。

[0048] 本申请中,由于目标神经网络是采用上述第一方面的方面构建得到的,比较符合或者贴近神经网络的应用需求,利用这样的神经网络进行图像分类,能够取得较好的图像分类效果(例如,分类结果更准确,等等)。

[0049] 第三方面,提供了一种图像处理方法,该方法包括:获取待处理图像;根据目标神经网络对待处理图像进行分类,得到待处理图像的分类结果。

[0050] 其中,目标神经网络由多个优化后的构建单元搭建而成,多个优化后的构建单元是通过对搜索网络中的构建单元的网络结构进行N个阶段的优化得到的,第i个阶段和第j个阶段为N个阶段中的任意两个阶段,搜索空间在第i个阶段的大小大于搜索网络在第j个阶段的大小,搜索网络在第i个阶段时包含的构建单元的数量小于搜索空间在第j个阶段时包含的构建单元的数量,搜索网络的搜索空间的减小和搜索网络的构建单元数量的增加使得优化过程中产生的显存消耗在预设范围内,搜索网络在第N个阶段包含的构建单元的数量与目标神经网络包含的构建单元的数量差异在预设范围内,目标神经网络包含的构建单元的数量是根据目标神经网络的应用需求确定的,N为大于1的正整数,i和j均为小于或者等于N的正整数,并且i小于j。

[0051] 本申请中,在目标神经网络构建之前的优化过程中,通过减小搜索空间的大小,增加构建单元的数量,能够尽可能的堆叠得到构建单元数量与最终要搭建的目标神经网络的构建单元数量比较接近的搜索网络。从而使得搜索网络优化后的构建单元能够更好地适用于目标神经网络的搭建,能够获得性能更好的目标神经网络,利用该目标神经网络进行图像分类能够取得较好的图像分类效果(例如,分类结果更准确,等等)。

[0052] 结合第三方面,在第三方面的某些实现方式中, $j=i+1$ 。

[0053] 当 $j=i+1$ 时,在优化过程中,任意两个相邻阶段之间都会满足搜索空间逐渐减小,搜索网络的构建单元逐渐增加,使得优化过程比较平稳。

[0054] 可选地,上述N个阶段中,搜索网络在任意两个相邻阶段的构建单元的数量变化值相同,搜索空间在任意两个相邻阶段的大小变化值也相同。

[0055] 在上述优化过程中,构建单元的数量变化以及搜索空间的大小变化都是均匀的,优化的过程更加平稳。

[0056] 可选地,上述目标神经网络是经过训练图片进行训练得到的神经网络。

[0057] 具体地,可以通过训练图片以及训练图片标记的类别信息对目标神经网络进行训练,训练完成的神经网络就可以用于进行图像分类了。

[0058] 第四方面,提供了一种图像处理方法,该方法包括:获取道路画面;根据目标神经网络对道路画面进行卷积处理,得到道路画面的多个卷积特征图;根据目标神经网络对道路画面的多个卷积特征图进行反卷积处理,获得该道路画面的语义分割结果。

[0059] 其中,上述目标神经网络是根据第一方面中的任意一种实现方式构建得到的神经网络。

[0060] 第五方面,提供了一种图像处理方法,该方法包括:获取人脸图像;根据目标神经

网络对人脸图像进行卷积处理,得到人脸图像的卷积特征图;将人脸图像的卷积特征图与身份证件图像的卷积特征图进行对比,得到人脸图像的验证结果。

[0061] 上述身份证件图像可的卷积特征图可以是预先获取的,并存储在相应的数据库中。例如,预先对身份证件图像进行卷积处理,将得到的卷积特征图存储到数据库中。

[0062] 另外,上述目标神经网络是根据第一方面中的任意一种实现方式构建得到的神经网络。

[0063] 应理解,在上述第一方面中对相关内容的扩展、限定、解释和说明也适用于第二方面、第三方面、第四方面和第五方面中相同的内容。

[0064] 第六方面,提供了一种神经网络构建装置,该装置包括:存储器,用于存储程序;处理器,用于执行所述存储器存储的程序,当所述存储器存储的程序被执行时,所述处理器用于执行第一方面中的任意一种实现方式中的方法。

[0065] 第七方面,提供了一种图像处理装置,该装置包括:存储器,用于存储程序;处理器,用于执行所述存储器存储的程序,当所述存储器存储的程序被执行时,所述处理器用于执行第二方面至第五方面中的任意一种实现方式中的方法。

[0066] 第八方面,提供一种计算机可读介质,该计算机可读介质存储用于设备执行的程序代码,该程序代码包括用于执行第一方面至第五方面中的任意一种实现方式中的方法。

[0067] 第九方面,提供一种包含指令的计算机程序产品,当该计算机程序产品在计算机上运行时,使得计算机执行上述第一方面至第五方面中的任意一种实现方式中的方法。

[0068] 第十方面,提供一种芯片,所述芯片包括处理器与数据接口,所述处理器通过所述数据接口读取存储器上存储的指令,执行上述第一方面至第五方面中的任意一种实现方式中的方法。

[0069] 可选地,作为一种实现方式,所述芯片还可以包括存储器,所述存储器中存储有指令,所述处理器用于执行所述存储器上存储的指令,当所述指令被执行时,所述处理器用于执行第一方面至第五方面中的任意一种实现方式中的方法。

## 附图说明

[0070] 图1是本申请实施例提供的一种人工智能主体框架示意图;

[0071] 图2为本申请实施例提供的一种应用环境的示意图;

[0072] 图3为本申请实施例提供的一种卷积神经网络结构示意图;

[0073] 图4为本申请实施例提供的一种卷积神经网络结构示意图;

[0074] 图5为本申请实施例提供的一种神经网络处理器的结构示意图;

[0075] 图6为本申请实施例提供的一种处理器的结构示意图;

[0076] 图7为本申请实施例提供的一种芯片的硬件结构示意图;

[0077] 图8为本申请实施例提供的一种系统架构的示意图;

[0078] 图9是本申请实施例的神经网络的构建方法的示意性流程图;

[0079] 图10是本申请实施例的构建单元的示意图;

[0080] 图11是本申请实施例的搜索网络的示意图;

[0081] 图12是本申请实施例的神经网络的构建方法的示意图;

[0082] 图13是本申请实施例的神经网络构建系统的示意图;

- [0083] 图14是本申请实施例的搜索网络的构建单元的网络结构优化过程的示意图；
- [0084] 图15是本申请实施例的操作数量规范模块的处理过程的示意图；
- [0085] 图16是本申请实施例的图像处理方法的示意性流程图；
- [0086] 图17是本申请实施例的神经网络构建装置的示意性框图；
- [0087] 图18是本申请实施例的图像处理装置的示意性框图；
- [0088] 图19是本申请实施例的神经网络训练装置的示意性框图。

### 具体实施方式

[0089] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0090] 图1示出一种人工智能主体框架示意图,该主体框架描述了人工智能系统总体工作流程,适用于通用的人工智能领域需求。

[0091] 下面从“智能信息链”(水平轴)和“信息技术(information technology,IT)价值链”(垂直轴)两个维度对上述人工智能主题框架进行详细的阐述。

[0092] “智能信息链”反映从数据的获取到处理的一列过程。举例来说,可以是智能信息感知、智能信息表示与形成、智能推理、智能决策、智能执行与输出的一般过程。在这个过程中,数据经历了“数据—信息—知识—智慧”的凝练过程。

[0093] “IT价值链”从人智能的底层基础设施、信息(提供和处理技术实现)到系统的产业生态过程,反映人工智能为信息技术产业带来的价值。

[0094] (1) 基础设施:

[0095] 基础设施为人工智能系统提供计算能力支持,实现与外部世界的沟通,并通过基础平台实现支撑。

[0096] 基础设施可以通过传感器与外部沟通,基础设施的计算能力可以由智能芯片提供。

[0097] 这里的智能芯片可以是中央处理器(central processing unit,CPU)、神经网络处理器(neural-network processing unit,NPU)、图形处理器(graphics processing unit,GPU)、专门应用的集成电路(application specific integrated circuit,ASIC)以及现场可编程门阵列(field programmable gate array,FPGA)等硬件加速芯片。

[0098] 基础设施的基础平台可以包括分布式计算框架及网络等相关的平台保障和支持,可以包括云存储和计算、互联互通网络等。

[0099] 例如,对于基础设施来说,可以通过传感器和外部沟通获取数据,然后将这些数据提供给基础平台提供的分布式计算系统中的智能芯片进行计算。

[0100] (2) 数据:

[0101] 基础设施的上一层的数据用于表示人工智能领域的数据来源。该数据涉及到图形、图像、语音、文本,还涉及到传统设备的物联网数据,包括已有系统的业务数据以及力、位移、液位、温度、湿度等感知数据。

[0102] (3) 数据处理:

[0103] 上述数据处理通常包括数据训练,机器学习,深度学习,搜索,推理,决策等处理方式。

[0104] 其中,机器学习和深度学习可以对数据进行符号化和形式化的智能信息建模、抽取、预处理、训练等。

[0105] 推理是指在计算机或智能系统中,模拟人类的智能推理方式,依据推理控制策略,利用形式化的信息进行机器思维和求解问题的过程,典型的功能是搜索与匹配。

[0106] 决策是指智能信息经过推理后进行决策的过程,通常提供分类、排序、预测等功能。

[0107] (4) 通用能力:

[0108] 对数据经过上面提到的数据处理后,进一步基于数据处理的结果可以形成一些通用的能力,比如可以是算法或者一个通用系统,例如,翻译,文本的分析,计算机视觉的处理,语音识别,图像的识别等等。

[0109] (5) 智能产品及行业应用:

[0110] 智能产品及行业应用指人工智能系统在各领域的产品和应用,是对人工智能整体解决方案的封装,将智能信息决策产品化、实现落地应用,其应用领域主要包括:智能制造、智能交通、智能家居、智能医疗、智能安防、自动驾驶,平安城市,智能终端等。

[0111] 本申请实施例可以应用在人工智能中的很多领域,例如,智能制造、智能交通、智能家居、智能医疗、智能安防、自动驾驶,平安城市等领域。

[0112] 具体地,本申请实施例可以具体应用在图像分类、图像检索、图像语义分割、图像超分辨率和自然语言处理等需要使用(深度)神经网络的领域。

[0113] 下面对相册图片分类和自动驾驶这两种应用场景进行简单的介绍。

[0114] 相册图片分类:

[0115] 具体地,当用户在终端设备(例如,手机)或者云盘上存储了大量的图片时,通过对相册中图像进行识别可以方便用户或者系统对相册进行分类管理,提升用户体验。

[0116] 利用本申请实施例的神经网络结构搜索方法能够搜索得到适用于相册分类的神经网络结构,然后再根据训练图片库中的训练图片对神经网络进行训练,就可以得到相册分类神经网络。接下来就可以利用该相册分类神经网络对图片进行分类,从而为不同的类别的图片打上标签,便于用户查看和查找。另外,这些图片的分类标签也可以提供给相册管理系统进行分类管理,节省用户的管理时间,提高相册管理的效率,提升用户体验。

[0117] 例如,如图2所示,可以通过神经网络构建系统(对应于本申请实施例的神经网络结构搜索方法)构建得到适用于相册分类的神经网络。在构建该神经网络时,可以利用训练图片库的对搜索网络中的构建单元的网络结构进行优化,得到优化后的构建单元,然后再利用该优化后的构建单元来搭建神经网络。在获得适用于相册分类的神经网络之后,可以再根据训练图片对该神经网络进行训练,得到相册分类神经网络。接下来,就可以利用相册分类神经网络对待处理图片进行分类。如图2所示,相册分类神经网络对输入的图片进行处理,得到图片的类别为郁金香。

[0118] 自动驾驶场景下的物体识别:

[0119] 自动驾驶中有大量的传感器数据需要处理,深度神经网络凭借着其强大的能力在自动驾驶中发挥着重要的作用。然而手工设计相应的数据处理网络费时费力。因此,通过采

用本申请实施例的神经网络结构搜索方法,能够构建得到适用于自动驾驶场景下进行数据处理的神经网络,接下来,通过自动驾驶场景下的数据对该神经网络进行训练,能够得到传感器数据处理网络,最后就可以利用该传感器处理网络对输入的道路画面进行处理,从而识别出道路画面中的不同物体。

[0120] 如图3所示,神经网络构建系统能够根据车辆检测任务构建出一个神经网络,在构建该神经网络时,可以利用传感器数据对搜索网络中的构建单元的网络结构进行优化,得到优化后的构建单元,然后再利用该优化后的构建单元来搭建神经网络。在得到了神经网络之后,就可以根据传感器数据对该神经网络进行训练,得到传感器数据处理网络。接下来,就可以利用该传感器数据处理网络对传感器数据进行处理。如图3所示,传感器数据处理网络对输入的道路画面进行处理,能够识别出道路画面中的车辆(如图3右下角矩形框部分所示)。

[0121] 由于本申请实施例涉及大量神经网络的应用,为了便于理解,下面先对本申请实施例可能涉及的神经网络的相关术语和概念进行介绍。

[0122] (1) 神经网络

[0123] 神经网络可以是由神经单元组成的,神经单元可以是指以  $x_s$  和截距1为输入的运算单元,该运算单元的输出可以为:

$$[0124] \quad h_{w,b}(x) = f(W^T x) = f\left(\sum_{s=1}^n W_s x_s + b\right) \quad (1-1)$$

[0125] 其中,  $s=1, 2, \dots, n$ ,  $n$  为大于1的自然数,  $W_s$  为  $x_s$  的权重,  $b$  为神经单元的偏置。  $f$  为神经单元的激活函数(activation functions),用于将非线性特性引入神经网络中,来将神经单元中的输入信号转换为输出信号。该激活函数的输出信号可以作为下一层卷积层的输入,激活函数可以是sigmoid函数。神经网络是将多个上述单一的神经单元联结在一起形成的网络,即一个神经单元的输出可以是另一个神经单元的输入。每个神经单元的输入可以与前一层的局部接受域相连,来提取局部接受域的特征,局部接受域可以由若干个神经单元组成的区域。

[0126] (2) 深度神经网络

[0127] 深度神经网络(deep neural network, DNN),也称多层神经网络,可以理解为具有多层隐含层的神经网络。按照不同层的位置对DNN进行划分,DNN内部的神经网络可以分为三类:输入层,隐含层,输出层。一般来说第一层是输入层,最后一层是输出层,中间的层数都是隐含层。层与层之间是全连接的,也就是说,第  $i$  层的任意一个神经元一定与第  $i+1$  层的任意一个神经元相连。

[0128] 虽然DNN看起来很复杂,但是就每一层的工作来说,其实并不复杂,简单来说就是如下线性关系表达式:  $\bar{y} = \alpha(W \cdot \bar{x} + \bar{b})$ , 其中,  $\bar{x}$  是输入向量,  $\bar{y}$  是输出向量,  $\bar{b}$  是偏移向量,  $W$  是权重矩阵(也称系数),  $\alpha(\cdot)$  是激活函数。每一层仅仅是对输入向量  $\bar{x}$  经过如此简单的操作得到输出向量  $\bar{y}$ 。由于DNN层数多,系数  $W$  和偏移向量  $\bar{b}$  的数量也比较多。这些参数在DNN中的定义如下所述:以系数  $W$  为例:假设在一个三层的DNN中,第二层的第4个神经元到第三层的第2个神经元的线性系数定义为  $W_{24}^3$ 。上标3代表系数  $W$  所在的层数,而下标对应的是输出的第三层索引2和输入的第二层索引4。

[0129] 综上,第L-1层的第k个神经元到第L层的第j个神经元的系数定义为 $W_{jk}^L$ 。

[0130] 需要注意的是,输入层是没有W参数的。在深度神经网络中,更多的隐含层让网络更能够刻画现实世界中的复杂情形。理论上而言,参数越多的模型复杂度越高,“容量”也就越大,也就意味着它能完成更复杂的学习任务。训练深度神经网络的也就是学习权重矩阵的过程,其最终目的是得到训练好的深度神经网络的所有层的权重矩阵(由很多层的向量W形成的权重矩阵)。

[0131] (3) 卷积神经网络

[0132] 卷积神经网络(convolutional neuron network,CNN)是一种带有卷积结构的深度神经网络。卷积神经网络包含了一个由卷积层和子采样层构成的特征抽取器,该特征抽取器可以看作是滤波器。卷积层是指卷积神经网络中对输入信号进行卷积处理的神经元层。在卷积神经网络的卷积层中,一个神经元可以只与部分邻层神经元连接。一个卷积层中,通常包含若干个特征平面,每个特征平面可以由一些矩形排列的神经单元组成。同一特征平面的神经单元共享权重,这里共享的权重就是卷积核。共享权重可以理解为提取图像信息的方式与位置无关。卷积核可以以随机大小的矩阵的形式初始化,在卷积神经网络的训练过程中卷积核可以通过学习得到合理的权重。另外,共享权重带来的直接好处是减少卷积神经网络各层之间的连接,同时又降低了过拟合的风险。

[0133] (4) 循环神经网络(recurrent neural networks,RNN)是用来处理序列数据的。在传统的神经网络模型中,是从输入层到隐含层再到输出层,层与层之间是全连接的,而对于每一层层内之间的各个节点是无连接的。这种普通的神经网络虽然解决了很多难题,但是却仍然对很多问题无能为力。例如,你要预测句子的下一个单词是什么,一般需要用前面的单词,因为一个句子中前后单词并不是独立的。RNN之所以称为循环神经网络,即一个序列当前的输出与前面的输出也有关。具体的表现形式为网络会对前面的信息进行记忆并应用于当前输出的计算中,即隐含层本层之间的节点不再无连接而是有连接的,并且隐含层的输入不仅包括输入层的输出还包括上一时刻隐含层的输出。理论上,RNN能够对任何长度的序列数据进行处理。对于RNN的训练和对传统的CNN或DNN的训练一样。

[0134] 既然已经有了卷积神经网络,为什么还要循环神经网络?原因很简单,在卷积神经网络中,有一个前提假设是:元素之间是相互独立的,输入与输出也是独立的,比如猫和狗。但现实世界中,很多元素都是相互连接的,比如股票随时间的变化,再比如一个人说了:我喜欢旅游,其中最喜欢的地方是云南,以后有机会一定要去。这里填空,人类应该都知道是填“云南”。因为人类会根据上下文的内容进行推断,但如何让机器做到这一步?RNN就应运而生了。RNN旨在让机器像人一样拥有记忆的能力。因此,RNN的输出就需要依赖当前的输入信息和历史的记忆信息。

[0135] (5) 损失函数

[0136] 在训练深度神经网络的过程中,因为希望深度神经网络的输出尽可能的接近真正想要预测的值,所以可以通过比较当前网络的预测值和真正想要的目标值,再根据两者之间的差异情况来更新每一层神经网络的权重向量(当然,在第一次更新之前通常会有初始化的过程,即为深度神经网络中的各层预先配置参数),比如,如果网络的预测值高了,就调整权重向量让它预测低一些,不断地调整,直到深度神经网络能够预测出真正想要的目标值或与真正想要的目标值非常接近的值。因此,就需要预先定义“如何比较预测值和目标值

之间的差异”，这便是损失函数(loss function)或目标函数(objective function)，它们是用来衡量预测值和目标值的差异的重要方程。其中，以损失函数举例，损失函数的输出值(loss)越高表示差异越大，那么深度神经网络的训练就变成了尽可能缩小这个loss的过程。

[0137] (6) 反向传播算法

[0138] 神经网络可以采用误差反向传播(back propagation, BP)算法在训练过程中修正初始的神经网络模型中参数的大小，使得神经网络模型的重建误差损失越来越小。具体地，前向传递输入信号直至输出会产生误差损失，通过反向传播误差损失信息来更新初始的神经网络模型中参数，从而使误差损失收敛。反向传播算法是以误差损失为主导的反向传播运动，旨在得到最优的神经网络模型的参数，例如权重矩阵。

[0139] 如图4所示，本申请实施例提供了一种系统架构100。在图4中，数据采集设备160用于采集训练数据。针对本申请实施例的图像处理方法来说，训练数据可以包括训练图像以及训练图像对应的分类结果，其中，训练图像的结果可以是人工预先标注的结果。

[0140] 在采集到训练数据之后，数据采集设备160将这些训练数据存入数据库130，训练设备120基于数据库130中维护的训练数据训练得到目标模型/规则101。

[0141] 下面对训练设备120基于训练数据得到目标模型/规则101进行描述，训练设备120对输入的原始图像进行处理，将输出的图像与原始图像进行对比，直到训练设备120输出的图像与原始图像的差值小于一定的阈值，从而完成目标模型/规则101的训练。

[0142] 上述目标模型/规则101能够用于实现本申请实施例的图像处理方法或者图像处理方法。本申请实施例中的目标模型/规则101具体可以为神经网络。需要说明的是，在实际的应用中，所述数据库130中维护的训练数据不一定都来自于数据采集设备160的采集，也有可能是从其他设备接收得到的。另外需要说明的是，训练设备120也不一定完全基于数据库130维护的训练数据进行目标模型/规则101的训练，也有可能从云端或其他地方获取训练数据进行模型训练，上述描述不应该作为对本申请实施例的限定。

[0143] 根据训练设备120训练得到的目标模型/规则101可以应用于不同的系统或设备中，如应用于图4所示的执行设备110，所述执行设备110可以是终端，如手机终端，平板电脑，笔记本电脑，增强现实(augmented reality, AR) AR/虚拟现实(virtual reality, VR)，车载终端等，还可以是服务器或者云端等。在图4中，执行设备110配置输入/输出(input/output, I/O)接口112，用于与外部设备进行数据交互，用户可以通过客户设备140向I/O接口112输入数据，所述输入数据在本申请实施例中可以包括：客户设备输入的待处理图像。

[0144] 预处理模块113和预处理模块114用于根据I/O接口112接收到的输入数据(如待处理图像)进行预处理，在本申请实施例中，也可以没有预处理模块113和预处理模块114(也可以只有其中的一个预处理模块)，而直接采用计算模块111对输入数据进行处理。

[0145] 在执行设备110对输入数据进行预处理，或者在执行设备110的计算模块111执行计算等相关的处理过程中，执行设备110可以调用数据存储系统150中的数据、代码等以用于相应的处理，也可以将相应处理得到的数据、指令等存入数据存储系统150中。

[0146] 最后，I/O接口112将处理结果，如上述得到的去噪处理后的图像返回给客户设备140，从而提供给用户。

[0147] 值得说明的是，训练设备120可以针对不同的目标或称不同的任务，基于不同的训

训练数据生成相应的目标模型/规则101,该相应的目标模型/规则101即可以用于实现上述目标或完成上述任务,从而为用户提供所需的结果。

[0148] 在图4中所示情况下,用户可以手动给定输入数据,该手动给定可以通过I/O接口112提供的界面进行操作。另一种情况下,客户设备140可以自动地向I/O接口112发送输入数据,如果要求客户设备140自动发送输入数据需要获得用户的授权,则用户可以在客户设备140中设置相应权限。用户可以在客户设备140查看执行设备110输出的结果,具体的呈现形式可以是显示、声音、动作等具体方式。客户设备140也可以作为数据采集端,采集如图所示输入I/O接口112的输入数据及输出I/O接口112的输出结果作为新的样本数据,并存入数据库130。当然,也可以不经过客户设备140进行采集,而是由I/O接口112直接将如图所示输入I/O接口112的输入数据及输出I/O接口112的输出结果,作为新的样本数据存入数据库130。

[0149] 值得注意的是,图4仅是本申请实施例提供的一种系统架构的示意图,图中所示设备、器件、模块等之间的位置关系不构成任何限制,例如,在图4中,数据存储系统150相对执行设备110是外部存储器,在其它情况下,也可以将数据存储系统150置于执行设备110中。

[0150] 如图4所示,根据训练设备120训练得到目标模型/规则101,该目标模型/规则101在本申请实施例中可以是本申请中的神经网络,具体的,本申请实施例提供的神经网络可以是CNN,深度卷积神经网络(deep convolutional neural networks,DCNN),循环神经网络(recurrent neural network,RNNS)等等。

[0151] 由于CNN是一种非常常见的神经网络,下面结合图5重点对CNN的结构进行详细的介绍。如上文的基础概念介绍所述,卷积神经网络是一种带有卷积结构的深度神经网络,是一种深度学习(deep learning)架构,深度学习架构是指通过机器学习的算法,在不同的抽象层级上进行多个层次的学习。作为一种深度学习架构,CNN是一种前馈(feed-forward)人工神经网络,该前馈人工神经网络中的各个神经元可以对输入其中的图像作出响应。

[0152] 本申请实施例的图像处理方法具体采用的神经网络的结构可以如图5所示。在图5中,卷积神经网络(CNN)200可以包括输入层210,卷积层/池化层220(其中池化层为可选的),以及神经网络层230。其中,输入层210可以获取待处理图像,并将获取到的待处理图像交由卷积层/池化层220以及后面的神经网络层230进行处理,可以得到图像的处理结果。下面对图5中的CNN 200中内部的层结构进行详细的介绍。

[0153] 卷积层/池化层220:

[0154] 卷积层:

[0155] 如图5所示卷积层/池化层220可以包括如示例221-226层,举例来说:在一种实现中,221层为卷积层,222层为池化层,223层为卷积层,224层为池化层,225为卷积层,226为池化层;在另一种实现方式中,221、222为卷积层,223为池化层,224、225为卷积层,226为池化层。即卷积层的输出可以作为随后的池化层的输入,也可以作为另一个卷积层的输入以继续进行卷积操作。

[0156] 下面将以卷积层221为例,介绍一层卷积层的内部工作原理。

[0157] 卷积层221可以包括很多个卷积算子,卷积算子也称为核,其在图像处理中的作用相当于一个从输入图像矩阵中提取特定信息的过滤器,卷积算子本质上可以是一个权重矩阵,这个权重矩阵通常被预先定义,在对图像进行卷积操作的过程中,权重矩阵通常在输入

图像上沿着水平方向一个像素接着一个像素(或两个像素接着两个像素……这取决于步长 stride 的取值)的进行处理,从而完成从图像中提取特定特征的工作。该权重矩阵的大小应该与图像的大小相关,需要注意的是,权重矩阵的纵深维度(depth dimension)和输入图像的纵深维度是相同的,在进行卷积运算的过程中,权重矩阵会延伸到输入图像的整个深度。因此,和一个单一的权重矩阵进行卷积会产生一个单一纵深维度的卷积化输出,但是大多数情况下不使用单一权重矩阵,而是应用多个尺寸(行×列)相同的权重矩阵,即多个同型矩阵。每个权重矩阵的输出被堆叠起来形成卷积图像的纵深维度,这里的维度可以理解为由上面所述的“多个”来决定。不同的权重矩阵可以用来提取图像中不同的特征,例如一个权重矩阵用来提取图像边缘信息,另一个权重矩阵用来提取图像的特定颜色,又一个权重矩阵用来对图像中不需要的噪点进行模糊化等。该多个权重矩阵尺寸(行×列)相同,经过该多个尺寸相同的权重矩阵提取后的卷积特征图的尺寸也相同,再将提取到的多个尺寸相同的卷积特征图合并形成卷积运算的输出。

[0158] 这些权重矩阵中的权重值在实际应用中需要经过大量的训练得到,通过训练得到的权重值形成的各个权重矩阵可以用来从输入图像中提取信息,从而使得卷积神经网络 200 进行正确的预测。

[0159] 当卷积神经网络 200 有多个卷积层的时候,初始的卷积层(例如 221)往往提取较多的一般特征,该一般特征也可以称之为低级别的特征;随着卷积神经网络 200 深度的加深,越往后的卷积层(例如 226)提取到的特征越来越复杂,比如高级别的语义之类的特征,语义越高的特征越适用于待解决的问题。

[0160] 池化层:

[0161] 由于常常需要减少训练参数的数量,因此卷积层之后常常需要周期性的引入池化层,在如图 5 中 220 所示例的 221-226 各层,可以是一层卷积层后面跟一层池化层,也可以是多层卷积层后面接一层或多层池化层。在图像处理过程中,池化层的唯一目的就是减少图像的空间大小。池化层可以包括平均池化算子和/或最大池化算子,以用于对输入图像进行采样得到较小尺寸的图像。平均池化算子可以在特定范围内对图像中的像素值进行计算产生平均值作为平均池化的结果。最大池化算子可以在特定范围内取该范围内值最大的像素作为最大池化的结果。另外,就像卷积层中用权重矩阵的大小应该与图像尺寸相关一样,池化层中的运算符也应该与图像的大小相关。通过池化层处理后输出的图像尺寸可以小于输入池化层的图像的尺寸,池化层输出的图像中每个像素点表示输入池化层的图像的对应子区域的平均值或最大值。

[0162] 神经网络层 230:

[0163] 在经过卷积层/池化层 220 的处理后,卷积神经网络 200 还不足以输出所需要的输出信息。因为如前所述,卷积层/池化层 220 只会提取特征,并减少输入图像带来的参数。然而为了生成最终的输出信息(所需要的类信息或其他相关信息),卷积神经网络 200 需要利用神经网络层 230 来生成一个或者一组所需要的类的数量的输出。因此,在神经网络层 230 中可以包括多层隐含层(如图 5 所示的 231、232 至 23n)以及输出层 240,该多层隐含层中所包含的参数可以根据具体的任务类型的相关训练数据进行预先训练得到,例如该任务类型可以包括图像识别,图像分类,图像超分辨率重建等等。

[0164] 在神经网络层 230 中的多层隐含层之后,也就是整个卷积神经网络 200 的最后层为

输出层240,该输出层240具有类似分类交叉熵的损失函数,具体用于计算预测误差,一旦整个卷积神经网络200的前向传播(如图5由210至240方向的传播为前向传播)完成,反向传播(如图5由240至210方向的传播为反向传播)就会开始更新前面提到的各层的权重值以及偏差,以减少卷积神经网络200的损失,及卷积神经网络200通过输出层输出的结果和理想结果之间的误差。

[0165] 本申请实施例的图像处理方法具体采用的神经网络的结构可以如图6所示。在图6中,卷积神经网络(CNN)200可以包括输入层110,卷积层/池化层120(其中池化层为可选的),以及神经网络层130。与图5相比,图6中的卷积层/池化层120中的多个卷积层/池化层并行,将分别提取的特征均输入给全神经网络层130进行处理。

[0166] 需要说明的是,图5和图6所示的卷积神经网络仅作为一种本申请实施例的图像处理方法的两种可能的卷积神经网络的示例,在具体的应用中,本申请实施例的图像处理方法所采用的卷积神经网络还可以以其他网络模型的形式存在。

[0167] 另外,采用本申请实施例的神经网络结构的搜索方法得到的卷积神经网络的结构可以如图5和图6中的卷积神经网络结构所示。

[0168] 图7为本申请实施例提供的一种芯片的硬件结构,该芯片包括神经网络处理器50。该芯片可以被设置在如图1所示的执行设备110中,用以完成计算模块111的计算工作。该芯片也可以被设置在如图1所示的训练设备120中,用以完成训练设备120的训练工作并输出目标模型/规则101。如图2所示的卷积神经网络中各层的算法均可在如图7所示的芯片中得以实现。

[0169] 神经网络处理器NPU 50NPU作为协处理器挂载到主中央处理器(central processing unit,CPU)(host CPU)上,由主CPU分配任务。NPU的核心部分为运算电路50,控制器504控制运算电路503提取存储器(权重存储器或输入存储器)中的数据并进行运算。

[0170] 在一些实现中,运算电路503内部包括多个处理单元(process engine,PE)。在一些实现中,运算电路503是二维脉动阵列。运算电路503还可以是一维脉动阵列或者能够执行例如乘法和加法这样的数学运算的其它电子线路。在一些实现中,运算电路503是通用的矩阵处理器。

[0171] 举例来说,假设有输入矩阵A,权重矩阵B,输出矩阵C。运算电路从权重存储器502中取矩阵B相应的数据,并缓存在运算电路中每一个PE上。运算电路从输入存储器501中取矩阵A数据与矩阵B进行矩阵运算,得到的矩阵的部分结果或最终结果,保存在累加器(accumulator)508中。

[0172] 向量计算单元507可以对运算电路的输出做进一步处理,如向量乘,向量加,指数运算,对数运算,大小比较等等。例如,向量计算单元507可以用于神经网络中非卷积/非FC层的网络计算,如池化(pooling),批归一化(batch normalization),局部响应归一化(local response normalization)等。

[0173] 在一些实现种,向量计算单元能507将经处理的输出的向量存储到统一缓存器506。例如,向量计算单元507可以将非线性函数应用到运算电路503的输出,例如累加值的向量,用以生成激活值。在一些实现中,向量计算单元507生成归一化的值、合并值,或二者均有。在一些实现中,处理过的输出的向量能够用作到运算电路503的激活输入,例如用于在神经网络中的后续层中的使用。

[0174] 统一存储器506用于存放输入数据以及输出数据。

[0175] 权重数据直接通过存储单元访问控制器505(direct memory access controller,DMAC)将外部存储器中的输入数据搬运到输入存储器501和/或统一存储器506、将外部存储器中的权重数据存入权重存储器502,以及将统一存储器506中的数据存入外部存储器。

[0176] 总线接口单元(bus interface unit,BIU)510,用于通过总线实现主CPU、DMAC和取指存储器509之间进行交互。

[0177] 与控制器504连接的取指存储器(instruction fetch buffer)509,用于存储控制器504使用的指令;

[0178] 控制器504,用于调用取指存储器509中缓存的指令,实现控制该运算加速器的工作过程。

[0179] 入口:可以根据实际发明说明这里的数据是说明数据,比如探测到车辆速度?障碍物距离等

[0180] 一般地,统一存储器506,输入存储器501,权重存储器502以及取指存储器509均为片上(On-Chip)存储器,外部存储器为该NPU外部的存储器,该外部存储器可以为双倍数据率同步动态随机存储器(double data rate synchronous dynamic random access memory,简称DDR SDRAM)、高带宽存储器(high bandwidth memory,HBM)或其他可读可写的存储器。

[0181] 其中,图2所示的卷积神经网络中各层的运算可以由运算电路303或向量计算单元307执行。

[0182] 上文中介绍的图4中的执行设备110能够执行本申请实施例的图像处理方法或者图像处理方法的各个步骤,图5和图6所示的CNN模型和图7所示的芯片也可以用于执行本申请实施例的图像处理方法或者图像处理方法的各个步骤。下面结合附图对本申请实施例的图像处理方法和本申请实施例的图像处理方法进行详细的介绍。

[0183] 如图8所示,本申请实施例提供了一种系统架构300。该系统架构包括本地设备301、本地设备302以及执行设备210和数据存储系统250,其中,本地设备301和本地设备302通过通信网络与执行设备210连接。

[0184] 执行设备210可以由一个或多个服务器实现。可选的,执行设备210可以与其它计算设备配合使用,例如:数据存储系统、路由器、负载均衡器等设备。执行设备210可以布置在一个物理站点上,或者分布在多个物理站点上。执行设备210可以使用数据存储系统250中的数据,或者调用数据存储系统250中的程序代码来实现本申请实施例的搜索神经网络结构的方法。

[0185] 具体地,执行设备210可以执行以下过程:确定搜索空间和多个构建单元;堆叠所述多个构建单元,以得到搜索网络,所述搜索网络是用于搜索神经网络结构的神经网络;在所述搜索空间内对所述搜索网络中的构建单元的网络结构进行优化,以得到优化后的构建单元,其中,在优化过程中搜索空间逐渐减小,构建单元数量逐渐增加,搜索空间的减小和构建单元数量的增加使得所述优化过程中产生的显存消耗在预设范围内;根据所述优化后的构建单元搭建所述目标神经网络。

[0186] 通过上述过程执行设备210能够搭建成一个目标神经网络,该目标神经网络可以

用于图像分类或者进行图像处理等等。

[0187] 用户可以操作各自的用户设备(例如本地设备301和本地设备302)与执行设备210进行交互。每个本地设备可以表示任何计算设备,例如个人计算机、计算机工作站、智能手机、平板电脑、智能摄像头、智能汽车或其他类型蜂窝电话、媒体消费设备、可穿戴设备、机顶盒、游戏机等。

[0188] 每个用户的本地设备可以通过任何通信机制/通信标准的通信网络与执行设备210进行交互,通信网络可以是广域网、局域网、点对点连接等方式,或它们的任意组合。

[0189] 在一种实现方式中,本地设备301、本地设备302从执行设备210获取到目标神经网络的相关参数,将目标神经网络部署在本地设备301、本地设备302上,利用该目标神经网络进行图像分类或者图像处理等等。

[0190] 在另一种实现中,执行设备210上可以直接部署目标神经网络,执行设备210通过从本地设备301和本地设备302获取待处理图像,并根据目标神经网络对待处理图像进行分类或者其他类型的图像处理。

[0191] 上述执行设备210也可以称为云端设备,此时执行设备210一般部署在云端。

[0192] 下面先结合图9对本申请实施例的神经网络的构建方法进行详细的介绍。图9所示的方法可以由神经网络构建装置来执行,该神经网络构建装置可以是电脑、服务器等运算能力足以用来神经网络构建装置。

[0193] 图9所示的方法包括步骤1001至1004,下面分别对这些步骤进行详细的描述。

[0194] 1001、确定搜索空间和多个构建单元。

[0195] 其中,上述搜索空间是根据待构建的目标神经网络的应用需求确定的。具体地,上述搜索空间可以是根据目标神经网络的处理数据的类型确定的。

[0196] 例如,当上述目标神经网络用于处理图像数据的神经网络时,上述搜索空间包含的操作的种类和数量要与图像数据的处理相适应;当上述目标神经网络用于处理语音数据时,上述搜索空间包含的操作的种类和数量要与语音数据的处理相适应。

[0197] 上述多个构建单元是根据搜索空间以及构建目标神经网络的设备的显存资源的大小确定的。另外,本申请中的构建单元是由多个节点之间通过神经网络的基本操作连接得到的一种网络结构,构建单元是用于构建神经网络的基础模块。

[0198] 如图10所示,位于虚线框内的3个节点(节点0,节点1和节点2)构成了一个构建单元,该构建单元可以接收节点 $c_{k-2}$ 和 $c_{k-1}$ 输出的数据( $c_{k-2}$ 和 $c_{k-1}$ 也可以是符合要求的特征图,例如, $c_{k-2}$ 和 $c_{k-1}$ 可以是输入图像经过一定的卷积处理后得到的特征图),并由节点0和1分别对输入的数据进行处理,其中,节点0输出的数据还会输入到节点1中进行处理,节点0和节点1输出的数据会送入到节点2中进行处理,节点2最终输出该构建单元处理完的数据。

[0199] 另外,上述节点 $c_{k-2}$ 和 $c_{k-1}$ 可以视为输入节点,这两个节点会向构建单元输入待处理的数据,而在构建单元内部,0和1是中间节点,节点2是输出节点。

[0200] 图10中的粗箭头表示一个或者多个基本操作,汇入同一个中间节点的基本操作运算结果在该中间节点处相加,图10中的细箭头表示通道维度的特征图连接,输出节点2输出的特征图由2个中间节点(节点0和节点1)的输出按照顺序在特征图通道维度连接而成。

[0201] 应理解,图10中的粗箭头和细箭头所对应的操作应是特定情况下涉及的操作,这

里的相加和通道维度连接在此处都是为卷积神经网络而设计的,在其他情况下,构建单元的节点之间所对应的操作也可以是其他类型的运算或操作。

[0202] 上述搜索空间包含的可以是预先设定好的卷积神经网络中的基础运算或者基础运算的组合,这些基础运算或者基础运算的组合可以统称为基本操作。

[0203] 上述搜索空间可以包含以下8种基本操作:

[0204] (1) 池化核大小为 $3 \times 3$ 的均值池化 (avg\_pool\_3x3);

[0205] (2) 池化核大小为 $3 \times 3$ 的最大值池化 (max\_pool\_3x3);

[0206] (3) 卷积核大小为 $3 \times 3$ 的分离卷积 (sep\_conv\_3x3);

[0207] (4) 卷积核大小为 $5 \times 5$ 的分离卷积 (sep\_conv\_5x5);

[0208] (5) 卷积核大小为 $3 \times 3$ 且空洞率为2的空洞卷积 (dil\_conv\_3x3);

[0209] (6) 卷积核大小为 $5 \times 5$ 且空洞率为2的空洞卷积 (dil\_conv\_5x5);

[0210] (7) 跳连接操作;

[0211] (8) 置零操作 (Zero, 相应位置所有神经元置零)。

[0212] 1002、堆叠多个构建单元,以得到搜索网络。

[0213] 上述搜索网络是用于搜索神经网络结构的神经网络。

[0214] 可选地,上述堆叠多个构建单元,以得到搜索网络,包括:按照预设的堆叠方式将所述多个构建单元依次堆叠起来,以得到搜索网络,其中,在该搜索网络中,位于搜索网络前面的构建单元的输出是位于搜索网络的后面的构建单元的输入。

[0215] 上述预设的堆叠方式可以包括在什么位置堆放什么类型的构建单元以及堆叠的数量等等。

[0216] 1003、在搜索空间内对搜索网络中的构建单元的网络结构进行优化,以得到优化后的构建单元。

[0217] 其中,对搜索网络中的构建单元的网络结构进行优化的优化过程可以包括N个阶段,第i个阶段和第j个阶段为N个阶段中的任意两个阶段,搜索空间在第i个阶段的大小大于搜索网络在第j个阶段的大小,搜索网络在第i个阶段时包含的构建单元的数量小于搜索空间在第j个阶段时包含的构建单元的数量,搜索网络的搜索空间的减小和搜索网络的构建单元数量的增加使得优化过程中产生的显存消耗在预设范围内。

[0218] 另外,在优化过程结束后,搜索网络在第N个阶段包含的构建单元的数量与目标神经网络包含的构建单元的数量差异在预设范围内,上述目标神经网络包含的构建单元的数量是根据目标神经网络的应用需求确定的,N为大于1的正整数,i和j均为小于或者等于N的正整数,并且i小于j。

[0219] 可选地,上述显存资源可以替换为缓存资源,该缓存资源是用于构建神经网络的设备在优化过程中用于存放运算数量的内存或者存储单元。

[0220] 上述缓存资源具体可以包括显存资源。

[0221] 可选地,上述目标神经网络包含的构建单元的数量是根据目标神经网络要处理的数据类型和/或计算的复杂度确定的。

[0222] 例如,当上述目标神经网络用于处理一些简单的文本数据时,目标神经网络包含较少数量的构建单元即可,当上述目标神经网络用于处理一些比较复杂的图像数据时,目标神经网络需要包含数量较多的构建单元。

[0223] 再如,当目标神经网络需要处理的数据复杂度较高时,目标神经网络需要包含数量较多的构建单元;当目标神经网络需要处理的数据复杂度较低时,目标神经网络需要较少数量的构建单元即可。

[0224] 可选地,上述N的大小是预先设置好的。

[0225] 上述N的大小可以根据目标神经网络的构建需求来确定。具体地,当目标神经网络需要在较短的时间内构建完成时,可以将N设置成一个较小的数值,当目标神经网络可以在较长的时间内构建完成时,可以将N设置成一个较大的数值。

[0226] 1004、根据优化后的构建单元搭建目标神经网络。

[0227] 本申请中,在对构建单元的网络结构进行优化的过程中,减少搜索空间节省的显存资源可以用来增加构建单元的数量,从而能够在在显存资源有限的情况下,尽可能的堆叠得到构建单元数量与最终要搭建的目标神经网络的构建单元数量比较接近的搜索网络。使得搜索网络优化后的构建单元能够更好地适用于目标神经网络的搭建,进而使得根据优化后的构建单元搭建成的目标神经网络能够更好地满足应用需求。

[0228] 具体而言,本申请在搜索网络的构建单元的网络结构的优化过程中,通过逐渐减少搜索空间的大小,并增加搜索网络的构建单元的数量,能够在构建出能够较好满足应用需求的目标神经网络的情况下,减少优化过程显存资源的依赖,使得在优化过程中仅仅依赖较少的显存资源就能够得到较好的满足应用需求的目标神经网络,也在一定程度上提高了显存资源的利用率。

[0229] 一般地,如果搜索网络的网络深度与待构建的目标神经网络的网络深度比较接近时,搜索网络中优化得到的构建单元比较适合用于搭建目标神经网络。神经网络的深度与包含的构建单元的数量是正相关的关系,因此,当搜索网络的构建单元的数量与目标神经网络的构建单元数量比较接近时,搜索网络的网络深度与目标神经网络的网络深度也比较接近。

[0230] 应理解,在上述优化过程中,从第i个阶段到第j个阶段,搜索空间的变小,搜索网络的构建单元的数量增加,并且,从第i个阶段到第j个阶段,搜索空间减小的幅度可以是构建单元数量增加的幅度是相同的。从第i个阶段到第j个阶段搜索空间减小的幅度可以根据第i个阶段到第j个阶段搜索网络的构建单元的数量增加的数量来确定,或者,从第i个阶段到第j个阶段搜索网络的构建单元的数量增加的数量可以根据第i个阶段到第j个阶段搜索空间减小的幅度来确定。

[0231] 或者,也可以结合显存资源的大小,来一起确定第i个阶段到第j个阶段搜索空间减小的幅度以及第i个阶段到第j个阶段搜索网络的构建单元的数量增加的数量。

[0232] 可选地,上述搜索空间在第i个阶段的大小为 $S_i$ ,上述搜索空间在第j个阶段的大小为 $S_j$ ,上述搜索网络在第i个阶段包含的构建单元数量为 $L_i$ 个,上述搜索网络在第j个阶段包含的构建单元数量为 $L_j$ 个,其中,上述 $L_j-L_i$ 大小是根据 $S_i-S_j$ 的大小确定的,或者,上述 $S_i-S_j$ 的大小是根据 $L_j-L_i$ 的大小确定的。

[0233] 具体地,在上述两个阶段中,可以预先设定 $S_i-S_j$ 的大小,然后再根据 $S_i-S_j$ 的大小确定 $L_j-L_i$ 的大小,使得由于搜索空间减小节省的显存资源与构建单元增加导致多消耗的显存资源的差值在一定阈值范围内(两者的差值越小越好)。

[0234] 在上述两个阶段中,也可以预先设定 $L_j-L_i$ 的大小,然后再根据 $L_j-L_i$ 的大小确定 $S_i-S_j$

$S_j$ 的大小,使得由于构建单元增加导致多消耗的显存资源与搜索空间减小节省的显存资源的差值在一定阈值范围内(两者的差值越小越好)。

[0235] 在本申请中,只要上述N个阶段中存在至少两个阶段满足搜索空间减小,构建单元数量增加即可,而不必使得每两个相邻的阶段都满足搜索空间减小,构建单元数量增加的要求。

[0236] 例如,上述 $N=4$ ,第2个阶段相比于第1阶段以及第4阶段相对于第3阶段均满足:搜索空间减小,搜索网络的构建单元数量增加。而第2个阶段和第3个阶段的搜索空间以及搜索网络包含的构建单元数量均没有发生变化。

[0237] 当然,如果上述N个阶段中的每两个阶段之间都满足搜索空间减小,构建单元数量增加,那么,将会使得N个阶段的优化过程中搜索空间与构建单元的数量变化更加平滑。

[0238] 可选地,上述 $i$ 和 $j$ 满足: $j=i+1$ 。

[0239] 当 $j=i+1$ 时,在优化过程中,任意两个相邻阶段之间都会满足搜索空间逐渐减小,搜索网络的构建单元逐渐增加,使得优化过程比较平稳。

[0240] 可选地,上述N个阶段中,搜索网络在任意两个相邻阶段的构建单元的数量变化值相同,搜索空间在任意两个相邻阶段的大小变化值也相同。

[0241] 在上述优化过程中,构建单元的数量变化以及搜索空间的大小变化都是均匀的,优化的过程更加平稳。

[0242] 可选地,上述第 $i+1$ 个阶段相对于第 $i$ 个阶段增加的构建单元的数量可以是根据上述数值N,以及搜索网络在优化前包含的构建单元的数量,以及目标神经网络中的构建单元的数量来确定的。

[0243] 例如,搜索网络在第 $i+1$ 个阶段相对于第 $i$ 个阶段增加的构建单元的数量为 $X$ ,优化开始前搜索网络包含的构建单元数量为 $U$ ,目标神经网络中的构建单元的数量为 $V$ ,那么, $X$ 可以根据公式 $X=(U-V)/N$ 计算得到。

[0244] 应理解,在优化过程中,搜索空间的大小降低的幅度以及搜索网络构建单元数量的增加幅度可以根据多种方式来确定,只要能够确保优化过程中搜索网络的搜索空间的减小和所述搜索网络的构建单元数量的增加使得所述优化过程中产生的显存消耗在预设范围内即可。

[0245] 在实际应用中可以先预先设定搜索空间大小降低的幅度,然后再确定搜索网络构建单元数量增加的幅度;也可以预先设定搜索网络的大小,再确定搜索空间大小降低的幅度。本申请对此不做限定,所有确保显存消耗在预设范围内的实现方式都在本申请的保护范围内。

[0246] 可选地,优化后的构建单元的各个节点之间的连接关系中包含的第一类操作的数量在预设范围内,第一类操作是不包含神经网络可训练参数的操作。

[0247] 本申请通过将第一类操作的数量限制在一定范围,使得最终搭建的目标神经网络的可训练参数保持在相对稳定的水平,进而使得目标神经网络的性能保持稳定。

[0248] 应理解,还可以将第一类操作的数量具体限制到某一个数值,这样使得最终得到的目标神经网络中包含固定数量的第一类操作,使得目标神经网络的性能更加稳定。

[0249] 具体地,上述第一类操作是不包含可训练参数的操作,如果此类操作过多会导致包含可训练参数的其他操作较少,从而神经网络总体的可训练参数较少,神经网络的特征

表达能力降低。

[0250] 由于在构建单元数量较多的搜索网络中进行结构搜索稳定性不足,会导致每次搜索得到的构建单元中第一类操作的数量具有一定的差异,搜索得到的神经网络结构(即构建单元)在相应任务上的性能表现波动。限制第一类操作的数量可以使得由搜索得到的神经网络结构搭建的测试网络的训练参数保持在相对稳定的水平,从而减小在相应任务上的性能波动。

[0251] 为了使得优化后的构建单元的各个节点之间的连接关系中包含的第一类操作的数量在预设范围内,可以在优化过程中对第一类操作的数量进行限制。

[0252] 假设将第一类操作的数量直接限制为第一数量,那么,在上述优化过程中,如果构建单元中第一类操作的数量为第一数量,那么,在优化过程中不更改第一类操作的数量;而如果构建单元中第一类操作的数量大于第一数量,那么,可以在优化过程中删减掉部分第一操作使得删减后第一类操作的数量等于第一数量;如果构建单元中第一类操作的数量小于第一类操作,那么,可以在优化过程中增加构建单元的数量,使得优化后构建单元的数量为第一数量。

[0253] 上述将第一类操作限定在固定数量的过程可以称为第一类操作的数量的规范流程,下面对第一类操作的数量的规范流程进行详细介绍。

[0254] 第一类操作的数量的规范流程可以根据预先制定的规范规则,在一种类型的构建单元中保留 $M_c$ 个第一类操作。

[0255] 具体地,如果该类型的构建单元中的第一类操作的数量等于 $M_c$ ,则直接将输入的构建单元结构输出;否则,执行如下流程:将该类型的构建单元所对应网络结构参数中与第一类操作所对应的相应网络结构参数进行降序排序,若第一类操作的数量小于 $M_c$ 则根据网络结构生成规则将不在构建单元中的权值最大且符合网络结构生成规则的第一类操作加入构建单元结构,并相应地根据网络结构生成规则以及网络结构参数删除被替换的相应基本操作;若第一类操作的数量大于 $M_c$ 则从构建单元结构中移除权值最小的第一类操作的,并根据网络结构生成规则以及网络结构参数加入对应的其他基本操作;重复执行本流程直到该种构建单元中的第一类操作的数量等于 $M_c$ 。

[0256] 上述第一类操作具体可以是跳连接(skip-connect)操作,也可以是置零操作。

[0257] 上述搜索网络可以包含多种类型的构建单元,下面简单介绍一下搜索网络包含的常见的构建单元。

[0258] 可选地,上述搜索网络中的构建单元包括第一类构建单元。

[0259] 其中,第一类构建单元是输入特征图的数量(具体可以是通道数)和大小分别与输出特征图的数量和大小相同的构建单元。

[0260] 例如,某个第一类构建单元的输入的是大小为 $C \times D_1 \times D_2$ ( $C$ 为通道数, $D_1$ 和 $D_2$ 分别是宽和高)的特征图,经过该第一类构建单元处理后输出的特征图的大小仍然是 $C \times D_1 \times D_2$ 。

[0261] 上述第一类构建单元具体可以是普通单元(normal cell)

[0262] 可选地,上述搜索网络中的构建单元包括第二类构建单元。

[0263] 其中,第二类构建单元的输出特征图的分辨率是输入特征图的 $1/M$ ,第二类构建单元的输出特图的数量是输入特征图的数量 $M$ 倍, $M$ 为大于1的正整数。

[0264] 上述M的取值一般可以是2、4、6和8等数值。

[0265] 例如,某个第二类构建单元的输入是1个大小为 $C \times D1 \times D2$  ( $C$ 为通道数, $D1$ 和 $D2$ 分别是宽和高, $C1$ 和 $C2$ 的乘积可以表示特征图的分辨率)的特征图,那么,经过该第二类构建单元处理后,得到的1个大小为 $4C \times (\frac{1}{2}D1 \times \frac{1}{2}D2)$ 的特征图。

[0266] 上述第二类构建单元具体可以是下采样单元(reduction cell)。

[0267] 当搜索网络由上述第一类构建单元和第二类构建单元组成时,搜索网络的结构可以如图11所示。

[0268] 如图11,搜索网络由5个构建单元依次堆叠而成,其中,位于搜索网络最前端和最后端的是第一类构建单元,每两个第一构建单元之间存在一个第二类构建单元。

[0269] 图11中的搜索网络中的第一个构建单元能够对输入的图像进行处理,第一类构建单元对图像进行处理后,将处理得到的特征图输入到第二类构建单元进行处理,这样依次向后传输,直到搜索网络中的最后一个第一类构建单元输出特征图。

[0270] 搜索网络的最后一个第一类构建单元输出的特征图送入到分类器中进行处理,由分类器根据特征图对图像进行分类。

[0271] 为了更好地理解本申请实施例的神经网络结构搜索方法,下面结合图12对本申请实施例的神经网络结构搜索方法的整体过程进行简单的介绍。

[0272] 如图12所示,可以根据待构建的神经网络的任务需求(也就是待构建的神经网络需要处理任务的任务类型)来确定构建何种类型的神经网络。接下来,再根据该神经网络处理的任务需求,确定搜索空间的大小和构建单元的数量,并对构建单元进行堆叠,得到搜索网络。在得到搜索网络之后就可以对搜索网络中的构建单元的网络结构进行优化(在优化过程中可以采用训练数据进行优化)了,对构建单元的网络结构进行优化可以分为渐进网络结构搜索和操作数量规范流程(就是将某一操作的数量限定在一定的范围内,在本申请中,主要是将第一类操作的数量限制在一定的范围内)。其中,渐进网络结构搜索就是在优化过程中逐渐减小搜索空间的大小,并逐渐增加构建单元的数量以获得与待构建的神经网络的构建单元数目比较接近的搜索网络(具体过程可以参见上文中的图9所示的方法中的相关描述)。操作数量规范流程可以用于确保优化后的构建单元中的第一类操作接的数量在一定的预设范围内。该渐进的网络结构搜索和操作数量规范流程相当于图9所示的方法中的步骤1003的优化过程。

[0273] 本申请实施例的神经网络的构建方法可以由神经网络构建系统来执行,图13示出了神经网络构建系统执行本申请实施例的神经网络结构搜索方法的过程。下面对图13所示的内容进行详细介绍。

[0274] 图13所示的神经网络构建系统主要包括操作仓库101、渐进网络结构搜索模块102,操作数量规范模块103构成。

[0275] 其中,操作仓库101可以包含预先设定好的卷积神经网络中的基本操作。渐进网络结构搜索模块102用于对搜索网络的构建单元的网络结构进行优化,在优化过程中,通过增加构建单元1021的堆叠数量,并减小搜索空间的大小来不断的更新搜索网络1022本身,从而实现对搜索网络的构建单元的网络结构的不断优化。

[0276] 操作数量规范模块103主要是将某一操作的数量限定在一定的范围内,在本申请

中,操作数量规范模块103主要是将第一类操作的数量限制在一定的范围内。

[0277] 具体地,可以根据目标任务确定操作仓库101(相当于上文中的搜索空间)的大小和初始数量的构建单元103,然后根据初始数量的构建单元103堆叠得到搜索网络。接下来,可以采用渐进结构搜索模块102对搜索网络进行优化,在优化过程中,逐渐减小搜索空间的大小,增加堆叠单元的数量,得到构建单元。接下来,再通过操作数量规范模块103将渐进网络结构搜索模块102得到的构建单元中的第一类操作限制在一定范围内,从而得到优化后的构建单元,这些优化后构建单元就可以用于搭建最终需要的目标神经网络了。

[0278] 在图13中,渐进网络结构搜索模块102和操作数量规范模块103处理的过程相当于图9所示的方法中的步骤1003中的优化过程。具体优化过程可参见步骤1003的相关描述。

[0279] 渐进网络结构搜索模块102进行优化操作的具体过程可以如图14所示。为了方便示意,图14对实际操作进行了一定的简化,只展示了第一类构建单元(具体可以是normal cell)的搜索过程,并且对第一类构建单元的具体示意图也进行了简化,只展示搜索过程,不代表具体结构。图中每一个箭头线代表一种基本操作,操作种类数量在示意图中做了简化;数字方框代表节点,本例中的节点为卷积神经网络的特征图。为了方便展示,我们将每一阶段的节点0和节点1及其对应的基本操作进行了特别展示,如各子图右上部分所示。

[0280] 在初始阶段,节点之间的连接由预先定义的搜索空间中所有可能的基本操作组成,图14中采用了5种基本操作,分别由5个带箭头线表示。由构建单元搭建而成的搜索网络拥有 $B_1=5$ 个构建单元,包括3个第一类构建单元和2个第二类构建单元。每种构建单元共享相同的操作和网络结构权值。经过网络参数和网络结构参数的优化,得到学习后的网络结构参数。第一类构建单元的节点0和节点1之间的相应基本操作的权值分别为0.21、0.26、0.18、0.03和0.32(图14中为示出这些权值)。根据预先设定的基本操作删除数量,可以删除权值最小的一个或多个操作。权值最小的一个箭头线(如图14中初始阶段所示,该权值最小的箭头线是节点0到节点1之间的第4个箭头线)所代表基本操作被删除,其余操作在本阶段输出的构建单元结构中被保留。注意,不同节点可以根据相应的网络结构权值进行操作,所保留的基本操作不一定相同。

[0281] 本例中的中间阶段只有一个。上一阶段(初始阶段)生成的构建单元结构,每一组节点对之间删去了1种基本操作,剩余4种基本操作。本阶段的搜索网络由上一阶段输出的构建单元搭建而成,拥有更多( $B_2=11$ )构建单元。上一阶段在构建单元中删除的基本操作所减少的显存消耗提供了搭建一个构建单元数量更多的搜索网络所需要的额外显存开销,在合理设计的情况下,可以保持稳定的显存使用率。本阶段继续执行与初始阶段相似的搜索网络搭建、网络结构参数优化与构建单元结构生成流程。在本阶段中,节点0和节点1之间的第三个和第四个(按照从左到右的顺序)带箭头线所代表的基本操作在生成的构建单元结构中被删除。

[0282] 在最终阶段,同样执行与之前阶段相似的搜索网络搭建( $B_2=17, O_3=2$ )、网络结构参数优化与构建单元生成流程。在本阶段的构建单元生成流程中,施加除与其他阶段相同的构建单元生成规则之外的额外规则网络结构生成规则,使得生成的构建单元结构具有与相应任务所匹配的结构特点。在本例中,此规则是每一个节点至多保留两种输入基本操作,即根据这一规则以及相应的网络结构参数,节点1和节点3之间的所有基本操作均不保留。最终生成的第一类构建单元如图14中的最终阶段中加粗的带箭头线与相应节点所示。

生成的构建单元结构和相应网络结构参数及其相对应的操作种类一起输出给后续模块或流程。

[0283] 上述图13中所示的操作数量规范流程模块103用于将第一类操作的数量约束在固定范围内(具体可以是将第一类操作的数量直接约束为某个数值),下面结合图15对操作数量规范流程模块103执行的具体过程进行描述。

[0284] 图15是本申请实施例的操作数量规范模块的处理过程的示意图。如图15所示,输入为渐进网络结构搜索模块输出的构建单元结构和相应网络结构参数及其对应操作种类,输出为操作数量规范后的构建单元结构。应理解,经过操作数量规范模块103处理后输出的构建单元结构中的第一类操作的数量限定在了一个固定数量。

[0285] 上述操作数量规范模块103的具体执行过程包括:

[0286] S1、判断输入的构建单元结构中的第一类操作的数量 $M_c$ 是否等于预先设定的固定数量 $M$ ,若 $M_c=M$ ,则直接输出该构建单元结构, $M_c \neq M$ ,则继续执行步骤S2。

[0287] S2、若 $M_c > M$ ,则根据相应的网络结构参数和网络结构生成规则,将该构建单元结构中权值最小的一个第一类操作替换成符合网络结构生成规则的其他基本操作;

[0288] 若 $M_c < M$ ,则根据网络结构参数和网络结构生成规则,将不属于该构建单元结构的符合网络结构生成规则的权值最大的第一类操作替换相应的其他种类基本操作。

[0289] S3、在步骤S2生成了构建单元之后,再将该构建单元送入到步骤S1中,继续进行判断。

[0290] 当S1中判断得到的结果是 $M_c=M$ 的话,则输出构建单元结构,否则继续执行步骤S2和S3。

[0291] 应理解,在上述图15所示的过程中,第一类操作具体可以是跳连接操作。

[0292] 为了与现有的神经网络的构建方法的性能进行对比,表1给出了相似约束下使用本申请实施例的神经网络的构建方法构建得到的神经网络和使用其他方法设计或搜索得到的神经网络在图像分类数据集上的分类准确率。

[0293] 为了对对不同方案的搜索效率,表1中同时给出了对比的神经网络结构的搜索时间。具体地,表1中的CIFAR10、CIFAR100、ImageNetTop1、ImageNetTop5分别表示分类准确率,其中,CIFAR10、CIFAR100和ImageNet分别是不同的数据集,Top1和Top5是子指标,指在前1个或者5个结果中出现正确结果的比例(准确率)。NASNet-A、AmoebaNet-B、ENAS、PNAS、和DARTS (2ND) 分别表示不同的网络结构,搜索开销的大小可以用在单个GPU运行时所需要的时间(这里时间一般用天表示)来表示。

[0294] 从表1中可以看出,使用本申请实施例的神经网络的构建方法构建得到的神经网络的分类准确率比其他方法设计或搜索得到的神经网络在图像分类数据集上的分类准确率更高,并且搜索开销更小,能够在搜索过程节省更多的资源。

[0295] 表1

[0296]

网络结构	CIFAR10	CIFAR100	ImageNetTop1	ImageNetTop5	搜索开销
NASNet-A	97.35	-	74.0	91.6	1800
AmoebaNet-B	97.45	-	74.0	91.5	3150
ENAS	97.11	-	-	-	0.5

PNAS	96.59	-	74.2	91.9	225
DARTS (2ND)	97.17	82.46	73.1	91.0	4
SNAS	97.15	-	72.3	90.8	1.5
本申请	97.45	83.48	75.6	92.6	0.2

[0297] 此外,本申请所提出的神经网络的构建方法中的操作数量规范流程能够有效地提高搜索的稳定性,提升构建得到的神经网络的性能。表2给出了使用操作数量规范流程前和使用操作数量规范流程后的得到的神经网络结构在公开数据集上的性能对比,其中,Run 1表示第一次试验的准确率,Run 2表示第二次试验的准确率,Run 3表示第三次试验的准确率。表2中的平均准确率表示的是第一次试验至第三次试验的平均准确率,表2中的标准差是第一次试验至第三次试验的准确率的标准差。从表2中可以看出,使用操作数量规范流程后得到的神经网络较使用前在性能和稳定性上有明显的提升。

[0298] 表2

[0299]

指标	使用前	使用后
Run 1	97.03	97.31
Run 2	97.42	97.42
Run 3	97.18	97.31
平均准确率	97.21	97.35
标准差	0.16	0.05

[0300] 上文结合附图对本申请实施例的神经网络的构建方法进行了详细的介绍,本申请实施例的神经网络的构建方法构建得到的神经网络可以用于图像处理(例如,图像分类)等,下面对这些具体应用进行介绍。

[0301] 图16是本申请实施例的图像处理方法的示意性流程图。图16所示的方法包括:

[0302] 2001、获取待处理图像;

[0303] 2002、根据目标神经网络对待处理图像进行分类,得到待处理图像的分类结果。

[0304] 其中,上述目标神经网络可以是根据图9所示的方法构建得到的。

[0305] 本申请中,在目标神经网络构建之前的优化过程中,通过减小搜索空间的大小,增加构建单元的数量,能够尽可能的堆叠得到构建单元数量与最终要搭建的目标神经网络的构建单元数量比较接近的搜索网络。从而使得搜索网络优化后的构建单元能够更好地适用于目标神经网络的搭建,能够获得性能更好的目标神经网络,利用该目标神经网络进行图像分类能够取得较好的图像分类效果(例如,分类结果更准确)。

[0306] 可选地,上述 $j=i+1$ 。

[0307] 当 $j=i+1$ 时,在优化过程中,任意两个相邻阶段之间都会满足搜索空间逐渐减小,搜索网络的构建单元逐渐增加,使得优化过程比较平稳。

[0308] 可选地,上述N个阶段中,搜索网络在任意两个相邻阶段的构建单元的数量变化值相同,搜索空间在任意两个相邻阶段的大小变化值也相同。

[0309] 在上述优化过程中,构建单元的数量变化以及搜索空间的大小变化都是均匀的,优化的过程更加平稳。

[0310] 可选地,上述目标神经网络是经过训练图片进行训练得到的神经网络。

[0311] 具体地,可以通过训练图片以及训练图片标记的类别信息对目标神经网络进行训练,训练完成的神经网络就可以用于进行图像分类了。

[0312] 图17是本申请实施例提供的神经网络构建装置的硬件结构示意图。图17所示的神经网络构建装置3000(该装置3000具体可以是一种计算机设备)包括存储器3001、处理器3002、通信接口3003以及总线3004。其中,存储器3001、处理器3002、通信接口3003通过总线3004实现彼此之间的通信连接。

[0313] 存储器3001可以是只读存储器(read only memory,ROM),静态存储设备,动态存储设备或者随机存取存储器(random access memory,RAM)。存储器3001可以存储程序,当存储器3001中存储的程序被处理器3002执行时,处理器3002用于执行本申请实施例的神经网络的构建方法的各个步骤。

[0314] 处理器3002可以采用通用的中央处理器(central processing unit,CPU),微处理器,应用专用集成电路(application specific integrated circuit,ASIC),图形处理器(graphics processing unit,GPU)或者一个或多个集成电路,用于执行相关程序,以实现本申请方法实施例的神经网络的构建方法。

[0315] 处理器3002还可以是一种集成电路芯片,具有信号的处理能力。在实现过程中,本申请的神经网络的构建方法的各个步骤可以通过处理器3002中的硬件的集成逻辑电路或者软件形式的指令完成。

[0316] 上述处理器3002还可以是通用处理器、数字信号处理器(digital signal processing,DSP)、专用集成电路(ASIC)、现成可编程门阵列(field programmable gate array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器3001,处理器3002读取存储器3001中的信息,结合其硬件完成本神经网络构建装置中包括的单元所需执行的功能,或者执行本申请方法实施例的神经网络的构建方法。

[0317] 通信接口3003使用例如但不限于收发器一类的收发装置,来实现装置3000与其他设备或通信网络之间的通信。例如,可以通过通信接口3003获取待构建的神经网络的信息以及构建神经网络过程中需要的训练数据。

[0318] 总线3004可包括在装置3000各个部件(例如,存储器3001、处理器3002、通信接口3003)之间传送信息的通路。

[0319] 图18是本申请实施例的图像处理装置的硬件结构示意图。图18所示的图像处理装置4000包括存储器4001、处理器4002、通信接口4003以及总线4004。其中,存储器4001、处理器4002、通信接口4003通过总线4004实现彼此之间的通信连接。

[0320] 存储器4001可以是ROM,静态存储设备和RAM。存储器4001可以存储程序,当存储器4001中存储的程序被处理器4002执行时,处理器4002和通信接口4003用于执行本申请实施例的图像处理方法的各个步骤。

[0321] 处理器4002可以采用通用的,CPU,微处理器,ASIC,GPU或者一个或多个集成电路,

用于执行相关程序,以实现本申请实施例的图像处理装置中的单元所需执行的功能,或者执行本申请方法实施例的图像处理方法。

[0322] 处理器4002还可以是一种集成电路芯片,具有信号的处理能力。在实现过程中,本申请实施例的图像处理方法的各个步骤可以通过处理器4002中的硬件的集成逻辑电路或者软件形式的指令完成。

[0323] 上述处理器4002还可以是通用处理器、DSP、ASIC、FPGA或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器4001,处理器4002读取存储器4001中的信息,结合其硬件完成本申请实施例的图像处理装置中包括的单元所需执行的功能,或者执行本申请方法实施例的图像处理方法。

[0324] 通信接口4003使用例如但不限于收发器一类的收发装置,来实现装置4000与其他设备或通信网络之间的通信。例如,可以通过通信接口4003获取待处理图像。

[0325] 总线4004可包括在装置4000各个部件(例如,存储器4001、处理器4002、通信接口4003)之间传送信息的通路。

[0326] 图19是本申请实施例的神经网络训练装置的硬件结构示意图。与上述装置3000和装置4000类似,图19所示的神经网络训练装置5000包括存储器5001、处理器5002、通信接口5003以及总线5004。其中,存储器5001、处理器5002、通信接口5003通过总线5004实现彼此之间的通信连接。

[0327] 在通过图17所示的神经网络构建装置构建得到了神经网络之后,可以通过图19所示的神经网络训练装置5000对该神经网络进行训练,训练得到的神经网络就可以用于执行本申请实施例的图像处理方法了。

[0328] 具体地,图19所示的装置可以通过通信接口5003从外界获取训练数据以及待训练的神经网络,然后由处理器根据训练数据对待训练的神经网络进行训练。

[0329] 应注意,尽管上述装置3000、装置4000和装置5000仅仅示出了存储器、处理器、通信接口,但是在具体实现过程中,本领域的技术人员应当理解,装置3000、装置4000和装置5000还可以包括实现正常运行所必须的其他器件。同时,根据具体需要,本领域的技术人员应当理解,装置3000、装置4000和装置5000还可包括实现其他附加功能的硬件器件。此外,本领域的技术人员应当理解,装置3000、装置4000和装置5000也可仅仅包括实现本申请实施例所必须的器件,而不必包括图17、图18和图19中所示的全部器件。

[0330] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0331] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统、

装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0332] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统、装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0333] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0334] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。

[0335] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(read-only memory,ROM)、随机存取存储器(random access memory,RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0336] 以上所述,仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应以所述权利要求的保护范围为准。

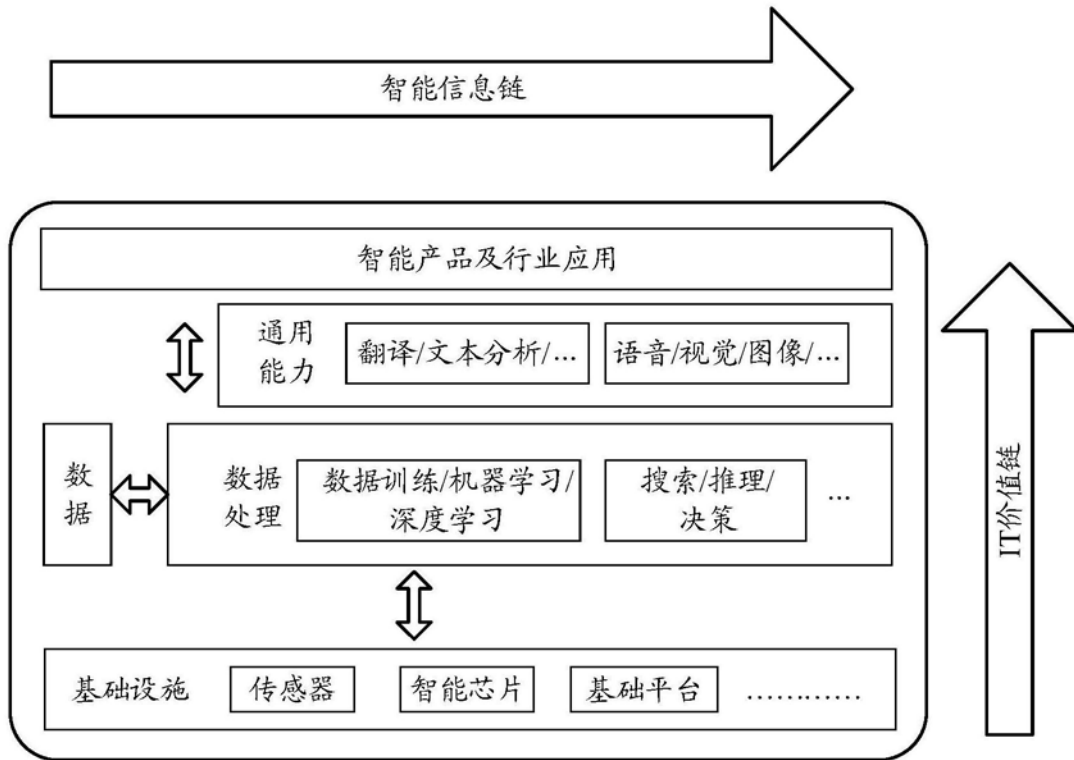


图1

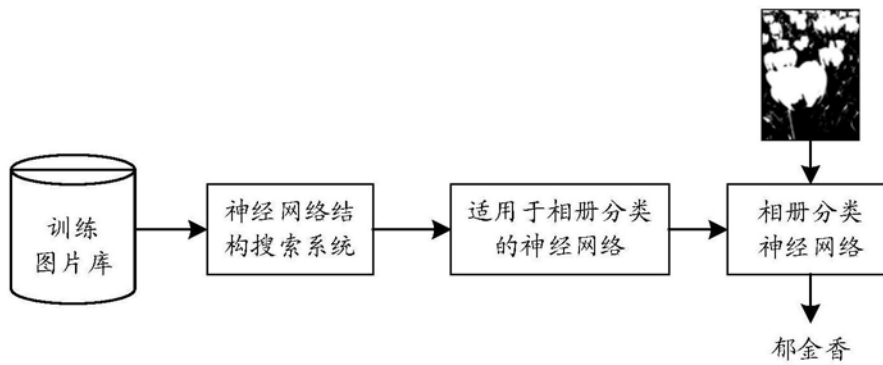


图2

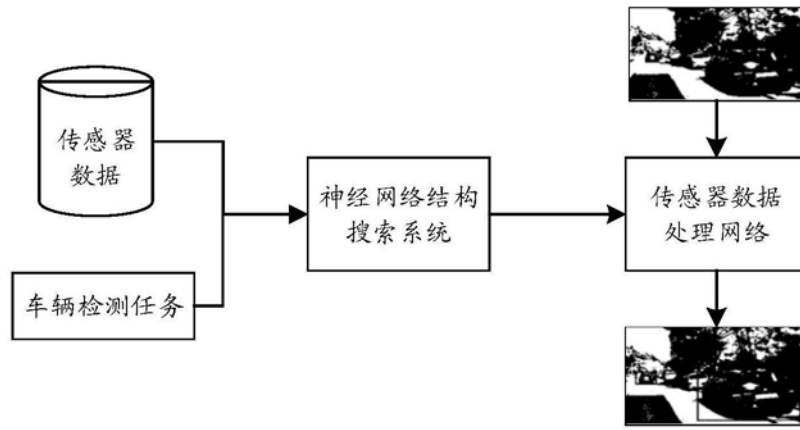


图3

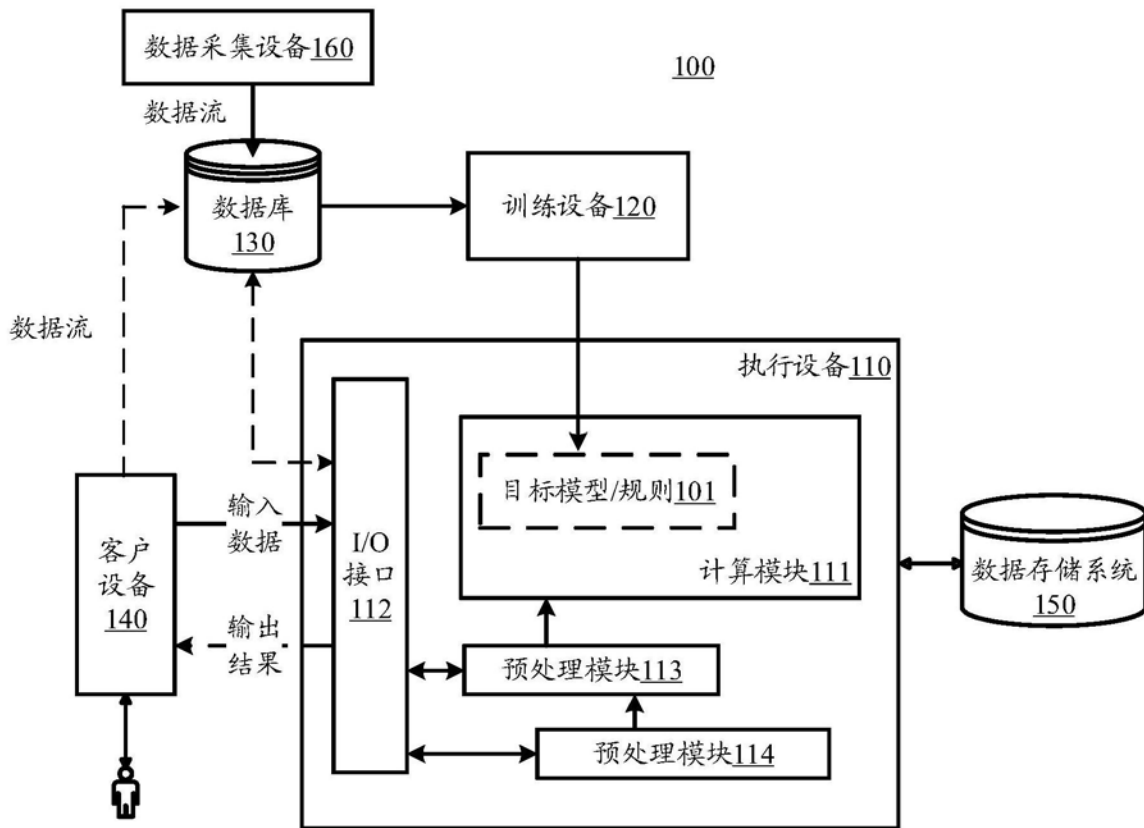


图4

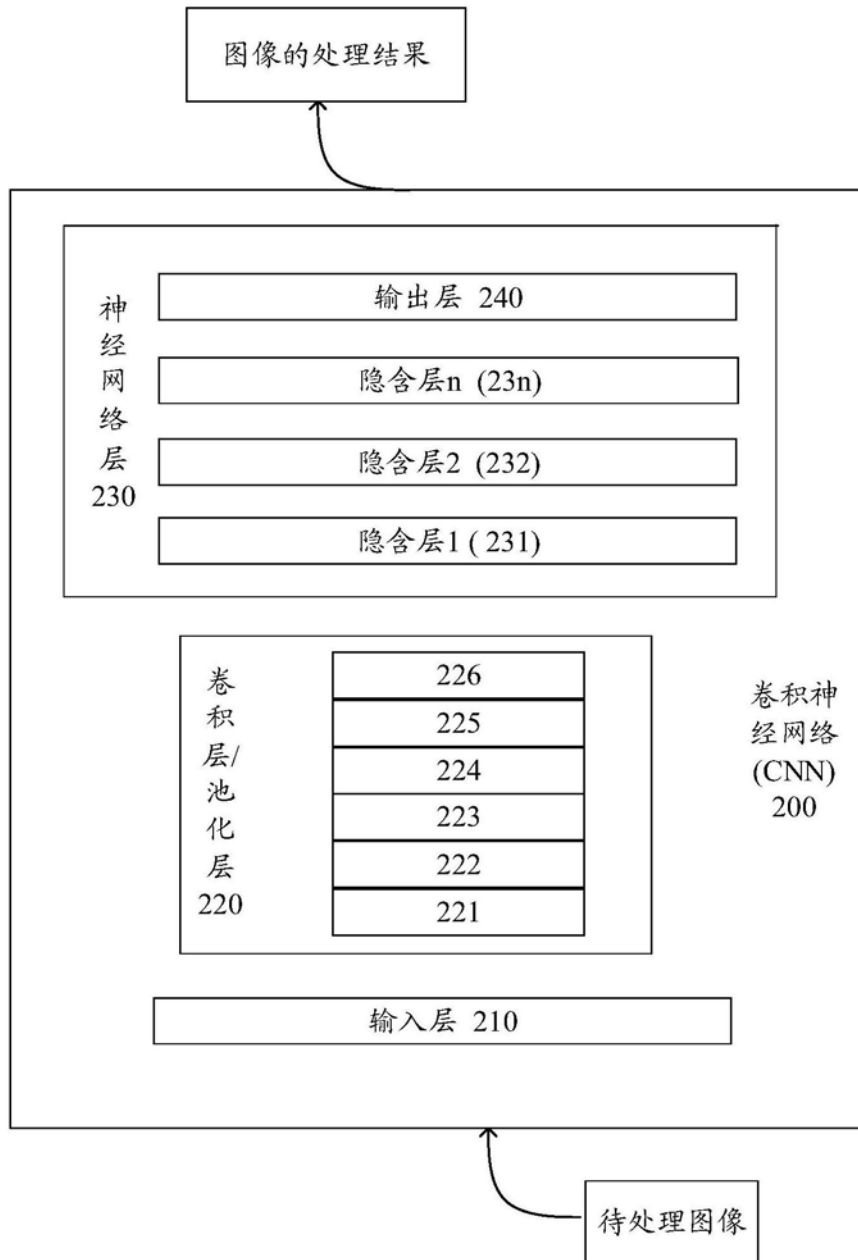


图5

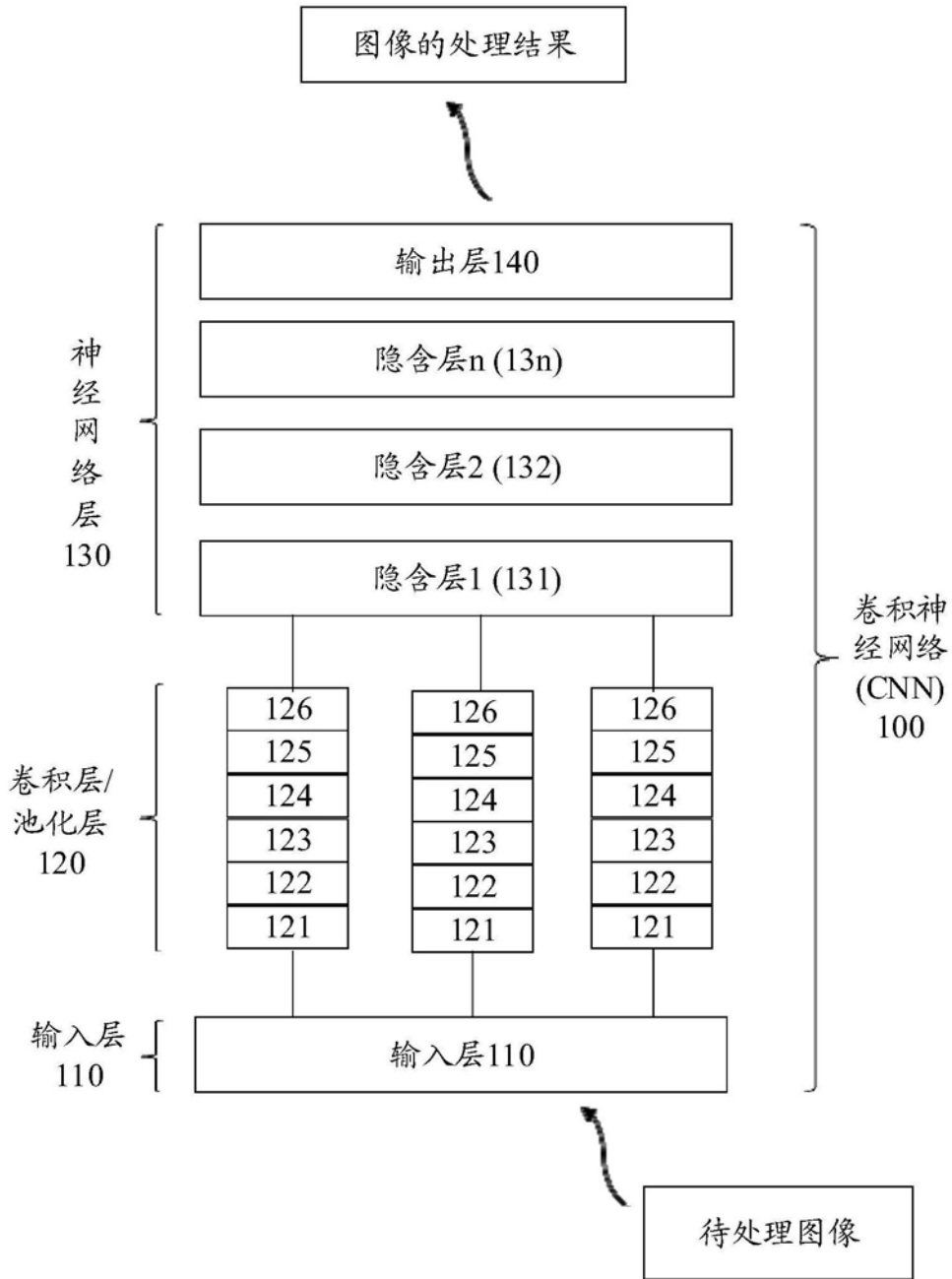


图6

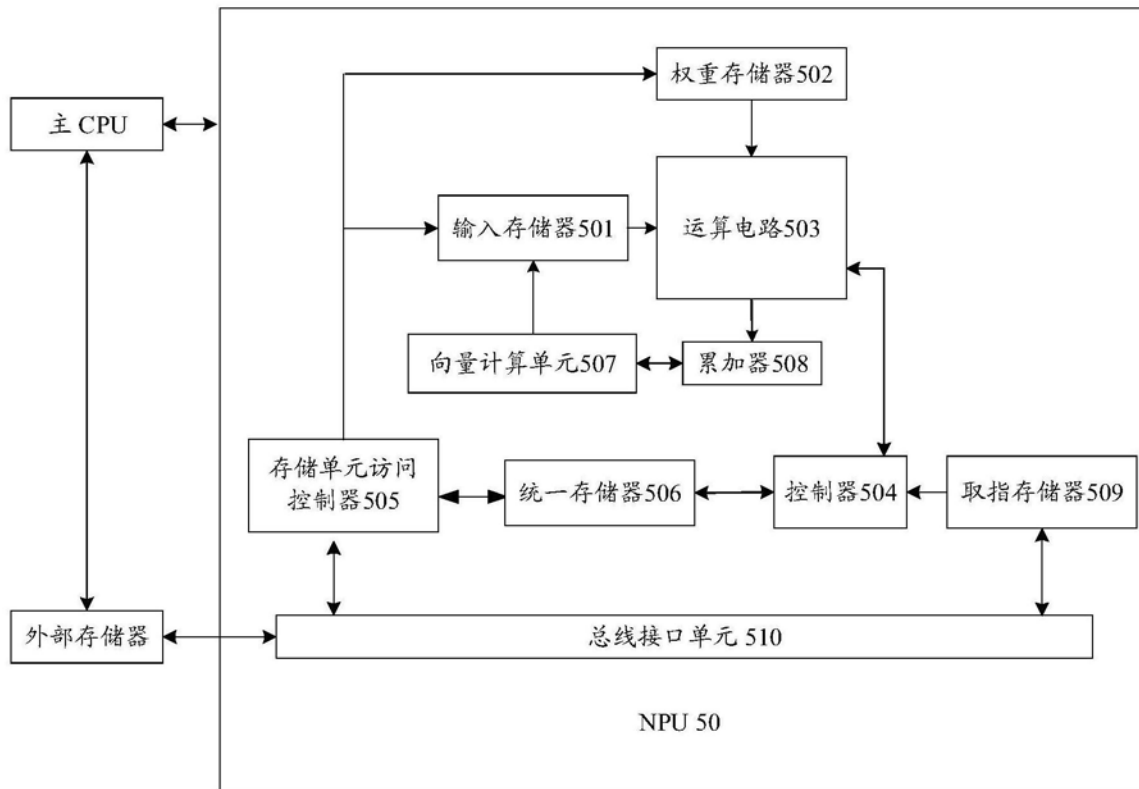


图7

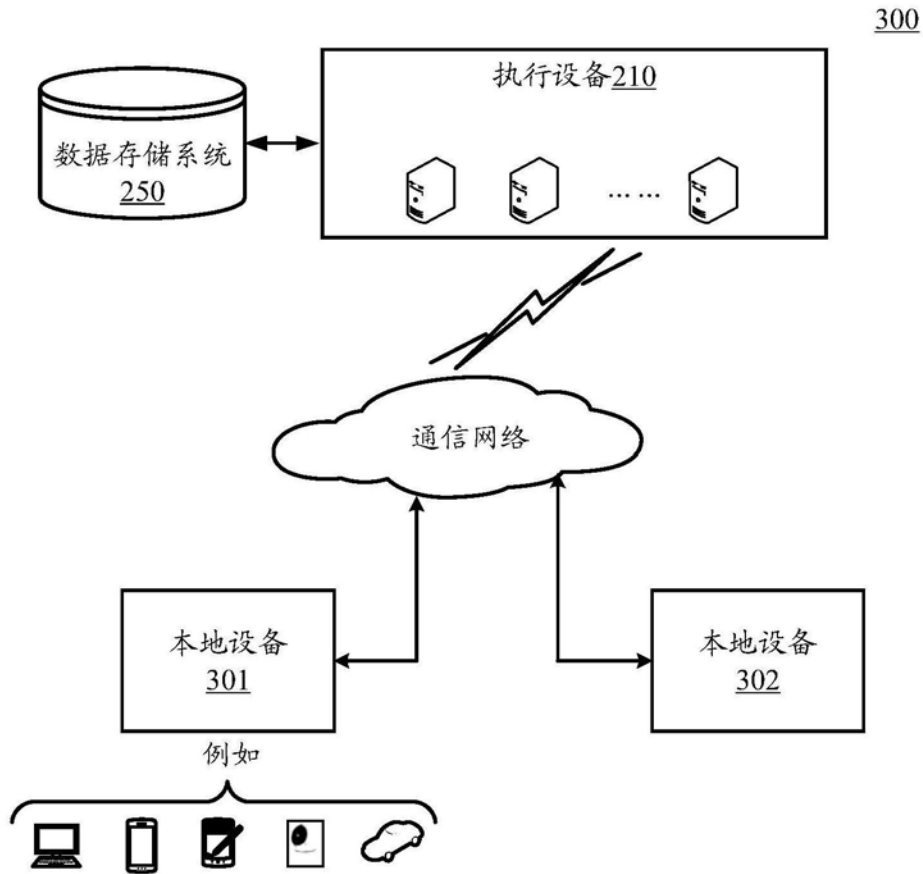


图8

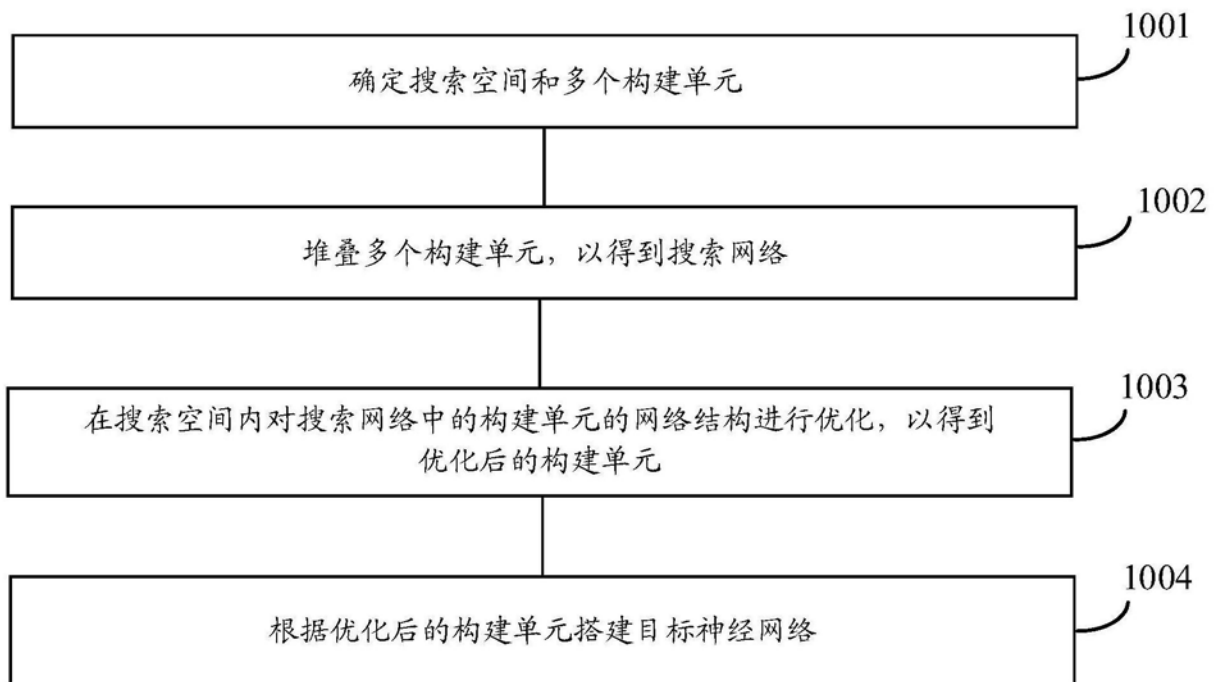


图9

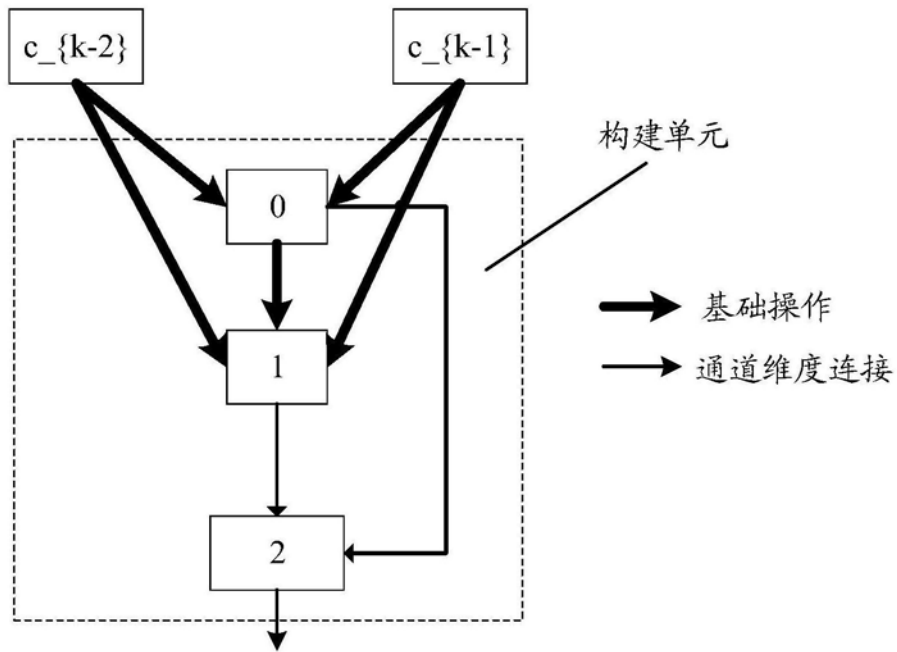


图10

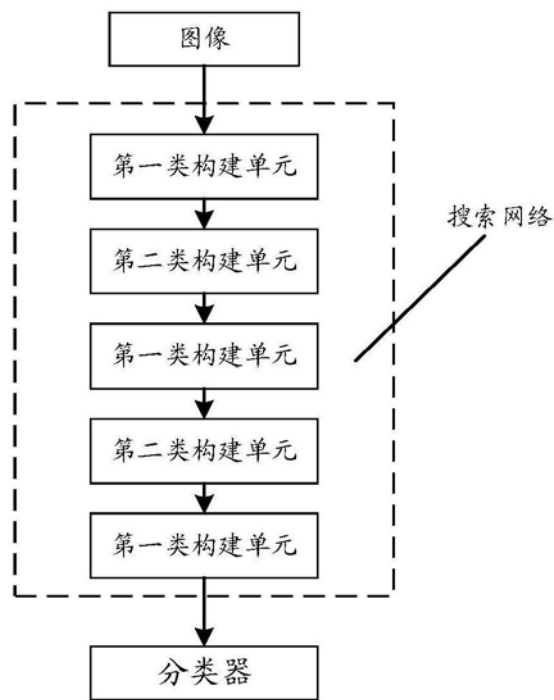


图11

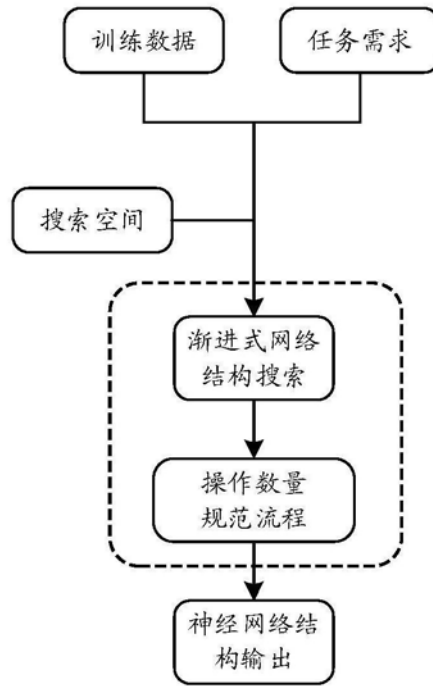


图12

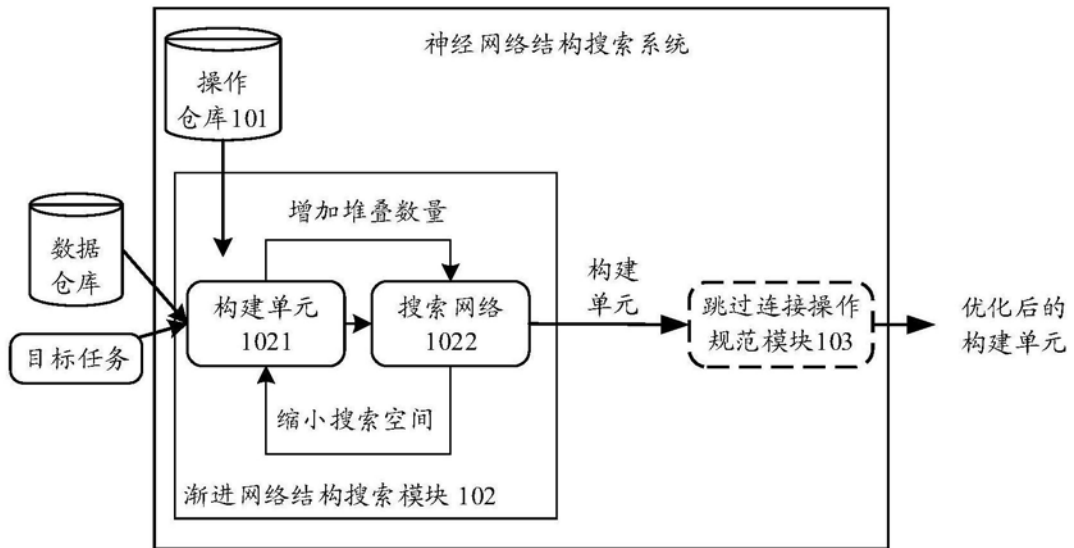


图13

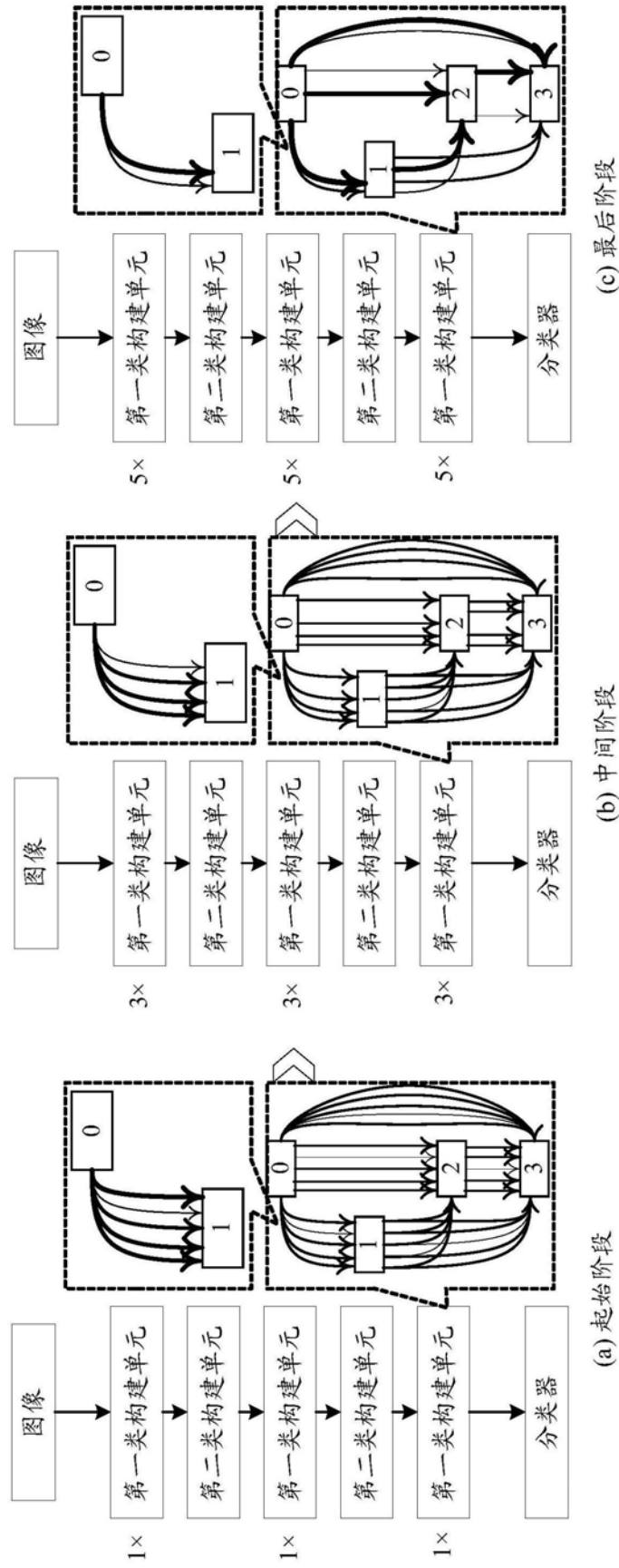


图14

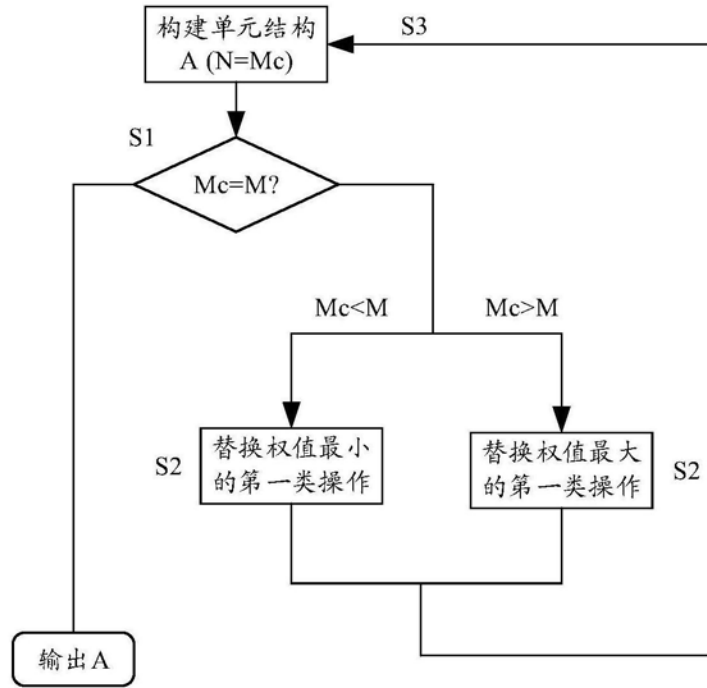


图15

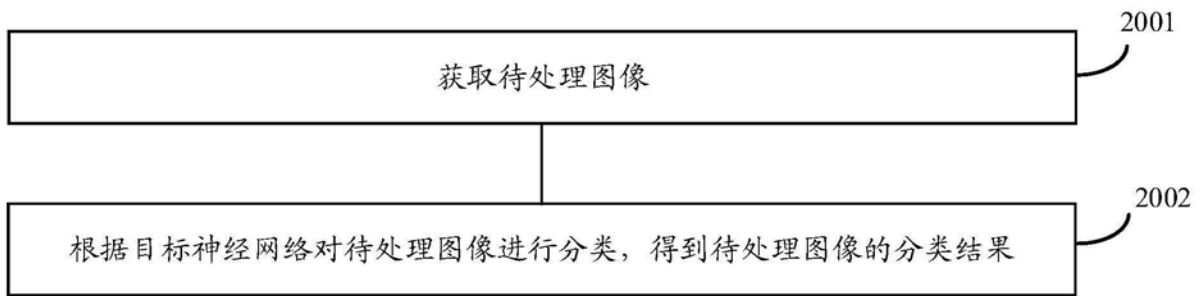


图16

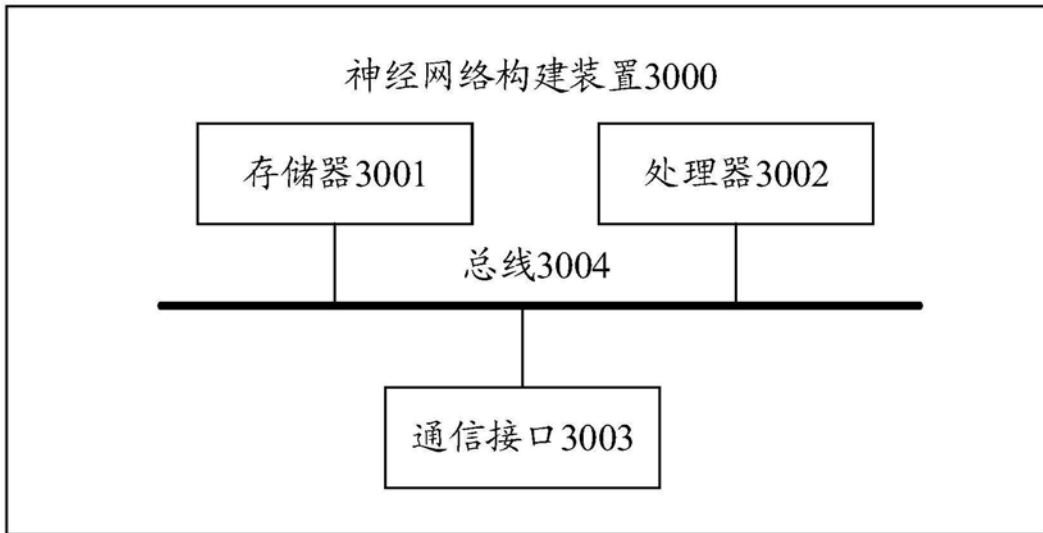


图17

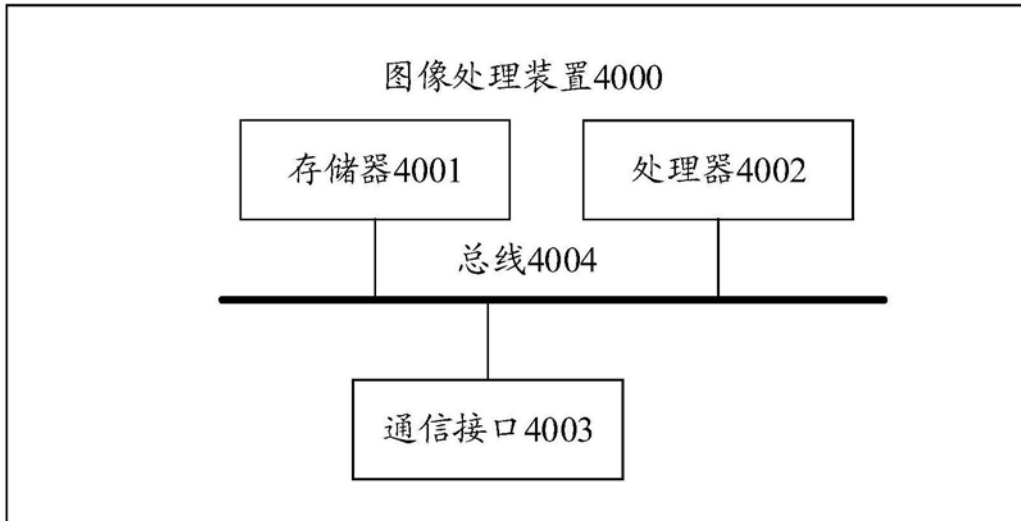


图18

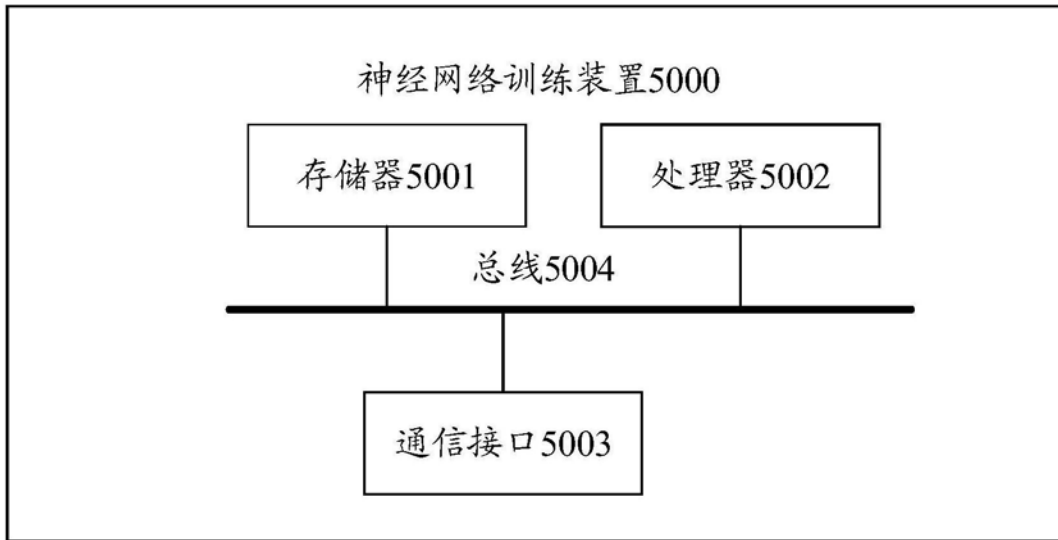


图19