



# (12)发明专利申请

(10)申请公布号 CN 111445420 A

(43)申请公布日 2020.07.24

(21)申请号 202010276826.6

(22)申请日 2020.04.09

(71)申请人 北京爱芯科技有限公司

地址 100082 北京市海淀区中关村大街1号  
14层1407室

(72)发明人 王艳龙

(74)专利代理机构 北京超凡宏宇专利代理事务  
所(特殊普通合伙) 11463

代理人 徐丽

(51) Int. Cl.

G06T 5/00(2006.01)

G06N 3/04(2006.01)

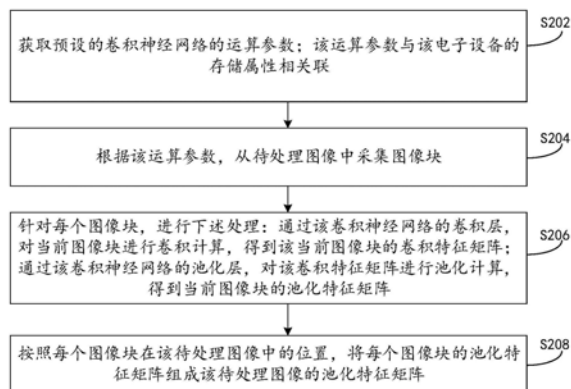
权利要求书3页 说明书14页 附图4页

## (54)发明名称

卷积神经网络的图像运算方法、装置和电子设备

## (57)摘要

本发明提供了一种卷积神经网络的图像运算方法、装置和电子设备,包括获取预设的卷积神经网络的运算参数;该运算参数与该电子设备的存储属性相关联;根据该运算参数,从待处理图像中采集图像块;针对每个图像块,通过该卷积神经网络计算其池化特征矩阵;根据每个图像块的池化特征矩阵组成该待处理图像的池化特征矩阵。本发明根据电子设备的内部缓存,推导池化层输出的特征矩阵的尺度,进而设置与该卷积神经网络的网络参数相匹配的窗口尺度参数,使得根据该窗口尺度参数采集的图像块,通过卷积运算得到的卷积特征矩阵中不出现冗余数据,从而可以对卷积层输出的卷积结果直接进行池化计算,无需将其缓存至DDR,节约了内存资源和网络宽带。



1. 一种卷积神经网络的图像运算方法,其特征在于,所述方法应用于实现卷积神经网络运算的电子设备;所述方法包括:

获取预设的卷积神经网络的运算参数;所述运算参数与所述电子设备的存储属性相关联;

根据所述运算参数,从待处理图像中采集图像块;

针对每个所述图像块,进行下述处理:通过所述卷积神经网络的卷积层,对当前图像块进行卷积计算,得到所述当前图像块的卷积特征矩阵;通过所述卷积神经网络的池化层,对所述卷积特征矩阵进行池化计算,得到所述当前图像块的池化特征矩阵;

按照每个所述图像块在所述待处理图像中的位置,将每个所述图像块的池化特征矩阵组成所述待处理图像的池化特征矩阵。

2. 根据权利要求1所述的方法,其特征在于,获取预设的卷积神经网络的运算参数的步骤,包括:

获取所述电子设备的硬件参数;

根据所述硬件参数,确定预设的卷积神经网络的运算参数。

3. 根据权利要求2所述的方法,其特征在于,根据所述硬件参数,确定预设的卷积神经网络的运算参数的步骤,包括:

根据所述硬件参数,确定预设的卷积神经网络的约束条件;其中,所述硬件参数包括所述电子设备的内部缓存大小;所述约束条件包括:所述卷积神经网络中卷积层输出的卷积特征矩阵的最大尺度,或者,所述卷积层输入的图像块的最大尺度;

根据所述约束条件,确定所述卷积神经网络的运算参数。

4. 根据权利要求3所述的方法,其特征在于,根据所述约束条件,确定所述卷积神经网络的运算参数的步骤,包括:

如果所述约束条件包括所述卷积层输出的卷积特征矩阵的最大尺度,根据所述卷积特征矩阵的最大尺度,确定所述卷积层输入的图像块的最大尺度;

根据所述卷积层输入的图像块的最大尺度,确定所述卷积神经网络中池化层输出的池化特征矩阵的尺度;

根据所述池化层输出的池化特征矩阵的尺度,确定所述卷积神经网络的运算参数。

5. 根据权利要求4所述的方法,其特征在于,根据所述卷积特征矩阵的最大尺度,确定所述卷积层输入的图像块的最大尺度的步骤,包括:

通过下述公式,确定所述卷积层输入的图像块的最大尺度:

$$i\_cov\_width\_valid\_max = (o\_cov\_width\_max - 1) * cs\_w + 1;$$

$$i\_cov\_height\_valid\_max = (o\_cov\_height\_max - 1) * cs\_h + 1;$$

其中, $i\_cov\_width\_valid\_max$ 为所述卷积层输入的图像块的最大宽度; $i\_cov\_height\_valid\_max$ 为所述卷积层输入的图像块的最大高度; $o\_cov\_width\_max$ 为所述卷积层输出的卷积特征矩阵的最大宽度; $o\_cov\_height\_max$ 为所述卷积层输出的卷积特征矩阵的最大高度; $cs\_w$ 为所述卷积层的水平步长; $cs\_h$ 为所述卷积层的垂直步长。

6. 根据权利要求4所述的方法,其特征在于,根据所述卷积层输入的图像块的最大尺度,确定所述卷积神经网络中池化层输出的池化特征矩阵的尺度的步骤,包括:

通过下述公式,确定所述卷积神经网络中池化层输出的池化特征矩阵的尺度:

$$(pw_w+(pl\_width-1)*ps_w-1)*cs_w+1=i\_cov\_width\_valid\_max;$$

$$(pw_h+(pl\_height-1)*ps_h-1)*cs_h+1=i\_cov\_height\_valid\_max;$$

其中,  $i\_cov\_width\_valid\_max$ 为所述卷积层输入的图像块的最大宽度;  $i\_cov\_height\_valid\_max$ 为所述卷积层输入的图像块的最大高度;  $cs_w$ 为所述卷积层的水平步长;  $cs_h$ 为所述卷积层的垂直步长;  $ps_w$ 为所述池化层的水平步长;  $ps_h$ 为所述池化层的垂直步长;  $pl\_width$ 为所述池化层输出的池化特征矩阵的宽度;  $pl\_height$ 为所述池化层输出的池化特征矩阵的高度;  $pw_w$ 为所述池化层的池化窗口宽度;  $pw_h$ 为所述池化层的池化窗口高度。

7. 根据权利要求4所述的方法, 其特征在于, 根据所述池化层输出的池化特征矩阵的尺度, 确定所述卷积神经网络的运算参数的步骤, 包括:

对所述池化层输出的池化特征矩阵的尺度进行取整处理;

根据取整后的所述池化层输出的池化特征矩阵的尺度, 确定所述卷积神经网络的运算参数。

8. 根据权利要求7所述的方法, 其特征在于, 根据取整后的所述池化层输出的池化特征矩阵的尺度, 确定所述卷积神经网络的运算参数的步骤, 包括:

通过下述公式, 计算所述运算参数中, 用于从待处理图像中采集图像块的窗口尺度:

$$i\_cov\_width=(pw_w+(o\_pl\_width-1)*ps_w-1)*cs_w+cw_w;$$

$$i\_cov\_height=(pw_h+(o\_pl\_height-1)*ps_h-1)*cs_h+cw_h;$$

其中,  $i\_cov\_width$ 为所述窗口的宽度;  $i\_cov\_height$ 为所述窗口的高度;  $pw_w$ 为所述池化层的池化窗口宽度;  $pw_h$ 为所述池化层的池化窗口高度;  $o\_pl\_width$ 为取整后的所述池化层输出的池化特征矩阵的宽度;  $o\_pl\_height$ 为取整后的所述池化层输出的池化特征矩阵的高度;  $ps_w$ 为所述池化层的水平步长;  $ps_h$ 为所述池化层的垂直步长;  $cs_w$ 为所述卷积层的水平步长;  $cs_h$ 为所述卷积层的垂直步长。

9. 根据权利要求8所述的方法, 其特征在于, 根据取整后的所述池化层输出的池化特征矩阵的尺度, 确定所述卷积神经网络的运算参数的步骤, 包括:

通过下述公式, 计算所述运算参数中, 用于从待处理图像中采集图像块的窗口位移:

$$cb\_ws=((pw_w+(o\_pl\_width-1)*ps_w-1)*cs_w+1-(pw_w-ps_w))*cs_w;$$

$$cb\_hs=((pw_h+(o\_pl\_height-1)*ps_h-1)*cs_h+1-(pw_h-ps_h))*cs_h;$$

其中,  $cb\_ws$ 为所述窗口的水平位移步长; 所述  $cb\_hs$ 为所述窗口的垂直位移步长。

10. 根据权利要求1所述的方法, 其特征在于, 所述运算参数包括: 用于从待处理图像中采集图像块的窗口尺度参数和窗口位移参数, 所述根据所述运算参数, 从待处理图像中采集图像块的步骤, 包括:

根据所述运算参数中的窗口尺度参数和预设的默认初始位置, 从待处理图像中确定第一个图像块;

根据所述窗口尺度参数和所述窗口位移参数, 按照预设顺序遍历所述待处理图像, 得到除所述第一个图像块以外的图像块。

11. 根据权利要求1所述的方法, 其特征在于, 按照每个所述图像块在所述待处理图像中的位置, 将每个所述图像块的池化特征矩阵组成所述待处理图像的池化特征矩阵的步骤, 包括:

按照每个所述图像块在所述待处理图像中的位置,将位置相邻的图像块的池化特性矩阵进行拼接处理,得到所述待处理图像的池化特征矩阵。

12. 一种卷积神经网络的图像运算装置,其特征在于,所述装置应用于实现卷积神经网络运算的电子设备;所述装置包括:

运算参数获取模块,用于获取预设的卷积神经网络的运算参数;所述运算参数与所述电子设备的存储属性相关联;

图像块采集模块,用于根据所述运算参数,从待处理图像中采集图像块;

图像块池化特征矩阵计算模块,用于针对每个所述图像块,进行下述处理:通过所述卷积神经网络的卷积层,对当前图像块进行卷积计算,得到所述当前图像块的卷积特征矩阵;通过所述卷积神经网络的池化层,对所述卷积特征矩阵进行池化计算,得到所述当前图像块的池化特征矩阵;

图像池化特征矩阵计算模块,用于按照每个所述图像块在所述待处理图像中的位置,将每个所述图像块的池化特征矩阵组成所述待处理图像的池化特征矩阵。

13. 一种电子设备,其特征在于,所述电子设备包括处理设备和存储装置;

所述存储装置上存储有计算机程序,所述计算机程序在被所述处理设备运行时执行1至11任一项所述的卷积神经网络的图像运算方法的步骤。

14. 一种计算机可读存储介质,所述计算机可读存储介质上存储有计算机程序,其特征在于,所述计算机程序被处理设备运行时执行1至11任一项所述的卷积神经网络的图像运算方法的步骤。

## 卷积神经网络的图像运算方法、装置和电子设备

### 技术领域

[0001] 本发明涉及图像处理技术领域,尤其是涉及一种卷积神经网络的图像运算方法、装置和电子设备。

### 背景技术

[0002] 卷积神经网络包括输入层、隐藏层和输出层;其中,隐藏层又包括卷积层和池化层,该池化层也称为采样层。相关技术中,卷积神经网络加速器通常采用固定尺度的窗口采集图像块,然后通过卷积层对采集到的图像块进行卷积计算后,并将卷积结果缓存至DDR(Double Data Rate Synchronous Dynamic Random Access Memory,双倍速率同步动态随机存储器),再从DDR中获取卷积结果,通过池化层进行池化计算,输出池化结果。该方式中,由于图像块的采集窗口尺度固定,如果相对于某个卷积层的卷积计算,该采集窗口尺度过大,则采集的图像块会较大,存在冗余数据;并且,卷积层输出的卷积结果相对于后续的池化层,也可能存在冗余数据。因此,需要采用上述将卷积结果缓存至DDR,再从DDR中读取卷积结果的方式解决冗余数据的问题。但是,在DDR中存取卷积结果的方式,会浪费大量的内存资源和网络带宽。

[0003] 整体而言,现有的卷积神经网络运算方式会导致大量内存资源和网络带宽的浪费。

### 发明内容

[0004] 有鉴于此,本发明的目的在于提供一种卷积神经网络的图像运算方法、装置和电子设备,可以减少内存资源的占用,并节约网络带宽。

[0005] 第一方面,本发明实施例提供了一种卷积神经网络的图像运算方法,该方法应用于实现卷积神经网络运算的电子设备;该方法包括:获取预设的卷积神经网络的运算参数;该运算参数与该电子设备的存储属性相关联;根据该运算参数,从待处理图像中采集图像块;针对每个图像块,进行下述处理:通过该卷积神经网络的卷积层,对当前图像块进行卷积计算,得到该当前图像块的卷积特征矩阵;通过该卷积神经网络的池化层,对该卷积特征矩阵进行池化计算,得到当前图像块的池化特征矩阵;按照每个图像块在该待处理图像中的位置,将每个图像块的池化特征矩阵组成该待处理图像的池化特征矩阵。

[0006] 在本发明较佳的实施例中,上述获取预设的卷积神经网络的运算参数的步骤,包括:获取该电子设备的硬件参数;根据该硬件参数,确定预设的卷积神经网络的运算参数。

[0007] 在本发明较佳的实施例中,上述根据该硬件参数,确定预设的卷积神经网络的运算参数的步骤,包括:根据该硬件参数,确定预设的卷积神经网络的约束条件;其中,该硬件参数包括该电子设备的内部缓存大小;该约束条件包括:该卷积神经网络中卷积层输出的卷积特征矩阵的最大尺度,或者,该卷积层输入的图像块的最大尺度;根据该约束条件,确定该卷积神经网络的运算参数。

[0008] 在本发明较佳的实施例中,上述根据该约束条件,确定该卷积神经网络的运算参

数的步骤,包括:如果该约束条件包括该卷积层输出的卷积特征矩阵的最大尺度,根据该卷积特征矩阵的最大尺度,确定该卷积层输入的图像块的最大尺度;根据该卷积层输入的图像块的最大尺度,确定该卷积神经网络中池化层输出的池化特征矩阵的尺度;根据该池化层输出的池化特征矩阵的尺度,确定该卷积神经网络的运算参数。

[0009] 在本发明较佳的实施例中,上述根据该卷积特征矩阵的最大尺度,确定该卷积层输入的图像块的最大尺度的步骤,包括:通过下述公式,确定该卷积层输入的图像块的最大尺度: $i\_cov\_width\_valid\_max=(o\_cov\_width\_max-1)*cs\_w+1$ ;  $i\_cov\_height\_valid\_max=(o\_cov\_height\_max-1)*cs\_h+1$ ;其中, $i\_cov\_width\_valid\_max$ 为该卷积层输入的图像块的最大宽度; $i\_cov\_height\_valid\_max$ 为该卷积层输入的图像块的最大高度; $o\_cov\_width\_max$ 为该卷积层输出的卷积特征矩阵的最大宽度; $o\_cov\_height\_max$ 为该卷积层输出的卷积特征矩阵的最大高度; $cs\_w$ 为该卷积层的水平步长; $cs\_h$ 为该卷积层的垂直步长。

[0010] 在本发明较佳的实施例中,上述根据该卷积层输入的图像块的最大尺度,确定该卷积神经网络中池化层输出的池化特征矩阵的尺度的步骤,包括:通过下述公式,确定该卷积神经网络中池化层输出的池化特征矩阵的尺度: $(pw\_w+(pl\_width-1)*ps\_w-1)*cs\_w+1=i\_cov\_width\_valid\_max$ ;  $(pw\_h+(pl\_height-1)*ps\_h-1)*cs\_h+1=i\_cov\_height\_valid\_max$ ;其中, $i\_cov\_width\_valid\_max$ 为该卷积层输入的图像块的最大宽度; $i\_cov\_height\_valid\_max$ 为该卷积层输入的图像块的最大高度; $cs\_w$ 为该卷积层的水平步长; $cs\_h$ 为该卷积层的垂直步长; $ps\_w$ 为该池化层的水平步长; $ps\_h$ 为该池化层的垂直步长; $pl\_width$ 为该池化层输出的池化特征矩阵的宽度; $pl\_height$ 为该池化层输出的池化特征矩阵的高度; $pw\_w$ 为该池化层的池化窗口宽度; $pw\_h$ 为该池化层的池化窗口高度。

[0011] 在本发明较佳的实施例中,上述根据该池化层输出的池化特征矩阵的尺度,确定该卷积神经网络的运算参数的步骤,包括:对该池化层输出的池化特征矩阵的尺度进行取整处理;根据取整后的该池化层输出的池化特征矩阵的尺度,确定该卷积神经网络的运算参数。

[0012] 在本发明较佳的实施例中,上述根据取整后的该池化层输出的池化特征矩阵的尺度,确定该卷积神经网络的运算参数的步骤,包括:通过下述公式,计算该运算参数中,用于从待处理图像中采集图像块的窗口尺度: $i\_cov\_width=(pw\_w+(o\_pl\_width-1)*ps\_w-1)*cs\_w+cw\_w$ ;  $i\_cov\_height=(pw\_h+(o\_pl\_height-1)*ps\_h-1)*cs\_h+cw\_h$ ;其中, $i\_cov\_width$ 为该窗口的宽度; $i\_cov\_height$ 为该窗口的高度; $pw\_w$ 为该池化层的池化窗口宽度; $pw\_h$ 为该池化层的池化窗口高度; $o\_pl\_width$ 为取整后的该池化层输出的池化特征矩阵的宽度; $o\_pl\_height$ 为取整后的该池化层输出的池化特征矩阵的高度; $ps\_w$ 为该池化层的水平步长; $ps\_h$ 为该池化层的垂直步长; $cs\_w$ 为该卷积层的水平步长; $cs\_h$ 为该卷积层的垂直步长。

[0013] 在本发明较佳的实施例中,上述根据取整后的该池化层输出的池化特征矩阵的尺度,确定该卷积神经网络的运算参数的步骤,包括:通过下述公式,计算该运算参数中,用于从待处理图像中采集图像块的窗口位移: $cb\_ws=((pw\_w+(o\_pl\_width-1)*ps\_w-1)*cs\_w+1-(pw\_w-ps\_w))*cs\_w$ ;  $cb\_hs=((pw\_h+(o\_pl\_height-1)*ps\_h-1)*cs\_h+1-(pw\_h-ps\_h))*cs\_h$ ;其中, $cb\_ws$ 为该窗口的水平位移步长; $cb\_hs$ 为该窗口的垂直位移步长。

[0014] 在本发明较佳的实施例中,上述运算参数包括:用于从待处理图像中采集图像块

的窗口尺度参数和窗口位移参数,上述根据该运算参数,从待处理图像中采集图像块的步骤,包括:根据该运算参数中的窗口尺度参数和预设的默认初始位置,从待处理图像中确定第一个图像块;根据该窗口尺度参数和该窗口位移参数,按照预设顺序遍历该待处理图像,得到除该第一个图像块以外的图像块。

[0015] 在本发明较佳的实施例中,上述按照每个图像块在待处理图像中的位置,将每个图像块的池化特征矩阵组成该待处理图像的池化特征矩阵的步骤,包括:按照每个该图像块在该待处理图像中的位置,将位置相邻的图像块的池化特性矩阵进行拼接处理,得到该待处理图像的池化特征矩阵。

[0016] 第二方面,本发明实施例还提供了一种卷积神经网络的图像运算装置,该装置应用于实现卷积神经网络运算的电子设备;该装置包括:运算参数获取模块,用于获取预设的卷积神经网络的运算参数;该运算参数与该电子设备的存储属性相关联;图像块采集模块,用于根据该运算参数,从待处理图像中采集图像块;图像块池化特征矩阵计算模块,用于针对每个图像块,进行下述处理:通过该卷积神经网络的卷积层,对当前图像块进行卷积计算,得到当前图像块的卷积特征矩阵;通过该卷积神经网络的池化层,对该卷积特征矩阵进行池化计算,得到该当前图像块的池化特征矩阵;图像池化特征矩阵计算模块,用于按照每个图像块在该待处理图像中的位置,将每个图像块的池化特征矩阵组成该待处理图像的池化特征矩阵。

[0017] 第四方面,本发明实施例还提供了一种用于实现卷积神经网络运算的电子设备,该电子设备包括处理设备和存储装置;该存储装置上存储有计算机程序,该计算机程序在被该处理设备运行时执行上述卷积神经网络的图像运算方法的步骤。

[0018] 第五方面,本发明实施例还提供了一种计算机可读存储介质,该计算机可读存储介质上存储有计算机程序,该计算机程序被处理设备运行时执行上述卷积神经网络的图像运算方法的步骤。

[0019] 本发明实施例带来了以下有益效果:

[0020] 本发明实施例提供的一种卷积神经网络的图像运算方法、装置和电子设备,首先获取预设的卷积神经网络的运算参数;该运算参数与该电子设备的存储属性相关联;然后根据该运算参数,从待处理图像中采集图像块;针对每个图像块,进行下述处理:通过该卷积神经网络的卷积层,对当前图像块进行卷积计算,得到该当前图像块的卷积特征矩阵;通过该卷积神经网络的池化层,对该卷积特征矩阵进行池化计算,得到当前图像块的池化特征矩阵;进而按照每个图像块在该待处理图像中的位置,将每个图像块的池化特征矩阵组成该待处理图像的池化特征矩阵。该方式中,基于下述原理:在卷积神经网络的卷积核参数和池化窗口参数确定的情况下,如果输入卷积层的图像块的尺寸一定,那么卷积层输出的卷积特征矩阵的尺度,以及池化层输出的池化特征矩阵的尺度也是确定的;因此,在卷积神经网络的卷积核参数和池化窗口参数确定的情况下,根据电子设备的存储属性,可以推导池化层输出的特征矩阵的尺度,进而设置恰当的窗口尺度参数(也即基于其关联关系预先推导出卷积神经网络的运算参数),使得采集得到的图像块的尺寸是与该卷积神经网络的卷积核参数和池化窗口参数相匹配的,进而使得该图像块在卷积运算后得到的卷积特征矩阵中不出现冗余数据,从而可以对卷积层输出的卷积结果直接进行池化计算,而无需将该卷积结果缓存至DDR,实现了卷积运算到池化运算的流水线设计,可以减少内存资源的占

用,并节约网络宽带。

[0021] 此外,由于该运算方法可以适用于具有不同卷积核参数及池化窗口参数的神经网络,因而该方法具有很好的通用性。

[0022] 本公开的其他特征和优点将在随后的说明书中阐述,或者,部分特征和优点可以从说明书推知或毫无疑义地确定,或者通过实施本公开的上述技术即可得知。

[0023] 为使本公开的上述目的、特征和优点能更明显易懂,下文特举较佳实施例,并配合所附附图,作详细说明如下。

## 附图说明

[0024] 为了更清楚地说明本发明具体实施方式或现有技术中的技术方案,下面将对具体实施方式或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施方式,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0025] 图1为本发明实施例提供的一种电子设备的结构示意图;

[0026] 图2为本发明实施例提供的一种卷积神经网络的图像运算方法的流程示意图;

[0027] 图3为本发明实施例提供的一种卷积神经网络图像运算中采集图像块的窗口位移示意图;

[0028] 图4为本发明实施例提供的一种采集图像块的示意图;

[0029] 图5为本发明实施例提供的另一种卷积神经网络的图像运算方法的流程示意图;

[0030] 图6为本发明实施例提供的一种流水线设计的卷积神经网络运算过程示意图;

[0031] 图7为本发明实施例提供的一种卷积神经网络的图像运算装置的结构示意图。

[0032] 图标:100-电子设备;102-处理设备;104-存储装置;106-输入装置;108-输出装置;110-图像采集设备;112-总线系统;71-运算参数获取模块;72-图像块采集模块;73-图像块池化特征矩阵计算模块;74-图像池化特征矩阵计算模块。

## 具体实施方式

[0033] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合附图对本发明的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,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,都属于本发明保护的范围。

[0034] 考虑到现有的卷积神经网络运算方式会导致大量内存资源和网络带宽的浪费的问题,本发明实施例提供的一种卷积神经网络的图像运算、装置和电子设备,该技术可以应用于实现卷积神经网络的硬件设计,以及涉及卷积神经网络运算的各种应用场景中。为便于对本实施例进行理解,首先对本发明实施例所公开的一种卷积神经网络的图像运算方法进行详细介绍。

[0035] 这里,参照图1来描述用于实现本发明实施例的卷积神经网络的图像运算、装置和电子设备的实例电子设备。

[0036] 如图1所示的一种电子设备的结构示意图,该电子设备100包括一个或多个处理设备102、一个或多个存储装置104、输入装置106、输出装置108以及一个或多个图像采集设备



110,这些组件通过总线系统112和/或其它形式的连接机构(未示出)互连。应当注意,图1所示的电子设备的组件和结构只是示例性的,而非限制性的,根据需要,该电子设备也可以具有其他组件和结构。

[0037] 处理设备102可以为智能终端,或者是包含中央处理单元(CPU)或者具有数据处理能力和/或指令执行能力的其它形式的处理单元的设备,可以对该电子设备中的其它组件的数据进行处理,还可以控制该电子设备中的其它组件以执行目标对象统计的功能。

[0038] 存储装置104可以包括一个或多个计算机程序产品,计算机程序产品可以包括各种形式的计算机可读存储介质,例如易失性存储器和/或非易失性存储器。易失性存储器例如可以包括随机存取存储器(RAM)和/或高速缓冲存储器(cache)等。非易失性存储器例如可以包括只读存储器(ROM)、硬盘、闪存等。在计算机可读存储介质上可以存储一个或多个计算机程序指令,处理设备102可以运行程序指令,以实现下文的本发明实施例中(由处理设备实现)的客户端功能以及/或者其他期望的功能。在计算机可读存储介质中还可以存储各种应用程序和各种数据,例如应用程序使用和/或产生的各种数据等。

[0039] 输入装置106可以是用户用来输入指令的装置,并且可以包括键盘、鼠标、麦克风和触摸屏等中的一个或多个。

[0040] 输出装置108可以向外部(例如,用户)输出各种信息(例如,图像或声音),并且可以包括显示器、扬声器等中的一个或多个。

[0041] 图像采集设备110可以采集训练图片,并且将采集到的预览视频帧或图像数据存储在存储装置104中以供其它组件使用。

[0042] 示例性地,用于实现根据本发明实施例的卷积神经网络的图像运算、装置和电子设备中的各器件可以集成设置,也可以分散设置,诸如将处理设备102、存储装置104、输入装置106和输出装置108集成设置于一体,而将图像采集设备110设置于可以采集到图片的指定位置。当上述电子设备中的各器件集成设置时,该电子设备可以被实现为诸如相机、智能手机、平板电脑、计算机、车载终端等智能终端。

[0043] 如图2所示,其为本发明实施例提供的一种卷积神经网络的图像运算方法的流程示意图,该方法应用于实现卷积神经网络运算的电子设备,由图2可见,该方法包括以下步骤:

[0044] 步骤S202:获取预设的卷积神经网络的运算参数;该运算参数与该电子设备的存储属性相关联。

[0045] 这里,上述实现卷积神经网络运算的电子设备可以是FPGA(Field Programmable Gate Array,现场可编程逻辑门阵列)芯片、Asic(Application Specific Integrated Circuit,专用集成电路)芯片等。

[0046] 其中,该卷积神经网络根据预设的运算参数构建,该运算参数与该电子设备的存储属性相关联,该存储属性包括硬件电路内部的缓存。在其中一种可能的实施方式中,该运算参数包括用于从待处理图像中采集图像块的窗口尺度参数和窗口位移参数。在实际操作中,根据窗口尺度参数进行图像块采集,通常窗口尺度参数包括窗口的宽度和高度。并且,窗口位移参数为采集图像时采集窗口的移动参数,通常包括位移方向和位移步长。这里,参见图3,所示为一种卷积神经网络图像运算中采集图像块的窗口位移示意图,在图3示出的实施方式中,展示了采集窗口的水平位移和垂直位移的情形。

[0047] 在本发明实施例中,上述窗口尺度参数与该卷积神经网络中池化层输出的特征矩阵的尺度相关联,其中,该卷积神经网络包括输入层、隐藏层和输出层,并且,隐藏层又包括卷积层和池化层。在实际操作中,根据上述窗口尺度参数先对待处理图像进行采样,得到图像块,然后将图像块输入卷积层进行卷积运算,得到卷积特征矩阵,再将该卷积特征输入到池化层,输出该图像块的池化特征矩阵。这里,由于在卷积核参数和池化窗口参数确定的情况下,如果输入图像块的尺寸一定,那么卷积层输出的卷积特征矩阵的尺度也是确定的,并且池化层输出的池化特征矩阵的尺度也是确定的;因此,可以如此反推,当卷积神经网络的卷积核参数和池化窗口参数确定的情况下,根据该电子设备的存储属性(例如, RAM缓存),可以推导池化层输出的特征矩阵的尺度,进而设置恰当的窗口尺度参数,使得采集得到的图像块的尺寸是与该卷积神经网络的卷积核参数和池化窗口参数相匹配的,例如,可以使符合尺寸图像块在卷积运算后得到的卷积特征矩阵中不出现冗余数据。

[0048] 这样,即可以对卷积运算后的卷积特征矩阵直接进行池化计算,得到图像块的池化特征矩阵,相比于现有获取图像块池化特征矩阵的技术,该方式不需要将卷积特征矩阵缓存至DDR,而可以实现卷积运算到池化运算的流水线设计。

[0049] 需要说明的是,在实际实现时,上述运算参数中的窗口尺度参数不仅仅与该卷积神经网络中池化层输出的特征矩阵的尺度相关联,还与该卷积神经网络各层参数相关联,比如卷积层卷积核大小、步长,池化层窗口、步长等等,同时窗口尺度参数还与系统的缓存容量相关联。

[0050] 步骤S204:根据该运算参数,从待处理图像中采集图像块。

[0051] 在实际操作中,从待处理图像中采集图像块的步骤可以通过下述步骤21-22实现:

[0052] (21) 根据该运算参数中的窗口尺度参数和预设的默认初始位置,从待处理图像中确定第一个图像块。

[0053] 其中,该初始位置可以是待处理图像中的任意位置,以采集窗口的左上角位置为例,可以设置初始位置为待处理图像的左上角所在像素点位置。

[0054] (22) 根据该窗口尺度参数和该窗口位移参数,按照预设顺序遍历该待处理图像,得到除该第一个图像块以外的图像块。

[0055] 这里,通过上述窗口尺度参数进行图像块采集,得到特定尺寸的图像块;并且,采集窗口根据上述窗口位移参数进行位移,其中预设的顺序即为窗口位移的方向路线,遍历该待处理图像,依次采集图像块,直至采集完待处理图像中的最后一个图像块。

[0056] 参见图4,所示为一种采集图像块的示意图,在图4示出的实施方式中,该窗口尺度参数包括窗口宽度和窗口高度,并且,该窗口位移参数包括水平位移步长和垂直位移步长,其中,采集窗口位移的顺序是从该图像的左上角依次向右水平移动直至图像的右边界,然后根据垂直位移步长换行,继续水平移动并采集图像块。

[0057] 步骤S206:针对每个图像块,进行下述处理:通过该卷积神经网络的卷积层,对当前图像块进行卷积计算,得到该当前图像块的卷积特征矩阵;通过该卷积神经网络的池化层,对该卷积特征矩阵进行池化计算,得到当前图像块的池化特征矩阵。

[0058] 在本实施例中,每个图像块经过卷积计算得到卷积特征矩阵后,直接对通过该卷积神经网络的池化层对该卷积特征矩阵进行池化计算,得到该图像块对应的池化特征矩阵,相比于现有的计算图像块池化特征矩阵的技术,该计算方式不需要将卷积特征矩阵先

缓存至DDR,然后再从DDR读取卷积结果以解决数据冗余的问题,因而可以减少内存资源的占用,并节约网络宽带。

[0059] 步骤S208:按照每个图像块在该待处理图像中的位置,将每个图像块的池化特征矩阵组成该待处理图像的池化特征矩阵。

[0060] 在其中一种可能的实施方式中,可以按照每个图像块在待处理图像中的位置,将位置相邻的图像块的池化特性矩阵进行拼接处理,得到该待处理图像的池化特征矩阵。

[0061] 本发明实施例提供的卷积神经网络的图像运算方法,首先获取预设的卷积神经网络的运算参数;该运算参数与该电子设备的存储属性相关联;然后根据该运算参数,从待处理图像中采集图像块;针对每个图像块,进行下述处理:通过该卷积神经网络的卷积层,对当前图像块进行卷积计算,得到该当前图像块的卷积特征矩阵;通过该卷积神经网络的池化层,对该卷积特征矩阵进行池化计算,得到当前图像块的池化特征矩阵;进而按照每个图像块在该待处理图像中的位置,将每个图像块的池化特征矩阵组成该待处理图像的池化特征矩阵。该方式中,基于下述原理:在卷积神经网络的卷积核参数和池化窗口参数确定的情况下,如果输入卷积层的图像块的尺寸一定,那么卷积层输出的卷积特征矩阵的尺度,以及池化层输出的池化特征矩阵的尺度也是确定的;因此,在卷积神经网络的卷积核参数和池化窗口参数确定的情况下,根据电子设备的存储属性,可以推导池化层输出的特征矩阵的尺度,进而设置恰当的窗口尺度参数(也即基于其关联关系预先推导出卷积神经网络的运算参数),使得采集得到的图像块的尺寸是与该卷积神经网络的卷积核参数和池化窗口参数相匹配的,进而使得图像块在卷积运算后得到的卷积特征矩阵中不出现冗余数据,从而可以对卷积层输出的卷积结果直接进行池化计算,而无需将该卷积结果缓存至DDR,实现了卷积运算到池化运算的流水线设计,可以减少内存资源的占用,并节约网络宽带。

[0062] 在图2所示卷积神经网络的图像运算方法的基础上,本实施例还提供了另一种卷积神经网络的图像运算方法,该方法应用于实现卷积神经网络运算的电子设备,其中,该方法重点描述了前述实施例中步骤S202(也即获取预设的卷积神经网络的运算参数)的具体实现过程。如图5所示为该卷积神经网络的运算方法的流程示意图,由图5可见,该方法包括以下步骤:

[0063] 步骤S502:获取该电子设备的硬件参数。

[0064] 这里,上述电子设备为用于实现卷积神经网络运算的电子设备,也即运行该卷积神经网络的运算方法的电子设备,其中,该电子设备可以是CNN硬件加速器。其中,硬件参数通常包括设备使用的处理器类型、主频和内存容量等。

[0065] 步骤S504:根据该硬件参数,确定预设的卷积神经网络的运算参数;该运算参数与该电子设备的存储属性相关联。

[0066] 在其中一种可能的实施方式中,可以通过下述步骤31-32确定预设的卷积神经网络的运算参数:

[0067] (31)根据该硬件参数,确定预设的卷积神经网络的约束条件;其中,该硬件参数包括该电子设备的内部缓存大小;该约束条件包括:该卷积神经网络中卷积层输出的卷积特征矩阵的最大尺度,或者,该卷积层输入的图像块的最大尺度。

[0068] 这里,对于预设的卷积神经网络,也即该卷积神经网络的卷积核参数和池化窗口参数确定的情况下,如果已知运行该卷积神经网络的电子设备的内部缓存大小,则可对应

的确定出该卷积神经网络中卷积层输出的卷积特征矩阵的最大尺度,并且,对于该卷积神经网络的卷积层,基于卷积层输出的卷积特征矩阵的尺度也可相应得到其输入的图像块的尺度,因而得到该卷积层输入的图像块的最大尺度。

[0069] (32) 根据该约束条件,确定该卷积神经网络的运算参数。

[0070] 在至少一种可能的实施方式中,如果该约束条件包括该卷积层输出的卷积特征矩阵的最大尺度,那么,可以通过下述步骤41-44来实现上述根据该约束条件,确定该卷积神经网络的运算参数的步骤:

[0071] (41) 根据该卷积特征矩阵的最大尺度,确定该卷积层输入的图像块的最大尺度。

[0072] 在本实施例中,通过下述公式确定该卷积层输入的图像块的最大尺度:

[0073]  $i\_cov\_width\_valid\_max = (o\_cov\_width\_max - 1) * cs\_w + 1$  (公式1.1)

[0074]  $i\_cov\_height\_valid\_max = (o\_cov\_height\_max - 1) * cs\_h + 1$  (公式1.2)

[0075] 其中, $i\_cov\_width\_valid\_max$ 为该卷积层输入的图像块的最大宽度; $i\_cov\_height\_valid\_max$ 为该卷积层输入的图像块的最大高度; $o\_cov\_width\_max$ 为该卷积层输出的卷积特征矩阵的最大宽度; $o\_cov\_height\_max$ 为该卷积层输出的卷积特征矩阵的最大高度; $cs\_w$ 为该卷积层的水平步长; $cs\_h$ 为该卷积层的垂直步长。

[0076] (42) 根据该卷积层输入的图像块的最大尺度,确定该卷积神经网络中池化层输出的池化特征矩阵的尺度。

[0077] 在本实施例中,通过下述公式,确定该卷积神经网络中池化层输出的池化特征矩阵的尺度:

[0078]  $(pw\_w + (pl\_width - 1) * ps\_w - 1) * cs\_w + 1 = i\_cov\_width\_valid\_max$  (公式1.3)

[0079]  $(pw\_h + (pl\_height - 1) * ps\_h - 1) * cs\_h + 1 = i\_cov\_height\_valid\_max$  (公式1.4)

[0080] 其中, $i\_cov\_width\_valid\_max$ 为该卷积层输入的图像块的最大宽度; $i\_cov\_height\_valid\_max$ 为该卷积层输入的图像块的最大高度; $cs\_w$ 为该卷积层的水平步长; $cs\_h$ 为该卷积层的垂直步长; $ps\_w$ 为该池化层的水平步长; $ps\_h$ 为该池化层的垂直步长; $pl\_width$ 为该池化层输出的池化特征矩阵的宽度; $pl\_height$ 为该池化层输出的池化特征矩阵的高度; $pw\_w$ 为该池化层的池化窗口宽度; $pw\_h$ 为该池化层的池化窗口高度。

[0081] (43) 根据该池化层输出的池化特征矩阵的尺度,确定该卷积神经网络的运算参数。

[0082] 在至少一种可能的实施方式中,可以首先对该池化层输出的池化特征矩阵的尺度进行取整处理;然后,根据取整后的该池化层输出的池化特征矩阵的尺度,确定该卷积神经网络的运算参数。

[0083] 其中,在对池化特征矩阵的尺度进行取整时,可以通过下述公式进行取整:

[0084]  $o\_pl\_width = \text{floor}(pl\_width)$  (公式1.5)

[0085]  $o\_pl\_height = \text{floor}(pl\_height)$  (公式1.6)

[0086] 上式中, $o\_pl\_width$ 为取整后的该池化层输出的池化特征矩阵的宽度; $o\_pl\_height$ 为取整后的该池化层输出的池化特征矩阵的高度; $\text{floor}$ 表示取整。在实际操作中,通常为向下取整,例如,如果计算得到的尺度为5.3,则取整为5。

[0087] 这里,将上述公式1.5和公式1.6代入上述公式1.3和公式1.4,可以得到下述公式:

[0088]  $i\_cov\_width\_valid = (pw\_w + (o\_pl\_width - 1) * ps\_w - 1) * cs\_w + 1$  (公式1.7)

[0089]  $i\_cov\_height\_valid = (pw\_h + (o\_pl\_height - 1) * ps\_h - 1) * cs\_h + 1$  (公式1.8)

[0090] 在实际操作中,由于卷积层输入的图像块的尺寸跟卷积核大小有关,考虑padding的情况下,在得到取整后的该池化层输出的池化特征矩阵的尺度之后,可以通过下述公式,计算该运算参数中用于从待处理图像中采集图像块的窗口尺度:

[0091]  $i\_cov\_width = (pw\_w + (o\_pl\_width - 1) * ps\_w - 1) * cs\_w + cw\_w$  (公式1.9)

[0092]  $i\_cov\_height = (pw\_h + (o\_pl\_height - 1) * ps\_h - 1) * cs\_h + cw\_h$  (公式1.10) 其中, $i\_cov\_width$ 为该窗口的宽度; $i\_cov\_height$ 为该窗口的高度; $pw\_w$ 为该池化层的池化窗口宽度; $pw\_h$ 为该池化层的池化窗口高度; $o\_pl\_width$ 为取整后的该池化层输出的池化特征矩阵的宽度; $o\_pl\_height$ 为取整后的该池化层输出的池化特征矩阵的高度; $ps\_w$ 为该池化层的水平步长; $ps\_h$ 为该池化层的垂直步长; $cs\_w$ 为该卷积层的水平步长; $cs\_h$ 为该卷积层的垂直步长。

[0093] 这里,上述采集图像块的窗口尺度也即是输入卷积层的图像块的尺度,相应地,也可以求得卷积层输出的卷积特征矩阵尺度如下:

[0094]  $o\_cov\_width = (i\_cov\_width\_valid - 1) / cs\_w + 1$  (公式1.11)

[0095]  $o\_cov\_height = (i\_cov\_height\_valid - 1) / cs\_h + 1$  (公式1.12)

[0096] 其中, $o\_cov\_width$ 为卷积特征矩阵的宽度; $o\_cov\_height$ 为卷积特征矩阵的高度。

[0097] 并且,在其中一种可能的实施方式中,还可以通过下述公式计算该运算参数中用于从待处理图像中采集图像块的窗口位移:

[0098]  $cb\_ws = ((pw\_w + (o\_pl\_width - 1) * ps\_w - 1) * cs\_w + 1 - (pw\_w - ps\_w)) * cs\_w$  (公式1.13)

[0099]  $cb\_hs = ((pw\_h + (o\_pl\_height - 1) * ps\_h - 1) * cs\_h + 1 - (pw\_h - ps\_h)) * cs\_h$  (公式1.14)

[0100] 其中, $cb\_ws$ 为该窗口的水平位移步长;该 $cb\_hs$ 为该窗口的垂直位移步长。

[0101] 由上述公式1.13和公式1.14可以看出,图像块的窗口位移除了与卷积参数和首个图像块尺寸有关之外,还与池化参数有关,当 $pw\_w$ 等于 $ps\_w$ ,且 $pw\_h$ 等于 $ps\_h$ 时,窗口位移仅与卷积参数和首个图像块的尺寸有关,由以上两个公式可以获得池化输出的相邻图像块,无任何重叠。如图6所示,为一种流水线设计的卷积神经网络运算过程示意图,由图6可以看出,图中四个图像块经过卷积和池化运算,完成了各个图像块的无缝对接,从而真正的实现卷积到池化的流水线设计。

[0102] 这样,即通过该池化层输出的池化特征矩阵的尺度,确定出该卷积神经网络的窗口尺度参数和窗口位移参数,进而可以根据该运行参数进行图像块采集。并且,根据上述步骤计算得到的该卷积神经网络的运行参数,是与运行该卷积神经网络的电子设备的硬件配置相匹配的,使得根据该运行参数采集得到的图像块,在卷积运算后得到的卷积特征矩阵中不出现冗余数据,从而可以对卷积层输出的卷积结果直接进行池化计算,而无需将该卷积结果缓存至DDR以解决冗余数据的问题。

[0103] 步骤S506:根据该运算参数,从待处理图像中采集图像块。

[0104] 步骤S508:针对每个图像块,进行下述处理:通过该卷积神经网络的卷积层,对当前图像块进行卷积计算,得到该当前图像块的卷积特征矩阵;通过该卷积神经网络的池化

层,对该卷积特征矩阵进行池化计算,得到当前图像块的池化特征矩阵。

[0105] 步骤S510:按照每个图像块在该待处理图像中的位置,将每个图像块的池化特征矩阵组成该待处理图像的池化特征矩阵。

[0106] 这里,本实施例中的步骤S506至步骤S510,对应上述实施例中的步骤S204至步骤S208,相应内容的描述可以参考前述实施例的对应部分,在此不再赘述。

[0107] 为了更清楚理解本实施例提供的卷积神经网络的图像运算方法中,获取卷积神经网络的运算参数的过程,下面介绍一个应用实例。

[0108] 假设根据某电子设备的内部RAM缓存大小,确定出卷积神经网络的约束条件如下表1和表2所示:

[0109] 表1图像块宽度硬件约束

cs_w	I_cov_width_valid_max
1	32
2	63
3	64

[0111] 表2图像块高度硬件约束

cs_h	i_cov_height_valid_max
1	16
2	31
3	46

[0113] 并且,该卷积神经网络中某层网络特性如下:pw\_w=5,ps\_w=4,cs\_w=2,cw\_w=3;pw\_h=5,ps\_h=4,cs\_h=2,cw\_h=3。

[0114] 由cs\_w=2查表可得:i\_cov\_width\_valid\_max=63。

[0115] 将i\_cov\_width\_valid\_max的值带入上述公式(1.3),得到:

[0116]  $(5+(p_l_w-1)*4-1)*2+1=63$ ,解出 $p_l\_width=7.75$ 。

[0117] 进而由公式(公式1.5)得到:o\_pl\_width=7。

[0118] 将o\_pl\_width的值带入(公式1.7),计算得到i\_cov\_width\_valid=57,可见,该值已经最接近理想值I\_cov\_width\_valid\_max。

[0119] 进一步地,由(公式1.9)计算得到i\_cov\_width=59;

[0120] 将i\_cov\_width的值带入(公式1.11)得到:o\_cov\_width=29。

[0121] 因此,输入卷积层的图像块宽度(也即窗口的宽度)i\_cov\_width为59,输出卷积层的卷积特征矩阵的宽度o\_cov\_widt为29,输出池化层的池化特征矩阵的宽度o\_pl\_w为7。同理,可求得输出卷积层的图像块高度(也即窗口的宽度)i\_cov\_height为27,输出卷积层的卷积特征矩阵的高度o\_cov\_height为13,输出池化层的池化特征矩阵的高度o\_pl\_height为3。

[0122] 这样,即实现了根据上述电子设备的硬件参数,确定预设的卷积神经网络的运算参数。在实际操作中,根据电子设备上卷积神经网络的网络参数,以及硬件设计存储资源大小的要求,根据上述公式1.1至公式1.14,即可计算出各个窗口信息(卷积输入、卷积输出、池化输出)和窗口的移位信息。由于这些窗口参数与卷积神经网络的通道数无关,因此,在硬件设计时可以通过cpu将上述运算参数配置到设计模块中的寄存器,再结合一定的控制

逻辑,即可实现卷积到池化的流水线设计。这里,由于该运算方法可以适用于具有不同卷积核参数及池化窗口参数的神经网络,因而该方法具有很好的通用性。

[0123] 本实施例提供的卷积神经网络的图像运算方法,根据实现该卷积神经网络运算的电子设备的硬件参数,确定出该卷积神经网络的约束条件,进而根据该约束条件确定该卷积神经网络的运算参数,并基于该运算参数采集图像块,使得图像块在卷积运算后得到的卷积特征矩阵中不出现冗余数据,从而可以对卷积层输出的卷积结果直接进行池化计算,而无需将该卷积结果缓存至DDR,实现了卷积运算到池化运算的流水线设计,可以减少内存资源的占用,进而减少CNN硬件实现面积,并且更加节约网络宽带。

[0124] 对应于图2中所示的卷积神经网络的图像运算方法,本发明实施例还提供了一种卷积神经网络的图像运算装置,该装置应用于实现卷积神经网络运算的电子设备,参见图7,所示为一种卷积神经网络的图像运算装置的结构示意图,由图7可见,该装置包括依次连接的运算参数获取模块71、图像块采集模块72、图像块池化特征矩阵计算模块73和图像池化特征矩阵计算模块74,其中,各个模块的功能如下:

[0125] 运算参数获取模块71,用于获取预设的卷积神经网络的运算参数;该运算参数与该电子设备的存储属性相关联;

[0126] 图像块采集模块72,用于根据该运算参数,从待处理图像中采集图像块;

[0127] 图像块池化特征矩阵计算模块73,用于针对每个图像块,进行下述处理:通过该卷积神经网络的卷积层,对当前图像块进行卷积计算,得到当前图像块的卷积特征矩阵;通过该卷积神经网络的池化层,对该卷积特征矩阵进行池化计算,得到该当前图像块的池化特征矩阵;

[0128] 图像池化特征矩阵计算模块74,用于按照每个图像块在该待处理图像中的位置,将每个图像块的池化特征矩阵组成该待处理图像的池化特征矩阵。

[0129] 本发明实施例提供了一种卷积神经网络的图像运算装置,首先获取预设的卷积神经网络的运算参数;该运算参数与该电子设备的存储属性相关联;然后根据该运算参数,从待处理图像中采集图像块;针对每个图像块,进行下述处理:通过该卷积神经网络的卷积层,对当前图像块进行卷积计算,得到该当前图像块的卷积特征矩阵;通过该卷积神经网络的池化层,对该卷积特征矩阵进行池化计算,得到当前图像块的池化特征矩阵;进而按照每个图像块在该待处理图像中的位置,将每个图像块的池化特征矩阵组成该待处理图像的池化特征矩阵。该装置中,基于下述原理:在卷积神经网络的卷积核参数和池化窗口参数确定的情况下,如果输入卷积层的图像块的尺寸一定,那么卷积层输出的卷积特征矩阵的尺度,以及池化层输出的池化特征矩阵的尺度也是确定的;因此,在卷积神经网络的卷积核参数和池化窗口参数确定的情况下,根据电子设备的存储属性,可以推导池化层输出的特征矩阵的尺度,进而设置恰当的窗口尺度参数(也即基于其关联关系预先推导出卷积神经网络的运算参数),使得采集得到的图像块的尺寸是与该卷积神经网络的卷积核参数和池化窗口参数相匹配的,进而使得图像块在卷积运算后得到的卷积特征矩阵中不出现冗余数据,从而可以对卷积层输出的卷积结果直接进行池化计算,而无需将该卷积结果缓存至DDR,实现了卷积运算到池化运算的流水线设计,可以减少内存资源的占用,并节约网络宽带。

[0130] 在其中一种可能的实施方式中,上述运算参数获取模块71还用于:获取该电子设备的硬件参数;根据该硬件参数,确定预设的卷积神经网络的运算参数。

[0131] 在另一种可能的实施方式中,上述运算参数获取模块71还用于:根据该硬件参数,确定预设的卷积神经网络的约束条件;其中,该硬件参数包括该电子设备的内部缓存大小;该约束条件包括:该卷积神经网络中卷积层输出的卷积特征矩阵的最大尺度,或者,该卷积层输入的图像块的最大尺度;根据该约束条件,确定该卷积神经网络的运算参数。

[0132] 在另一种可能的实施方式中,上述运算参数获取模块71还用于:如果该约束条件包括该卷积层输出的卷积特征矩阵的最大尺度,根据该卷积特征矩阵的最大尺度,确定该卷积层输入的图像块的最大尺度;根据该卷积层输入的图像块的最大尺度,确定该卷积神经网络中池化层输出的池化特征矩阵的尺度;根据该池化层输出的池化特征矩阵的尺度,确定该卷积神经网络的运算参数。

[0133] 在另一种可能的实施方式中,上述运算参数获取模块71还用于:通过下述公式,确定该卷积层输入的图像块的最大尺度: $i\_cov\_width\_valid\_max = (o\_cov\_width\_max - 1) * cs\_w + 1$ ;  $i\_cov\_height\_valid\_max = (o\_cov\_height\_max - 1) * cs\_h + 1$ ;其中, $i\_cov\_width\_valid\_max$ 为该卷积层输入的图像块的最大宽度; $i\_cov\_height\_valid\_max$ 为该卷积层输入的图像块的最大高度; $o\_cov\_width\_max$ 为该卷积层输出的卷积特征矩阵的最大宽度; $o\_cov\_height\_max$ 为该卷积层输出的卷积特征矩阵的最大高度; $cs\_w$ 为该卷积层的水平步长; $cs\_h$ 为该卷积层的垂直步长。

[0134] 在另一种可能的实施方式中,上述运算参数获取模块71还用于:通过下述公式,确定该卷积神经网络中池化层输出的池化特征矩阵的尺度: $(pw\_w + (pl\_width - 1) * ps\_w - 1) * cs\_w + 1 = i\_cov\_width\_valid\_max$ ;  $(pw\_h + (pl\_height - 1) * ps\_h - 1) * cs\_h + 1 = i\_cov\_height\_valid\_max$ ;其中, $i\_cov\_width\_valid\_max$ 为该卷积层输入的图像块的最大宽度; $i\_cov\_height\_valid\_max$ 为该卷积层输入的图像块的最大高度; $cs\_w$ 为该卷积层的水平步长; $cs\_h$ 为该卷积层的垂直步长; $ps\_w$ 为该池化层的水平步长; $ps\_h$ 为该池化层的垂直步长; $pl\_width$ 为该池化层输出的池化特征矩阵的宽度; $pl\_height$ 为该池化层输出的池化特征矩阵的高度; $pw\_w$ 为该池化层的池化窗口宽度; $pw\_h$ 为该池化层的池化窗口高度。

[0135] 在另一种可能的实施方式中,上述运算参数获取模块71还用于:对该池化层输出的池化特征矩阵的尺度进行取整处理;根据取整后的该池化层输出的池化特征矩阵的尺度,确定该卷积神经网络的运算参数。

[0136] 在另一种可能的实施方式中,上述运算参数获取模块71还用于:通过下述公式,计算该运算参数中,用于从待处理图像中采集图像块的窗口尺度: $i\_cov\_width = (pw\_w + (o\_pl\_width - 1) * ps\_w - 1) * cs\_w + cw\_w$ ;  $i\_cov\_height = (pw\_h + (o\_pl\_height - 1) * ps\_h - 1) * cs\_h + cw\_h$ ;其中, $i\_cov\_width$ 为该窗口的宽度; $i\_cov\_height$ 为该窗口的高度; $pw\_w$ 为该池化层的池化窗口宽度; $pw\_h$ 为该池化层的池化窗口高度; $o\_pl\_width$ 为取整后的该池化层输出的池化特征矩阵的宽度; $o\_pl\_height$ 为取整后的该池化层输出的池化特征矩阵的高度; $ps\_w$ 为该池化层的水平步长; $ps\_h$ 为该池化层的垂直步长; $cs\_w$ 为该卷积层的水平步长; $cs\_h$ 为该卷积层的垂直步长。

[0137] 在另一种可能的实施方式中,上述运算参数获取模块71还用于:通过下述公式,计算该运算参数中,用于从待处理图像中采集图像块的窗口位移: $cb\_ws = ((pw\_w + (o\_pl\_width - 1) * ps\_w - 1) * cs\_w + 1 - (pw\_w - ps\_w)) * cs\_w$ ;  $cb\_hs = ((pw\_h + (o\_pl\_height - 1) * ps\_h - 1) * cs\_h + 1 - (pw\_h - ps\_h)) * cs\_h$ ;其中, $cb\_ws$ 为该窗口的水平位移步长;该 $cb\_hs$ 为该窗口



的垂直位移步长。

[0138] 在另一种可能的实施方式中,上述运算参数包括:用于从待处理图像中采集图像块的窗口尺度参数和窗口位移参数,上述图像块采集模块72还用于:根据该运算参数中的窗口尺度参数和预设的默认初始位置,从待处理图像中确定第一个图像块;根据该窗口尺度参数和该窗口位移参数,按照预设顺序遍历该待处理图像,得到除该第一个图像块以外的图像块。

[0139] 在另一种可能的实施方式中,上述图像池化特征矩阵计算模块74还用于:按照每个该图像块在该待处理图像中的位置,将位置相邻的图像块的池化特性矩阵进行拼接处理,得到该待处理图像的池化特征矩阵。

[0140] 本发明实施例提供的卷积神经网络的图像运算装置,其实现原理及产生的技术效果和前述卷积神经网络的图像运算方法实施例相同,为简要描述,卷积神经网络的图像运算装置的实施例部分未提及之处,可参考前述卷积神经网络的图像运算方法实施例中相应内容。

[0141] 本发明实施例还提供了一种机器可读存储介质,该机器可读存储介质存储有机器可执行指令,该机器可执行指令在被处理器调用和执行时,该机器可执行指令促使处理器实现上述卷积神经网络的图像运算方法,具体实现可参见前述方法实施例,在此不再赘述。

[0142] 本发明实施例所提供的卷积神经网络的图像运算方法、卷积神经网络的图像运算装置和电子设备的计算机程序产品,包括存储了程序代码的计算机可读存储介质,所述程序代码包括的指令可用于执行前面方法实施例中所述的卷积神经网络的图像运算方法,具体实现可参见方法实施例,在此不再赘述。

[0143] 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,可以存储在一个处理器可执行的非易失的计算机可读取存储介质中。基于这样的理解,本发明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,或者网络设备等)执行本发明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(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0144] 另外,在本发明实施例的描述中,除非另有明确的规定和限定,术语“安装”、“相连”、“连接”应做广义理解,例如,可以是固定连接,也可以是可拆卸连接,或一体地连接;可以是机械连接,也可以是电连接;可以是直接相连,也可以通过中间媒介间接相连,可以是两个元件内部的连通。对于本领域的普通技术人员而言,可以根据具体情况理解上述术语在本发明中的具体含义。

[0145] 在本发明的描述中,需要说明的是,术语“中心”、“上”、“下”、“左”、“右”、“竖直”、“水平”、“内”、“外”等指示的方位或位置关系为基于附图所示的方位或位置关系,仅是为了便于描述本发明和简化描述,而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作,因此不能理解为对本发明的限制。此外,术语“第一”、“第二”、“第三”仅用于描述目的,而不能理解为指示或暗示相对重要性。

[0146] 最后应说明的是:以上所述实施例,仅为本发明的具体实施方式,用以说明本发明

的技术方案,而非对其限制,本发明的保护范围并不局限于此,尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,其依然可以对前述实施例所记载的技术方案进行修改或可轻易想到变化,或者对其中部分技术特征进行等同替换;而这些修改、变化或者替换,并不使相应技术方案的本质脱离本发明实施例技术方案的精神和范围,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以所述权利要求的保护范围为准。

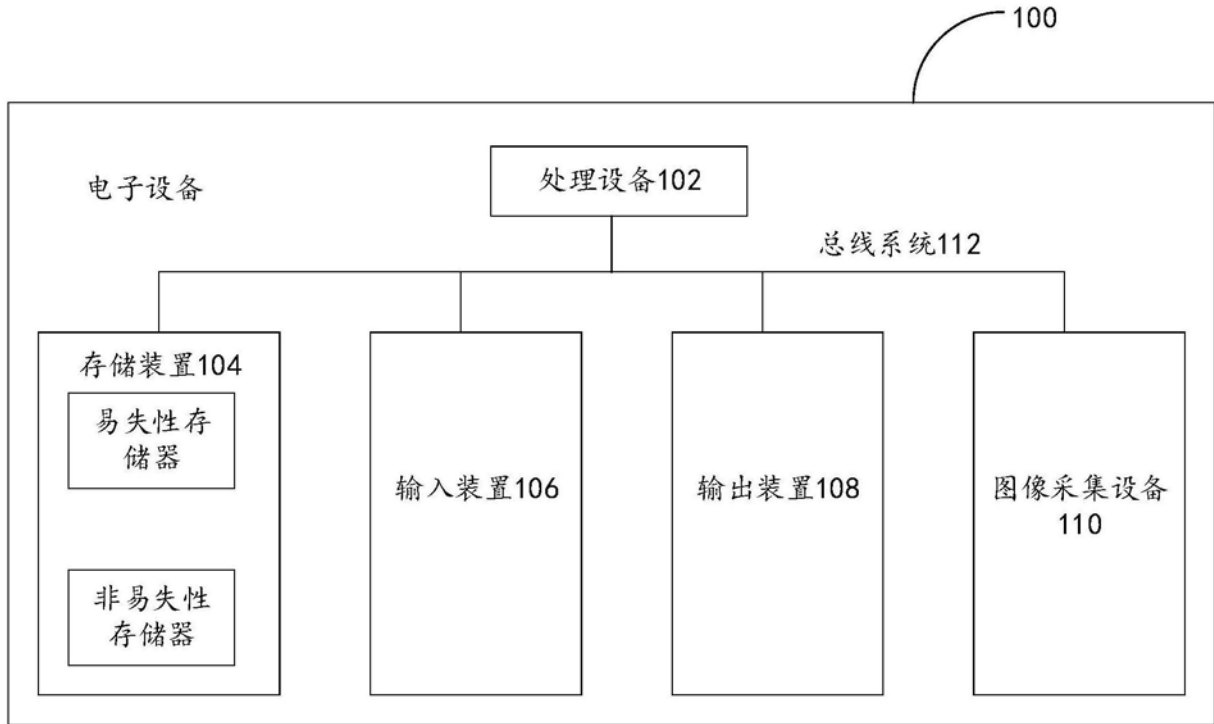


图1

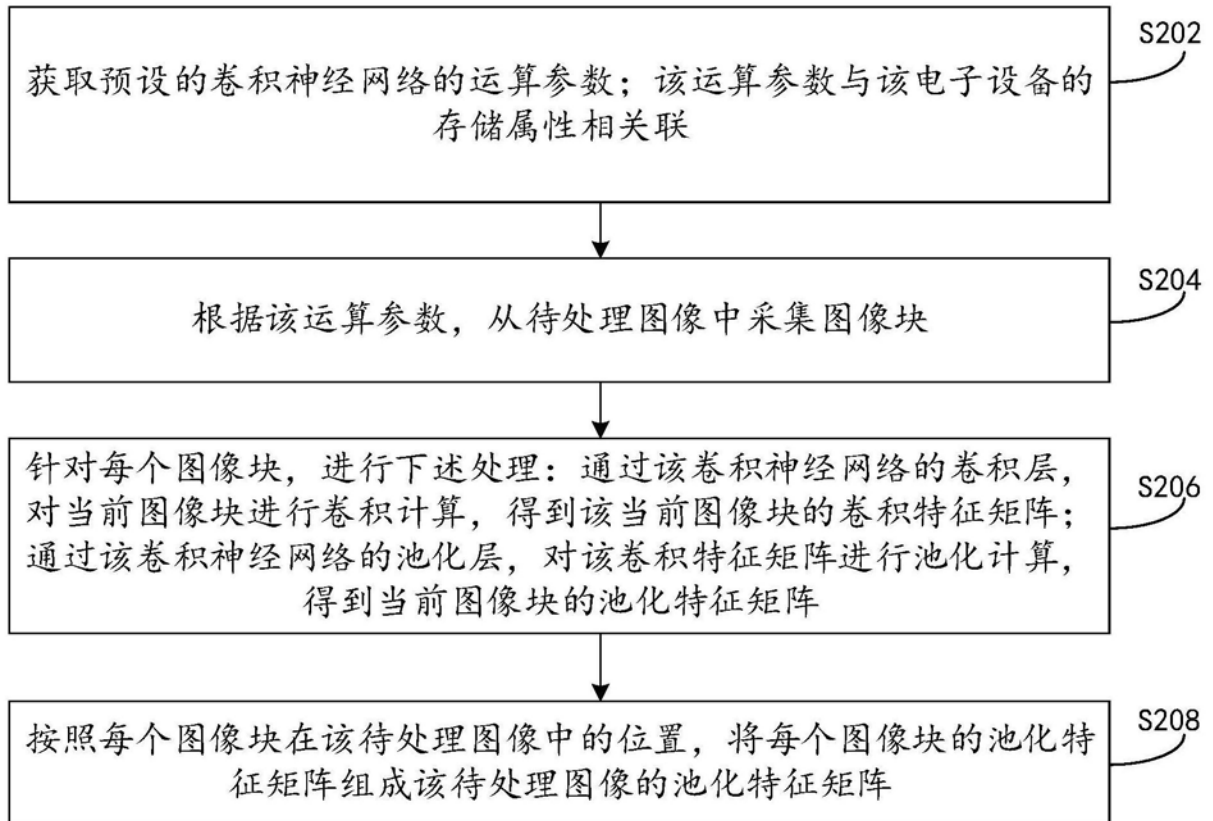


图2

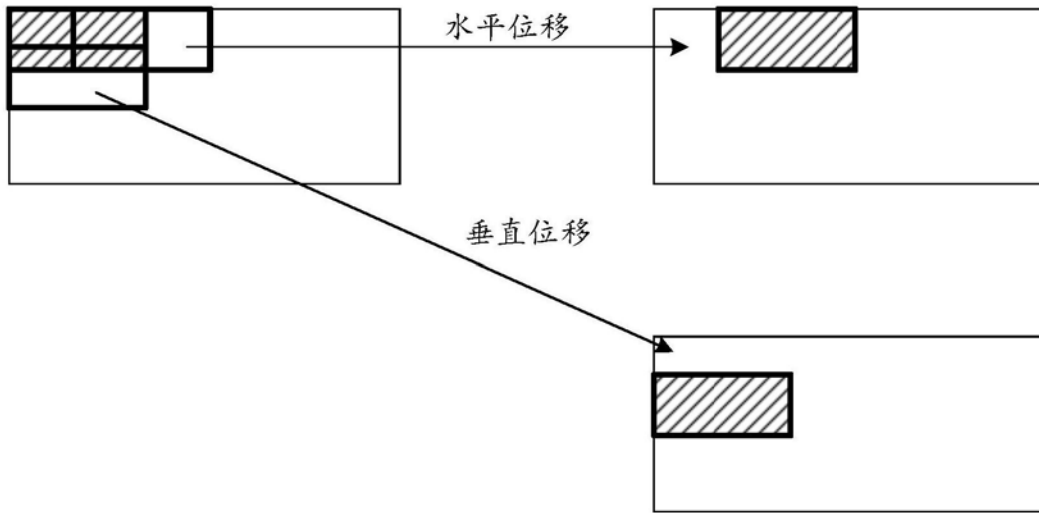


图3

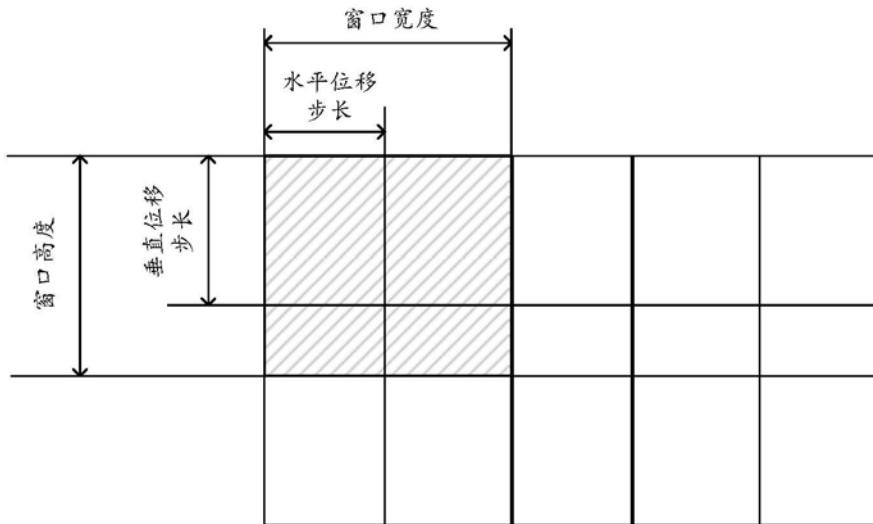


图4

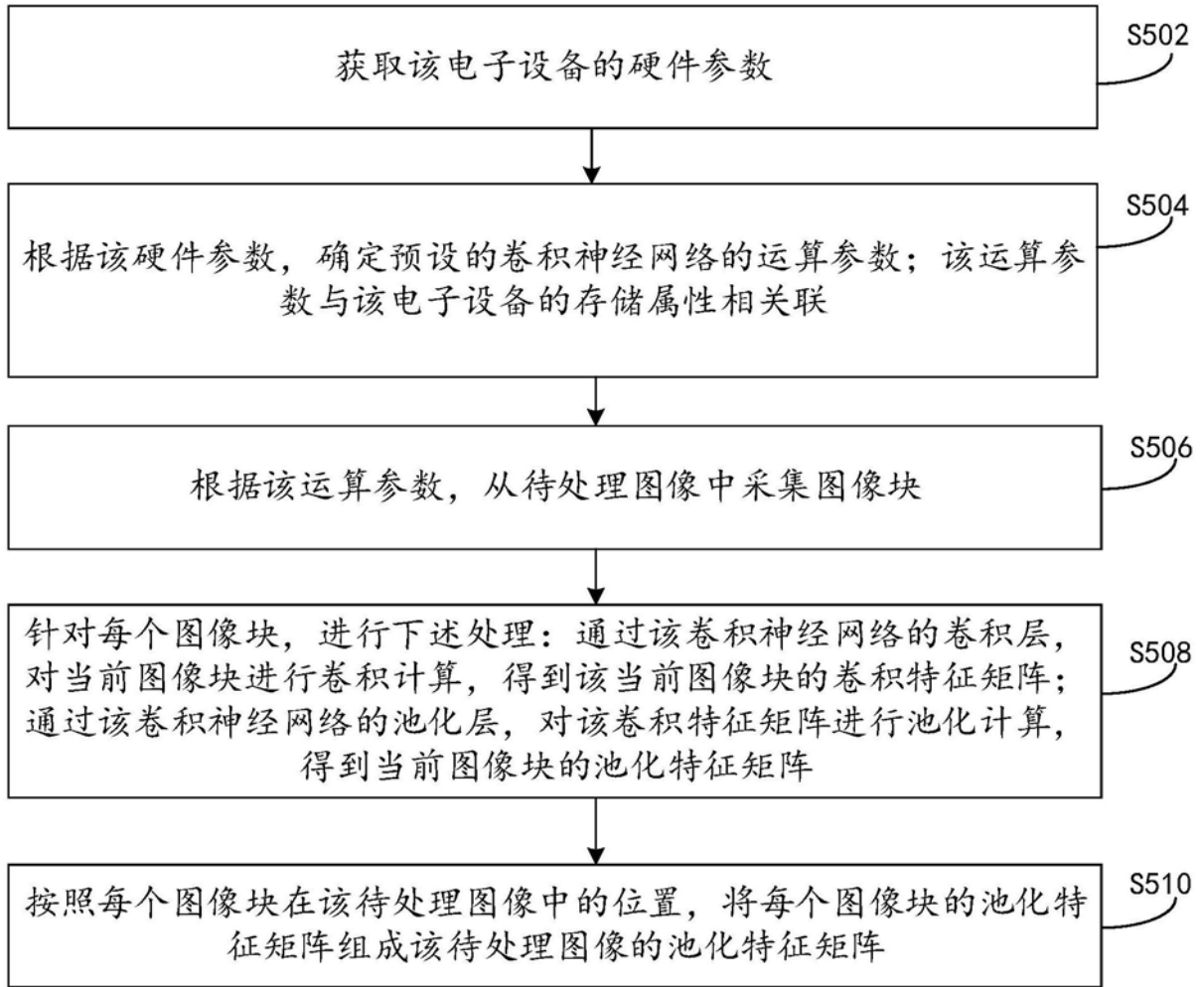


图5

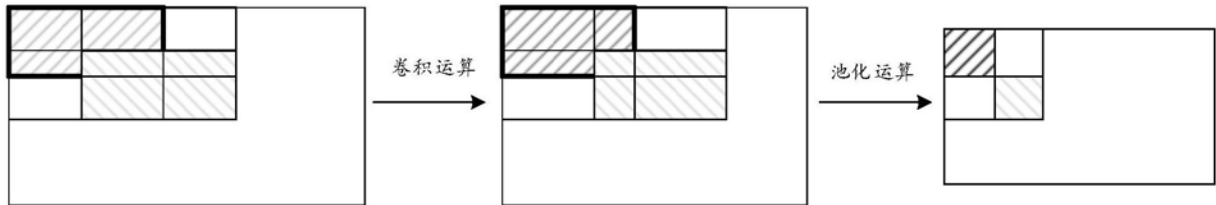


图6

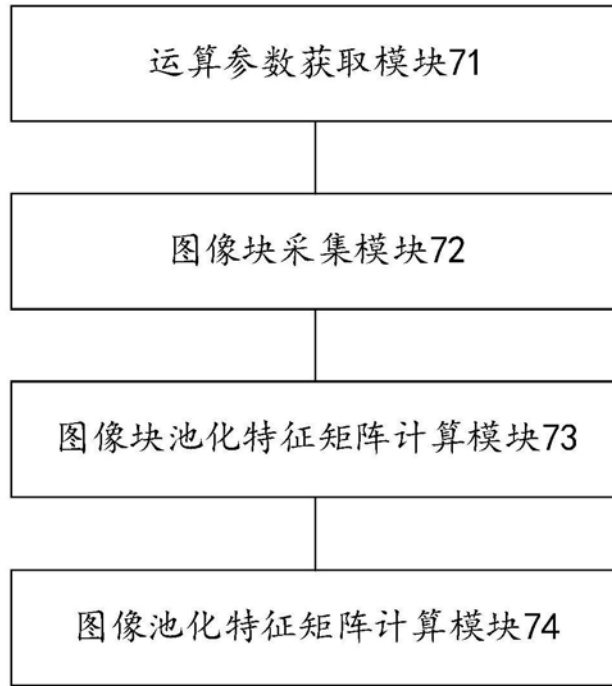


图7