



(12)发明专利

(10)授权公告号 CN 106295707 B

(45)授权公告日 2019.07.02

(21)申请号 201610683578.0

G06N 3/08(2006.01)

(22)申请日 2016.08.17

(56)对比文件

(65)同一申请的已公布的文献号
申请公布号 CN 106295707 A

CN 105512676 A,2016.04.20,

CN 104899579 A,2015.09.09,

CN 105160361 A,2015.12.16,

CN 104899579 A,2015.09.09,

CN 104992167 A,2015.10.21,

CN 105551047 A,2016.05.04,

(43)申请公布日 2017.01.04

(73)专利权人 北京小米移动软件有限公司
地址 100085 北京市海淀区清河中街68号
华润五彩城购物中心二期9层01房间

Kaiming He 等.Deep Residual Learning

for Image Recognition.《<https://arxiv.org/abs/1512.03385>》.2015,

(72)发明人 万韶华 杨松 陈志军

Ross Girshick.Fast R-CNN.《2015 IEEE International Conference on Computer Vision》.2015,

(74)专利代理机构 北京同立钧成知识产权代理有限公司 11205

代理人 杨泽 刘芳

审查员 黄文琪

(51)Int.Cl.

G06K 9/62(2006.01)

G06K 9/00(2006.01)

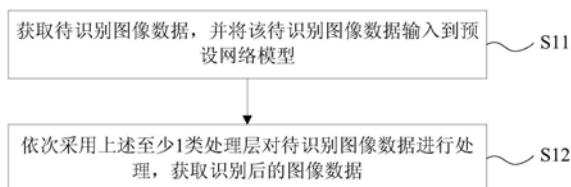
权利要求书2页 说明书13页 附图4页

(54)发明名称

图像识别方法和装置

(57)摘要

本公开是关于一种图像识别方法和装置,其中,该方法包括:获取待识别图像数据,并将待识别图像数据输入到预设网络模型,该预设网络模型包括至少1类处理层,每类处理层包含至少3个按序依次排列子处理层,依次采用上述至少1类处理层对待识别图像数据进行处理,获取识别后的图像数据,其中,每类处理层中的至少子3个子处理层依次对待识别图像数据进行处理,且每类处理层中从第3个子处理层开始,每个子处理层的输入包括至少2个在先子处理层的输出。本公开的技术方案,能够更快速的获取到识别后的图像数据,提高该预设网络模型的精度,进而提高了图像识别的精度。



1. 一种图像识别方法,其特征在于,包括:

获取待识别图像数据,并将所述待识别图像数据输入到预设网络模型,所述预设网络模型包括至少1类处理层,每类处理层包含至少3个子处理层,所述至少3个子处理层按序依次排列;

依次采用所述至少1类处理层对所述待识别图像数据进行处理,获取识别后的图像数据;

其中,每类处理层中的所述至少3个子处理层依次对所述待识别图像数据进行处理,且每类处理层中从第3个所述子处理层开始,每个所述子处理层的输入包括:至少2个在先子处理层的输出;所述待识别图像数据对应的矩阵向量与第一类处理层的第一个子处理层对应的矩阵向量相匹配。

2. 根据权利要求1所述的方法,其特征在于,所述依次采用所述至少1类处理层对所述待识别图像数据进行处理,获取识别后的图像数据,包括:

根据随机梯度下降法,依次采用所述至少1类处理层对所述待识别图像数据进行处理,获取识别后的图像数据。

3. 根据权利要求1或2所述的方法,其特征在于,所述至少1类处理层包括:卷积层和全连接层;

其中,所述卷积层包含至少3个子卷积层,所述全连接层包含至少3个子全连接层。

4. 根据权利要求3所述的方法,其特征在于,所述依次采用所述至少1类处理层对所述待识别图像数据进行处理,获取识别后的图像数据,包括:

采用至少3个子卷积层依次对所述待识别图像数据进行卷积处理,得到卷积后的图像数据;

采用至少3个子全连接层依次对所述卷积后的图像数据进行全连接处理,获取识别后的图像数据。

5. 根据权利要求4所述的方法,其特征在于,所述方法还包括:

对每个子全连接层的总处理数据量进行压缩。

6. 根据权利要求5所述的方法,其特征在于,所述对每个子全连接层的总处理数据量进行压缩,包括:

分别对每个子全连接层的参数矩阵进行奇异值分解,获取每个子全连接层的参数矩阵对应的第一参数子矩阵以及第二参数子矩阵;

则所述采用至少3个子全连接层依次对所述卷积后的图像数据进行全连接处理,获取识别后的图像数据,包括:

采用所述至少3个子全连接层中每个子全连接层的参数矩阵对应的第一参数子矩阵以及第二参数子矩阵依次对所述卷积后的图像数据对应的二维矩阵进行全连接处理,获取识别后的图像数据。

7. 一种图像识别装置,其特征在于,包括:

获取模块,被配置为获取待识别图像数据,并将所述待识别图像数据输入到预设网络模型,所述预设网络模型包括至少1类处理层,每类处理层包含至少3个子处理层,所述至少3个子处理层按序依次排列;

处理模块,被配置为依次采用所述至少1类处理层对所述获取模块获取到的所述待识

别图像数据进行处理,获取识别后的图像数据;

其中,每类处理层中的所述至少3个子处理层依次对所述待识别图像数据进行处理,且每类处理层中从第3个所述子处理层开始,每个所述子处理层的输入包括:至少2个在先子处理层的输出;所述待识别图像数据对应的矩阵向量与第一类处理层的第一个子处理层对应的矩阵向量相匹配。

8.根据权利要求7所述的装置,其特征在于,所述处理模块,还被配置为根据随机梯度下降法,依次采用所述至少1类处理层对所述待识别图像数据进行处理,获取识别后的图像数据。

9.根据权利要求7或8所述的装置,其特征在于,所述至少1类处理层包括:卷积层和全连接层;

其中,所述卷积层包含至少3个子卷积层,所述全连接层包含至少3个子全连接层。

10.根据权利要求9所述的装置,其特征在于,所述处理模块,包括:第一处理子模块和第二处理子模块;

所述第一处理子模块,被配置为采用至少3个子卷积层依次对所述获取模块获取到的所述待识别图像数据进行卷积处理,得到卷积后的图像数据;

所述第二处理子模块,被配置为采用至少3个子全连接层依次对所述第一处理子模块得到的所述卷积后的图像数据进行全连接处理,获取识别后的图像数据。

11.根据权利要求10所述的装置,其特征在于,所述装置还包括:压缩模块;

所述压缩模块,被配置为对每个子全连接层的总处理数据量进行压缩。

12.根据权利要求11所述的装置,其特征在于,所述压缩模块,包括:分解子模块;

所述分解子模块,被配置为分别对每个子全连接层的参数矩阵进行奇异值分解,获取每个子全连接层的参数矩阵对应的第一参数子矩阵以及第二参数子矩阵;

则所述第二处理子模块,还被配置为采用所述至少3个子全连接层中所述分解子模块获取到的每个子全连接层的参数矩阵对应的第一参数子矩阵以及第二参数子矩阵依次对所述卷积后的图像数据对应的二维矩阵进行全连接处理,获取识别后的图像数据。

13.一种图像识别装置,其特征在于,所述装置包括:

处理器;

用于存储所述处理器的可执行指令的存储器;

其中,所述处理器被配置为:

获取待识别图像数据,并将所述待识别图像数据输入到预设网络模型,所述预设网络模型包括至少1类处理层,每类处理层包含至少3个子处理层,所述至少3个子处理层按序依次排列;

依次采用所述至少1类处理层对所述待识别图像数据进行处理,获取识别后的图像数据;

其中,每类处理层中的所述至少3个子处理层依次对所述待识别图像数据进行处理,且每类处理层中从第3个所述子处理层开始,每个所述子处理层的输入包括:至少2个在先子处理层的输出;所述待识别图像数据对应的矩阵向量与第一类处理层的第一个子处理层对应的矩阵向量相匹配。

图像识别方法和装置

技术领域

[0001] 本公开涉及图像处理技术领域,尤其涉及一种图像识别方法和装置。

背景技术

[0002] 图像识别是利用计算机对图像进行处理、分析和理解,以识别各种不同模式的目标和对象的技术。目前,图像识别技术已广泛应用于生物医学、货物检测、目标跟踪、自主车导航等多个领域。

[0003] 现有技术中,通常对某个特定领域的内容进行图像识别或者对图像中的不同区域进行识别以得到不同的识别结果。对于某一区域的图像,往往只能给出单一的识别结果。

发明内容

[0004] 本公开实施例提供了一种图像识别方法和装置。所述技术方案如下:

[0005] 根据本公开实施例的第一方面,提供一种图像识别方法,该方法包括:

[0006] 获取待识别图像数据,并将所述待识别图像数据输入到预设网络模型,所述预设网络模型包括至少1类处理层,每类处理层包含至少3个子处理层,所述至少3个子处理层按序依次排列;

[0007] 依次采用所述至少1类处理层对所述待识别图像数据进行处理,获取识别后的图像数据;

[0008] 其中,每类处理层中的所述至少3个子处理层依次对所述待识别图像数据进行处理,且每类处理层中从第3个所述子处理层开始,每个所述子处理层的输入包括:至少2个在先子处理层的输出。

[0009] 根据本公开实施例的第二方面,提供了一种图像识别装置,该装置包括:

[0010] 获取模块,被配置为获取待识别图像数据,并将所述待识别图像数据输入到预设网络模型,所述预设网络模型包括至少1类处理层,每类处理层包含至少3个子处理层,所述至少3个子处理层按序依次排列;

[0011] 处理模块,被配置为依次采用所述至少1类处理层对所述获取模块获取到的所述待识别图像数据进行处理,获取识别后的图像数据;

[0012] 其中,每类处理层中的所述至少3个子处理层依次对所述待识别图像数据进行处理,且每类处理层中从第3个所述子处理层开始,每个所述子处理层的输入包括:至少2个在先子处理层的输出。

[0013] 根据本公开实施例的第三方面,提供了一种图像识别装置,该装置包括:

[0014] 处理器;

[0015] 用于存储所述处理器的可执行指令的存储器;

[0016] 其中,所述处理器被配置为:

[0017] 获取待识别图像数据,并将所述待识别图像数据输入到预设网络模型,所述预设网络模型包括至少1类处理层,每类处理层包含至少3个子处理层,所述至少3个子处理层按

序依次排列；

[0018] 依次采用所述至少1类处理层对所述待识别图像数据进行处理，获取识别后的图像数据；

[0019] 其中，每类处理层中的所述至少子3个子处理层依次对所述待识别图像数据进行处理，且每类处理层中从第3个所述子处理层开始，每个所述子处理层的输入包括：至少2个在先子处理层的输出。

[0020] 本公开实施例提供的技术方案可以包括以下有益效果：

[0021] 通过将获取到的待识别图像数据输入到预设网络模型，并依次采用预设网络模型包括的至少1类处理层对待识别图像数据进行处理，获取识别后的图像数据，由于该预设模块每类处理层中的至少子3个子处理层依次对待识别图像数据进行处理，且每类处理层中从第3个子处理层开始，每个子处理层的输入包括至少2个在先子处理层的输出，其能够更迅速的获取到识别后的图像数据，提高了该预设网络模型的精度，进而提高了图像识别的精度。

[0022] 应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，并不能限制本公开。

附图说明

[0023] 此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本公开的实施例，并与说明书一起用于解释本公开的原理。

[0024] 图1是根据一示例性实施例示出的一种图像识别方法的流程图；

[0025] 图2为根据一示例性实施例示出的预设网络模型的结构框图；

[0026] 图3是根据另一示例性实施例示出的一种图像识别方法的流程图；

[0027] 图4是根据再一示例性实施例示出的一种图像识别方法的流程图；

[0028] 图5为Alex Net深度卷积模型网络的一种框图；

[0029] 图6是根据又一示例性实施例示出的一种图像识别方法的流程图；

[0030] 图7是根据一示例性实施例示出的一种图像识别装置的框图；

[0031] 图8是根据另一示例性实施例示出的一种图像识别装置的框图；

[0032] 图9是根据再一示例性实施例示出的一种图像识别装置的框图；

[0033] 图10是根据一示例性实施例示出的一种图像识别装置的实体的框图；

[0034] 图11是根据一示例性实施例示出的一种图像识别装置的框图。

[0035] 通过上述附图，已示出本公开明确的实施例，后文中将有更详细的描述。这些附图和文字描述并不是为了通过任何方式限制本公开构思的范围，而是通过参考特定实施例为本领域技术人员说明本公开的概念。

具体实施方式

[0036] 这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本公开相一致的所有实施方式。相反，它们仅是与如所附权利要求书中所详述的、本公开的一些方面相一致的装置和方法的例子。

[0037] 图1是根据一示例性实施例示出的一种图像识别方法的流程图。图2为根据一示例性实施例示出的预设网络模型的结构框图。本实施例以图像识别方法应用于终端设备或服务器中来举例说明。其中,该终端设备例如是手机、PDA(Personal Digital Assistant,个人数字助理,简称PDA)等,也可以是具有图像识别功能的其他设备。如图1所示,该图像识别方法可以包括如下几个步骤:

[0038] 在步骤S11中,获取待识别图像数据,并将该待识别图像数据输入到预设网络模型。

[0039] 其中,预设网络模型包括至少1类处理层,每类处理层包含至少3个子处理层,该至少3个子处理层按序依次排列。

[0040] 作为一种示例,如图2所示,假设预设网络模型包括两类处理层,分别是第一类处理层21和第二类处理层22,该第一类处理层21包括五个子处理层,分别为第一类第一子处理层211、第一类第二子处理层212、第一类第三子处理层213、第一类第四子处理层214和第一类第五子处理层215,该第二类处理层22包括三个子处理层,分别为第二类第一子处理层221、第二类第二子处理层222和第二类第三子处理层223,而且,第一类处理层21和第二类处理层22依次排列,且第一类处理层21中的五个子处理层和第二类处理层22中的三个子处理层依次排列。也即,待识别图像数据首先输入到第一类处理层21的第一类第一子处理层211中,且待识别图像数据满足第一类第一子处理层211的输入大小要求。

[0041] 可选的,在本实施例中,该图像识别方法可适用于对某一图像中目标物的识别应用中,例如,图像中建筑物或动植物类别的识别等。

[0042] 在对待识别图像进行识别时,首先需要获取到待识别图像数据,该待识别图像数据,例如是像素图像数据或深度图像数据,其中,像素图像数据也即RGB(红、绿、蓝)图像数据,指的是图像分辨率的大小,可利用RGB传感器获得,而深度图像数据也即像素深度图像数据,是指存储每个像素所用的位数,用来度量图像的分辨率,可以通过深度传感器获取。

[0043] 可选的,在本公开的实施例中,可以通过摄像设备对目标场景进行图像数据采集计算来获取待识别图像数据,进而在获取到待识别图像数据之后,将待识别图像数据输入到预设网络模型中进行处理。

[0044] 在本实施例中,选用的预设网络模型包括至少1类处理层,每类处理层包含至少3个子处理层,并且至少3个子处理层按序依次排列。可选的,该预设模块可以是Alex Net深度卷积网络。

[0045] 值得说明的是,该待识别图像数据需要与预设网络模型中第一类处理层的第一个子处理层相对应,也即,待识别图像数据对应的矩阵向量与第一类处理层的第一个子处理层对应的矩阵向量相匹配。

[0046] 在步骤S12中,依次采用上述至少1类处理层对待识别图像数据进行处理,获取识别后的图像数据。

[0047] 其中,每类处理层中的至少子3个子处理层依次对待识别图像数据进行处理,且每类处理层中从第3个子处理层开始,每个子处理层的输入包括:至少2个在先子处理层的输出。

[0048] 在本实施例中,终端设备采用预设网络模型的至少1类处理层对获取到的待识别图像数据进行处理,具体的,使待识别图像数据依次经过每类处理层内的至少3个子处理

层,利用该至少3个子处理层依次对待识别图像数据进行处理,而且,为了加快预设网络模型对待识别图像数据的处理速度,提高图像识别的精度,每类处理层中从第3个子处理层开始,通过增加在先子处理层与该子处理层的快捷链接,使得每个子处理层的输入包括至少2个在先子处理层的输出,这样从第3个子处理层开始,每个子处理层的运算数据量增加了一倍,进而加快了整个预设网络模型的处理速度,提高了图像识别的精度。

[0049] 需要说明的是,从第3个子处理层开始,当通过快捷链接的方式将在先子处理层的输出作为该子处理层的输入时,在先子处理层的输出需要与该子处理层的输入相匹配,若在先子处理层的输出与子处理层的输入不匹配,则需要按照该子处理层的输入对在先子处理层的输出进行上采样或下采样,进而使在先子处理层的输出经过采样之后,其大小能够满足该子处理层的输入要求。

[0050] 可选的,如图2所示,第一类处理层的第4个子处理层的输入包括在先3个子处理层的输出,则分别在第一类第一子处理层211、第一类第二子处理层212和第一类第四子处理层214之间建立快捷链接,将第一类第一子处理层211、第一类第二子处理层212的输出分别经过上采样和下采样之后和第一类第三子处理层213的输出一起输入到第一类第四子处理层214中。

[0051] 本实施例提供的图像识别方法,通过将获取到的待识别图像数据输入到预设网络模型,并依次采用预设网络模型包括的至少1类处理层对待识别图像数据进行处理,获取识别后的图像数据,由于该预设模块每类处理层中的至少子3个子处理层依次对待识别图像数据进行处理,且每类处理层中从第3个子处理层开始,每个子处理层的输入包括至少2个在先子处理层的输出,其能够更容易获取到识别后的图像数据,提高该预设网络模型的精度,进而提高图像识别的精度。

[0052] 在上述实施例的基础上,本实施例涉及采用至少1类处理层对待识别图像数据进行处理的具体方法。即,对上述步骤S12(依次采用至少1类处理层对待识别图像数据进行处理,获取识别后的图像数据)的详细说明,详细请参照图3所示实施例。

[0053] 图3是根据另一示例性实施例示出的一种图像识别方法的流程图。本公开实施例是在上述实施例的基础上对图像识别方法技术方案的进一步说明。如图3所示,本实施例涉及的图像识别方法用于终端设备中,该终端设备例如可以是手机、PDA等,也可以是具有通话功能的其他设备。在本实施例涉及的图像识别方法中,上述步骤S12(依次采用至少1类处理层对待识别图像数据进行处理,获取识别后的图像数据)可通过如下步骤实现。

[0054] 在步骤S31中,根据随机梯度下降法,依次采用至少1类处理层对待识别图像数据进行处理,获取识别后的图像数据。

[0055] 可选的,随机梯度下降法(stochastic gradient descent,简称SGD)也叫增量梯度下降法,具体是根据随机选取的某个单独样例的误差增量计算权值更新,进而得到近似解的梯度下降搜索方法。随机梯度下降法与梯度下降法相比,收敛速度快,因此,本公开实施例根据随机梯度下降法,依次采用预设网络模型的至少1类处理层对待识别图像数据进行处理,获取识别后的图像数据,可有效加快待识别图像数据的识别速度,提高图像识别的精度。

[0056] 值得说明的是,本公开实施例以预设网络模型为Alex Net深度卷积网络为例进行说明。下面简要描述随机梯度下降法应用于该预设网络模型的实现原理:

[0057] 首先,利用公式(1)表示预设网络模型对图像识别的总误差:

$$[0058] \quad E = \frac{1}{2} \sum_i (y_i - a_i)^2 \quad (1)$$

[0059] 其中,E表示预设网络模型的总误差数据, y_i 表示第*i*个图像的真实图像数据, a_i 表示第*i*个图像利用预设网络模型识别后的图像数据。

[0060] 在利用该预设网络模型进行图像识别时,该预设网络模型中每类处理层的每一子处理层的权值系数矩阵用W表示,那么,本公开实施例的目的就是通过调节权值系数矩阵W来使总误差数据达到最小。众所周知,由于权值系数矩阵W的变化量可用公式(2)所示的偏导公式表示,因此,对于预设网络模型中每类处理层的每一子处理层的权值系数可通过公式(3)进行更新。

$$[0061] \quad \Delta W \propto -\frac{\partial E}{\partial W} \quad (2)$$

$$[0062] \quad \theta_{ij}^{(l+1)} = \theta_{ij}^{(l)} + \Delta \theta_{ij}^{(l)} = \theta_{ij}^{(l)} - \alpha \frac{\partial E(\theta)}{\partial \theta_{ij}^{(l)}} \quad (3)$$

[0063] 其中, $\theta_{ij}^{(l+1)}$ 是第*i*个图像的第*j*个待识别图像数据在该预设网络模型的第*l*+1层的权值系数, $\theta_{ij}^{(l)}$ 是第*i*个图像的第*j*个待识别图像数据在该预设网络模型的第*l*层的权值系数, $\Delta \theta_{ij}^{(l)}$ 是第*i*个图像的第*j*个待识别图像数据在该预设网络模型的第*l*层的误差权值系数, α 为该预设网络模型中第*l*层的误差权值系数的权重。

[0064] 对于该预设网络模型的最后一层(L),第*i*个图像经过上述公式(1)至公式(3)的计算之后,可直接计算出该预设网络模型识别后的图像数据 $a_i^{(L)}$ 与该第*i*个图像的真实图像数据之间的差值,本实施例中,将其定义为 $\delta_i^{(L)}$ 。对于该预设网络模型中间各层中的节点,可通过计算各层节点差值的加权平均值来计算该节点的差值。

[0065] 在该预设网络模型的最后一层中,识别后的图像数据 $a_i^{(L)}$ 与该第*i*个图像的真实图像数据之间的差值 $\delta_i^{(L)}$ 用公式(4)表示:

$$[0066] \quad \delta_i^{(L)} = \frac{\partial E}{\partial z_i^{(L)}} = \frac{\partial \frac{1}{2} (y_i - a_i)^2}{\partial z_i^{(L)}} = \frac{\partial \left[\frac{1}{2} (y - g(z_i^{(L)}))^2 \right]}{\partial z_i^{(L)}} = (a_i^{(L)} - y) \cdot g'(z_i^{(L)}) \quad (4)$$

[0067] 其中, $z_i^{(L)}$ 为第L层中第*z*个节点的权值系数, $g(z_i^{(L)})$ 为第*i*个图像经过第L层的第*z*个节点后得到的图像数据, $g'(z_i^{(L)})$ 为第L层的第*z*个节点的差值。

[0068] 对于该预设网络模型最后一层之前的每一层(1),识别后的图像数据 $a_i^{(l)}$ 与该第*i*个图像的真实图像数据之间的差值 $\delta_i^{(l)}$ 用公式(5)表示:

$$\begin{aligned}
\delta_i^{(l)} &= \frac{\partial E}{\partial z_i^{(l)}} = \sum_j^{N^{(l+1)}} \frac{\partial E}{\partial z_j^{(l+1)}} \cdot \frac{\partial z_j^{(l+1)}}{\partial z_i^{(l)}} \\
[0069] \quad &= \sum_j^{N^{(l+1)}} \delta_j^{(l+1)} \cdot \frac{\partial \left[\sum_k^{N^{(l)}} \theta_{jk}^{(l)} \cdot g(z_k^{(l)}) \right]}{\partial z_i^{(l)}}, \quad k \in i \\
&= \sum_j^{N^{(l+1)}} (\delta_j^{(l+1)} \cdot \theta_{ji}^{(l)}) \cdot g'(z_i^{(l)})
\end{aligned} \tag{5}$$

[0070] 其中, $z_i^{(l)}$ 为第1层中第z个节点的权值系数, $g(z_k^{(l)})$ 为第k个图像经过第1层的第z个节点后得到的图像数据, $g'(z_i^{(l)})$ 为第1层的第z个节点的差值, $N^{(l+1)}$ 为第i个图像的待识别图像数据总个数, j为1与 $N^{(l+1)}$ 之间的任一待识别图像数据, $\theta_{ji}^{(l)}$ 是第i个图像的第j个待识别图像数据在该预设网络模型的第1层的权值系数。

[0071] 因此, 第i个图像经过第1层后得到的图像数据 $a_i^{(l)}$ 与该第i个图像的真实图像数据之间的差值 $\delta_i^{(l)}$ 可用公式 (6) 表示:

$$[0072] \quad \delta_i^{(l)} = \sum_j^{N^{(l+1)}} (\delta_j^{(l+1)} \cdot \theta_{ji}^{(l)}) \cdot g'(z_i^{(l)}) \tag{6}$$

[0073] 由前述可知, 本公开实施例的目的是计算出 $\frac{\partial E(\theta)}{\partial \theta_{ij}^{(l)}}$, 由于 $\frac{\partial E}{\partial \theta'_{ji}}$ 可用公式 (7) 表示, 因此, 本实施例中预设网络模型的权重系数的随机下降方程可用公式 (8) 表示。

$$[0074] \quad \frac{\partial E}{\partial \theta'_{ji}} = \frac{\partial E}{\partial z_i^{(l+1)}} \cdot \frac{\partial z_i^{(l+1)}}{\partial \theta'_{ji}} = \delta_i^{(l+1)} \cdot a_j^{(l)} \tag{7}$$

$$[0075] \quad \theta'_{ji} = \theta_{ji} - \alpha \cdot \delta_i^{(l+1)} \cdot a_j^{(l)} \tag{8}$$

[0076] 其中, θ'_{ji} 为该预设网络模型中权值系数矩阵W中各参数值的变化量, α 为该预设网络模型中误差权值系数 $\delta_i^{(l+1)}$ 的权重, $a_j^{(l)}$ 为第i个图像经过预设网络模型的第1层后的图像数据。

[0077] 值得说明的是, 上述给出的随机梯度下降法只是给出了一种快速计算收敛的方法, 对于具体的计算细节本公开实施例并不对其进行限定。

[0078] 可选的, 在图1和图3所示实施例提供的图像识别方法中, 上述至少1类处理层包括: 卷积层和全连接层。

[0079] 其中, 该卷积层包含至少3个子卷积层, 该全连接层包含至少3个子全连接层。

[0080] 作为一种示例, 当预设网络模型具体采用Alex Net深度卷积网络实现时, 该Alex Net深度卷积网络包括卷积层 (Convolution Layer)、激活层 (Activation Layer)、池化层 (Pooling Layer)、全连接层 (Fully Connected Layer) 等, 其中, 卷积层包括5个子卷积层, 全连接层包括3个子全连接层等, 而且, 卷积层与卷积层之间包括激活层和池化层。

[0081] 在上述实施例的基础上, 当上述至少1类处理层包括: 卷积层和全连接层, 且, 卷积层包含至少3个子卷积层, 全连接层包含至少3个子全连接层时, 上述步骤S12 (依次采用至

少1类处理层对待识别图像数据进行处理,获取识别后的图像数据)的可能实现方式请参照图4所示实施例。

[0082] 图4是根据再一示例性实施例示出的一种图像识别方法的流程图。本公开实施例是在上述实施例的基础上对图像识别方法技术方案的进一步说明。本实施例涉及的图像识别方法同样用于终端设备中,该终端设备例如可以是手机、PDA等,也可以是具有通话功能的其他设备。如图4所示,上述步骤S12包括如下步骤。

[0083] 在步骤S41中,采用至少3个子卷积层依次对待识别图像数据进行卷积处理,得到卷积后的图像数据。

[0084] 如前所述,预设网络模型的至少1类处理层包括卷积层和全连接层等,作为一种示例,一般情况下,卷积层包括的至少3个子卷积层均位于全连接层之前,那么,当将获取到的待识别图像数据对应的矩阵输入到该预设网络模型后,则首先采用该至少3个子卷积层依次对待识别图像数据进行卷积处理,从最后一个子卷积层输出卷积后的图像数据。

[0085] 在步骤S42中,采用至少3个子全连接层依次对卷积后的图像数据进行全连接处理,获取识别后的图像数据。

[0086] 当待识别图像数据经过卷积层中各子卷积层的卷积处理后,得到尺寸及像素点信息都在预想范围内的二维矩阵(也即,卷积后的图像数据),将该二维矩阵再输入到全连接层,经过全连接层的至少3个子全连接层全连接处理后,就可以得到识别后的图像数据,根据该识别后的图像数据,便可确定待识别图像所属的图像类别或领域等信息。

[0087] 本实施例提供的图像识别方法,当预设网络模型的处理层包括卷积层和全连接层,且卷积层包含至少3个子卷积层,全连接层包含至少3个子全连接层时,则采用该至少3个子卷积层和至少3个子全连接层依次对待识别图像数据进行卷积处理和全连接处理来获取识别后的图像数据,不仅能够识别出该待识别图像所属的领域,还能识别出该图像所属的类别,因此,大大提高了图像识别的效率和准确率。

[0088] 可选的,在上述各实施例的基础上,本公开又一实施例提供的图像识别方法中,该图像识别方法,还包括如下步骤。

[0089] 对每个子全连接层的总处理数据量进行压缩。

[0090] 如前所述,目前最流行的深度卷积网络由各类计算层叠加而成,因此,在实际应用中的预设网络模型均包括卷积层、激活层、池化层和全连接层等组成,使得整个预设网络模型参数众多,预设网络模型的大小通常在几百MB至1GB,其虽然能够很容易地部署在高性能服务器上,却难以部署到内存资源紧张的嵌入式设备中。

[0091] 作为一种示例,图5为Alex Net深度卷积模型网络的一种框图。在Alex Net深度卷积模型网络中,第一类处理层21为卷积层,且该卷积层包括5个子卷积层(C1、C2、C3、C4、C5),第二类处理层22为全连接层,且全连接层包括3个子全连接层(FC6、FC7、FC8),在Alex Net深度卷积模型网络中,全连接层的大小在整个Alex Net深度卷积模型网络中的比例极大。如图5所示,首先分别对三个子全连接层(FC6、FC7、FC8)的大小分析如下:

[0092] 从图5可以看出,FC6层的输入是一个大小为 $256*13*13$ 的张量,输出是一个大小为4096的向量,则FC6层处理的参数总数等于FC6层的输入乘以FC6层的输出,也即, $256*13*13*4096=177209344$ 。若以float类型存储这些参数,那么FC6层的大小为 $177209344*4/1024/1024=676\text{MB}$ 。

[0093] 类似的,FC7层的输入是一个大小为4096的向量,输出是一个大小为4096的向量,FC7层处理的参数总数为 $4096*4096=16777216$,FC7层的大小为 $16777216*4/1024/1024=64\text{MB}$ 。

[0094] FC8层的输入是一个大小为4096的向量,输出是一个大小为1000的向量,FC8层处理的参数总数为 $4096*1000=4096000$,FC8层的大小为 $4096000*4/1024/1024=15\text{MB}$ 。

[0095] 从上述分析可知,FC6、FC7、FC8的总大小达到了755MB。

[0096] 进一步的,如图5所示,当C1层是一个大小为 $96*55*55$ 的张量,C2层是一个大小为 $256*27*27$ 的张量,C3层是一个大小为 $384*13*13$ 的张量,C4层是一个大小为 $384*13*13$ 的张量,C5层是一个大小为 $256*13*13$ 的张量时,第一类处理层21为卷积层,按照计算FC6、FC7、FC8的总大小的方法依次对卷积层的5个子卷积层(C1、C2、C3、C4、C5)进行计算,经过计算得知该5个子卷积层(C1、C2、C3、C4、C5)的总大小仅仅为10MB左右,具体的计算方法此处不再赘述。所以,本公开实施例主要对全连接层的3个子全连接层(FC6、FC7、FC8)的总处理数据量进行压缩以降低整个Alex Net深度卷积模型网络的大小。

[0097] 值得说明的是,本公开实施例并不对Alex Net深度卷积模型网络中每一子卷积层和每一子连接层的大小进行限定,其可根据实际需要进行设置,而且不同的预设网络模型中处理层的种类、每类处理层包括子处理层的多少以及每个子处理层的大小在不同的实施例中也可能互不相同,本公开实施例均不对其进行限定。

[0098] 可选的,上述对每个子全连接层的总处理数据量进行压缩的一种可能实现方式包括图6所示实施例的步骤。

[0099] 图6是根据又一示例性实施例示出的一种图像识别方法的流程图。本公开实施例是在上述实施例的基础上对图像识别方法技术方案的进一步说明。本实施例涉及的图像识别方法同样用于终端设备中,该终端设备例如可以是手机、PDA等,也可以是具有通话功能的其他设备。

[0100] 可选的,本公开实施例是在图4所示实施例的基础上,将对每个子全连接层的总处理数据量进行压缩应用在步骤S41之后进行举例说明。如图6所示,对每个子全连接层的总处理数据量进行压缩包括如下步骤:

[0101] 在步骤S61中,分别对每个子全连接层的参数矩阵进行奇异值分解,获取每个子全连接层的参数矩阵对应的第一参数子矩阵以及第二参数子矩阵;

[0102] 可选的,预设网络模型中每个子全连接层的参数矩阵可以用一个 $u \times v$ 的矩阵 W 表示, W 经过奇异值分解可以近似的表示为 $W=U\Sigma_m V^T$,其中, U 是一个 $u \times m$ 的矩阵, Σ_m 是一个 $m \times m$ 的对角矩阵, V 是一个 $v \times m$ 的矩阵。也即,在本实施例中,子全连接层的参数矩阵 W 经过奇异值分解后可以得到其对应的第一参数子矩阵 U 和第二参数子矩阵 V 。

[0103] 则上述步骤S42(采用至少3个子全连接层依次对卷积后的图像数据进行全连接处理,获取识别后的图像数据)的一种可能替代方案如步骤S62所示。

[0104] 在步骤S62中,采用至少3个子全连接层中每个子全连接层的参数矩阵对应的第一参数子矩阵以及第二参数子矩阵依次对卷积后的图像数据对应的二维矩阵进行全连接处理,获取识别后的图像数据。

[0105] 在实际应用中,基于奇异值分解的加速原理来降低预设网络模型的总大小,首先将参数矩阵为 W 的子全连接层,替换为两个参数子矩阵,第一参数子矩阵的参数设置为 Σ

mV^T , 第二参数子矩阵的参数设置为U, 进而采用该第一参数子矩阵 $\Sigma_m V^T$ 和第二参数子矩阵U依次对经过卷积层卷积处理后的图像数据对应的二维矩阵进行全连接处理, 获取到识别后的图像数据。

[0106] 在本公开实施例中的待识别图像数据量较大时, 经过上述奇异值分解则可将子全连接层替换为两个参数子矩阵, 进而使子全连接层参数的个数从 uv 降低到了 $m(u+v)$, 从而极大地降低了运算次数, 极大地加速了图像识别效率。

[0107] 具体的, 对于如图5所示的Alex Net深度卷积模型网络, 子全连接层FC6的总处理参数用 $u \times v$ 的矩阵W表示, 其中, $u=256 \times 13 \times 13=43264$, $v=4096$ 。若W用奇异值SVD分解的方法近似的表示为 $W=U \Sigma_m V^T$, 当对角矩阵 Σ_m 为 $m \times m=1024$ 的矩阵时, 子全连接层FC6经过压缩处理后的大小由原来的676MB降为了185MB (也即, $m(u+v)=1024 \times (43264+4096) \times 4/1024/1024=185\text{MB}$), 而当对角矩阵 Σ_m 为 $m \times m=512$ 的矩阵时, 子全连接层FC6经过压缩处理后的大小由原来的676MB降为了92.5MB。由此可知, m 值越小, 压缩比例越大。

[0108] 类似的, 当对角矩阵 Σ_m 为 $m \times m=512$ 的矩阵时, 子全连接层FC7经过奇异值分解SVD分解压缩处理后的大小由原来的64MB降为了16MB (也即, $m(u+v)=512 \times (4096+4096) \times 4/1024/1024=16\text{MB}$)。当对角矩阵 Σ_m 为 $m \times m=128$ 的矩阵时, 子全连接层FC8经过奇异值分解SVD分解压缩处理后的大小由原来的15MB降为了2.5MB (也即, $m(u+v)=128 \times (4096+1000) \times 4/1024/1024=2.5\text{MB}$)。

[0109] 本实施例中的Alex Net深度卷积模型网络经过上述压缩处理 (假设对FC6和FC7进行 $m=512$ 的SVD分解, 对FC8进行 $m=128$ 的SVD分解), Alex Net深度卷积模型网络由原来的总大小765MB降为了111M, 压缩效果非常明显。

[0110] 值得说明的是, 本公开实施例还可对预设网络模型的卷积层进行压缩处理, 其并不对预设网络模型中处理层是否进行压缩处理进行限定, 其需要根据实际情况进行选择。例如, 对于上述Alex Net深度卷积模型网络中卷积层包括的子卷积层 (C1、C2、C3、C4、C5), 由于卷积层的总大小仅仅为10MB左右, 所以, 在使用预设网络模型对图像进行处理时, 并不需要对卷积层进行压缩处理。

[0111] 综上所述, 本公开实施例提供的图像识别方法, 通过对预设网络模型的每个子全连接层的总处理数据量进行压缩, 也即, 通过对每个子全连接层的参数矩阵进行奇异值分解, 获取每个子全连接层的参数矩阵对应的第一参数子矩阵以及第二参数子矩阵, 并采用至少3个子全连接层中每个子全连接层的参数矩阵对应的第一参数子矩阵以及第二参数子矩阵依次对卷积后的图像数据对应的二维矩阵进行全连接处理, 获取识别后的图像数据, 能够极大的缩减预设网络模型的参数个数, 极大的降低预设网络模型的尺寸, 为将预设网络模型部署到内存资源紧张的设备上提供了可能。

[0112] 以上描述了图像识别方法的实现过程, 该过程可以由图像识别装置来实现, 以下将对图像识别装置的内部功能和结构进行说明。对于本公开装置实施例中未披露的细节, 请参照本公开方法实施例。

[0113] 图7是根据一示例性实施例示出的一种图像识别装置的框图, 该图像识别装置可以通过软件、硬件或者两者的结合实现成为包含图像识别功能的电子设备的部分或者全部。该图像识别装置包括: 获取模块71和处理模块72。

[0114] 该获取模块71, 被配置为获取待识别图像数据, 并将该待识别图像数据输入到预

设网络模型。

[0115] 其中,该预设网络模型包括至少1类处理层,每类处理层包含至少3个子处理层,所述至少3个子处理层按序依次排列。

[0116] 该处理模块72,被配置为依次采用上述至少1类处理层对获取模块71获取到的待识别图像数据进行处理,获取识别后的图像数据。

[0117] 其中,每类处理层中的至少子3个子处理层依次对待识别图像数据进行处理,且每类处理层中从第3个子处理层开始,每个子处理层的输入包括:至少2个在先子处理层的输出。

[0118] 关于本公开实施例提供的图像识别装置,其中各个模块执行操作的具体方式已经在图1所示图像识别方法实施例中进行了详细描述,此处将不做详细阐述说明。

[0119] 可选的,在本公开的另一实施例中,上述处理模块72,还被配置为根据随机梯度下降法,依次采用预设网络模型的至少1类处理层对获取模块71获取到的待识别图像数据进行处理,获取识别后的图像数据。

[0120] 可选的,在上述任一实施例中,上述至少1类处理层包括:卷积层和全连接层。

[0121] 其中,卷积层包含至少3个子卷积层,全连接层包含至少3个子全连接层。

[0122] 图8是根据另一示例性实施例示出的一种图像识别装置的框图。本公开实施例是在上述实施例的基础上对图像识别装置的进一步说明。如图8所示,在本实施例提供的图像识别装置中,上述处理模块72,包括:第一处理子模块81和第二处理子模块82。

[0123] 该第一处理子模块81,被配置为采用至少3个子卷积层依次对获取模块71获取到的待识别图像数据进行卷积处理,得到卷积后的图像数据。

[0124] 该第二处理子模块82,被配置为采用至少3个子全连接层依次对第一处理子模块81得到的卷积后的图像数据进行全连接处理,获取识别后的图像数据。

[0125] 关于本公开实施例提供的图像识别装置,其中各个模块执行操作的具体方式已经在图3所示图像识别方法实施例中进行了详细描述,此处将不做详细阐述说明。

[0126] 图9是根据再一示例性实施例示出的一种图像识别装置的框图,在图9所示实施例的基础上,该装置还包括:压缩模块91。

[0127] 可选的,该压缩模块91,被配置为对每个子全连接层的总处理数据量进行压缩。

[0128] 作为一种示例,该压缩模块91在应用在第一处理子模块81和第二处理子模块82之间。

[0129] 可选的,如图9所示,该压缩模块91,包括:分解子模块911。

[0130] 该分解子模块911,被配置为分别对每个子全连接层的参数矩阵进行奇异值分解,获取每个子全连接层的参数矩阵对应的第一参数子矩阵以及第二参数子矩阵。

[0131] 相应的,上述第二处理子模块82,还被配置为采用至少3个子全连接层中分解子模块911获取到的每个子全连接层的参数矩阵对应的第一参数子矩阵以及第二参数子矩阵依次对卷积后的图像数据对应的二维矩阵进行全连接处理,获取识别后的图像数据。

[0132] 关于本公开实施例提供的图像识别装置,其中各个模块执行操作的具体方式已经在图6所示图像识别方法实施例中进行了详细描述,此处将不做详细阐述说明。

[0133] 以上描述了图像识别装置的内部功能模块和结构示意图,图10是根据一示例性实施例示出的一种图像识别装置的实体的框图,如图10所示,该图像识别装置包括:存储器101

和处理器102。

[0134] 存储器101,用于存储处理器可执行指令。

[0135] 处理器102,用于:

[0136] 获取待识别图像数据,并将该待识别图像数据输入到预设网络模型,该预设网络模型包括至少1类处理层,每类处理层包含至少3个子处理层,该至少3个子处理层按序依次排列;

[0137] 依次采用上述至少1类处理层对所述待识别图像数据进行处理,获取识别后的图像数据;

[0138] 其中,每类处理层中的至少子3个子处理层依次对待识别图像数据进行处理,且每类处理层中从第3个子处理层开始,每个子处理层的输入包括:至少2个在先子处理层的输出。

[0139] 值得说明的是,在图10所示的图像识别装置的实施例中,应理解,该处理器可以是中央处理单元(英文:Central Processing Unit,简称:CPU),还可以是其他通用处理器、数字信号处理器(英文:Digital Signal Processor,简称:DSP)、专用集成电路(英文:Application Specific Integrated Circuit,简称:ASIC)等。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,而前述的存储器可以是只读存储器(英文:read-only memory,缩写:ROM)、随机存取存储器(英文:random access memory,简称:RAM)、快闪存储器、硬盘或者固态硬盘。SIM卡也称为用户身份识别卡、智能卡,数字移动电话机必须装上此卡方能使用。即在电脑芯片上存储了数字移动电话客户的信息,加密的密钥以及用户的电话簿等内容。结合本发明实施例所公开的方法的步骤可以直接体现为硬件处理器执行完成,或者用处理器中的硬件及软件模块组合执行完成。

[0140] 图11是根据一示例性实施例示出的一种图像识别装置的框图。例如,图像识别装置1000可以是移动电话,计算机,数字广播终端,消息收发设备,游戏控制台,平板设备,医疗设备,健身设备,个人数字助理等。

[0141] 参照图11,图像识别装置1000可以包括以下一个或多个组件:处理组件1002,存储器1004,电源组件1006,多媒体组件1008,音频组件1010,输入/输出(I/O)接口1012,传感器组件1014,以及通信组件1016。

[0142] 处理组件1002通常控制图像识别装置1000的整体操作,诸如与显示,电话呼叫,数据通信,相机操作和记录操作相关联的操作。处理组件1002可以包括一个或多个处理器1020来执行指令,以完成上述的方法的全部或部分步骤。此外,处理组件1002可以包括一个或多个模块,便于处理组件1002和其他组件之间的交互。例如,处理组件1002可以包括多媒体模块,以方便多媒体组件1008和处理组件1002之间的交互。

[0143] 存储器1004被配置为存储各种类型的数据以支持在图像识别装置1000的操作。这些数据的示例包括用于在图像识别装置1000上操作的任何应用程序或方法的指令,联系人数据,电话簿数据,消息,图片,视频等。存储器1004可以由任何类型的易失性或非易失性存储设备或者它们的组合实现,如静态随机存取存储器(SRAM),电可擦除可编程只读存储器(EEPROM),可擦除可编程只读存储器(EPROM),可编程只读存储器(PROM),只读存储器(ROM),磁存储器,快闪存储器,磁盘或光盘。

[0144] 电源组件1006为图像识别装置1000的各种组件提供电力。电源组件1006可以包括

电源管理系统,一个或多个电源,及其他与为图像识别装置1000生成、管理和分配电力相关联的组件。

[0145] 多媒体组件1008包括在图像识别装置1000和用户之间的提供一个输出接口的屏幕。在一些实施例中,屏幕可以包括液晶显示器(LCD)和触摸面板(TP)。如果屏幕包括触摸面板,屏幕可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中,多媒体组件1008包括一个前置摄像头和/或后置摄像头。当图像识别装置1000处于操作模式,如拍摄模式或视频模式时,前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[0146] 音频组件1010被配置为输出和/或输入音频信号。例如,音频组件1010包括一个麦克风(MIC),当图像识别装置1000处于操作模式,如呼叫模式、记录模式和语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器1004或经由通信组件1016发送。在一些实施例中,音频组件1010还包括一个扬声器,用于输出音频信号。

[0147] I/O接口1012为处理组件1002和外围接口模块之间提供接口,上述外围接口模块可以是键盘,点击轮,按钮等。这些按钮可包括但不限于:主页按钮、音量按钮、启动按钮和锁定按钮。

[0148] 传感器组件1014包括一个或多个传感器,用于为图像识别装置1000提供各个方面的状态评估。例如,传感器组件1014可以检测到图像识别装置1000的打开/关闭状态,组件的相对定位,例如所述组件为图像识别装置1000的显示器和小键盘,传感器组件1014还可以检测图像识别装置或图像识别装置1000一个组件的位置改变,用户与图像识别装置1000接触的存在或不存在,图像识别装置1000方位或加速/减速和图像识别装置1000的温度变化。传感器组件1014可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件1014还可以包括光传感器,如CMOS或CCD图像传感器,用于在成像应用中使用。在一些实施例中,该传感器组件1014还可以包括加速度传感器,陀螺仪传感器,磁传感器,压力传感器或温度传感器。

[0149] 通信组件1016被配置为便于图像识别装置1000和其他设备之间有线或无线方式的通信。图像识别装置1000可以接入基于通信标准的无线网络,如WiFi,2G或3G,或它们的组合。在一个示例性实施例中,通信组件1016经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中,所述通信组件1016还包括近场通信(NFC)模块,以促进短程通信。例如,在NFC模块可基于射频识别(RFID)技术,红外数据协会(IrDA)技术,超宽带(UWB)技术,蓝牙(BT)技术和其他技术来实现。

[0150] 在示例性实施例中,图像识别装置1000可以被一个或多个应用专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑器件(PLD)、现场可编程门阵列(FPGA)、控制器、微控制器、微处理器或其他电子元件实现,用于执行上述方法。

[0151] 在示例性实施例中,还提供了一种包括指令的非临时性计算机可读存储介质,例如包括指令的存储器1004,上述指令可由图像识别装置1000的处理器1020执行以完成上述

方法。例如,所述非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0152] 一种非临时性计算机可读存储介质,当所述存储介质中的指令由图像识别装置1000的处理器执行时,使得图像识别装置1000能够执行上述各个实施例提供的图像识别方法。

[0153] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本公开的其它实施方案。本申请旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的真正范围和精神由下面的权利要求书指出。

[0154] 应当理解的是,本公开并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本公开的范围仅由所附的权利要求书来限制。

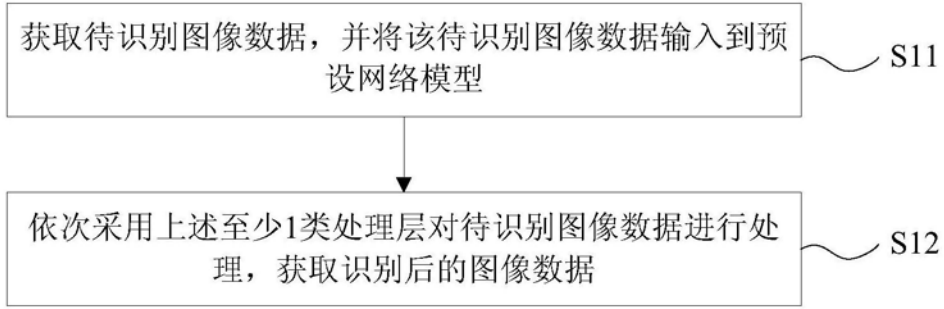


图1

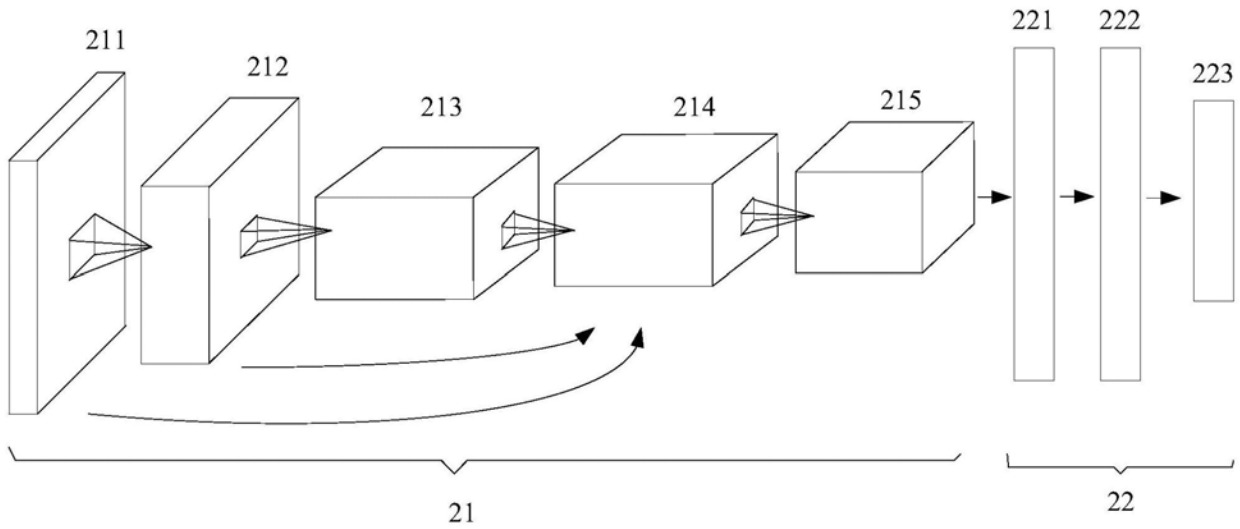


图2

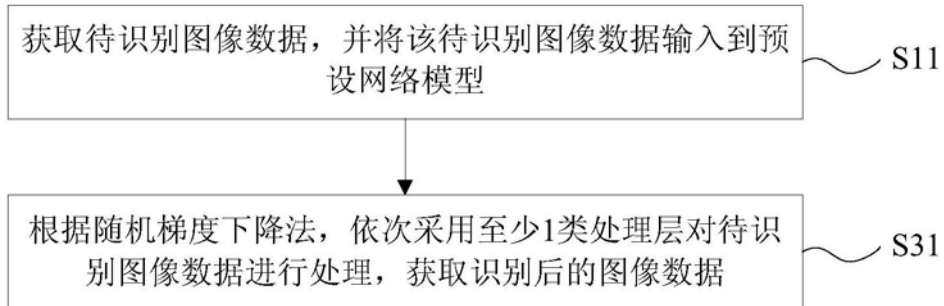


图3

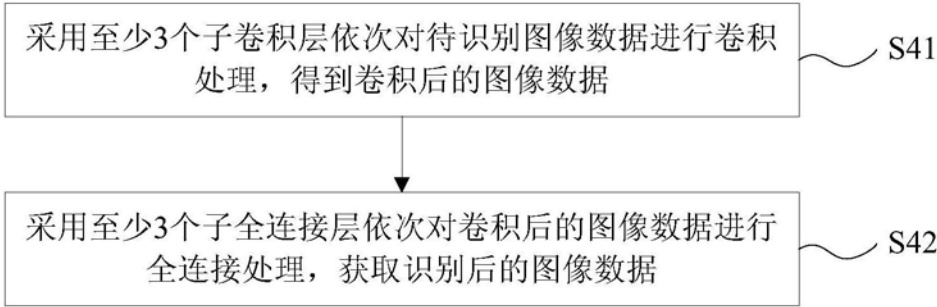


图4

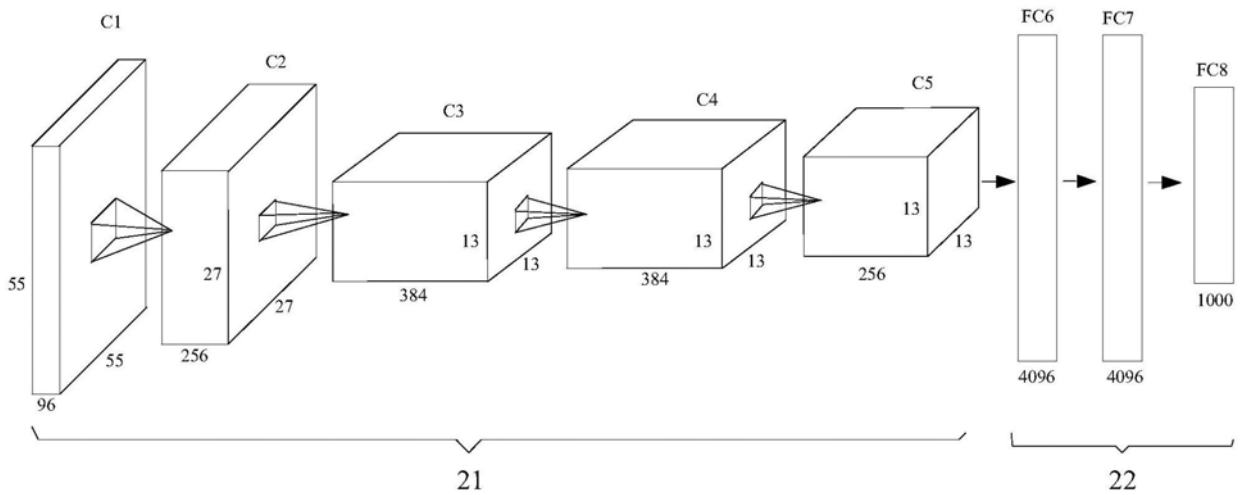


图5

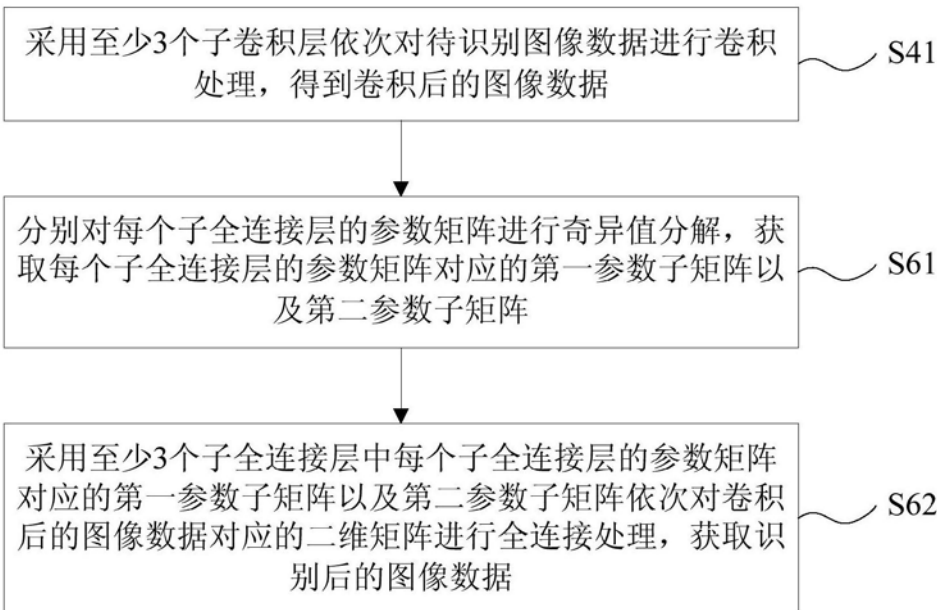


图6

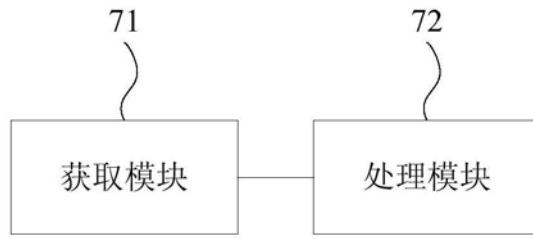


图7

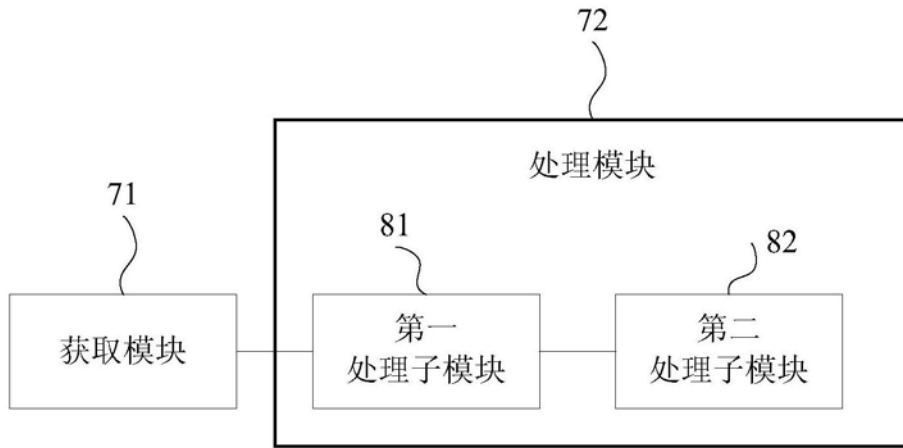


图8

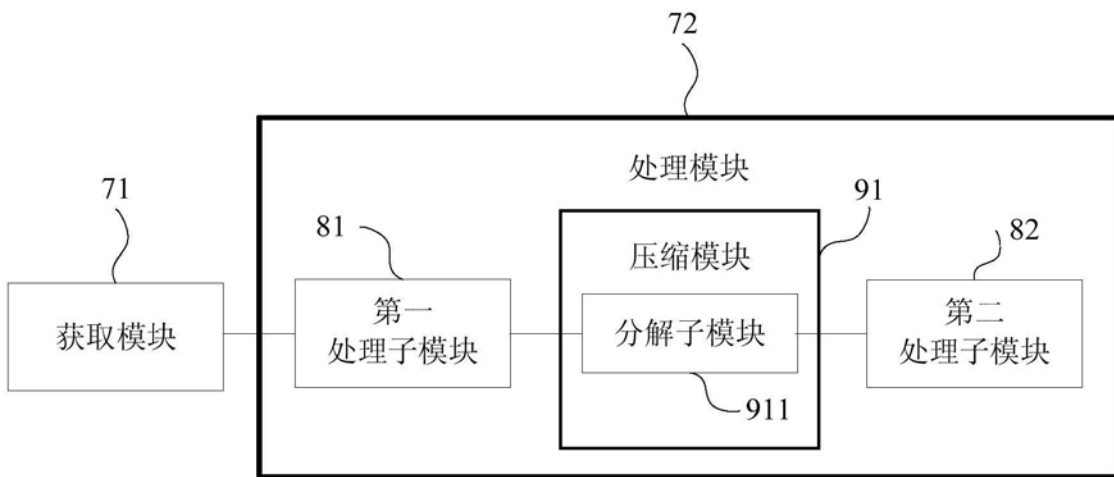


图9

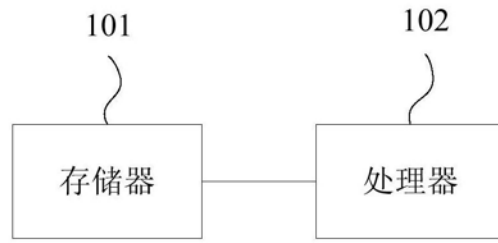


图10

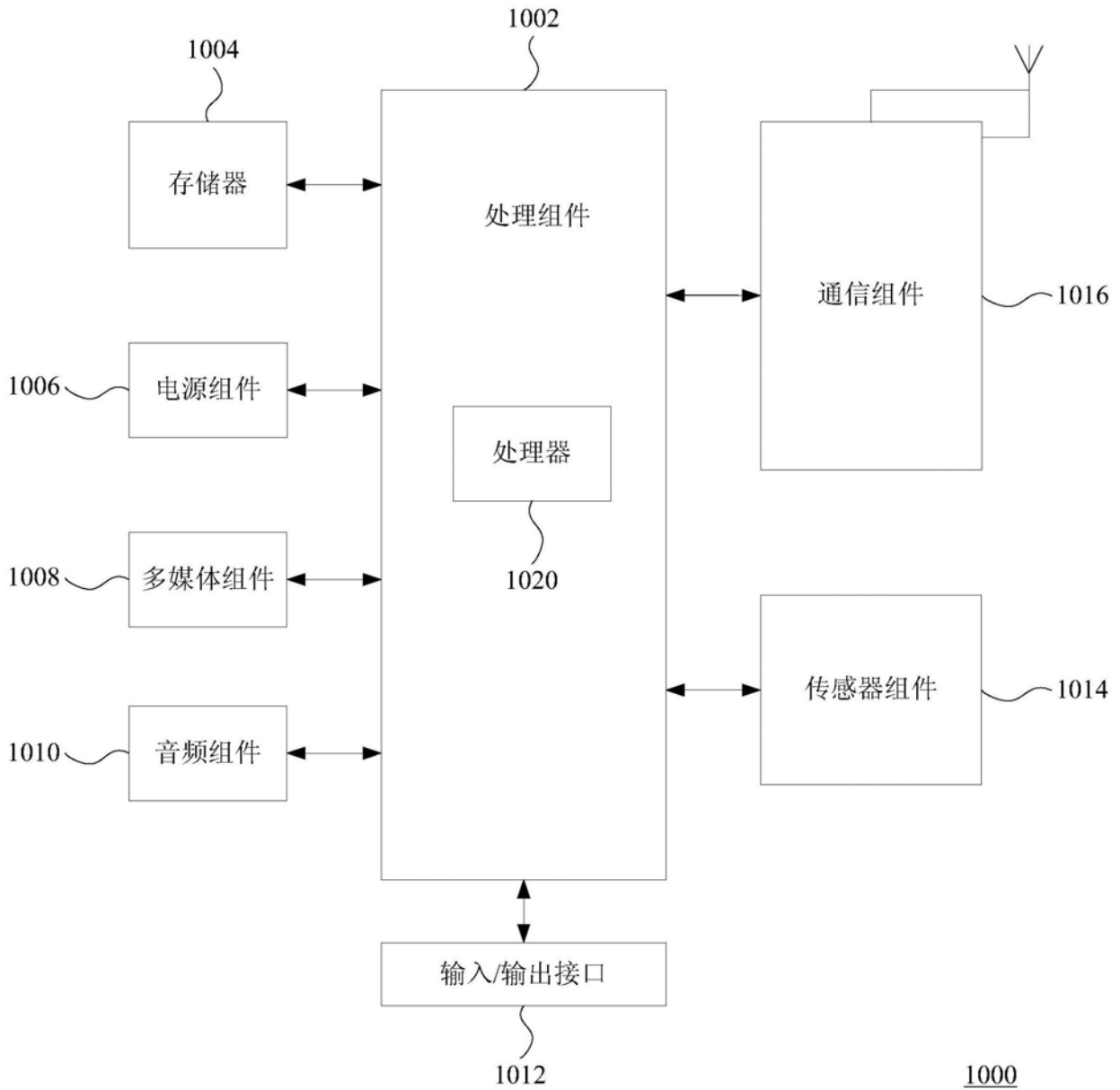


图11