(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 108960053 A
(43)申请公布日 2018.12.07

(21)申请号 201810525499.6

(22)申请日 2018.05.28

(71)申请人 北京陌上花科技有限公司
地址 100080 北京市海淀区丹棱街6号中关村金融大厦SOHO3Q

(72)发明人 张默

(74)专利代理机构 北京卓唐知识产权代理有限公司 11541
代理人 唐海力 李志刚

(51)Int.Cl.
G06K 9/00(2006.01)

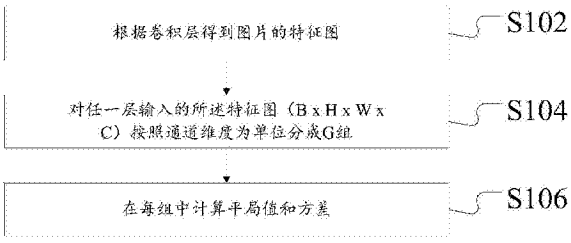
权利要求书2页 说明书10页 附图3页

(54)发明名称

归一化处理方法及装置、客户端

(57)摘要

本申请公开了一种归一化处理方法及装置、客户端。该归一化处理方法包括根据卷积层得到图片的特征图;对任一层输入的所述特征图(B x H x W x C)按照通道维度为单位分成G组;在每组中计算平局值和方差;本申请解决了加快训练速度和准确率无法达到均衡的技术问题。本申请达到了减少了训练网络时所需要批量大小的目的同时,能够确保了准确率相比大批量的同样网络不会下降。本申请的方法能够很好地解决模型训练时需要大批量和预测时运行速度慢的问题,同时保证了准确率。



CN 108960053 A

1.一种归一化处理方法,其特征在于,用于卷积神经网络,所述方法包括:

根据卷积层得到图片的特征图;

对任一层输入的所述特征图(B x H x W x C)按照通道维度为单位分成G组;

在每组中计算平局值和方差;

其中,B为图片的数目,C为通道数目,H为特征图的长,W为特征图的宽,G为组的个数。

2.根据权利要求1所述的归一化处理方法,其特征在于,在每组中计算平局值和方差包括:根据图片的数目将每个图片的数目内的均值和方差相加并取均值作为最后的均值和方差。

3.根据权利要求1所述的归一化处理方法,其特征在于,在每组中计算平局值和方差包括:

结合前后图片的数目的信息将多次迭代的信息结合;

对于第N-1次迭代的结果,通过加权平均法将迭代结果结合并计算最终的平均值。

4.根据权利要求1所述的归一化处理方法,其特征在于,在每组中计算平局值和方差包括:将组归一化处理操作和批量归一化处理操作结合。

5.根据权利要求1所述的归一化处理方法,其特征在于,根据卷积层得到图片的特征图包括:

在输入层输入待识别图片;

建立多个分组卷积模块;

通过多个所述分组卷积模块输出所述待识别图片的特征图;以及

根据所述特征图在输出层输出图像识别结果;

其中,所述分组卷积模块中至少包括:一深度可分离卷积单元和一预设卷积核大小的卷积单元。

6.一种归一化处理装置,其特征在于,包括:

特征图输入模块,用于根据卷积层得到图片的特征图;

分组模块,用于对任一层输入的所述特征图(B x H x W x C)按照通道维度为单位分成G组;

计算模块,用于在每组中计算平局值和方差;

其中,B为图片的数目,C为通道数目,H为特征图的长,W为特征图的宽,G为组的个数。

7.根据权利要求6所述的归一化处理装置,其特征在于,计算模块包括:第一计算单元,

所述第一计算单元,用于根据图片的数目将每个图片的数目内的均值和方差相加并取均值作为最后的均值和方差。

8.根据权利要求6所述的归一化处理装置,其特征在于,计算模块包括:第二计算单元,

所述第二计算单元,用于结合前后图片的数目的信息将多次迭代的信息结合;

以及对于第N-1次迭代的结果,通过加权平均法将迭代结果结合并计算最终的平均值。

9.根据权利要求6所述的归一化处理装置,其特征在于,所述特征图输入模块包括:

输入单元,用于在输入层输入待识别图片;

建立单元,用于建立多个分组卷积模块;

第一输出单元,用于通过多个所述分组卷积模块输出所述待识别图片的特征图;以及

第二输出单元,用于根据所述特征图在输出层输出图像识别结果;

其中，所述分组卷积模块中至少包括：一深度可分离卷积单元和一预设卷积核大小的卷积单元。

10.一种客户端，其特征在于，包括如权利要求6至9任一项所述的归一化处理装置。

归一化处理方法及装置、客户端

技术领域

[0001] 本申请涉及计算机视觉领域，具体而言，涉及一种归一化处理方法及装置、客户端。

背景技术

[0002] 随着计算机视觉的快速发展，人脸识别，物体检测等领域已经有了很大的进展，尤其在准确率上有了很大的提升，很多深层次网络的出现更是加快了人脸识别，物体检测等领域的进展，在很多计算机视觉的公开数据集上，都有了很大的飞跃。

[0003] 比如，人脸识别LFW数据集，准确率已经达到99.83％，远超人眼准确度，如Pascal VOC数据集，物体检测也将近90％的准确率，再如COCO数据集，物体检测达到50％以上的准确率，由此可见很多准确率高的方法都是基于很大的网络。然而上述方法中也有着运行速度慢的弊端，同时训练时间长。

[0004] 发明人发现，加快训练速度和准确率无法达到均衡。进一步地，无法在移动端，服务器端等多种平台上流畅运行。

[0005] 针对相关技术中加快训练速度和准确率无法达到均衡的问题，目前尚未提出有效的解决方案。

发明内容

[0006] 本申请的主要目的在于提供一种归一化处理方法，以解决加快训练速度和准确率无法达到均衡的问题。

[0007] 为了实现上述目的，根据本申请的一个方面，提供了一种归一化处理方法。

[0008] 根据本申请的归一化处理方法包括：

[0009] 根据卷积层得到图片的特征图；对任一层输入的所述特征图(B x H x W xC)按照通道维度为单位分成G组；在每组中计算平局值和方差；其中，B为图片的数目，C为通道数目，H为特征图的长，W为特征图的宽，G为组的个数。

[0010] 进一步地，在每组中计算平局值和方差包括：根据图片的数目将每个图片的数目内的均值和方差相加并取均值作为最后的均值和方差。

[0011] 进一步地，在每组中计算平局值和方差包括：结合前后图片的数目的信息将多次迭代的信息结合；对于第N-1次迭代的结果，通过加权平均法将迭代结果结合并计算最终的平均值。

[0012] 进一步地，在每组中计算平局值和方差包括：将组归一化处理操作和批量归一化处理操作结合。

[0013] 进一步地，根据卷积层得到图片的特征图包括：在输入层输入待识别图片；建立多个分组卷积模块；通过多个所述分组卷积模块输出所述待识别图片的特征图；以及根据所述特征图在输出层输出图像识别结果；其中，所述分组卷积模块中至少包括：一深度可分离卷积单元和一预设卷积核大小的卷积单元。

[0014] 为了实现上述目的，根据本申请的另一方面，提供了一种归一化处理装置。

[0015] 根据本申请的归一化处理装置包括：特征图输入模块，用于根据卷积层得到图片的特征图；分组模块，用于对任一层输入的所述特征图(B x H x W xC)按照通道维度为单位分成G组；计算模块，用于在每组中计算平局值和方差；其中，B为图片的数目，C为通道数目，H为特征图的长，W为特征图的宽，G为组的个数。

[0016] 进一步地，计算模块包括：第一计算单元，所述第一计算单元，用于根据图片的数目将每个图片的数目内的均值和方差相加并取均值作为最后的均值和方差。

[0017] 进一步地，计算模块包括：第二计算单元，所述第二计算单元，用于结合前后图片的数目的信息将多次迭代的信息结合；以及对于第N-1次迭代的结果，通过加权平均法将迭代结果结合并计算最终的平均值。

[0018] 进一步地，所述特征图输入模块包括：输入单元，用于在输入层输入待识别图片；建立单元，用于建立多个分组卷积模块；第一输出单元，用于通过多个所述分组卷积模块输出所述待识别图片的特征图；以及第二输出单元，用于根据所述特征图在输出层输出图像识别结果；其中，所述分组卷积模块中至少包括：一深度可分离卷积单元和一预设卷积核大小的卷积单元。

[0019] 在本申请实施例中，采用对任一层输入的所述特征图(B x H x W x C)按照通道维度为单位分成G组的方式，通过在每组中计算平局值和方差，达到了减少了训练网络时所需要批量大小的目的，从而实现了准确率相比大批量的同样网络不会下降的技术效果，进而解决了加快训练速度和准确率无法达到均衡的技术问题。

附图说明

[0020] 构成本申请的一部分的附图用来提供对本申请的进一步理解，使得本申请的其它特征、目的和优点变得更明显。本申请的示意性实施例附图及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：

[0021] 图1是根据本申请第一实施例的归一化处理方法示意图；

[0022] 图2是根据本申请第二实施例的归一化处理方法示意图；

[0023] 图3是根据本申请第三实施例的归一化处理方法示意图；

[0024] 图4是根据本申请第一实施例的归一化处理装置示意图；

[0025] 图5是根据本申请第二实施例的归一化处理装置示意图；以及

[0026] 图6是根据本申请第三实施例的归一化处理装置示意图。

具体实施方式

[0027] 为了使本技术领域的人员更好地理解本申请方案，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清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分的实施例，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，都应当属于本申请保护的范围。

[0028] 需要说明的是，本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用

的数据在适当情况下可以互换，以便这里描述的本申请的实施例。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0029] 需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

[0030] 如图1所示，该方法包括如下的步骤S102至步骤S106：

[0031] 步骤S102，根据卷积层得到图片的特征图；

[0032] 假设输入特征图大小为Sf x Sf x IN，经过一次卷积，输出的特征图大小为Sf x Sf x OUT，

[0033] 按照传统的卷积操作，卷积核K的大小为：Sk x Sk x IN x OUT，

[0034] 其中，Sf是特征图的尺寸，Sk是卷积核的尺寸，IN是输入特征图的通道数，OUT是输出特征图的通道数，一次卷积操作的过程如下：

[0035] IN个Sk x Sk个卷积核与IN个输入特征图做卷积，得到的结果相加，得到一张输出特征图，同理，一共OUT次操作，得到OUT个输出特征图，用公式表示这个过程如下：

[0036] $$G_{k,l,n}=\sum_{i,j,m}K_{i,j,m,n}\cdot F_{k+i-1,l+j-1,m}$$

[0037] 根据公式，可以计算出传统卷积层的计算量为：

[0038] Sk x Sk x IN x OUT x Sf x Sf

[0039] 参数量为：

[0040] Sk x Sk x IN x OUT。

[0041] 步骤S104，对任一层输入的所述特征图(B x H x W x C)按照通道维度为单位分成G组；

[0042] 其中，B为图片的数目，C为通道数目，H为特征图的长，W为特征图的宽，G为组的个数。

[0043] 步骤S106，在每组中计算平局值和方差；

[0044] 结合了组归一化(Group Normalization)和批归一化(Batch Normalization)的优点，针对第n次迭代，任一层输入的特征图(B x H x W x C)，B代表Batch Size，指的是图片的数目；G代表组的个数，C代表通道数目，H，W代表特征图的尺寸(长宽)。

[0045] 在本申请中以通道维度为单位，分成G组。

[0046] 对于G组中的每组计算平局值和方差。

[0047] $$\mu_i=\frac{1}{m}\sum_{k\in S_i}x_k$$

公式一

[0048] $$\sigma_i=\sqrt{\frac{1}{m}\sum_{k\in S_i}\left(x_k-\mu_i\right)^2+e}$$

公式二

[0049] $$S_i=\left\{K\left\|\left\lfloor\frac{K_C}{C/G}\right\rfloor=\left\lfloor\frac{i_C}{C/G}\right\rfloor\right\}\right.$$ 公式三

[0050] 其中，公式一中μ_i是计算的平均值，i的取值范围是[0,C/G]，公式二中σ_i是计算的方差，S_i是用来计算均值和方差的像素集合，m指的是集合S_i的大小，e是一个很小的常量(本申请实施例中取值为1e-6)，i_C指的是通道方向的下标，k_C指的是像素的下标，所以公式三是指所有在一个组内的像素的下标的集合。

[0051] 具体地，针对第n次迭代，任一层输入的特征图(B x H x W x C)，在本申请中以通道维度为单位，分成G组，每组计算平局值和方差。优选地，作为本实施例中的优选，结合批量的信息，将每个批量内的均值和方差相加并取均值作为最后的均值和方差，对于批量大小Batch Size为1的情况，则本申请中退化成组归一化。

[0052] 优选地，作为本实施例中的优选，本申请中结合前后批量的信息，将多次迭代的信息结合在一起，对于第n-1次迭代的结果，本申请通过加权平均法将其结合，权重参数为w_{n-1}和w_n，所以最终的平均值mean等于w_{n-1}x $mean_{n-1}$+w_n x $mean_n$，其中$w_{n-1}+w_n=1$，通常设置w_{n-1}为0.99，w_n为0.01。

[0053] 在本申请中提出的组批归一化方法，结合了组归一化和批归一化的优点，一方面减少了训练网络时所需要的批量大小Batch Size的大小，另一方面，结合了批量维度和时间维度的信息，保证了在小批量的情况下，准确率相比大批量的同样网络不会下降。

[0054] 为了解决基于大网络如等网络训练速度慢的问题，本申请提出了全新的网络GBCNN，全称Group Batch Convolution Network，可以用于人脸识别，物体检测等深度学习领域，本申请一方面提出组批归一化Group Batch Normalization的方法(在每组中计算平局值和方差包括：将组归一化处理操作和批量归一化处理操作结合)，加速训练速度，另一方面使用分组卷积Group Convolution的策略，实现了网络的加速。

[0055] 通过具体实验，本网络可以用于多项基于深度学习的任务中，在批数量等于1的情况下，本申请提出的组批归一化实现了与批数量等于32情况下，Batch Normalization相近的准确度。同时，优选地在移动端，服务器端等多种平台都可以流畅运行，对应网络运行速度可以提高将近20倍。

[0056] 从以上的描述中，可以看出，本申请实现了如下技术效果：

[0057] 在本申请实施例中，采用对任一层输入的所述特征图(B x H x W x C)按照通道维度为单位分成G组的方式，通过在每组中计算平局值和方差，达到了减少了训练网络时所需要批量大小的目的，从而实现了准确率相比大批量的同样网络不会下降的技术效果，进而解决了加快训练速度和准确率无法达到均衡的技术问题。

[0058] 根据本申请实施例，作为本申请实施例的优选，如图2所示，在每组中计算平局值和方差包括：根据图片的数目将每个图片的数目内的均值和方差相加并取均值作为最后的均值和方差。优选地，作为本实施例中的优选，结合批量的信息，将每个批量内的均值和方差相加并取均值作为最后的均值和方差，对于批量大小Batch Size为1的情况，则本申请中退化成组归一化。

[0059] 和/或在每组中计算平局值和方差包括：

[0060] 步骤S202，结合前后图片的数目的信息将多次迭代的信息结合；

[0061] 步骤S204，对于第N-1次迭代的结果，通过加权平均法将迭代结果结合并计算最终的平均值。

[0062] 优选地，作为本实施例中的优选，本申请中结合前后批量的信息，将多次迭代的信息结合在一起，对于第N-1次迭代的结果，本申请通过加权平均法将其结合，权重参数为w_{n-1}和w_n，所以最终的平均值mean等于w_{n-1}x $mean_{n-1}$+w_n x $mean_n$，其中$w_{n-1}+w_n=1$，通常设置w_{n-1}为0.99，w_n为0.01。

[0063] 优选地，将在上述每组中计算平局值和方差包括：将组归一化处理操作和批量归一化处理操作结合。

[0064] 根据本申请实施例，作为本申请实施例的优选，如图3所示，根据卷积层得到图片的特征图包括：

[0065] 步骤S302，在输入层输入待识别图片；

[0066] 待识别图片可以用于人脸识别或者物体检测。

[0067] 特别地，在输入层输入待识别图片可以使用于无人车，安防等实时检测和识别。

[0068] 步骤S304，建立多个分组卷积模块；

[0069] 其中，所述分组卷积模块中至少包括：一深度可分离卷积单元和一预设卷积核大小的卷积单元。

[0070] 在本步骤中提供了分组卷积模块，每个分组卷积模块中至少包括：一深度可分离卷积单元和一预设卷积核大小的卷积单元。由于在一个标准的CNN网络中按顺序可包括：卷积层、批归一化层(Group Batch Normalization)，激活函数层(Sigmoid)，在本实施例中提出的分组卷积模块可以替换现有CNN网络中的卷积层。

[0071] 优选地，每个所述分组卷积模块中至少包括：一深度可分离卷积单元和一卷积核大小为1*1的卷积单元。可以理解，上述的深度可分离卷积单元为本领域技术人员公知的一种卷积单元结构，由于能够有效利用参数，因此深度可分离卷积单元也可以用于移动设备中。此外，采用卷积核大小为1*1的卷积单元，可减少模型参数。

[0072] 每个组模块指的是深度分离卷积单元+1*1的卷积单元，1*1的卷积单元接在深度分离单元之后，可建立通道间的相关性。

[0073] 步骤S306，通过多个所述分组卷积模块输出所述待识别图片的特征图；

[0074] 本申请中的分组卷积模块与传统的卷积层相比，具有如下的特点：

[0075] 假设输入特征图大小为Sf x Sf x IN，经过一次卷积，输出的特征图大小为Sf x Sf x OUT，

[0076] 按照传统的卷积操作，卷积核K的大小为：Sk x Sk x IN x OUT，

[0077] 其中，Sf是特征图的尺寸，Sk是卷积核的尺寸，IN是输入特征图的通道数，OUT是输出特征图的通道数，一次卷积操作的过程如下：

[0078] IN个Sk x Sk个卷积核与IN个输入特征图做卷积，得到的结果相加，得到一张输出特征图，同理，一共OUT次操作，得到OUT个输出特征图，用公式表示这个过程如下：

[0079] $$G_{k,l,n}=\sum_{i,j,m}K_{i,j,m,n}\cdot F_{k+i-1,l+j-1,m}$$

[0080] 根据公式，可以计算出传统卷积层的计算量为：

[0081] Sk x Sk x IN x OUT x Sf x Sf

[0082] 参数量为：

[0083] Sk x Sk x IN x OUT。

[0084] 对应一次传统卷积的是个分组卷积模块(即深度可分离卷积单元+卷积核大小为1*1的卷积单元)。其中，深度可分离卷积单元的具体实现如下：其卷积核K的大小为：Sk x Sk x IN，卷积核只跟对应通道的输入特征图做卷积，得到输出特征图，所以输出的特征图大小为Sf x Sf x IN，用公式表示这个过程如下：

[0085] $$\hat{G}_{k,l,m}=\sum_{i,j}\hat{K}_{i,j,m}\cdot F_{k+i-1,l+j-1,m}$$

[0086] 根据公式，可以计算出深度分离卷积的计算量为：

[0087] Sk x Sk x IN x Sf x Sf

[0088] 参数量为：

[0089] Sk x Sk x IN。

[0090] 深度分离卷积之后，进入一层传统的批归一化层和激活层，然后是1x1卷积层，1x1卷积层的卷积核大小为1x 1x IN x OUT，操作跟传统卷积一致，计算量为1x 1x IN x OUT x Sf x Sf，参数量为1x 1x IN x OUT。

[0091] 优选地，通过多个所述分组卷积模块输出所述待识别图片的特征图之后还依次通过：批归一化层和激活函数层。

[0092] 步骤S308，根据所述特征图在输出层输出图像识别结果；

[0093] 具体地，分组卷积模块总的计算量为：

[0094] Sk x Sk x IN x Sf x Sf+1x 1x IN x OUT x Sf x Sf＝(Sk x Sk+OUT)x IN x Sf x Sf，

[0095] 参数量：Sk x Sk x IN+1x 1x IN x OUT＝(Sk x Sk+OUT)x IN，

[0096] 相比传统卷积，

[0097] 计算量：(Sk x Sk+OUT)/Sk x Sk x OUT＝1/OUT+1/(Sk x Sk)，

[0098] 参数量：(Sk x Sk+OUT)/Sk x Sk x OUT＝1/OUT+1/(Sk x Sk)，

[0099] 可知，根据特征图在输出层输出图像识别结果时采用分组卷积模块，每个分组卷积模块中至少包括：一深度可分离卷积单元和一预设卷积核大小的卷积单元，可以将计算量和参数量减少。

[0100] 需要说明的是，在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行，并且，虽然在流程图中示出了逻辑顺序，但是在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤。

[0101] 根据本申请实施例，还提供了一种用于实施上述归一化处理方法的装置，如图4所示，该装置包括：特征图输入模块10，用于根据卷积层得到图片的特征图；分组模块20，用于对任一层输入的所述特征图(B x H x W x C)按照通道维度为单位分成G组；计算模块30，用于在每组中计算平局值和方差；其中，B为图片的数目，C为通道数目，H为特征图的长，W为特征图的宽，G为组的个数。

[0102] 在本申请实施例的特征图输入模块10中假设输入特征图大小为Sf x Sf x IN，经过一次卷积，输出的特征图大小为Sf x Sf x OUT，

[0103] 按照传统的卷积操作，卷积核K的大小为：Sk x Sk x IN x OUT，

[0104] 其中，Sf是特征图的尺寸，Sk是卷积核的尺寸，IN是输入特征图的通道数，OUT是输出特征图的通道数，一次卷积操作的过程如下：

[0105] IN个Sk x Sk个卷积核与IN个输入特征图做卷积，得到的结果相加，得到一张输出特征图，同理，一共OUT次操作，得到OUT个输出特征图，用公式表示这个过程如下：

[0106] $$G_{k,l,n}=\sum_{i,j,m}K_{i,j,m,n}\cdot F_{k+i-1,l+j-1,m}$$

[0107] 根据公式，可以计算出传统卷积层的计算量为：

[0108] Sk x Sk x IN x OUT x Sf x Sf

[0109] 参数量为：

[0110] Sk x Sk x IN x OUT。

[0111] 在本申请实施例的分组模块20中其中，B为图片的数目，C为通道数目，H为特征图的长，W为特征图的宽，G为组的个数。

[0112] 在本申请实施例的计算模块30中结合了组归一化(Group Normalization)和批归一化(Batch Normalization)的优点，针对第n次迭代，任一层输入的特征图(B x H x W x C)，B代表Batch SIze，指的是图片的数目；G代表组的个数，C代表通道数目，H，W代表特征图的尺寸(长宽)。

[0113] 在本申请中以通道维度为单位，分成G组。

[0114] 对于G组中的每组计算平局值和方差。

[0115] $$\mu_i=\frac{1}{m}\sum_{k\in S_i}x_k$$

公式一

[0116] $$\sigma_i=\sqrt{\frac{1}{m}\sum_{k\in S_i}\left(x_k-\mu_i\right)^2+e}$$

公式二

[0117] $$S_i=\left\{K\left\|\left\lfloor\frac{K_C}{C/G}\right\rfloor=\left\lfloor\frac{i_C}{C/G}\right\rfloor\right.\right\}$$

公式三

[0118] 其中，公式一中μ_i是计算的平均值，i的取值范围是[0,C/G]，公式二中σ_i是计算的方差，S_i是用来计算均值和方差的像素集合，m指的是集合S_i的大小，e是一个很小的常量(本申请实施例中取值为1e-6)，i_C指的是通道方向的下标，k_C指的是像素的下标，所以公式三是指所有在一个组内的像素的下标的集合。

[0119] 具体地，针对第n次迭代，任一层输入的特征图(B x H x W x C)，在本申请中以通道维度为单位，分成G组，每组计算平局值和方差。优选地，作为本实施例中的优选，结合批量的信息，将每个批量内的均值和方差相加并取均值作为最后的均值和方差，对于批量大小Batch Size为1的情况，则本申请中退化成组归一化。

[0120] 优选地，作为本实施例中的优选，本申请中结合前后批量的信息，将多次迭代的信息结合在一起，对于第n-1次迭代的结果，本申请通过加权平均法将其结合，权重参数为w_{n-1}和w_n，所以最终的平均值mean等于w_{n-1}x $mean_{n-1}$+w_n x $mean_n$，其中$w_{n-1}+w_n=1$，通常设置w_{n-1}为0.99，w_n为0.01。

[0121] 在本申请中提出的组批归一化方法，结合了组归一化和批归一化的优点，一方面减少了训练网络时所需要的批量大小Batch Size的大小，另一方面，结合了批量维度和时间维度的信息，保证了在小批量的情况下，准确率相比大批量的同样网络不会下降。

[0122] 根据本申请实施例，作为本申请实施例的优选，如图5所示，计算模块30包括：第一计算单元301，所述第一计算单元301，用于根据图片的数目将每个图片的数目内的均值和方差相加并取均值作为最后的均值和方差。作为本实施例中的优选，结合批量的信息，将每个批量内的均值和方差相加并取均值作为最后的均值和方差，对于批量大小Batch Size为1的情况，则本申请中退化成组归一化。

[0123] 计算模块30包括：第二计算单元302，所述第二计算单元302，用于结合前后图片的数目的信息将多次迭代的信息结合；以及对于第N-1次迭代的结果，通过加权平均法将迭代结果结合并计算最终的平均值。

[0124] 优选地，作为本实施例中的优选，本申请中结合前后批量的信息，将多次迭代的信息结合在一起，对于第N-1次迭代的结果，本申请通过加权平均法将其结合，权重参数为w_{n-1}和w_n，所以最终的平均值mean等于w_{n-1}x $mean_{n-1}$+w_n x $mean_n$，其中$w_{n-1}+w_n=1$，通常设置w_{n-1}为0.99，w_n为0.01。

[0125] 优选地，将在上述每组中计算平局值和方差包括：将组归一化处理操作和批量归一化处理操作结合。

[0126] 根据本申请实施例，作为本申请实施例的优选，如图6所示，所述特征图输入模块10包括：输入单元101，用于在输入层输入待识别图片；建立单元102，用于建立多个分组卷积模块；第一输出单元103，用于通过多个所述分组卷积模块输出所述待识别图片的特征图；以及第二输出单元104，用于根据所述特征图在输出层输出图像识别结果；其中，所述分组卷积模块中至少包括：一深度可分离卷积单元和一预设卷积核大小的卷积单元。

[0127] 本申请实施例的输入单元101中待识别图片可以用于人脸识别或者物体检测。

[0128] 特别地，在输入层输入待识别图片可以使用于无人车，安防等实时检测和识别。

[0129] 本申请实施例的建立单元102中其中，所述分组卷积模块中至少包括：一深度可分离卷积单元和一预设卷积核大小的卷积单元。

[0130] 在本步骤中提供了分组卷积模块，每个分组卷积模块中至少包括：一深度可分离卷积单元和一预设卷积核大小的卷积单元。由于在一个标准的CNN网络中按顺序可包括：卷积层、批归一化层(Group Batch Normalization)，激活函数层(Sigmoid)，在本实施例中提出的分组卷积模块可以替换现有CNN网络中的卷积层。

[0131] 优选地，每个所述分组卷积模块中至少包括：一深度可分离卷积单元和一卷积核大小为1*1的卷积单元。可以理解，上述的深度可分离卷积单元为本领域技术人员公知的一种卷积单元结构，由于能够有效利用参数，因此深度可分离卷积单元也可以用于移动设备中。此外，采用卷积核大小为1*1的卷积单元，可减少模型参数。

[0132] 每个组模块指的是深度分离卷积单元+1*1的卷积单元，1*1的卷积单元接在深度分离单元之后，可建立通道间的相关性。

[0133] 本申请实施例的第一输出单元103中本申请中的分组卷积模块与传统的卷积层相比，具有如下的特点：

[0134] 假设输入特征图大小为Sf x Sf x IN，经过一次卷积，输出的特征图大小为Sf x

Sf x OUT，

[0135] 按照传统的卷积操作，卷积核K的大小为：Sk x Sk x IN x OUT，

[0136] 其中，Sf是特征图的尺寸，Sk是卷积核的尺寸，IN是输入特征图的通道数，OUT是输出特征图的通道数，一次卷积操作的过程如下：

[0137] IN个Sk x Sk个卷积核与IN个输入特征图做卷积，得到的结果相加，得到一张输出特征图，同理，一共OUT次操作，得到OUT个输出特征图，用公式表示这个过程如下：

[0138] $$G_{k,l,n}=\sum_{i,j,m}K_{i,j,m,n}\cdot F_{k+i-1,l+j-1,m}$$

[0139] 根据公式，可以计算出传统卷积层的计算量为：

[0140] Sk x Sk x IN x OUT x Sf x Sf

[0141] 参数量为：

[0142] Sk x Sk x IN x OUT。

[0143] 对应一次传统卷积的是个分组卷积模块(即深度可分离卷积单元+卷积核大小为1*1的卷积单元)。其中，深度可分离卷积单元的具体实现如下：其卷积核K的大小为：Sk x Sk x IN，卷积核只跟对应通道的输入特征图做卷积，得到输出特征图，所以输出的特征图大小为Sf x Sf x IN，用公式表示这个过程如下：

[0144] $$\hat{G}_{k,l,m}=\sum_{i,j}\hat{K}_{i,j,m}\cdot F_{k+i-1,l+j-1,m}$$

[0145] 根据公式，可以计算出深度分离卷积的计算量为：

[0146] Sk x Sk x IN x Sf x Sf

[0147] 参数量为：

[0148] Sk x Sk x IN。

[0149] 深度分离卷积之后，进入一层传统的批归一化层和激活层，然后是1x1卷积层，1x1卷积层的卷积核大小为1x 1x IN x OUT，操作跟传统卷积一致，计算量为1x 1x IN x OUT x Sf x Sf，参数量为1x 1x IN x OUT。

[0150] 优选地，通过多个所述分组卷积模块输出所述待识别图片的特征图之后还依次通过：批归一化层和激活函数层。

[0151] 本申请实施例的第二输出单元104中具体地，分组卷积模块总的计算量为：

[0152] Sk x Sk x IN x Sf x Sf+1x 1x IN x OUT x Sf x Sf＝(Sk x Sk+OUT)x IN x Sf x Sf，

[0153] 参数量：Sk x Sk x IN+1x 1x IN x OUT＝(Sk x Sk+OUT)x IN，

[0154] 相比传统卷积，

[0155] 计算量：(Sk x Sk+OUT)/Sk x Sk x OUT＝1/OUT+1/(Sk x Sk)，

[0156] 参数量：(Sk x Sk+OUT)/Sk x Sk x OUT＝1/OUT+1/(Sk x Sk)，

[0157] 可知，根据特征图在输出层输出图像识别结果时采用分组卷积模块，每个分组卷积模块中至少包括：一深度可分离卷积单元和一预设卷积核大小的卷积单元，可以将计算量和参数量减少。

[0158] 在本申请另一实施例中还提供了一种客户端，包括所述的归一化处理装置。所述归一化处理装置的实现原理和有益效果如上描述，不再进行赘述。

[0159] 显然，本领域的技术人员应该明白，上述的本申请的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个的计算装置上，或者分布在多个计算装置所组成的网络上，可选地，它们可以用计算装置可执行的程序代码来实现，从而，可以将它们存储在存储装置中由计算装置来执行，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样，本申请不限制于任何特定的硬件和软件结合。

[0160] 以上所述仅为本申请的优选实施例而已，并不用于限制本申请，对于本领域的技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

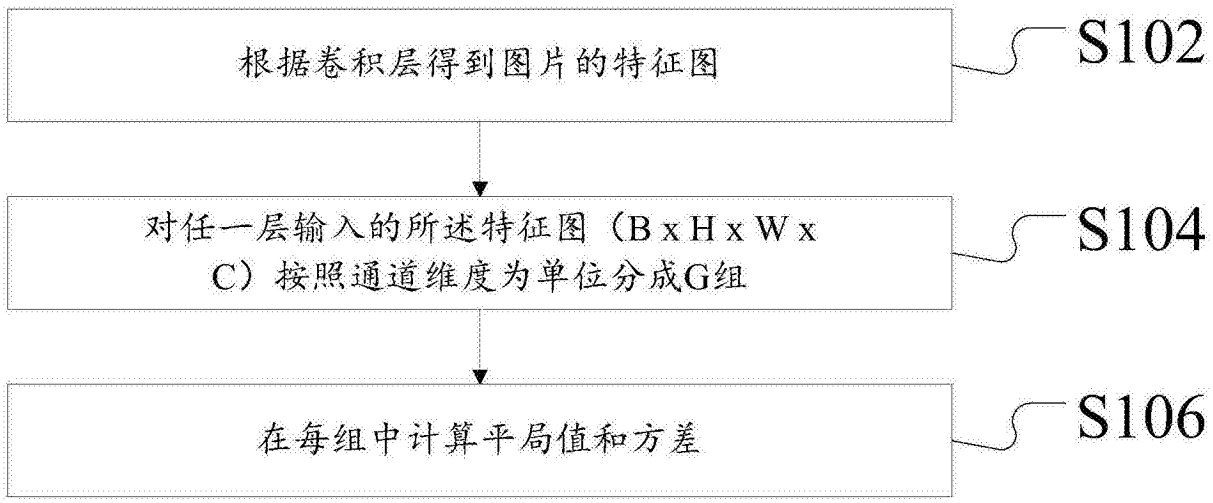

根据卷积层得到图片的特征图
S102
对任一层输入的所述特征图（B x H x W x C）按照通道维度为单位分成G组
S104
在每组中计算平局值和方差
S106

图1

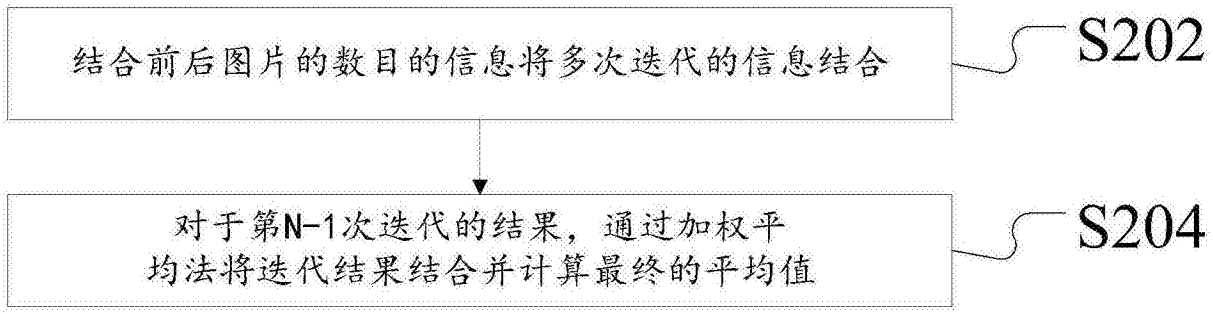

结合前后图片的数目的信息将多次迭代的信息结合
S202
对于第N-1次迭代的结果，通过加权平均法将迭代结果结合并计算最终的平均值
S204

图2

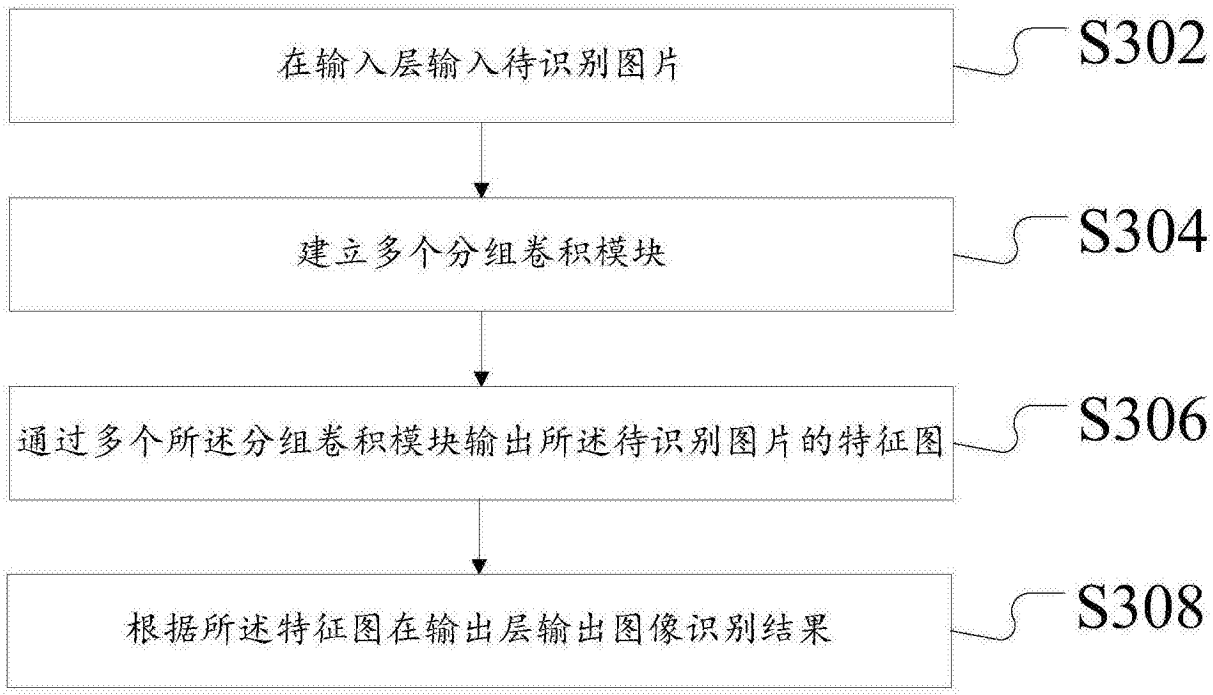

在输入层输入待识别图片
S302
建立多个分组卷积模块
S304
通过多个所述分组卷积模块输出所述待识别图片的特征图
S306
根据所述特征图在输出层输出图像识别结果
S308

图3


10
特征图输入模块
20
分组模块
30
计算模块

图4

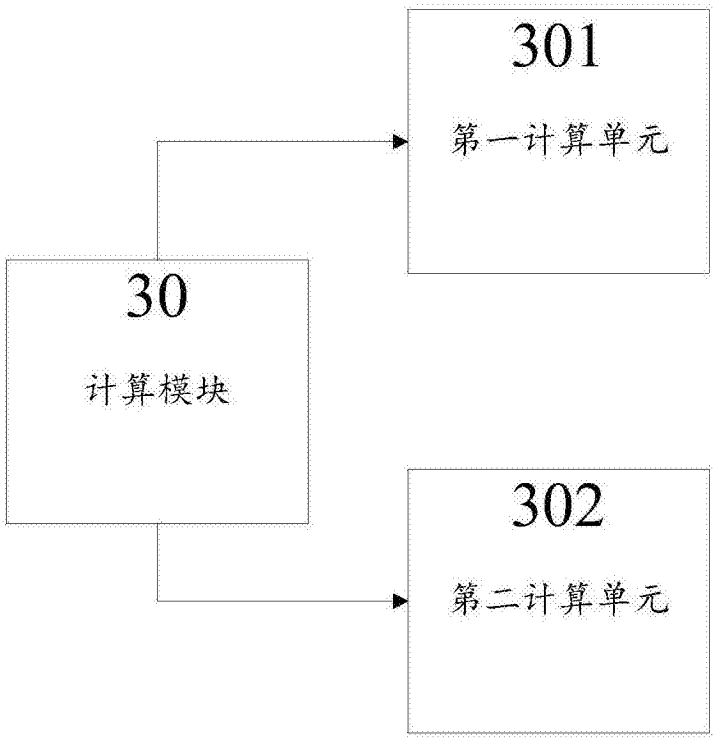

301
第一计算单元
30
计算模块
302
第二计算单元

图5

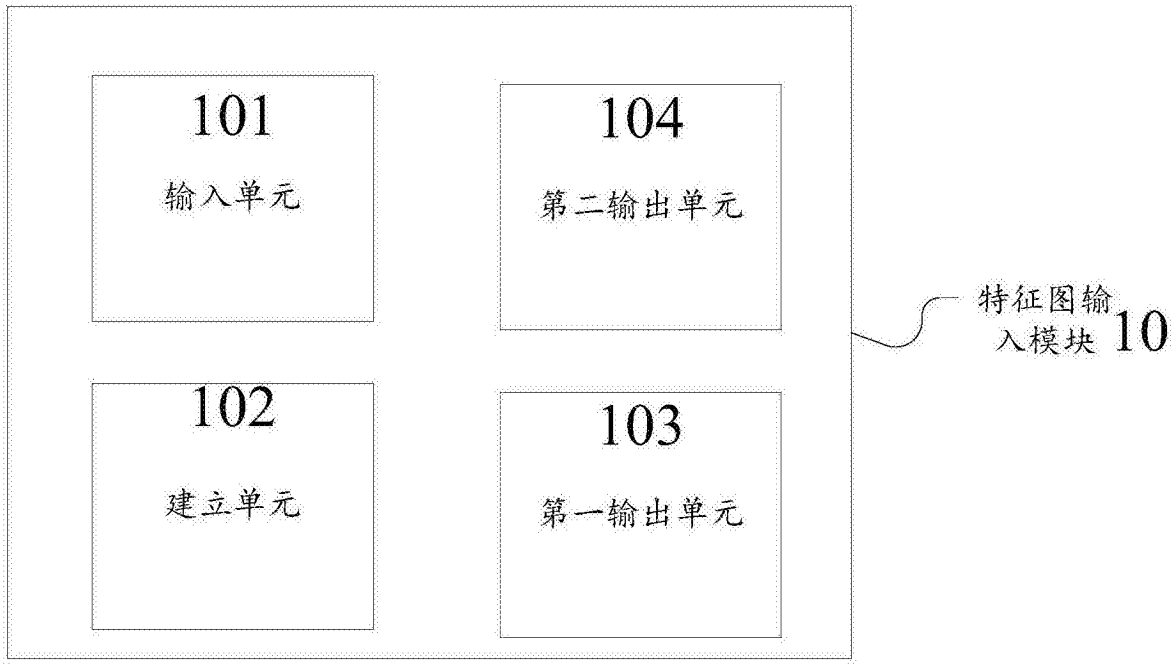

101
输入单元
104
第二输出单元
特征图输入模块
10
102
建立单元
103
第一输出单元

图6