



(12)发明专利申请

(10)申请公布号 CN 106874913 A

(43)申请公布日 2017.06.20

(21)申请号 201611243913.1

(22)申请日 2016.12.29

(71)申请人 南京江南博睿高新技术研究院有限公司

地址 210019 江苏省南京市建邺区奥体大街69号

(72)发明人 吴斌 欧烈川 刘潇

(74)专利代理机构 南京正联知识产权代理有限公司 32243

代理人 黄智明

(51)Int.Cl.

G06K 9/46(2006.01)

G06K 9/62(2006.01)

G06N 3/04(2006.01)

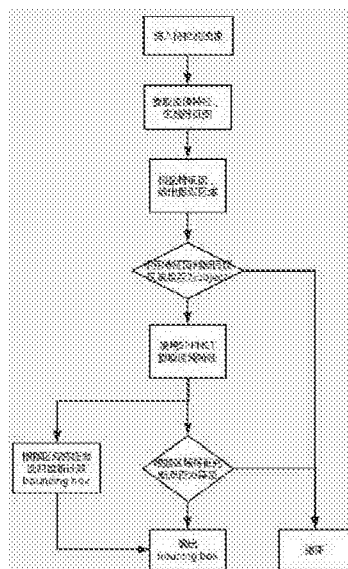
权利要求书1页 说明书4页 附图2页

(54)发明名称

一种菜品检测方法

(57)摘要

本发明提出了一种菜品检测方法,包括以下步骤:1)输入待检测的图像;2)使用预先训练的卷积神经网络提取图像特征;3)以特征图中每个元素在原图中对应位置为中心,给出提议区域;4)根据每个提议区域中心对应的特征向量判断该区域是否为物体;5)如果提议区域是物体,则使用SPPNET提取该区域的特征;6)根据区域特征做回归,计算更准确的bounding box,同时判断该区域是否为菜品。本发明的方法可以有效降低输出的错误率,获得更好的使用体验。



1. 一种菜品检测方法,其特征在于,所述方法包括以下步骤:1)输入待检测的图像;2)使用预先训练的卷积神经网络提取图像特征;3)以特征图中每个元素在原图中对应位置为中心,给出提议区域;4)根据每个提议区域中心对应的特征向量判断该区域是否为物体;5)如果提议区域是物体,则使用SPPNET提取该区域的特征;6)根据区域特征做回归,计算更准确的bounding box,同时判断该区域是否为菜品。

2. 如权利要求1所述的菜品检测方法,其特征在于,步骤2)提取图像特征的步骤包括:首先将待检测图像转换为224*224像素作为特征提取网络的输入,然后做5层卷积,每层卷积核分别为7*7、5*5、3*3、3*3、3*3,并且对每层提取的特征图的局部区域做最大值池化,得到大小为51*39,每个位置256维的特征图,该256维向量即为该位置对应原图局部区域的特征向量。

3. 如权利要求2所述的菜品检测方法,其特征在于,步骤3)给出提议区域的步骤包括:以特征图中每个元素在原图中的对应位置为中心,以原图大小为1000*600为标准,给出三种尺度128、256、512,三种比例1:1、1:2、2:1,共九个提议区域。

4. 如权利要求3所述的菜品检测方法,其特征在于,步骤4)判断是否为物体的步骤包括:如果原图中的bounding box和提议区域重合部分超过70%则认为该区域包含物体,否则认为不包含,对图像作标签;使用提议区域中心对应的特征向量作为特征;使用特征和标签训练一个全连接的神经网络来判断该区域是否为物体。

5. 如权利要求4所述的菜品检测方法,其特征在于,步骤5)包括:在提议区域是物体的情况下,在特征图上找到待提取特征的局部图像对应的区域,将该区域分别以4*4、2*2、1*1三种比例将特征图进行分割,对分割好的每一块做最大值池化,提取该区块的256维特征,将各区块提取出的特征连接到一起,最终得到 $4*4*256+2*2*256+1*1*256=5376$ 维特征。

6. 如权利要求5所述的菜品检测方法,其特征在于,步骤6)包括:将通过SPPNET提取出的区域特征向量作为输入,以该区域属于菜品和背景的可能性作为输出,训练一个神经网络,通过该神经网络进行判断。

7. 如权利要求6所述的菜品检测方法,其特征在于,判断是否属于菜品时,当为菜品的可能性为90%时将该物体判断为菜品。

一种菜品检测方法

技术领域

[0001] 本发明属于计算机图像处理领域,具体来说涉及一种快速检测图像中的菜品区域的技术。

背景技术

[0002] 在使用图像识别菜品种类的任务中,通常识别系统的返回结果只限于菜品种类,并且局限于菜品这一大类,对于并不包含菜品的图片其返回结果也仅限于识别程序所包含的种类。如果提交给识别系统的图像并不包含菜品,系统并不能对图像是否包含菜品做出判断,仍会返回最有可能的菜品种类,从而产生错误的结果。

发明内容

[0003] 因此,本发明提供一种从图像中检测菜品所在区域的方案,可检测出图像中包含的一个或多个菜品,并分别给出每个菜品所在区域的左上角坐标 x, y 以及宽度 w 和高度 h ,在图像中不包含菜品的情况下则无返回结果。

[0004] 具体来说,本发明采用了以下技术方案:

一种菜品检测方法,其特征在于,所述方法包括以下步骤:1)输入待检测的图像;2)使用预先训练的卷积神经网络提取图像特征;3)以特征图中每个元素在原图中对应位置为中心,给出提议区域;4)根据每个提议区域中心对应的特征向量判断该区域是否为物体;5)如果提议区域是物体,则使用SPPNET提取该区域的特征;6)根据区域特征做回归,计算更准确的bounding box,同时判断该区域是否为菜品。

[0005] 在以上方法中,步骤2)提取图像特征的步骤包括:首先将待检测图像转换为 $224*224$ 像素作为特征提取网络的输入,然后做5层卷积,每层卷积核分别为 $7*7, 5*5, 3*3, 3*3, 3*3$,并且对每层提取的特征图的局部区域做最大值池化,得到大小为 $51*39$,每个位置256维的特征图,该256维向量即为该位置对应原图局部区域的特征向量。

[0006] 进一步,步骤3)给出提议区域的步骤包括:以特征图中每个元素在原图中的对应位置为中心,以原图大小为 $1000*600$ 为标准,给出三种尺度128、256、512,三种比例1:1、1:2、2:1,共九个提议区域。

[0007] 更进一步,步骤4)判断是否为物体的步骤包括:如果原图中的bounding box和提议区域重合部分超过70%则认为该区域包含物体,否则认为不包含,对图像作标签;使用提议区域中心对应的特征向量作为特征;使用特征和标签训练一个全连接的神经网络来判断该区域是否为物体。

[0008] 接着,步骤5)包括:在提议区域是物体的情况下,在特征图上找到待提取特征的局部图像对应的区域,将该区域分别以 $4*4, 2*2, 1*1$ 三种比例将特征图进行分割,对分割好的每一块做最大值池化,提取该区块的256维特征,将各区块提取出的特征连接到一起,最终得到 $4*4*256+2*2*256+1*1*256=5376$ 维特征。

[0009] 更进一步,步骤6)包括:将通过SPPNET提取出的区域特征向量作为输入,以该区域

属于菜品和背景的可能性作为输出,训练一个神经网络,通过该神经网络进行判断。在以上步骤时,判断是否属于菜品时,当为菜品的可能性为90%时将该物体判断为菜品。

附图说明

[0010] 图1为本发明方法的流程示意图;

图2为本发明方法步骤中所用到的提议区域示意图。

具体实施方式

[0011] 在使用图像识别菜品种类的任务中,通常识别系统的返回结果只限于菜品种类。如果提交给识别系统的图像并不包含菜品,系统并不能对图像是否包含菜品做出判断,仍会返回最有可能的菜品种类,从而产生错误的结果。

[0012] 因此,本发明提供一种从图像中检测菜品所在区域的方案,可检测出图像中包含的一个或多个菜品,并分别给出每个菜品所在区域的左上角坐标 x, y 以及宽度 w 和高度 h ,在图像中不包含菜品的情况下则无返回结果。

[0013] 在本文采用的技术之前,传统的物体检测领域的主流方法是DPM (Deformable parts models)。DPM在VOC2007数据集上达到了43%的mAP (Mean Average Precision,平均正确率均值,在物体检测领域为多个不同物体的检测的准确率的平均值),运算效率可以接近实时(30fps~100fps)。

[0014] 随后出现的RCNN引入了卷积神经网络作为特征提取方法,并提出了一套物体检测框架,即提议区域→特征提取→分类和bounding box调整。该方法在准确率上有所突破,在VOC2007上达到了53.5%的mAP。然而由于需要对每个提议区域做特征提取,该算法非常耗时,根据提议区域数量的不同,该算法需要几十秒到上百秒不等的运算时间。

[0015] 由于RCNN提取的提议区域存在大量重叠部分,因此提议区域的特征提取存在大量重复计算。针对这一点,fast RCNN引入了SPPNET,使得只需要对原图像整体提取一次特征,就可以计算出所有提议区域的特征向量。该算法极大地降低了运行时间,处理一张图片需要的时间被缩短到2s左右,并且检测的mAP也被提高到了70%。

[0016] fast RCNN虽然降低了RCNN的运行时间,但仍然难以满足实时性的要求,而其运行效率的瓶颈在于提议可能为物体的区域这一步骤。本发明最终采用的faster RCNN在这一方面做出了改进,提出了RPN(region proposal network)来给出提议区域,结合fast RCNN的方法,在保证准确率不受影响的情况下,将运行时间压缩至可以满足实时性的要求。根据faster RCNN选取的特征提取网络的不同,其运行效率可达到5-20fps不等,虽然不及DPM的运行效率,但该方法在准确率上有极大的优势,并且在本发明的应用场景,即菜品检测问题中,faster RCNN的运行效率已经足够。

[0017] 本发明提出的图像中菜品区域的检测技术是为了配合菜品识别技术而开发。常见的菜品识别技术可分类的种类有限,并且局限于菜品这一大类,对于并不包含菜品的图片其返回结果也仅限于识别程序所包含的种类。也就是对于并不包含菜品的图片,识别程序仍然会返回某个菜品的名称,从而造成令人费解的结果。本发明提出的菜品检测方法可以在识别程序执行之前判断图像中是否包含菜品,筛选掉不包含菜品的图像;同时可以给出菜品所在区域的bounding box,可以让菜品识别程序针对更准确的菜品区域进行识别,提

高识别的准确率。

[0018] 本发明解决的技术问题属于计算机视觉中的物体检测问题。物体检测问题的主要任务是在图像中求出指定物体所在区域的bounding box。本发明采用的技术方案即为物体检测领域的RCNN技术。

[0019] RCNN经历了三个阶段的发展,即RCNN, fast RCNN, faster RCNN三个发展阶段。三个发展阶段均采用RCNN的技术框架,逐步将各模块的功能使用神经网络实现,期间算法的准确率和运行效率都有明显提升。

[0020] RCNN主要由四个功能模块组成:使用第三方工具给出可能为物体的区域(region proposal);使用卷积神经网络提取提议区域的图像特征;使用SVM,根据提取出的提议区域的特征向量判断该区域的分类;同时使用提议区域的特征向量对物体的bounding box做回归,得出更准确的bounding box。

[0021] fast RCNN对RCNN的改进主要有:

(1) 引入SPPNET提取提议区域的特征。SPPNET是一项用于从不同尺寸和比例的图像中提取出相同维度特征的技术。RCNN对每个提议区域都要通过卷积神经网络提取特征,运算量很大,非常耗时。引入SPPNET技术后,只需对整幅图像提取一次特征,然后使用SPPNET分别对每个提议区域提取特征即可。

[0022] (2) 判断提议区域的种类,使用神经网络替代SVM作为分类器。

[0023] 最后,faster RCNN将提议区域也使用神经网络实现,将整个RCNN框架都使用神经网络技术完成。本发明采用faster RCNN作为技术框架,采用人工标注的345张包含菜品的图像和9663张不包含菜品的图像作为训练集,训练得到一个神经网络来求出可能的菜品所在区域。其具体检测过程如下:

(1) 输入待检测图像。

[0024] (2) 使用预先训练的卷积神经网络提取图像特征。

[0025] 首先将待检测图像转换为224*224像素作为特征提取网络的输入。

[0026] 特征提取网络具有五层结构,每层采用不同尺度的卷积核(7*7, 5*5, 3*3, 3*3, 3*3)做卷积,并对该层提取的特征图的局部区域做最大值池化操作(max pooling,即取该区域的极大值作为区域的特征值)。经过五层卷积和池化操作,得到大小为51*39,每个位置256维的特征图。该256维向量即为该位置对应原图局部区域的特征向量。

[0027] (3) 以特征图中每个元素在原图中的对应位置为中心,以原图大小为1000*600为标准,给出三种尺度(128、256、512)、三种比例(1:1、1:2、2:1),共九个提议区域。参见图1,提议区域即可能为物体的区域,需要后续步骤来判断。本发明采用的三种尺度和比例在运用于菜品识别时,可以涵盖大多数物体可能的大小,相比于现有技术,运算量降低,促进了实时反馈。

[0028] (4) 根据每个提议区域中心对应的特征向量判断该区域是否为物体,如果不是则退出。具体实现方法为:如果原图中的bounding box和提议区域重合部分超过一定阈值,如70%则认为该区域包含物体,否则认为不包含,对图像作标签;使用提议区域中心对应的特征向量作为特征;使用特征和标签训练一个全连接的神经网络来判断该区域是否为物体。特征向量作为该网络的输入,网络输出结果为0~1的浮点数,即该区域包含物体的可能性。最后设置阈值,在可能性大于阈值的情况下则认为该区域包含物体。

[0029] (5) 如果提议区域是物体,使用SPPNET提取该区域的特征。SPPNET是一种用于提取不同尺度区域的特征的方法。具体方法为:首先在特征图上找到待提取特征的局部图像对应的区域;将该区域分别以 $4*4$ 、 $2*2$ 、 $1*1$ 三种比例将特征图进行分割;对分割好的每一块做最大值池化操作,提取该区块的256维特征;将各区块提取出的特征连接到一起,最终得到 $4*4*256+2*2*256+1*1*256=5376$ 维特征。

[0030] (6) 根据区域特征做回归,计算更准确的bounding box,同时判断该区域是否为菜品。通过回归计算bounding box同样是通过训练一个神经网络来完成的。将(5)中SPPNET提取出的区域特征向量作为输入,物体所在的实际位置(左上角坐标 x,y 以及宽度和高度 w,h)作为输出,训练一个神经网络,该神经网络即可求得更准确的bounding box。判断采用另一个神经网络,输入和用于回归的神经网络一样,即SPPNET提取出的特征,输出为该区域属于某一类的可能性,在本项目中即为菜品和背景(非菜品)两类。实际应用时在是菜品的可能性大于一定的阈值,如90%时才认为是菜品。

[0031] (7) 如果不是菜品则结束,如果是菜品则输出更准确的bounding box作为最终结果。

[0032] 上面结合附图和具体实施例对本发明的实施方式作了详细的说明,但是本发明不限于上述实施方式,在所属技术领域普通技术人员所具备的知识范围内,还可以在不脱离本发明宗旨的前提下做出各种变化。

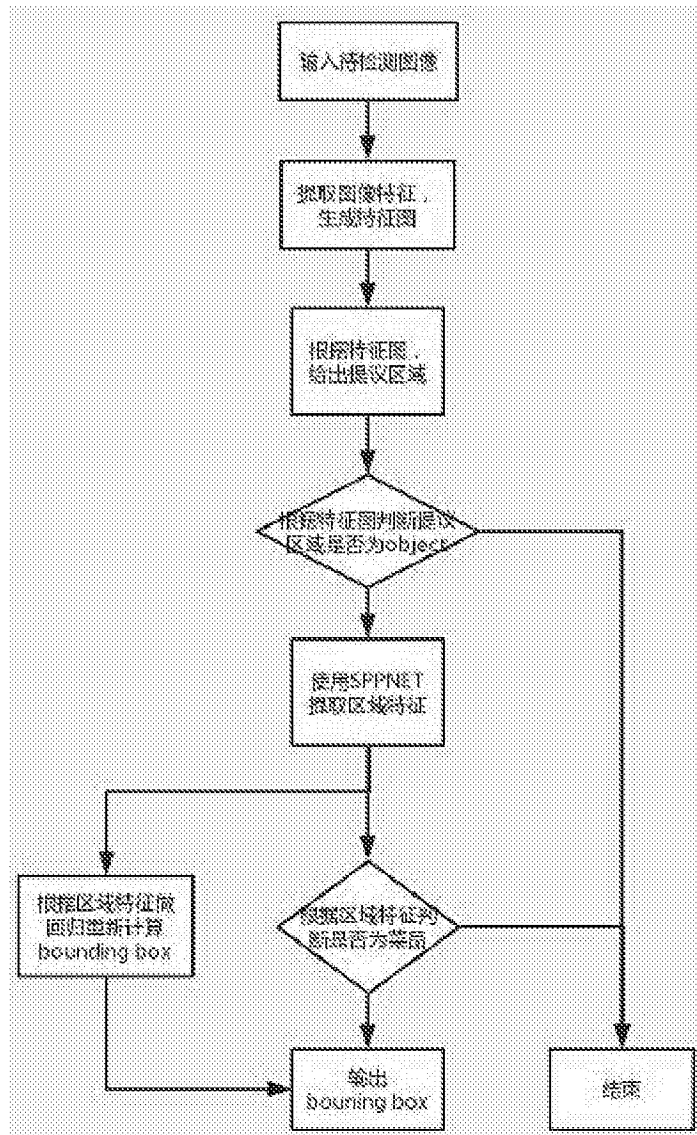


图1

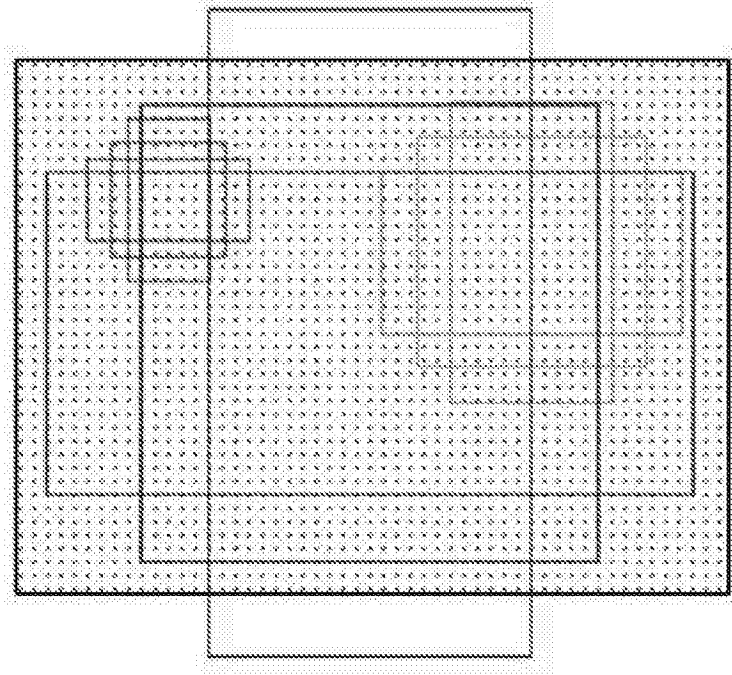


图2