



# (12)发明专利申请

(10)申请公布号 CN 111199224 A

(43)申请公布日 2020.05.26

(21)申请号 201811379524.0

(22)申请日 2018.11.20

(71)申请人 中国电信股份有限公司

地址 100033 北京市西城区金融大街31号

(72)发明人 龚靖 沈雷 王铮 杨迪 汪少敏

(74)专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 许蓓

(51)Int.Cl.

G06K 9/32(2006.01)

G06N 3/04(2006.01)

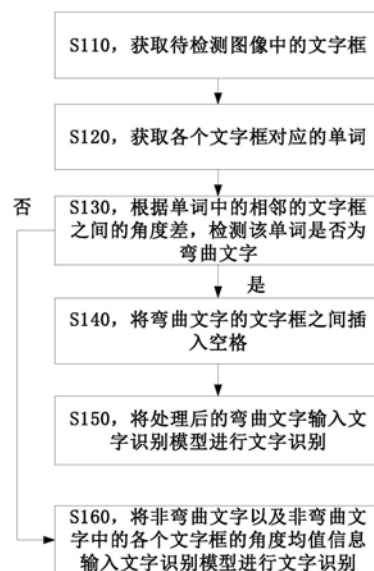
权利要求书1页 说明书4页 附图1页

## (54)发明名称

弯曲文字识别方法和装置

## (57)摘要

本公开提出一种弯曲文字识别方法和装置，涉及文字识别领域。获取待检测图像中的文字框以及各个文字框对应的单词，根据单词中的相邻的文字框之间的角度差，检测该单词是否为弯曲文字，将弯曲文字的文字框之间插入空格，将处理后的弯曲文字输入文字识别模型进行文字识别。从而，实现弯曲文字的检测和识别。



1. 一种弯曲文字识别方法,包括:  
获取待检测图像中的文字框以及各个文字框对应的单词;  
根据单词中的相邻的文字框之间的角度差,检测该单词是否为弯曲文字;  
将弯曲文字的文字框之间插入空格;  
将处理后的弯曲文字输入文字识别模型进行文字识别。
2. 如权利要求1所述的方法,其中,  
待检测图像中的文字框通过将待检测图像输入卷积神经网络CNN算法来获得,  
其中,卷积神经网络算法预先利用文字样本进行训练。
3. 如权利要求1所述的方法,其中,  
各个文字框对应的单词通过将各个文字框输入深度优先搜索DFS算法来获得。
4. 如权利要求1所述的方法,其中,  
如果单词中的相邻的文字框之间的角度差处于最小阈值和最大阈值之间,该单词被判定为弯曲文字。
5. 如权利要求4所述的方法,  
如果单词中的相邻的文字框之间的角度差小于或等于最小阈值,该单词被判定为非弯曲文字;  
如果单词中的相邻的文字框之间的角度差大于或等于最大阈值,拆分该单词。
6. 如权利要求5所述的方法,还包括:  
将非弯曲文字以及非弯曲文字中的各个文字框的角度均值信息输入文字识别模型进行文字识别。
7. 如权利要求1或6所述的方法,其中,  
所述文字识别模型为联结主义时间分类CTC文字识别模型。
8. 如权利要求3所述的方法,其中,  
把每个文字框作为节点来建立图模型,用DFS算法从图模型中找连通分量,每个连通分量为一个单词。
9. 一种弯曲文字识别装置,包括:  
存储器;以及  
耦接至所述存储器的处理器,所述处理器被配置为基于存储在所述存储器中的指令,执行权利要求1-7中任一项所述的弯曲文字识别方法。
10. 一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现权利要求1-7中任一项所述的弯曲文字识别方法。

## 弯曲文字识别方法和装置

### 技术领域

[0001] 本公开涉及文字识别领域,特别涉及一种弯曲文字识别方法和装置。

### 背景技术

[0002] 目前人工智能领域,单发多框检测(single shot multibox detection,简称SSD)方法仅能够对水平文字进行检测,其延伸出的Seglink方法也仅能对同一直线上的倾斜文字进行检测。相关技术均无法对弯曲文字进行检测。

### 发明内容

[0003] 本公开提出一种能够检测和识别弯曲文字的方案。

[0004] 本公开的一些实施例提出一种弯曲文字识别方法,包括:

[0005] 获取待检测图像中的文字框以及各个文字框对应的单词;

[0006] 根据单词中的相邻的文字框之间的角度差,检测该单词是否为弯曲文字;

[0007] 将弯曲文字的文字框之间插入空格;

[0008] 将处理后的弯曲文字输入文字识别模型进行文字识别。

[0009] 在一些实施例中,待检测图像中的文字框通过将待检测图像输入卷积神经网络CNN算法来获得,

[0010] 其中,卷积神经网络算法预先利用文字样本进行训练。

[0011] 在一些实施例中,各个文字框对应的单词通过将各个文字框输入深度优先搜索DFS算法来获得。

[0012] 在一些实施例中,如果单词中的相邻的文字框之间的角度差处于最小阈值和最大阈值之间,该单词被判定为弯曲文字。

[0013] 在一些实施例中,如果单词中的相邻的文字框之间的角度差小于或等于最小阈值,该单词被判定为非弯曲文字;

[0014] 如果单词中的相邻的文字框之间的角度差大于或等于最大阈值,拆分该单词。

[0015] 在一些实施例中,还包括:

[0016] 将非弯曲文字以及非弯曲文字中的各个文字框的角度均值信息输入文字识别模型进行文字识别。

[0017] 在一些实施例中,所述文字识别模型为联结主义时间分类CTC文字识别模型。

[0018] 在一些实施例中,把每个文字框作为节点来建立图模型,用DFS算法从图模型中找连通分量,每个连通分量为一个单词。

[0019] 本公开的一些实施例提出一种弯曲文字识别装置,包括:

[0020] 存储器;以及

[0021] 耦接至所述存储器的处理器,所述处理器被配置为基于存储在所述存储器中的指令,执行前述任一个实施例的弯曲文字识别方法。

[0022] 本公开的一些实施例提出一种计算机可读存储介质,其上存储有计算机程序,该

程序被处理器执行时实现前述任一个实施例的弯曲文字识别方法。

### 附图说明

[0023] 下面将对实施例或相关技术描述中所需要使用的附图作简单地介绍。根据下面参照附图的详细描述,可以更加清楚地理解本公开,

[0024] 显而易见地,下面描述中的附图仅仅是本公开的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0025] 图1为本公开一些实施例的弯曲文字识别方法的流程示意图。

[0026] 图2为本公开一些实施例的弯曲文字识别装置的结构示意图。

### 具体实施方式

[0027] 下面将结合本公开实施例中的附图,对本公开实施例中的技术方案进行清楚、完整地描述。

[0028] 图1为本公开一些实施例的弯曲文字识别方法的流程示意图。

[0029] 如图1所示,该实施例的方法包括:

[0030] S110,获取待检测图像中的文字框(设为segment,可简写为seg)。

[0031] 其中,文字框也称“段”,是覆盖单词的一部分的边界框。

[0032] 在一些实施例中,将待检测图像输入卷积神经网络(CNN,Convolutional Neural Networks)算法,输出待检测图像中的文字框。其中,卷积神经网络算法预先利用文字样本进行训练。文字框的描述参数例如包括 $(x, y, w, h, \theta)$ ,其中, $(x, y)$ 为位置坐标, $(w, h)$ 为宽和高,以及角度为 $\theta$ 。

[0033] S120,获取各个文字框对应的单词(设为word)。

[0034] 在一些实施例中,将各个文字框的信息输入深度优先搜索(Depth-First Search, DFS)算法来获得各个文字框对应的单词。

[0035] 具体地,把每个文字框作为节点来建立图模型,用DFS算法从图模型中找连通分量,每个连通分量为一个单词。

[0036] S130,根据单词中的相邻的文字框之间的角度差,检测该单词是否为弯曲文字。

[0037] 检测规则例如为:

[0038] 将相邻的文字框之间的角度差与预设的最小阈值和最大阈值进行比较,根据比较结果进行判定。

[0039] 如果单词中的相邻的文字框之间的角度差处于最小阈值和最大阈值之间,说明该相邻的文字框处于曲线上,该单词被判定为弯曲文字。

[0040] 如果单词中的相邻的文字框之间的角度差小于或等于最小阈值,说明该相邻的文字框处于直线上,该单词被判定为非弯曲文字,例如可以是水平文字或同一直线上的倾斜文字。

[0041] 如果单词中的相邻的文字框之间的角度差大于或等于最大阈值,说明这两个文字框属于不同的单词,拆分该单词。

[0042] S140,将弯曲文字的文字框之间插入空格。

[0043] S150,将处理后的弯曲文字输入文字识别模型进行文字识别。

[0044] 在一些实施例中,所述文字识别模型为联结主义时间分类(CTC,Connectionist Temporal Classification)文字识别模型。

[0045] S160,将非弯曲文字以及非弯曲文字中的各个文字框的角度均值信息输入文字识别模型进行文字识别。

[0046] 在一些实施例中,所述文字识别模型为CTC文字识别模型。

[0047] 上述实施例实现了弯曲文字的检测和识别,可以应用于商标标识、广告、艺术字及其他弯曲文字的检测。

[0048] 下面用算法的方式描述上述方案。

[0049] 首先,通过CNN算法获取待检测图像中的文字框 $(x,y,w,h,\theta)$ 。

[0050] 接着,利用DFS算法获得各个文字框对应的单词,一个由n个文字框组成的单词word例如可以表示为:

[0051]  $word = (seg_1, seg_2, seg_3, \dots, seg_i, \dots, seg_n)$

[0052] 令循环中各计数变量 $i=0, j=1, k=0$ 。

[0053] 令 $k=k+1, i=i+1$ ,若 $i < n$ 成立,则循环执行以下操作1-3):

[0054] 1)  $word_i[k] = seg_i, curve_j = false, sum\_theta_j = \theta_i$

[0055] 其中, $curve$ 表示弯曲文字标识,如果是 $false$ ,说明不是弯曲文字,如果是 $true$ ,说明是弯曲文字, $\theta_i$ 表示文字框 $seg_i$ 的角度。

[0056] 2) 计算同一连通分量中相邻文字框的角度差

[0057]  $diff\_theta_i = |\theta_{i+1} - \theta_i|$

[0058] 3) 将角度差 $diff\_theta_i$ 与最小阈值 $diff\_min$ 和最大阈值 $diff\_max$ 进行比较:

[0059] 3-1) 如果 $diff\_theta_i \leq diff\_min$ ,说明该相邻的文字框处于直线上,计算这些文字框的角度均值 $aver_j: sum\_theta_j = sum\_theta_j + \theta_{i+1}, aver_j = sum\_theta_j / k$ ;

[0060] 3-2) 如果 $diff\_min < diff\_theta_i < diff\_max$ ,说明该相邻的文字框处于曲线上, $curve_j = true$ ,文字框 $seg_i$ 和 $seg_{i+1}$ 之间插入空格;

[0061] 3-3) 如果 $diff\_theta_i \geq diff\_max$ ,说明这两个文字框属于不同的单词,拆分该单词,令 $j = j+1, k=0$ 。

[0062] 通过循环执行操作1-3),可以得到各个单词 $word_j$ 。

[0063] 最后,将 $word_j$ 输入CTC模型进行文字识别。

[0064] 图2为本公开一些实施例的弯曲文字识别装置的结构示意图。

[0065] 如图2所示,该实施例的装置200包括:

[0066] 存储器210以及耦接至该存储器210的处理器220,处理器220被配置为基于存储在存储器210中的指令,执行前述任意一些实施例中的网络性能监控方法。

[0067] 其中,存储器210例如可以包括系统存储器、固定非易失性存储介质等。系统存储器例如存储有操作系统、应用程序、引导装载程序(Boot Loader)以及其他程序等。

[0068] 本公开的一些实施例还提出一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现前述任一个实施例的弯曲文字识别方法。

[0069] 本领域内的技术人员应当明白,本公开的实施例可提供为方法、系统、或计算机程序产品。因此,本公开可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本公开可采用在一个或多个其中包含有计算机可用程序代码的计算

机可用非瞬时性存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0070] 本公开是参照根据本公开实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解为可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0071] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0072] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0073] 以上所述仅为本公开的较佳实施例,并不用以限制本公开,凡在本公开的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本公开的保护范围之内。

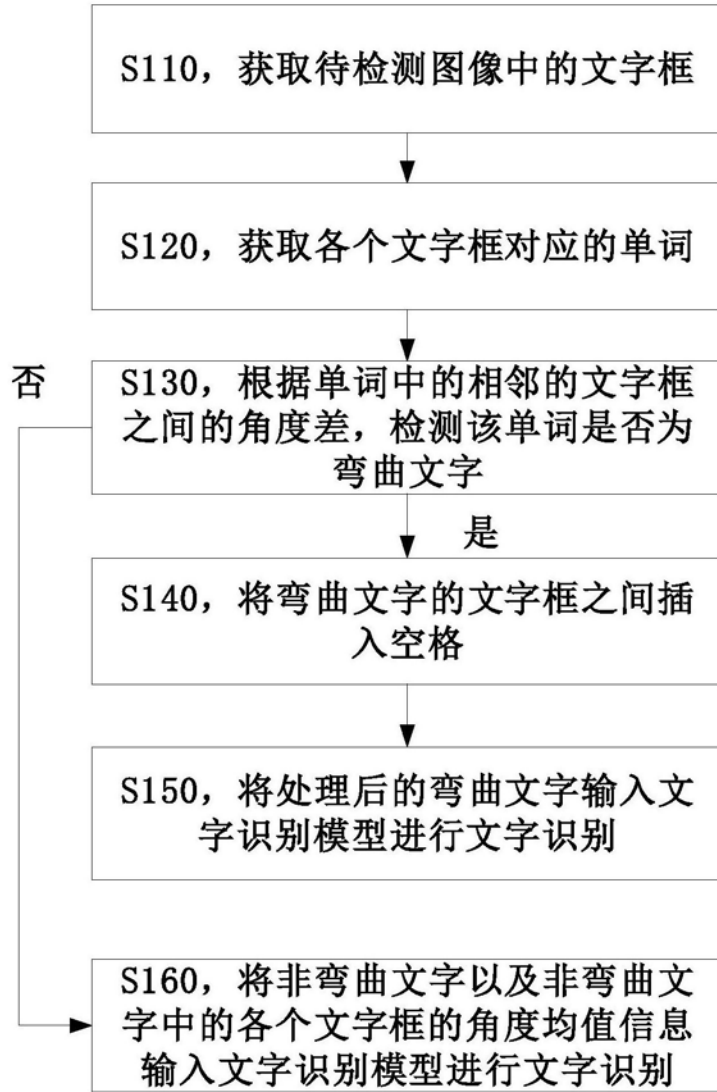


图1



图2