



(12) 发明专利

(10) 授权公告号 CN 113807336 B

(45) 授权公告日 2023. 06. 30

(21) 申请号 202110906651.7

(22) 申请日 2021.08.09

(65) 同一申请的已公布的文献号

申请公布号 CN 113807336 A

(43) 申请公布日 2021.12.17

(73) 专利权人 华南理工大学

地址 510640 广东省广州市天河区五山路
381号

专利权人 人工智能与数字经济广东省实验
室(广州)

(72) 发明人 黄双萍 刘宗昊 王庆丰

(74) 专利代理机构 广州市华学知识产权代理有
限公司 44245

专利代理师 李君

(51) Int.Cl.

G06V 30/14 (2022.01)

G06V 10/82 (2022.01)

(56) 对比文件

CN 108549893 A, 2018.09.18

CN 108805131 A, 2018.11.13

CN 110147786 A, 2019.08.20

CN 110837835 A, 2020.02.25

CN 110929665 A, 2020.03.27

CN 111898411 A, 2020.11.06

审查员 洪汇隆

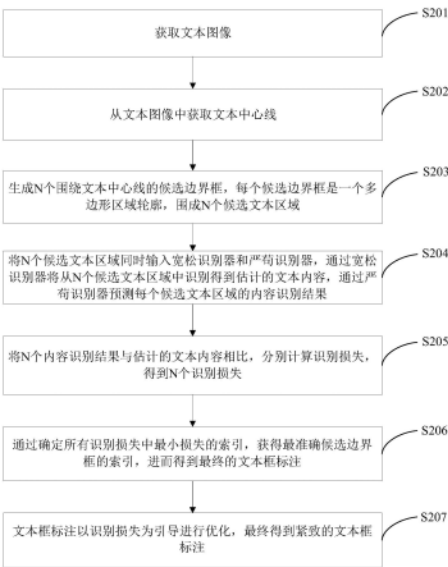
权利要求书3页 说明书8页 附图5页

(54) 发明名称

图像文本检测半自动标注方法、系统、计算机设备及介质

(57) 摘要

本发明公开了一种图像文本检测半自动标注方法、系统、计算机设备及介质,所述方法包括:获取文本图像;从文本图像中获取文本中心线;生成N个围绕文本中心线的候选边界框,每个候选边界框是一个多边形区域轮廓,围成N个候选文本区域;将N个候选文本区域同时输入宽松识别器和严苛识别器,通过宽松识别器将从N个候选文本区域中识别得到估计的文本内容,通过严苛识别器预测每个候选文本区域的内容识别结果;将N个内容识别结果与估计的文本内容相比,分别计算识别损失,得到N个识别损失;通过确定所有识别损失中最小损失的索引,获得最准确候选边界框的索引,进而得到最终的文本框标注;将文本框标注以识别损失为引导进行优化,最终得到紧致的文本框标注。本发明能够提高文本检测标注效率和标注效果。



1. 一种图像文本检测半自动标注方法,其特征在于,所述方法包括:

获取文本图像;

从文本图像中获取文本中心线,所述文本中心线是一条贯穿文本中心的弯曲折线,由依次连接K+1个点构成K条直线段构成;

生成N个围绕文本中心线的候选边界框,每个候选边界框是一个多边形区域轮廓,围成N个候选文本区域;

将N个候选文本区域同时输入宽松识别器和严苛识别器,通过宽松识别器将从N个候选文本区域中识别得到估计的文本内容,通过严苛识别器预测每个候选文本区域的内容识别结果;

将N个内容识别结果与估计的文本内容相比,分别计算识别损失,得到N个识别损失;

通过确定所有识别损失中最小损失的索引,获得最准确候选边界框的索引,进而得到最终的文本框标注;

将文本框标注以识别损失为引导进行优化,最终得到紧致的文本框标注;

所述生成N个围绕文本中心线的候选边界框,具体包括:

确定K+1条法线,每条法线 n_i 与文本中心线相交于点 c_i ,同时与文本中心线在点 c_i 的切线垂直;

在每条法线 n_i 上,确定一条长度为 h_j 的线段 l_i ,线段 l_i 被点 c_i 二等分,将所有线段的端点视为多边形的顶点,依次连接所有顶点获得一个候选边界框 B_j ;

通过确定N个不同的变量 h_j ,以获得N个候选边界框;

所述变量 h_j 的确定如下式:

$$h_j = \frac{9j^2}{4N^2}$$

其中, $j=1,2,\dots,N$;

所述通过宽松识别器将从N个候选文本区域中识别得到估计的文本内容,具体包括:

通过宽松识别器对N个候选文本区域进行识别,得到N个识别结果 $\{T_j | j=1,2,\dots,N\}$,识别结果 $T_j = (t_j^{u,v}) \in [0,1]^{L \times C}$ 是一个形状为 $L \times C$ 的矩阵;

计算相邻识别结果 T_j 和 T_{j-1} 的差异 d_j ,从差异 d_j 最小的识别结果 T_{j^*} 中获得估计文本内容 T^* ,其中 $j^* = \underset{j}{\operatorname{argmin}} d_j$;

所述从差异 d_j 最小的识别结果 T_{j^*} 中获得估计文本内容 T^* ,如下式:

$$t_j^{u,v} = \begin{cases} 1, & \text{if } t_j^{u,v} = \max_v t_j^{u,v} \\ 0, & \text{others} \end{cases}$$

$$T^* = (t_j^{u,v}) \in \{0,1\}^{L \times C}$$

其中,分量 $t_j^{u,v}$ 表示识别结果 T_j 第u个字符属于第v类的概率。

2. 根据权利要求1所述的图像文本检测半自动标注方法,其特征在于,所述宽松识别器的结构是一个基于卷积神经网络的图像文本识别器,包括矫正器、第一编码器、第一序列模型和第一解码器;

所述矫正器,用于矫正输入图像区域 R_j 的文本图像形状;

所述第一编码器,用于对矫正后的文本图像进行特征提取;

所述第一序列模型,用于提取上下文依赖特征;

所述第一解码器,用于对上下文依赖特征进行转译输出识别结果 T_j ;

所述宽松识别器使用合成图像的宽松文本区域进行训练,所述宽松文本区域是区域图像内除文本外,引入适量背景干扰的区域。

3.根据权利要求1所述的图像文本检测半自动标注方法,其特征在于,严苛识别器的结构是一个基于卷积神经网络的图像文本识别器,包括第二编码器、第二序列模型和第二解码器;

所述第二编码器,用于对输入图像区域 R_j 进行特征提取;

所述第二序列模型,用于提取上下文依赖特征;

所述第二解码器,用于对上下文依赖特征进行解码输出识别结果 s_j ;

所述严苛识别器使用合成图像的紧致文本区域进行训练,所述紧致文本区域是区域图像内除文本外,不存在背景干扰的图像区域。

4.根据权利要求1所述的图像文本检测半自动标注方法,其特征在于,所述将文本框标注以识别损失为引导进行优化,如下式:

$$B'_{j*} = B_{j*} - \mu \times \nabla_{B_{j*}} l_{j*}$$

其中, $\nabla_{B_{j*}} l_{j*}$ 表示识别损失 l_{j*} 对 B_{j*} 的梯度, μ 则表示更新步长。

5.一种图像文本检测半自动标注系统,其特征在于,所述系统包括:

文本图像获取模块,用于获取文本图像;

文本中心线获取模块,用于从文本图像中获取文本中心线,所述文本中心线是一条贯穿文本中心的弯曲折线,由依次连接 $K+1$ 个点构成 K 条直线段构成;

候选边界框生成模块,用于生成 N 个围绕文本中心线的候选边界框,每个候选边界框是一个多边形区域轮廓,围成 N 个候选文本区域;

识别模块,用于将 N 个候选文本区域同时输入宽松识别器和严苛识别器,通过宽松识别器将从 N 个候选文本区域中识别得到估计的文本内容,通过严苛识别器预测每个候选文本区域的内容识别结果;

识别损失计算模块,用于将 N 个内容识别结果与估计的文本内容相比,分别计算识别损失,得到 N 个识别损失;

文本框标注获取模块,用于通过确定所有识别损失中最小损失的索引,获得最准确候选边界框的索引,进而得到最终的文本框标注;

优化模块,用于将文本框标注以识别损失为引导进行优化,最终得到紧致的文本框标注;

所述生成 N 个围绕文本中心线的候选边界框,具体包括:

确定 $K+1$ 条法线,每条法线 n_i 与文本中心线相交于点 c_i ,同时与文本中心线在点 c_i 的切线垂直;

在每条法线 n_i 上,确定一条长度为 h_j 的线段 l_i ,线段 l_i 被点 c_i 二等分,将所有线段的端点视为多边形的顶点,依次连接所有顶点获得一个候选边界框 B_j ;

通过确定N个不同的变量 h_j ,以获得N个候选边界框;

所述变量 h_j 的确定如下式:

$$h_j = \frac{9j^2}{4N^2}$$

其中, $j=1, 2, \dots, N$;

所述通过宽松识别器将从N个候选文本区域中识别得到估计的文本内容,具体包括:

通过宽松识别器对N个候选文本区域进行识别,得到N个识别结果 $\{T_j | j=1, 2, \dots, N\}$,识别结果 $T_j = (t_j^{u,v}) \in [0,1]^{L \times C}$ 是一个形状为 $L \times C$ 的矩阵;

计算相邻识别结果 T_j 和 T_{j-1} 的差异 d_j ,从差异 d_j 最小的识别结果 T_{j^*} 中获得估计文本内容 T^* ,其中 $j^* = \underset{j}{\operatorname{argmin}} d_j$;

所述从差异 d_j 最小的识别结果 T_{j^*} 中获得估计文本内容 T^* ,如下式:

$$t_j^{u,v} = \begin{cases} 1, & \text{if } t_j^{u,v} = \max_v t_j^{u,v} \\ 0, & \text{others} \end{cases}$$

$$T^* = (t_j^{u,v}) \in \{0,1\}^{L \times C}$$

其中,分量 $t_j^{u,v}$ 表示识别结果 T_j 第u个字符属于第v类的概率。

6. 一种计算机设备,包括处理器以及用于存储处理器可执行程序存储器,其特征在于,所述处理器执行存储器存储的程序时,实现权利要求1-4任一项所述的图像文本检测半自动标注方法。

7. 一种存储介质,存储有程序,其特征在于,所述程序被处理器执行时,实现权利要求1-4任一项所述的图像文本检测半自动标注方法。

图像文本检测半自动标注方法、系统、计算机设备及介质

技术领域

[0001] 本发明涉及一种图像文本检测半自动标注方法、系统、计算机设备及存储介质,属于人工智能技术领域、OCR领域。

背景技术

[0002] 随着人工智能技术的发展,文本检测技术作为一项基础的计算机视觉任务得到了较大发展。文本检测指的是定位一张图像中的文本区域,该技术可广泛应用于无人驾驶、机器人导航、盲人辅助等行业中。随着深度学习等大数据驱动的机器学习算法在包括自然语言处理,计算机视觉,语音识别等领域内取得巨大成功,基于深度学习的图像文本检测技术也得到较大发展,图像文本检测算法的性能显著提升。然而,这些方法依赖大量的检测标注数据。

[0003] 到目前为止,检测标注数据的获得基本依赖人工,耗时耗力且昂贵。尤其对于不规则的文本区域位置,通常需要比较多的点数才能标注,效率极低,且因人为主观性导致精度较低。为此,需要半自动或者自动标注算法代替人工标注方法,提高标注效率和精准度。自动算法到目前为止都是采用检测算法完成,即所谓的预标注。由于检测算法性能受限,用检测算法预标注的自动算法尚不能产生高质量标注,仍需较多人工校验才能得到真正可用的标注,没有实质性解决图像文本检测标注难题。

发明内容

[0004] 有鉴于此,本发明提供了一种图像文本检测半自动标注方法、系统、计算机设备及存储介质,其能够提高文本检测标注效率和标注效果。

[0005] 本发明的第一个目的在于提供一种图像文本检测半自动标注方法。

[0006] 本发明的第二个目的在于提供一种图像文本检测半自动标注系统。

[0007] 本发明的第三个目的在于提供一种计算机设备。

[0008] 本发明的第四个目的在于提供一种存储介质。

[0009] 本发明的第一个目的可以通过采取如下技术方案达到:

[0010] 一种图像文本检测半自动标注方法,所述方法包括:

[0011] 获取文本图像;

[0012] 从文本图像中获取文本中心线,所述文本中心线是一条贯穿文本中心的弯曲折线,由依次连接 $K+1$ 个点构成 K 条直线段构成;

[0013] 生成 N 个围绕文本中心线的候选边界框,每个候选边界框是一个多边形区域轮廓,围成 N 个候选文本区域;

[0014] 将 N 个候选文本区域同时输入宽松识别器和严苛识别器,通过宽松识别器将从 N 个候选文本区域中识别得到估计的文本内容,通过严苛识别器预测每个候选文本区域的内容识别结果;

[0015] 将 N 个内容识别结果与估计的文本内容相比,分别计算识别损失,得到 N 个识别损

失；

[0016] 通过确定所有识别损失中最小损失的索引,获得最准确候选边界框的索引,进而得到最终的文本框标注；

[0017] 将文本框标注以识别损失为引导进行优化,最终得到紧致的文本框标注。

[0018] 进一步的,所述生成N个围绕文本中心线的候选边界框,具体包括:

[0019] 确定K+1条法线,每条法线 n_i 与文本中心线相交于点 c_i ,同时与文本中心线在点 c_i 的切线垂直；

[0020] 在每条法线 n_i 上,确定一条长度为 h_j 的线段 l_i ,线段 l_i 被点 c_i 二等分,将所有线段的端点视为多边形的顶点,依次连接所有顶点获得一个候选边界框 B_j ；

[0021] 通过确定N个不同的变量 h_j ,以获得N个候选边界框。

[0022] 进一步的,所述变量 h_j 的确定如下式:

$$[0023] \quad h_j = \frac{9j^2}{4N^2}$$

[0024] 其中, $j=1,2,\dots,N$ 。

[0025] 进一步的,所述通过宽松识别器将从N个候选文本区域中识别得到估计的文本内容,具体包括:

[0026] 通过宽松识别器对N个候选文本区域进行识别,得到N个识别结果 $\{T_j | j=1, 2, \dots, N\}$,识别结果 $T_j = (t_j^{u,v}) \in [0,1]^{L \times C}$ 是一个形状为 $L \times C$ 的矩阵；

[0027] 计算相邻识别结果 T_j 和 T_{j-1} 的差异 d_j ,从差异 d_j 最小的识别结果 T_{j^*} 中获得估计文本内容 T^* ,其中 $j^* = \underset{j}{\operatorname{argmin}} d_j$ ；

[0028] 所述从差异 d_j 最小的识别结果 T_{j^*} 中获得估计文本内容 T^* ,如下式:

$$[0029] \quad t_j^{u,v} = \begin{cases} 1, & \text{if } t_j^{u,v} = \max_v t_j^{u,v} \\ 0, & \text{others} \end{cases}$$

$$[0030] \quad T^* = (t_j^{u,v}) \in \{0,1\}^{L \times C}$$

[0031] 其中,分量 $t_j^{u,v}$ 表示识别结果 T_j 第u个字符属于第v类的概率。

[0032] 进一步的,所述宽松识别器的结构是一个基于卷积神经网络的图像文本识别器,包括矫正器、第一编码器、第一序列模型和第一解码器；

[0033] 所述矫正器,用于矫正输入图像区域 R_j 的文本图像形状；

[0034] 所述第一编码器,用于对矫正后的文本图像进行特征提取；

[0035] 所述第一序列模型,用于提取上下文依赖特征；

[0036] 所述第一解码器,用于对上下文依赖特征进行转译输出识别结果 T_j ；

[0037] 所述宽松识别器使用合成图像的宽松文本区域进行训练,所述宽松文本区域是区域图像内除文本外,引入适量背景干扰的区域。

[0038] 进一步的,严苛识别器的结构是一个基于卷积神经网络的图像文本识别器,包括第二编码器、第二序列模型和第二解码器；

[0039] 所述第二编码器,用于对输入图像区域 R_j 进行特征提取；

[0040] 所述第二序列模型,用于提取上下文依赖特征;

[0041] 所述第二解码器,用于对上下文依赖特征进行解码输出识别结果 s_j ;

[0042] 所述严苛识别器使用合成图像的紧致文本区域进行训练,所述紧致文本区域是区域图像内除文本外,不存在背景干扰的图像区域。

[0043] 进一步的,所述将文本框标注以识别损失为引导进行优化,如下式:

$$[0044] \quad B_{j^*}' = B_{j^*} - \mu \times \nabla_{B_{j^*}} l_{j^*}$$

[0045] 其中, $\nabla_{B_{j^*}} l_{j^*}$ 表示识别损失 l_{j^*} 对 B_{j^*} 的梯度, μ 则表示更新步长。

[0046] 本发明的第二个目的可以通过采取如下技术方案达到:

[0047] 一种图像文本检测半自动标注系统,所述系统包括:

[0048] 文本图像获取模块,用于获取文本图像;

[0049] 文本中心线获取模块,用于从文本图像中获取文本中心线,所述文本中心线是一条贯穿文本中心的弯曲折线,由依次连接 $K+1$ 个点构成 K 条直线段构成;

[0050] 候选边界框生成模块,用于生成 N 个围绕文本中心线的候选边界框,每个候选边界框是一个多边形区域轮廓,围成 N 个候选文本区域;

[0051] 识别模块,用于将 N 个候选文本区域同时输入宽松识别器和严苛识别器,通过宽松识别器将从 N 个候选文本区域中识别得到估计的文本内容,通过严苛识别器预测每个候选文本区域的内容识别结果;

[0052] 识别损失计算模块,用于将 N 个内容识别结果与估计的文本内容相比,分别计算识别损失,得到 N 个识别损失;

[0053] 文本框标注获取模块,用于通过确定所有识别损失中最小损失的索引,获得最准确候选边界框的索引,进而得到最终的文本框标注;

[0054] 优化模块,用于将文本框标注以识别损失为引导进行优化,最终得到紧致的文本框标注。

[0055] 本发明的第三个目的可以通过采取如下技术方案达到:

[0056] 一种计算机设备,包括处理器以及用于存储处理器可执行程序的存储器,所述处理器执行存储器存储的程序时,实现上述的图像文本检测半自动标注方法。

[0057] 本发明的第四个目的可以通过采取如下技术方案达到:

[0058] 一种存储介质,存储有程序,所述程序被处理器执行时,实现上述的图像文本检测半自动标注方法。

[0059] 本发明相对于现有技术具有如下的有益效果:

[0060] 本发明通过从文本图像中获取文本中心线,生成围绕文本中心线的候选边界框,然后将候选边界框输入宽松识别器和严苛识别器,通过宽松识别器识别得到估计的文本内容,通过严苛识别器预测内容识别结果,进而计算识别损失,并通过确定所有识别损失中最小损失的索引,获得最准确候选边界框的索引,进而得到最终的文本框标注,将文本框标注以识别损失为引导进行优化,得到紧致的文本框标注,实现了半自动标注,半自动标注通过介于人工标注和自动标注算法之间,能兼顾标注效率和标注效果。

附图说明

[0061] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例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,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图示出的结构获得其他的附图。

[0062] 图1为本发明实施例1的图像文本检测半自动标注方法的简易流程图。

[0063] 图2为本发明实施例1的图像文本检测半自动标注方法的具体流程图。

[0064] 图3为本发明实施例1的图像文本检测半自动标注方法中候选边界生成的流程示意图。

[0065] 图4为本发明实施例1的图像文本检测半自动标注方法中紧致边界估计的流程示意图。

[0066] 图5为本发明实施例2的图像文本检测半自动标注系统的结构框图。

[0067] 图6为本发明实施例3的计算机设备的结构框图。

具体实施方式

[0068] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明的一部分实施例,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作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。

[0069] 实施例1:

[0070] 如图1和图2所示,本实施例提供了一种图像文本检测半自动标注方法,该方法包括以下步骤:

[0071] S201、获取文本图像。

[0072] 本实施例的文本图像为场景文本图像,可以通过采集获取,例如通过摄像头拍摄场景文本图像,也可以从数据库查找获取,例如预先在数据库内存储场景文本图像,从数据库中搜索场景文本图像即可得到,以获取的文本图像为输入。

[0073] S202、从文本图像中获取文本中心线。

[0074] 本实施例中,文本中心线是一条贯穿文本中心的弯折线,由依次连接 $K+1$ 个点构成 K 条直线段构成, $K+1$ 个点记为 $\{c_i | i=1,2,\dots,K+1\}$,以文本中心线为输入。

[0075] S203、生成 N 个围绕文本中心线的候选边界框,每个候选边界框是一个多边形区域轮廓,围成 N 个候选文本区域。

[0076] 本实施例中, N 个候选边界框记为 $\{B_j | j=1,2,\dots,N\}$, N 个候选文本区域记为 $\{R_j | j=1,2,\dots,N\}$, $N=30$ 。

[0077] 结合图3,该步骤S203为候选边界框步骤,具体包括:

[0078] S2031、确定 $K+1$ 条法线,每条法线 n_i 与文本中心线相交于点 c_i ,同时与文本中心线在点 c_i 的切线垂直。

[0079] S2032、在每条法线 n_i 上,确定一条长度为 h_j 的线段 l_i ,线段 l_i 被点 c_i 二等分,将所有线段的端点视为多边形的顶点,依次连接所有顶点获得一个候选边界框 B_j 。

[0080] S2033、通过确定 N 个不同的变量 h_j ,以获得 N 个候选边界框。

[0081] 本实施例中,不同的变量 h_j 可以确定不同的多边形边界框,候选 N 个不同的 h_j 可以获得 N 个候选边界框,变量 h_j 的确定如下式:

$$[0082] \quad h_j = \frac{9j^2}{4N^2}$$

[0083] 其中, $j=1,2,\dots,N$ 。

[0084] 接下来的步骤S204~S206为语义边界决策步骤:

[0085] S204、将 N 个候选文本区域同时输入宽松识别器和严苛识别器,通过宽松识别器将从 N 个候选文本区域中识别得到估计的文本内容,通过严苛识别器预测每个候选文本区域的内容识别结果。

[0086] 本实施例中,通过宽松识别器将从 N 个候选文本区域中识别得到估计的文本内容,具体包括:

[0087] 1) 通过宽松识别器对 N 个候选文本区域 $\{R_j | j=1,2,\dots,N\}$ 进行识别,得到 N 个识别结果 $\{T_j | j=1,2,\dots,N\}$,识别结果 $T_j = (t_j^{u,v}) \in [0,1]^{L \times C}$ 是一个形状为 $L \times C$ 的矩阵。

[0088] 2) 计算相邻识别结果 T_j 和 T_{j-1} 的差异 d_j ,从差异 d_j 最小的识别结果 T_{j^*} 中获得估计文本内容 T^* ,其中 $j^* = \underset{j}{\operatorname{argmin}} d_j$ 。

[0089] 计算相邻识别结果 T_j 和 T_{j-1} 的差异 d_j 包括但不限于交叉熵损失函数、CTC (Connectionist Temporal Classification) 损失函数,以及编辑距离,从差异 d_j 最小的识别结果 T_{j^*} 中获得估计文本内容 T^* ,如下式:

$$[0090] \quad t_j^{u,v} = \begin{cases} 1, & \text{if } t_j^{u,v} = \max_v t_j^{u,v} \\ 0, & \text{others} \end{cases}$$

$$[0091] \quad T^* = (t_j^{u,v}) \in \{0,1\}^{L \times C}$$

[0092] 其中,分量 $t_j^{u,v}$ 表示识别结果 T_j 第 u 个字符属于第 v 类的概率。

[0093] 本实施例的宽松识别器的结构是一个基于卷积神经网络的图像文本识别器,包括矫正器、第一编码器、第一序列模型和第一解码器,各个部分的说明如下:

[0094] 矫正器,用于矫正输入图像区域 R_j 的文本图像形状;

[0095] 第一编码器,用于对矫正后的文本图像进行特征提取;

[0096] 第一序列模型,用于提取上下文依赖特征;

[0097] 第一解码器,用于对上下文依赖特征进行转译输出识别结果 T_j ;

[0098] 宽松识别器使用合成图像的宽松文本区域进行训练,宽松文本区域是区域图像内除文本外,引入适量背景干扰的区域。

[0099] 本实施例的严苛识别器的结构是一个基于卷积神经网络的图像文本识别器,包括第二编码器、第二序列模型和第二解码器,各个部分的说明如下:

[0100] 第二编码器,用于对输入图像区域 R_j 进行特征提取;

[0101] 第二序列模型,用于提取上下文依赖特征;

[0102] 第二解码器,用于对上下文依赖特征进行解码输出识别结果 s_j ;

[0103] 严苛识别器使用合成图像的紧致文本区域进行训练,紧致文本区域是区域图像内除文本外,不存在背景干扰的图像区域。

[0104] S205、将N个内容识别结果与估计的文本内容相比,分别计算识别损失,得到N个识别损失。

[0105] 本实施例中,识别损失记为 $\{l_j | j=1,2,\dots,N\}$,计算识别损失包括但不限于交叉熵损失函数、CTC(Connectionist Temporal Classification)损失函数,以及编辑距离。

[0106] S206、通过确定所有识别损失中最小损失的索引,获得最准确候选边界框的索引,进而得到最终的文本框标注。

[0107] S207、将文本框标注以识别损失为引导进行优化,最终得到紧致的文本框标注。

[0108] 结合图4,该步骤S107为紧致边界估计步骤,将文本框标注以识别损失为引导进行优化,如下式:

$$[0109] \quad B'_{j^*} = B_{j^*} - \mu \times \nabla_{B_{j^*}} l_{j^*}$$

[0110] 其中, $\nabla_{B_{j^*}} l_{j^*}$ 表示识别损失 l_{j^*} 对 B_{j^*} 的梯度, μ 则表示更新步长。

[0111] 上述实施例中,由于严苛识别器使用合成图像的紧致文本区域进行训练,对文本图像背景区域敏感,因而得到的文本检测框紧致且准确率较高。

[0112] 应当注意,尽管在附图中以特定顺序描述了上述实施例的方法操作,但是这并非要求或者暗示必须按照该特定顺序来执行这些操作,或是必须执行全部所示的操作才能实现期望的结果。相反,描绘的步骤可以改变执行顺序。附加地或备选地,可以省略某些步骤,将多个步骤合并为一个步骤执行,和/或将一个步骤分解为多个步骤执行。

[0113] 实施例2:

[0114] 如图5所示,本实施例提供了一种图像文本检测半自动标注系统,该系统包括文本图像获取模块501、文本中心线获取模块502、候选边界框生成模块503、识别模块504、识别损失计算模块505、文本框标注获取模块506和优化模块507,各个模块的具体功能如下:

[0115] 文本图像获取模块501,用于获取文本图像。

[0116] 文本中心线获取模块502,用于从文本图像中获取文本中心线,所述文本中心线是一条贯穿文本中心的弯曲折线,由依次连接K+1个点构成K条直线段构成。

[0117] 候选边界框生成模块503,用于生成N个围绕文本中心线的候选边界框,每个候选边界框是一个多边形区域轮廓,围成N个候选文本区域。

[0118] 识别模块504,用于将N个候选文本区域同时输入宽松识别器和严苛识别器,通过宽松识别器将从N个候选文本区域中识别得到估计的文本内容,通过严苛识别器预测每个候选文本区域的内容识别结果。

[0119] 识别损失计算模块505,用于将N个内容识别结果与估计的文本内容相比,分别计算识别损失,得到N个识别损失。

[0120] 文本框标注获取模块506,用于通过确定所有识别损失中最小损失的索引,获得最准确候选边界框的索引,进而得到最终的文本框标注。

[0121] 优化模块507,用于将文本框标注以识别损失为引导进行优化,最终得到紧致的文本框标注。

[0122] 本实施例中各个模块的具体实现可以参见上述实施例1,在此不再一一赘述;需要

说明的是,本实施例提供的系统仅以上述各功能单元的划分进行举例说明,在实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即将内部结构划分成不同的功能模块,以完成以上描述的全部或者部分功能。

[0123] 实施例3:

[0124] 本实施例提供了一种计算机设备,该计算机设备可以是计算机,如图6所示,其包括通过系统总线601、连接的处理器602、存储器、输入装置603、显示装置604和网络接口605,该处理器用于提供计算和控制能力,该存储器包括非易失性存储介质606和内存存储607,该非易失性存储介质606存储有操作系统、计算机程序和数据库,该内存存储607为非易失性存储介质中的操作系统和计算机程序的运行提供环境,处理器602执行存储器存储的计算机程序时,实现上述实施例1的图像文本检测半自动标注方法,如下:

[0125] 获取文本图像;

[0126] 从文本图像中获取文本中心线,所述文本中心线是一条贯穿文本中心的弯曲折线,由依次连接 $K+1$ 个点构成 K 条直线段构成;

[0127] 生成 N 个围绕文本中心线的候选边界框,每个候选边界框是一个多边形区域轮廓,围成 N 个候选文本区域;

[0128] 将 N 个候选文本区域同时输入宽松识别器和严苛识别器,通过宽松识别器将从 N 个候选文本区域中识别得到估计的文本内容,通过严苛识别器预测每个候选文本区域的内容识别结果;

[0129] 将 N 个内容识别结果与估计的文本内容相比,分别计算识别损失,得到 N 个识别损失;

[0130] 通过确定所有识别损失中最小损失的索引,获得最准确候选边界框的索引,进而得到最终的文本框标注;

[0131] 将文本框标注以识别损失为引导进行优化,最终得到紧致的文本框标注。

[0132] 实施例4:

[0133] 本实施例提供了一种存储介质,该存储介质为计算机可读存储介质,其存储有计算机程序,所述计算机程序被处理器执行时,实现上述实施例1的图像文本检测半自动标注方法,如下:

[0134] 获取文本图像;

[0135] 从文本图像中获取文本中心线,所述文本中心线是一条贯穿文本中心的弯曲折线,由依次连接 $K+1$ 个点构成 K 条直线段构成;

[0136] 生成 N 个围绕文本中心线的候选边界框,每个候选边界框是一个多边形区域轮廓,围成 N 个候选文本区域;

[0137] 将 N 个候选文本区域同时输入宽松识别器和严苛识别器,通过宽松识别器将从 N 个候选文本区域中识别得到估计的文本内容,通过严苛识别器预测每个候选文本区域的内容识别结果;

[0138] 将 N 个内容识别结果与估计的文本内容相比,分别计算识别损失,得到 N 个识别损失;

[0139] 通过确定所有识别损失中最小损失的索引,获得最准确候选边界框的索引,进而得到最终的文本框标注;

[0140] 将文本框标注以识别损失为引导进行优化,最终得到紧致的文本框标注。

[0141] 需要说明的是,本实施例的计算机可读存储介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。

[0142] 综上所述,本发明通过从文本图像中获取文本中心线,生成围绕文本中心线的候选边界框,然后将候选边界框输入宽松识别器和严苛识别器,通过宽松识别器识别得到估计的文本内容,通过严苛识别器预测内容识别结果,进而计算识别损失,并通过确定所有识别损失中最小损失的索引,获得最准确候选边界框的索引,进而得到最终的文本框标注,将文本框标注以识别损失为引导进行优化,得到紧致的文本框标注,实现了半自动标注,半自动标注通过介于人工标注和自动标注算法之间,能兼顾标注效率和标注效果。

[0143] 以上所述,仅为本发明专利较佳的实施例,但本发明专利的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明专利所公开的范围内,根据本发明专利的技术方案及其发明构思加以等同替换或改变,都属于本发明专利的保护范围。

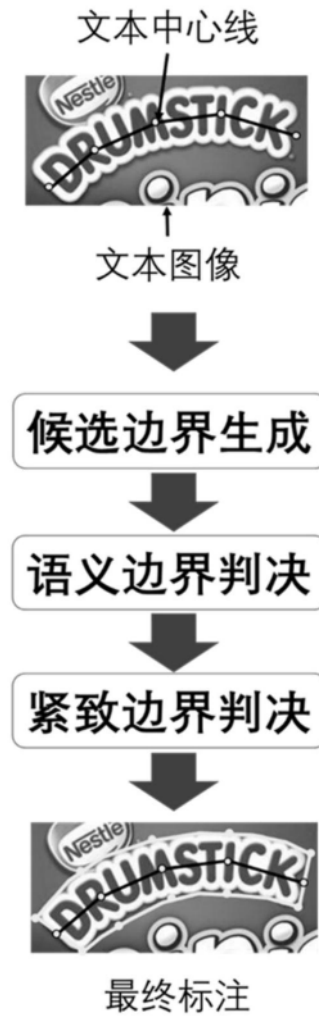


图1

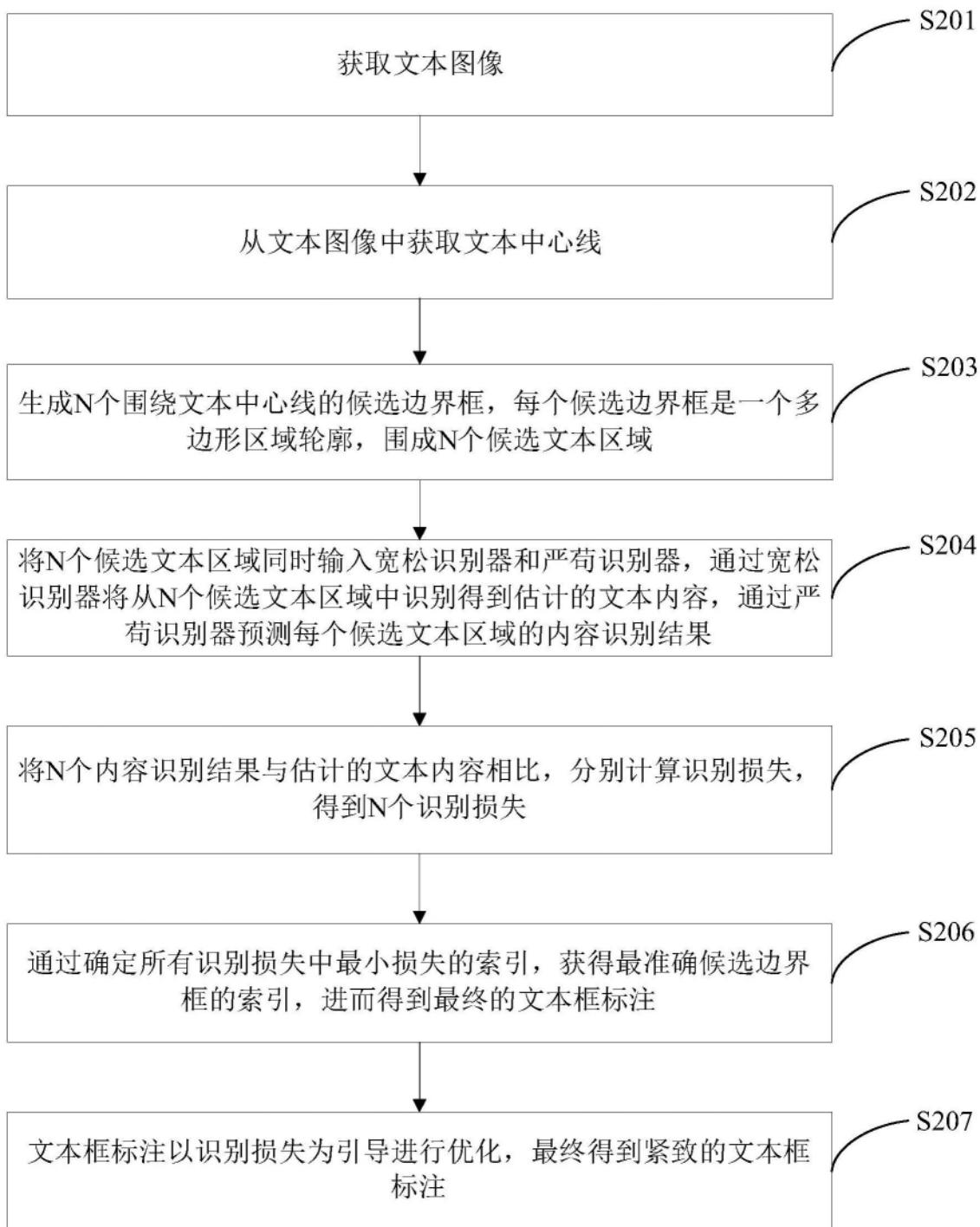


图2

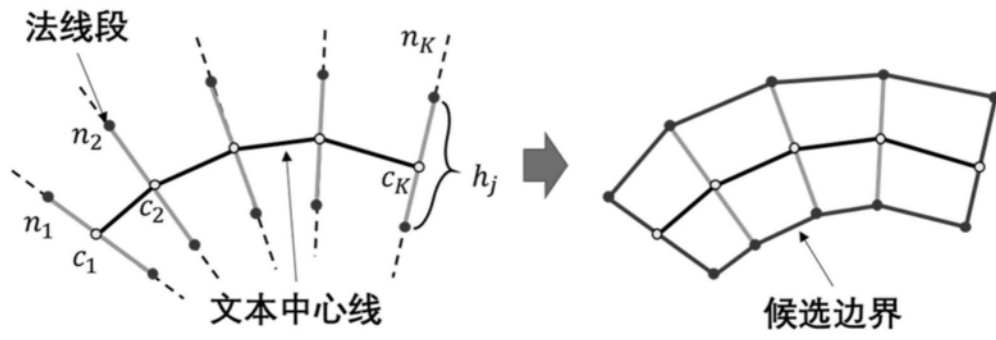


图3

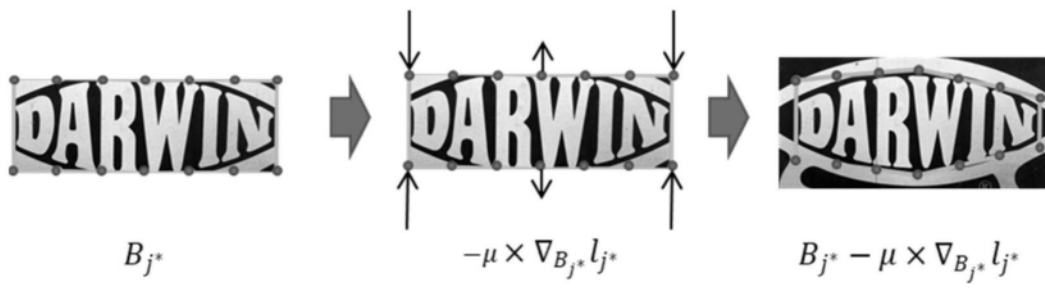


图4

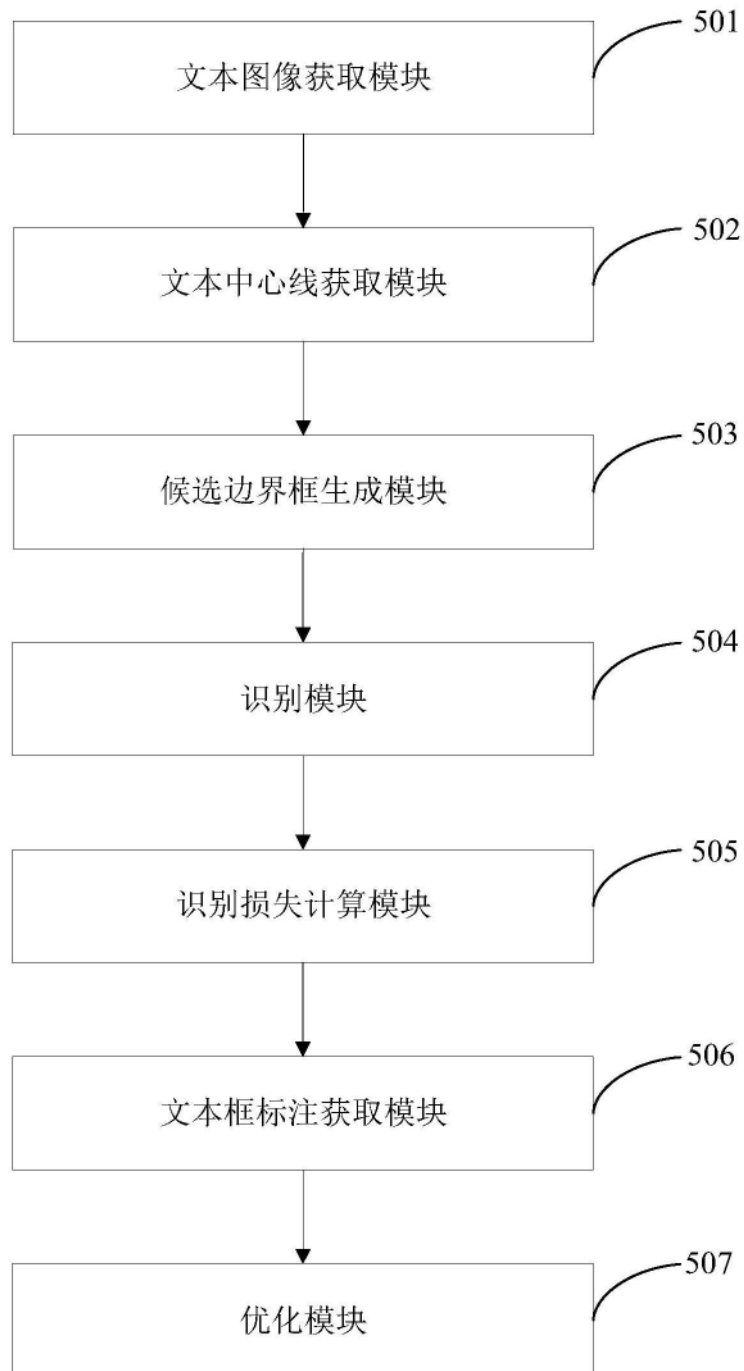


图5

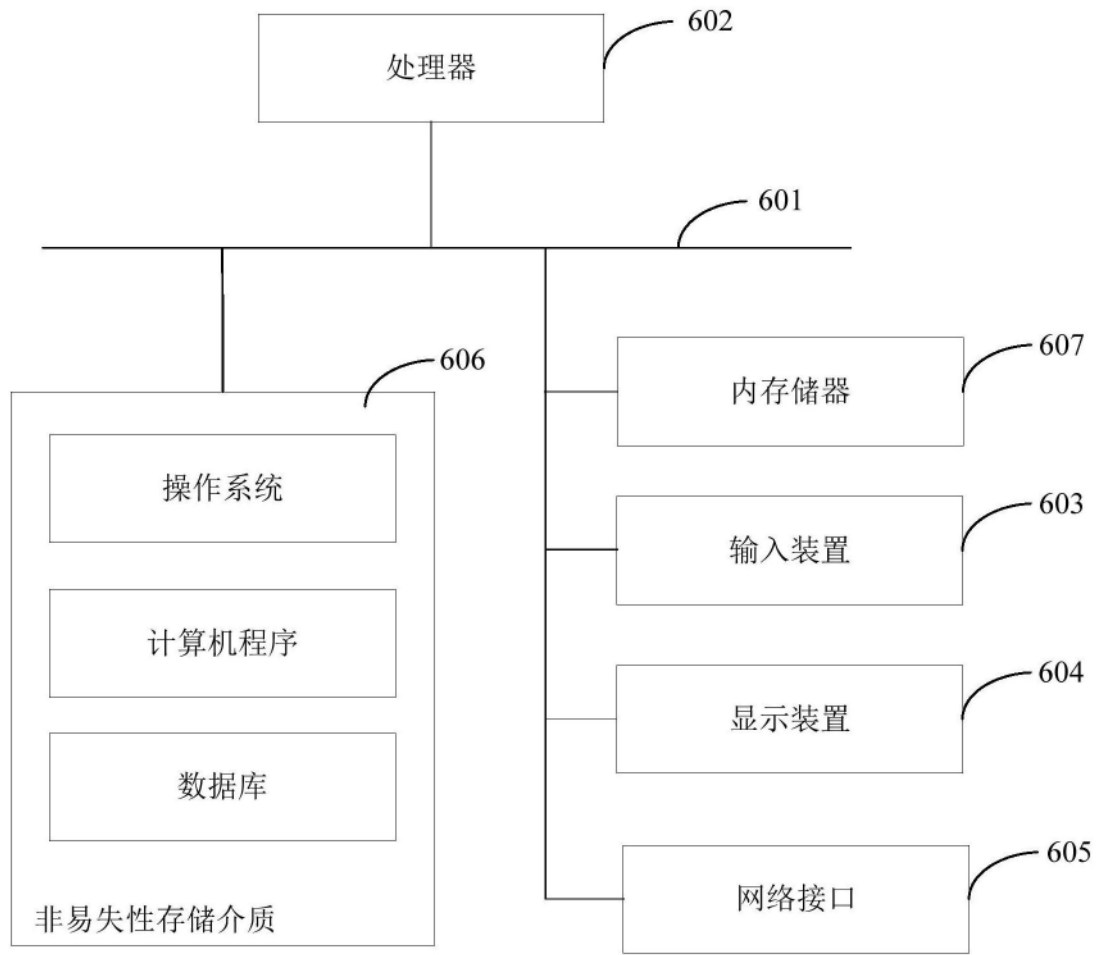


图6