



(12)发明专利

(10)授权公告号 CN 108304835 B

(45)授权公告日 2019.12.06

(21)申请号 201810090633.4

G06K 9/34(2006.01)

(22)申请日 2018.01.30

G06K 9/62(2006.01)

(65)同一申请的已公布的文献号

G06N 3/08(2006.01)

申请公布号 CN 108304835 A

审查员 崔芳婷

(43)申请公布日 2018.07.20

(73)专利权人 百度在线网络技术(北京)有限公司

地址 100085 北京市海淀区上地十街10号  
百度大厦三层

(72)发明人 刘家铭 章成全 韩钧宇 丁二锐

(74)专利代理机构 北京英赛嘉华知识产权代理  
有限责任公司 11204

代理人 王达佐 马晓亚

(51)Int.Cl.

G06K 9/20(2006.01)

权利要求书4页 说明书17页 附图7页

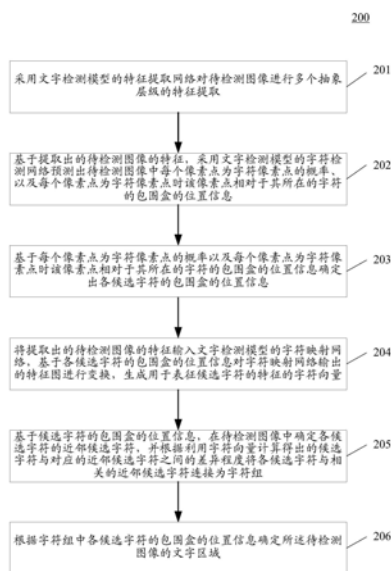
(54)发明名称

文字检测方法和装置

(57)摘要

本申请实施例公开了文字检测方法和装置。该方法的一具体实施方式包括:采用特征提取网络对待检测图像进行多个抽象层级的特征提取;采用字符检测网络预测出待检测图像中每个像素点为字符像素点的概率、以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息;基于字符检测网络的预测结果确定出各候选字符的包围盒的位置信息;将提取出特征输入字符映射网络,对字符映射网络输出的特征图进行变换,生成字符向量;在待检测图像中确定各候选字符的近邻候选字符,并将各候选字符与相关的近邻候选字符连接为字符组;根据字符组中各候选字符的包围盒的位置信息确定待检测图像的文字区域。该实施方式提升了不规则文字检测的准确度。

CN 108304835 B



1. 一种文字检测方法,包括:

采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取;

基于提取出的所述待检测图像的特征,采用文字检测模型的字符检测网络预测出所述待检测图像中每个像素点为字符像素点的概率、以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息;

基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息;

将提取出的所述待检测图像的特征输入文字检测模型的字符映射网络,基于各候选字符的包围盒的位置信息对所述字符映射网络输出的特征图进行变换,生成用于表征所述候选字符的特征的字符向量;

基于所述候选字符的包围盒的位置信息,在所述待检测图像中确定各候选字符的近邻候选字符,并根据利用字符向量计算得出的所述候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组;

根据字符组中各候选字符的包围盒的位置信息确定所述待检测图像的文字区域;

其中,所述基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息,包括:采用非极大值抑制法滤除各字符像素点定位出的字符的包围盒的坐标中被重复定位的字符的包围盒的坐标,得到各候选字符的包围盒的坐标。

2. 根据权利要求1所述的方法,其中,所述采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取,包括:

将所述待检测图像输入所述特征提取网络中,抽取所述特征提取网络的多个不同卷积层的输出作为多个抽象层级的特征;

拼接多个抽象层级的特征,或利用特征金字塔网络对多个抽象层次的特征进行处理,生成所述待检测图像的特征。

3. 根据权利要求1所述的方法,其中,所述每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息,包括:每个像素点为字符像素点时该像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点坐标的偏移量。

4. 根据权利要求3所述的方法,其中,在采用非极大值抑制法滤除各字符像素点定位出的字符的包围盒的坐标中被重复定位的字符的包围盒的坐标,得到各候选字符的包围盒的坐标之前,所述基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息,还包括:

确定概率高于预设概率阈值的像素点为字符像素点;

基于确定出的各字符像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点的坐标的偏移量,确定由所述字符像素点定位出的字符的包围盒的坐标。

5. 根据权利要求1所述的方法,其中,所述基于所述候选字符的包围盒的位置信息,在所述待检测图像中确定各候选字符的近邻候选字符,包括:

采用K近邻算法,基于各候选字符的包围盒的位置信息对候选字符进行分类,根据分类结果确定各候选字符的近邻候选字符;以及

所述根据利用字符向量计算得出的所述候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组,包括:

计算所述候选字符的字符向量与对应的各近邻候选字符的字符向量之间的欧式距离,作为所述候选字符与对应的各近邻候选字符之间的差异程度;

将与所述候选字符之间的差异程度小于预设差异度阈值的近邻候选字符作为与所述候选字符相关的近邻候选字符,将各候选字符与相关的近邻候选字符连接为字符组。

6. 根据权利要求1所述的方法,其中,所述根据字符组中各候选字符的包围盒的位置信息确定所述待检测图像的文字区域,包括:

基于所述字符组中各候选字符的包围盒的位置信息划定环绕所述字符组中所有字符的包围线作为所述待检测图像的文字区域。

7. 根据权利要求1-6任一项所述的方法,其中,所述方法还包括:

采用机器学习方法,基于样本图像训练所述文字检测模型。

8. 根据权利要求7所述的方法,其中,所述采用机器学习方法,基于样本图像训练所述文字检测模型,包括:

获取已采用包围盒方式标记所包含的字符的样本图像;

将样本图像输入所述文字检测模型,以对样本图像中的文字区域进行预测,得到所述样本图像中的像素点是否为字符像素点的预测结果、所述样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果以及对所述样本图像中的字符组的预测结果;

计算预设损失函数的值,计算文字检测模型中每个参数相对于所述预设损失函数的梯度,利用反向传播算法更新模型参数,直到所述预设损失函数的值满足预设的收敛条件;

所述预设损失函数包括分类损失函数、包围盒回归损失函数以及字符连接损失函数;

其中,所述分类损失函数的值用于表征所述字符检测网络对所述样本图像中的像素点是否为字符像素点的预测结果与所述样本图像中的像素点是否为字符像素点的标记结果之间的差异;

所述包围盒回归损失函数的值用于表征所述字符检测网络对所述样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果与所述样本图像中的对应字符像素点所在字符的包围盒的位置信息的标记结果之间的差异;

所述字符连接损失函数的值用于表征所述文字检测模型对样本图像中的字符组的预测结果与预测出的字符组中的候选字符的字符向量之间的差异度的一致性程度。

9. 一种文字检测装置,包括:

提取单元,用于采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取;

检测单元,用于基于提取出的所述待检测图像的特征,采用文字检测模型的字符检测网络预测出所述待检测图像中每个像素点为字符像素点的概率、以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息;

定位单元,基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息;

映射单元,用于将提取出的所述待检测图像的特征输入文字检测模型的字符映射网络,基于各候选字符的包围盒的位置信息对所述字符映射网络输出的特征图进行变换,生

成用于表征所述候选字符的特征的字符向量；

连接单元，用于基于所述候选字符的包围盒的位置信息，在所述待检测图像中确定各候选字符的近邻候选字符，并根据利用字符向量计算得出的所述候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组；

确定单元，用于根据字符组中各候选字符的包围盒的位置信息确定所述待检测图像的文字区域；

其中，所述定位单元用于按照如下方式确定出各候选字符的包围盒的位置信息：采用非极大值抑制法滤除各字符像素点定位出的字符的包围盒的坐标中被重复定位的字符的包围盒的坐标，得到各候选字符的包围盒的坐标。

10. 根据权利要求9所述的装置，其中，所述提取单元用于采用文字检测模型的特征提取网络，按照如下方式对待检测图像进行多个抽象层级的特征提取：

将所述待检测图像输入所述特征提取网络中，抽取所述特征提取网络的多个不同卷积层的输出作为多个抽象层级的特征；

拼接多个抽象层级的特征，或利用特征金字塔网络对多个抽象层次的特征进行处理，生成所述待检测图像的特征。

11. 根据权利要求9所述的装置，其中，所述每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息，包括：每个像素点为字符像素点时该像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点坐标的偏移量。

12. 根据权利要求11所述的装置，其中，所述定位单元还用于：

确定概率高于预设概率阈值的像素点为字符像素点；

基于确定出的各字符像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点的坐标的偏移量，确定由所述字符像素点定位出的字符的包围盒的坐标。

13. 根据权利要求9所述的装置，其中，所述连接单元用于基于所述候选字符的包围盒的位置信息，按照如下方式在所述待检测图像中确定各候选字符的近邻候选字符：

采用K近邻算法，基于各候选字符的包围盒的位置信息对候选字符进行分类，根据分类结果确定各候选字符的近邻候选字符；以及

所述连接单元用于按照如下方式将各候选字符与相关的近邻候选字符连接为字符组：

计算所述候选字符的字符向量与对应的各近邻候选字符的字符向量之间的欧式距离，作为所述候选字符与对应的各近邻候选字符之间的差异程度；

将与所述候选字符之间的差异程度小于预设差异度阈值的近邻候选字符作为与所述候选字符相关的近邻候选字符，将各候选字符与相关的近邻候选字符连接为字符组。

14. 根据权利要求9所述的装置，其中，所述确定单元用于按照如下方式确定所述待检测图像的文字区域：

基于所述字符组中各候选字符的包围盒的位置信息划定环绕所述字符组中所有字符的包围线作为所述待检测图像的文字区域。

15. 根据权利要求9-14任一项所述的装置，其中，所述装置还包括：

训练单元，用于采用机器学习方法，基于样本图像训练所述文字检测模型。

16. 根据权利要求15所述的装置，其中，所述训练单元进一步用于按照如下方式训练所述文字检测模型：

获取已采用包围盒方式标记所包含的字符的样本图像；

将样本图像输入所述文字检测模型，以对样本图像中的文字区域进行预测，得到所述样本图像中的像素点是否为字符像素点的预测结果、所述样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果以及对所述样本图像中的字符组的预测结果；

计算预设损失函数的值，计算文字检测模型中每个参数相对于所述预设损失函数的梯度，利用反向传播算法更新模型参数，直到所述预设损失函数的值满足预设的收敛条件；

所述预设损失函数包括分类损失函数、包围盒回归损失函数以及字符连接损失函数；

其中，所述分类损失函数的值用于表征所述字符检测网络对所述样本图像中的像素点是否为字符像素点的预测结果与所述样本图像中的像素点是否为字符像素点的标记结果之间的差异；

所述包围盒回归损失函数的值用于表征所述字符检测网络对所述样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果与所述样本图像中的对应字符像素点所在字符的包围盒的位置信息的标记结果之间的差异；

所述字符连接损失函数的值用于表征所述文字检测模型对样本图像中的字符组的预测结果与预测出的字符组中的候选字符的字符向量之间的差异度的一致性程度。

17. 一种服务器，包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求1-8中任一所述的方法。

18. 一种计算机可读存储介质，其上存储有计算机程序，其中，所述程序被处理器执行时实现如权利要求1-8中任一所述的方法。

## 文字检测方法和装置

### 技术领域

[0001] 本申请实施例涉及计算机技术领域,具体涉及图像处理技术领域,尤其涉及文字检测方法和装置。

### 背景技术

[0002] 随着计算机视觉技术的发展,基于图像的文字检测和文字识别技术得到了越来越广泛的应用,例如在文档处理,地理位置读取、图像检索等方面,文字检测和文字识别具有很多实际的应用。

[0003] 在光学字符识别(Optical Character Recognition,OCR)领域中,文字边框通常被定义为矩形的检测框,文字之间的连接通常呈直线连接。然而实际采集数据中存在大量扭曲、变形等不规则的文字组合,例如由于图像畸变或拍摄角度造成的图像中文字发生扭曲、或在一些诸如商标、徽章等的设计中会采用一些变形的字符。

[0004] 现有的文字检测算法中要求图像中的基本组件之间的连接呈直线,或者需要像素级的标注精度,不易于应用在更广泛的场景中。

### 发明内容

[0005] 本申请实施例提出了文字检测方法和装置。

[0006] 第一方面,本申请实施例提供了一种文字检测方法,包括:采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取;基于提取出的待检测图像的特征,采用文字检测模型的字符检测网络预测出待检测图像中每个像素点为字符像素点的概率、以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息;基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息;将提取出的待检测图像的特征输入文字检测模型的字符映射网络,基于各候选字符的包围盒的位置信息对字符映射网络输出的特征图进行变换,生成用于表征候选字符的特征的字符向量;基于候选字符的包围盒的位置信息,在待检测图像中确定各候选字符的近邻候选字符,并根据利用字符向量计算得出的候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组;根据字符组中各候选字符的包围盒的位置信息确定待检测图像的文字区域。

[0007] 在一些实施例中,上述采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取,包括:将待检测图像输入特征提取网络中,抽取特征提取网络的多个不同卷积层的输出作为多个抽象层级的特征;拼接多个抽象层级的特征,或利用特征金字塔网络对多个抽象层次的特征进行处理,生成待检测图像的特征。

[0008] 在一些实施例中,上述每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息,包括:每个像素点为字符像素点时该像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点坐标的偏移量。

[0009] 在一些实施例中,上述基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息,包括:确定概率高于预设概率阈值的像素点为字符像素点;基于确定出的各字符像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点的坐标的偏移量,确定由字符像素点定位出的字符的包围盒的坐标;采用非极大值抑制法滤除各字符像素点定位出的字符的包围盒的坐标中被重复定位的字符的包围盒的坐标,得到各候选字符的包围盒的坐标。

[0010] 在一些实施例中,上述基于候选字符的包围盒的位置信息,在待检测图像中确定各候选字符的近邻候选字符,包括:采用K近邻算法,基于各候选字符的包围盒的位置信息对候选字符进行分类,根据分类结果确定各候选字符的近邻候选字符;以及上述根据利用字符向量计算得出的候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组,包括:计算候选字符的字符向量与对应的各近邻候选字符的字符向量之间的欧式距离,作为候选字符与对应的各近邻候选字符之间的差异程度;将与候选字符之间的差异程度小于预设差异度阈值的近邻候选字符作为与候选字符相关的近邻候选字符,将各候选字符与相关的近邻候选字符连接为字符组。

[0011] 在一些实施例中,上述根据字符组中各候选字符的包围盒的位置信息确定待检测图像的文字区域,包括:基于字符组中各候选字符的包围盒的位置信息划定环绕字符组中所有字符的包围线作为待检测图像的文字区域。

[0012] 在一些实施例中,上述方法还包括:采用机器学习方法,基于样本图像训练文字检测模型。

[0013] 在一些实施例中,上述采用机器学习方法,基于样本图像训练文字检测模型,包括:获取已采用包围盒方式标记所包含的字符的样本图像;将样本图像输入文字检测模型,以对样本图像中的文字区域进行预测,得到样本图像中的像素点是否为字符像素点的预测结果、样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果以及对样本图像中的字符组的预测结果;计算预设损失函数的值,计算文字检测模型中每个参数相对于预设损失函数的梯度,利用反向传播算法更新模型参数,直到预设损失函数的值满足预设的收敛条件。预设损失函数包括分类损失函数、包围盒回归损失函数以及字符连接损失函数;其中,分类损失函数的值用于表征字符检测网络对样本图像中的像素点是否为字符像素点的预测结果与样本图像中的像素点是否为字符像素点的标记结果之间的差异;包围盒回归损失函数的值用于表征字符检测网络对样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果与样本图像中的对应字符像素点所在字符的包围盒的位置信息的标记结果之间的差异;字符连接损失函数的值用于表征文字检测模型对样本图像中的字符组的预测结果与预测出的字符组中的候选字符的字符向量之间的差异度的一致性程度。

[0014] 第二方面,本申请实施例提供了一种文字检测装置,包括:提取单元,用于采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取;检测单元,用于基于提取出的待检测图像的特征,采用文字检测模型的字符检测网络预测出待检测图像中每个像素点为字符像素点的概率、以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息;定位单元,基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字

符的包围盒的位置信息;映射单元,用于将提取出的待检测图像的特征输入文字检测模型的字符映射网络,基于各候选字符的包围盒的位置信息对字符映射网络输出的特征图进行变换,生成用于表征候选字符的特征的字符向量;连接单元,用于基于候选字符的包围盒的位置信息,在待检测图像中确定各候选字符的近邻候选字符,并根据利用字符向量计算得出的候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组;确定单元,用于根据字符组中各候选字符的包围盒的位置信息确定待检测图像的文字区域。

[0015] 在一些实施例中,上述提取单元用于采用文字检测模型的特征提取网络,按照如下方式对待检测图像进行多个抽象层级的特征提取:将待检测图像输入特征提取网络中,抽取特征提取网络的多个不同卷积层的输出作为多个抽象层级的特征;拼接多个抽象层级的特征,或利用特征金字塔网络对多个抽象层次的特征进行处理,生成待检测图像的特征。

[0016] 在一些实施例中,上述每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息,包括:每个像素点为字符像素点时该像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点坐标的偏移量。

[0017] 在一些实施例中,上述定位单元用于基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息,按照如下方式确定出各候选字符的包围盒的位置信息:确定概率高于预设概率阈值的像素点为字符像素点;基于确定出的各字符像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点的坐标的偏移量,确定由字符像素点定位出的字符的包围盒的坐标;采用非极大值抑制法滤除各字符像素点定位出的字符的包围盒的坐标中被重复定位的字符的包围盒的坐标,得到各候选字符的包围盒的坐标。

[0018] 在一些实施例中,上述连接单元用于基于候选字符的包围盒的位置信息,按照如下方式在待检测图像中确定各候选字符的近邻候选字符:采用K近邻算法,基于各候选字符的包围盒的位置信息对候选字符进行分类,根据分类结果确定各候选字符的近邻候选字符;以及上述连接单元用于按照如下方式将各候选字符与相关的近邻候选字符连接为字符组:计算候选字符的字符向量与对应的各近邻候选字符的字符向量之间的欧式距离,作为候选字符与对应的各近邻候选字符之间的差异程度;将与候选字符之间的差异程度小于预设差异度阈值的近邻候选字符作为与候选字符相关的近邻候选字符,将各候选字符与相关的近邻候选字符连接为字符组。

[0019] 在一些实施例中,上述确定单元用于按照如下方式确定待检测图像的文字区域:基于字符组中各候选字符的包围盒的位置信息划定环绕字符组中所有字符的包围线作为待检测图像的文字区域。

[0020] 在一些实施例中,上述装置还包括:训练单元,用于采用机器学习方法,基于样本图像训练文字检测模型。

[0021] 在一些实施例中,上述训练单元进一步用于按照如下方式训练文字检测模型:获取已采用包围盒方式标记所包含的字符的样本图像;将样本图像输入文字检测模型,以对样本图像中的文字区域进行预测,得到样本图像中的像素点是否为字符像素点的预测结果、样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果以及对样本图像中的字符组的预测结果;计算预设损失函数的值,计算文字检测模型中每个参数相对于预



设损失函数的梯度,利用反向传播算法更新模型参数,直到预设损失函数的值满足预设的收敛条件。预设损失函数包括分类损失函数、包围盒回归损失函数以及字符连接损失函数;其中,分类损失函数的值用于表征字符检测网络对样本图像中的像素点是否为字符像素点的预测结果与样本图像中的像素点是否为字符像素点的标记结果之间的差异;包围盒回归损失函数的值用于表征字符检测网络对样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果与样本图像中的对应字符像素点所在字符的包围盒的位置信息的标记结果之间的差异;字符连接损失函数的值用于表征文字检测模型对样本图像中的字符组的预测结果与预测出的字符组中的候选字符的字符向量之间的差异度的一致性程度。

[0022] 第三方面,本申请实施例提供了一种服务器,包括:一个或多个处理器;存储装置,用于存储一个或多个程序,当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现如第一方面提供的文字检测方法。

[0023] 第四方面,本申请实施例提供了一种计算机可读存储介质,其上存储有计算机程序,其中,程序被处理器执行时实现第一方面提供的文字检测方法。

[0024] 本申请上述实施例的文字检测方法和装置,通过采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取,随后基于提取出的待检测图像的特征,采用文字检测模型的字符检测网络预测出待检测图像中每个像素点为字符像素点的概率、以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息;而后基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息;然后将提取出的待检测图像的特征输入文字检测模型的字符映射网络,基于各候选字符的包围盒的位置信息对字符映射网络输出的特征图进行变换,生成用于表征候选字符的特征的字符向量;之后基于候选字符的包围盒的位置信息,在待检测图像中确定各候选字符的近邻候选字符,并根据利用字符向量计算得出的候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组;最后根据字符组中各候选字符的包围盒的位置信息确定待检测图像的文字区域,实现了通用场景的文字检测,提升了不规则文字的检测精度。

## 附图说明

[0025] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述,本申请的其它特征、目的和优点将会变得更明显:

[0026] 图1是本申请可以应用于其中的示例性系统架构图;

[0027] 图2是根据本申请的文字检测方法的一个实施例的流程图;

[0028] 图3是文字检测模型中的字符检测网络的检测结果示意图;

[0029] 图4是文字检测模型中确定出候选字符的包围盒的位置信息的原理示意图;

[0030] 图5是根据本申请的文字检测方法的流程的一个实现原理示意图;

[0031] 图6是根据本申请的文字检测方法中训练文字检测模型的一种具体实现方式的流程图;

[0032] 图7是根据本申请实施例的文字检测装置的一个结构示意图;

[0033] 图8是适于用来实现本申请实施例的服务器的计算机系统的结构示意图。

## 具体实施方式

[0034] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是,此处所描述的具体实施例仅仅用于解释相关发明,而非对该发明的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与有关发明相关的部分。

[0035] 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。

[0036] 图1示出了可以应用本申请的文字检测方法或文字检测装置的示例性系统架构100。

[0037] 如图1所示,系统架构100可以包括终端设备101、102、103,网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0038] 用户110可以使用终端设备101、102、103通过网络104与服务器105交互,以接收或发送消息等。终端设备101、102、103上可以安装有各种客户端应用,例如文档处理类应用、图像处理类应用、搜索类应用、生活服务类应用、购物类应用等。

[0039] 终端设备101、102、103可以是支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、智能手表、膝上型便携计算机和台式计算机等等。在一些场景中,终端设备101、102、103还可以具有图像采集功能,例如终端设备101、102、103可以具有摄像头。

[0040] 服务器105可以是提供各种服务的服务器,例如对终端设备101、102、103发送的图像处理请求进行处理的后台服务器。后台服务器可以对接收到的图像处理请求进行解析、提取图像特征、识别图像中的目标等处理,并将处理结果反馈给终端设备101、102、103。

[0041] 需要说明的是,本申请实施例所提供的文字检测方法可以由服务器105执行,相应地,文字检测装置可以设置于服务器105中。

[0042] 需要指出的是,服务器105的本地也可以直接存储待检测的图像,服务器105可以直接提取本地的待检测图像进行检测,此时,示例性系统架构100可以不存在终端设备101、102、103和网络104。

[0043] 还需要指出的是,终端设备101、102、103可以具有处理器,终端设备101、102、103可以安装有图像处理类应用,终端设备101、102、103也可以基于图像处理类应用对待检测图像进行文字检测,此时,本申请实施例的文字检测方法也可以由终端设备101、102、103执行,相应地,文字检测装置也可以设置于终端设备101、102、103中。此时,示例性系统架构100可以不存在服务器105和网络104。

[0044] 应该理解,图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的智能终端设备、网络和服务器的。例如服务器可以为集群式的服务器,包括部署了不同进程的多台服务器。

[0045] 继续参考图2,其示出了根据本申请的文字检测方法的一个实施例的流程200。该文字检测方法,包括以下步骤:

[0046] 步骤201,采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取。

[0047] 在本实施例中,上述文字检测方法运行于其上的电子设备(例如图1所示的服务器105)可以首先获取待检测图像。待检测图像可以为包含文字字符的图像,例如,可以为包含

道路指示牌、文字商标、或者文档中的文字的图像。这些图像可以为在实际场景中对道路指示牌、商标或者其他包含文字字符的标志物进行图像采集而获取的，也可以是将包含文字的文档进行格式转换而生成的。上述电子设备可以向存储待检测图像的设备发出请求，并接收存储待检测图像的设备响应于该请求而发送的待检测图像。上述电子设备也可以直接存储待检测图像，在执行文字检测的时候可以从存储器中调取待检测图像。

[0048] 之后，上述电子设备可以将待检测图像输入文字检测模型的特征提取网络中，来提取待检测图像的特征。其中，文字检测模型可以是基于卷积神经网络构建的，包含多个卷积层，多个池化层，以及全连接层。在这里，文字检测模型可以包括特征提取网络、字符检测网络以及字符映射网络。特征提取网络、字符检测网络以及字符映射网络均分别可以包括多个卷积层、池化层以及全连接层。每个卷积层可以具有至少一个卷积核，每个卷积层利用卷积核对输入该卷积层的图像进行卷积运算，输出特征图。

[0049] 可选地，上述特征提取网络可以采用诸如VGG卷积神经网络、深度ResNet (Residual Networks, 残差网络) 等深度卷积神经网络架构。

[0050] 上述特征提取网络中的每个卷积层输出的特征图为待检测图像的一个抽象层级的特征，不同的卷积层输出的特征图为不同抽象层级的特征。在本实施例中，可以抽取多个不同卷积层输出的特征图，形成多个抽象层级的特征。在这里，抽象层级为特征的抽象层级，低层的卷积层可以提取出抽象层级较低的特征，例如纹理、线条等，高层的卷积层可以提取出抽象层级较高的特征，例如在高层级的卷积层可以提取出人脸图像的眼睛、鼻子、嘴巴等特征。其中低层的卷积层为靠近文字检测模型的输入层的卷积层，高层的卷积层为靠近特征提取网络的输出层的卷积层。

[0051] 在本实施例的一些可选的实现方式中，可以按照如下方式对待检测图像进行多个抽象层级的特征提取：首先将待检测图像输入特征提取网络中，抽取特征提取网络的多个不同卷积层的输出作为多个抽象层级的特征，即可以同时抽取特征提取网络的低层、中层以及高层输出的特征图。然后可以拼接多个抽象层级的特征，或利用特征金字塔网络 (Feature Pyramid Network, FPN) 对多个抽象层次的特征进行处理，生成待检测图像的特征。具体来说，可以直接将抽取出的多个抽象层级的特征进行拼接，或者可以采用特征金字塔将抽取出的特征融合，生成抽象层次更丰富的特征，拼接或融合生成的特征可以是包含了多个抽象层级的特征的整图特征。这样使得生成的特征融合了语义信息较少、位置特征较明确的低抽象层级的特征，和语义信息丰富、位置特征较粗略的高抽象层级的特征，增加了特征维度。

[0052] 步骤202，基于提取出的待检测图像的特征，采用文字检测模型的字符检测网络预测出待检测图像中每个像素点为字符像素点的概率、以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息。

[0053] 在本实施例中，可以基于提取出的特征进行候选字符位置的预测。具体来说，可以将提取出的待检测图像的特征输入文字检测模型的字符检测网络，输出多个通道的特征图。该字符检测网络可以是全卷积网络，包括多个卷积层。字符检测网络可以预测出每个像素点为字符像素点的概率，并且预测出每个像素点为字符像素点时该像素点相对于其所在的字符的相对位置。

[0054] 字符像素点为表示字符上的一点的像素点。通常若待检测图像包含字符，该字符

由多个字符像素点组合形成。像素点为字符像素点的概率可以表征该像素点为字符像素点的可能性大小。在本实施例中,可以通过字符检测网络的多个卷积层对提取出的待检测图像的特征进行卷积运算,得出每个像素点为字符像素点的概率。并且,上述字符检测网络通过对待检测图像的特征进行多层卷积运算可以计算出该字符像素点在字符中的位置。在这里,字符的位置信息可以由字符的包围盒的位置信息来定义。包围盒可以是矩形包围盒,可以是图像中的字符的最小外接矩形。包围盒的位置信息可以用包围盒的四个顶点的坐标表示,或者可以由包围盒的其中两个顶点的坐标来表示,或者可以由包围盒的一个顶点的坐标及包围盒的宽度和高度来表示。字符像素点在字符中的位置可以由字符像素点的坐标与字符的包围盒的顶点坐标之间的距离表示。

[0055] 在一些可选的实现方式中,可以将待检测图像的各像素点按照其位置依次进行编号,例如左上角的像素点编号为1,然后找出字符的包围盒的顶点所在像素点的编号,根据编号与像素位置的一一对应关系,可以得出每个像素点与其所在字符的包围盒的相对位置关系。

[0056] 在一些可选的实现方式中,上述每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息可以包括:每个像素点为字符像素点时该像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点坐标的偏移量。

[0057] 图3示出了文字检测模型中的字符检测网络的检测结果的一个示意图。如图3所示,字符检测网络对待检测图像IM中的像素点Q的检测结果为一个五元组 $(p, \Delta x1, \Delta y1, \Delta x2, \Delta y2)$ ,其中, $p$ 为该像素点Q为字符“大”的一个像素点的概率, $\Delta x1, \Delta y1$ 为像素点Q相对于其所在的字符“大”的包围盒的左上角顶点A沿x方向和y方向的坐标偏移, $\Delta x2, \Delta y2$ 为像素点Q相对于其所在的字符“大”的包围盒的右下角顶点C沿x方向和y方向的坐标偏移。

[0058] 从图3可以看出,字符检测网络可以得出待检测图像的每个像素点的多元组,从而字符检测网络可以输出多个通道的特征图,其中一个通道的特征图为各像素点为字符像素点的概率,其余通道表示每个像素点为字符像素点时其与所在的字符的包围盒之间的相对位置信息。

[0059] 步骤203,基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息。

[0060] 确定待检测图像中的各个像素点为字符像素点时该像素点与其所在的字符的包围盒的相对位置信息之后,可以确定出各个字符所在的字符的包围盒信息。之后可以将所在的包围盒的位置信息相同或差异小于预设的差异值的所有像素点聚类,判断同一聚类内的像素点为字符像素点的概率是否满足预设的条件,例如可以判断同一聚类内的各像素点为字符像素点的概率的平均值是否大于第一预设值,或者可以判断同一聚类内是否存在像素点为字符像素点的概率大于第二预设值,若是,则可以确定该聚类内的像素点所在的字符为候选字符,该聚类内的像素点所在的字符的包围盒的位置信息即为对应的候选字符的包围盒的位置信息。若同一聚类内的像素点为字符像素点的概率不满足预设的条件,则可以确定该聚类内的像素点不是字符像素点。

[0061] 在本实施例的一些可选的实现方式中,字符检测网络预测出的每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息为:每个像素点为字符像素点

时该像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点坐标的偏移量。这时,上述基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息的步骤可以包括:首先确定概率高于预设概率阈值的像素点为字符像素点;然后基于确定出的各字符像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点的坐标的偏移量,确定由字符像素点定位出的字符的包围盒的坐标;采用非极大值抑制法滤除各字符像素点定位出的字符的包围盒的坐标中被重复定位的字符的包围盒的坐标,得到各候选字符的包围盒的坐标。

[0062] 具体地,在上述实现方式中,可以根据字符检测网络的预测结果,筛选出概率高于预设概率阈值的像素点为字符像素点。然后根据各字符像素点相对于其所在的字符的包围盒的对角线上的两个顶点坐标的偏移量,以及字符像素点的坐标计算得出该字符像素点所在字符的矩形包围盒的对角线上的两个顶点的坐标。在这里,各字符的矩形包围盒的两条相互垂直的边分别平行于待检测图像的图像坐标系的两条坐标轴。这样,可以根据包围盒的对角线上的两个顶点的坐标计算得出其他两个顶点的坐标,从而定位出包围盒的坐标。

[0063] 每个字符可能包含多个字符像素点,所以在确定出各字符像素点多定位出的字符的包围盒的坐标之后,还需要将所定位出的位置重叠的包围盒合并,被合并的包围盒所对应的多个字符像素点为同一候选字符的多个字符像素点。可以采用非极大值抑制法来滤除重复定位出的包围盒,具体可以对定位出各包围盒的字符的概率(定位出各包围盒的像素点为字符像素点的概率)进行降序排序,确定第一个包围盒为第一个候选字符的包围盒,然后依次确定当前的包围盒与之前的包围盒的重叠面积是否大于预设的面积,若是,则将当前的包围盒滤除,否则确定当前的包围盒为一个新的候选字符的包围盒。这样,通过搜索最大值,滤除非极大值,能够滤除重复定位出的字符的包围盒的坐标,经过滤除操作后得到的包围盒即为候选字符的包围盒,从而得到了候选字符的包围盒的坐标。

[0064] 请参考图4,其示出了文字检测模型中确定出候选字符的包围盒的位置信息的一种可选方式的原理示意图。如图4所示,在待检测图像IM中,字符像素点Q1所定位出的字符的包围盒为BoxQ1,字符像素点Q2所定位出的字符的包围盒为BoxQ2,字符像素点Q3所定位出的字符的包围盒为BoxQ3。其中包围盒BoxQ1与BoxQ3的交叠面积超过包围盒BoxQ1的面积的一半,且超过包围盒BoxQ3的面积的一半,包围盒BoxQ1与BoxQ2的交叠面积未超过包围盒BoxQ1的面积的一半,且未超过包围盒BoxQ2的面积的一半,则可以将BoxQ1和BoxQ2归为一类,BoxQ2为另一类。可以进一步比较定位出包围盒BoxQ1的像素点Q1为字符像素点的概率和定位出包围盒BoxQ2的像素点Q2为字符像素点的概率,滤除概率较小的包围盒Q1,则得到两个候选字符的包围盒分别为BoxQ2和BoxQ3。

[0065] 步骤204,将提取出的待检测图像的特征输入文字检测模型的字符映射网络,基于各候选字符的包围盒的位置信息对字符映射网络输出的特征图进行变换,生成用于表征候选字符的特征的字符向量。

[0066] 在本实施例中,可以通过文字检测模型的字符映射网络将步骤201提取出的特征映射至高维空间。该字符映射网络可以包括多个卷积层,待检测图像的特征经过字符映射网络的多个卷积层运算之后,得到包含多维特征的特征图。在这里,字符映射网络的输出可以例如为128个特征图,则字符映射网络将待检测图像的特征映射至128维的特征空间中。

[0067] 之后可以将各候选字符的包围盒的位置信息所指示的区域作为感兴趣区域 (Region of Interest, ROI), 将ROI映射至字符映射网络输出的特征图中, 对字符映射网络输出的特征图中的ROI进行池化操作, 并利用全连接层对池化操作的结果进行分类, 得到对应于各ROI的向量, 即为各ROI对应的候选字符的字符向量。

[0068] 步骤205, 基于候选字符的包围盒的位置信息, 在待检测图像中确定各候选字符的近邻候选字符, 并根据利用字符向量计算得出的候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组。

[0069] 在确定候选字符的包围盒的位置以及候选字符的字符向量之后, 可以对由多个候选字符所在的区域连接形成的文字区域进行预测。具体地, 可以根据候选字符的包围盒的位置信息, 计算候选字符之间的距离, 根据所有的距离的统计结果确定相邻候选字符的距离阈值, 然后将每个候选字符作为目标候选字符, 将与目标候选字符之间的距离小于该距离阈值的候选字符确定为该目标候选字符的近邻候选字符, 这样, 可以查找出每个候选字符的近邻候选字符。可选地, 在计算候选字符之间的距离时, 可以选定包围盒的其中一个顶点 (例如左上角) 计算相邻候选字符的包围盒的该顶点之间的距离, 也根据候选字符的包围盒的位置确定出候选字符的中心位置, 计算候选字符的中心位置之间的距离。

[0070] 然后, 可以基于各候选字符的字符向量计算候选字符与近邻候选字符之间的差异程度, 根据差异程度确定与候选字符相关的近邻候选字符。具体可以将差异程度小于预设差异度阈值的近邻候选字符作为与候选字符相关的近邻候选字符。最后可以将候选字符与相关的近邻候选字符连接为字符组。

[0071] 通常在图像中一个文字区域中的字符在旋转角度、畸变程度、字体类型等方面的差异程度较小, 本实施例通过提取字符向量并计算字符向量之间的差异度来量化各候选字符与近邻候选字符之间的差异程度, 可以有效地提取出在旋转角度、畸变程度、字体类型等方面相类似的候选字符, 从而预测出由多个候选字符所在区域连接形成的文字区域。

[0072] 在本实施例的一些可选的实现方式中, 上述基于所述候选字符的包围盒的位置信息, 在待检测图像中确定各候选字符的近邻候选字符的步骤可以按照如下方式执行: 采用K近邻算法, 基于各候选字符的包围盒的位置信息对候选字符进行分类, 根据分类结果确定各候选字符的近邻候选字符。在这里, K可以为正整数, 可以为根据经验设定的值, 也可以是基于已标记的样本数据对K近邻分类器训练得到的值。得到的分类结果中, 同一分类的候选字符之间互为近邻候选字符。由此可以确定出各候选字符的近邻候选字符。

[0073] 进一步地, 上述根据利用字符向量计算得出的所述候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组的步骤可以按照如下方式实现: 计算候选字符的字符向量与对应的各近邻候选字符的字符向量之间的欧式距离, 作为候选字符与对应的各近邻候选字符之间的差异程度; 将与候选字符之间的差异程度小于预设差异度阈值的近邻候选字符作为与所述候选字符相关的近邻候选字符, 将各候选字符与相关的近邻候选字符连接为字符组。即可以将字符向量之间的欧式距离作为候选字符与近邻候选字符之间的差异度的度量, 欧式距离小于设定的欧式距离阈值, 则可以确定对应的两个候选字符相关, 可以将候选字符与相关的近邻候选字符连接为字符组。在这里, 预设的差异度阈值可以为设定的欧式距离阈值。

[0074] 需要说明的是, 在本实施例的其他可选实现方式中, 候选字符与近邻候选字符之

间的差异度还可以利用二者的字符向量,采用其他差异度计算方式得出,例如可以计算余弦相似度、皮尔森相关系数相似度度量参数的倒数,作为差异度,此处不做赘述。

[0075] 步骤206,根据字符组中各候选字符的包围盒的位置信息确定待检测图像的文字区域。

[0076] 最后,可以将字符组中的各候选字符的包围盒连通,形成环绕整个字符组的包围盒,该包围盒即为该字符组对应的文字区域。如果步骤205检测出多个字符组,则在步骤206中,可以形成多个文字区域。可选地,还可以将多个文字区域相连接,形成最终检测得到的文字区域。

[0077] 在本实施例的一些可选的实现方式中,可以按照如下方式确定待检测图像的文字区域:基于字符组中各候选字符的包围盒的位置信息划定环绕字符组中所有字符的包围线作为待检测图像的文字区域。

[0078] 上述包围线可以是由字符组中各候选字符的包围盒的顶点连接形成的,因此,可以根据字符组中各候选字符的包围盒的位置信息确定出包围线上的各连接点的位置,进而确定出包围线所环绕的区域为文字区域。

[0079] 图5示出了根据本申请的字符检测方法的流程的一个实现原理示意图。如图5所示,待检测图像首先输入特征提取网络进行多个抽象层级的特征提取,然后将提取出的特征输入字符检测网络,输出五个通道的特征图,其中一个通道为概率图,表示各像素点为字符像素点的概率,另外四个通道为位置偏移图,分别表示各像素点相对于其所在的字符的包围盒的左上角顶点的坐标和右下角顶点的坐标的坐标偏移量。随后可以根据概率图和位置偏移图滤除重复的包围盒(即滤除重复定位的字符),得到候选字符区域。

[0080] 将特征提取网络提取出的特征输入字符映射网络,得到高维的字符映射图,然后将候选字符区域作为ROI在字符映射图上进行池化操作,得出各候选字符的字符向量。之后根据候选字符区域筛选出各候选字符的近邻候选字符,并计算候选字符与近邻候选字符的字符向量之间的差异度,以确定相关的近邻候选字符,最后连接相关的近邻候选字符,形成字符组,并计算包围字符组的包围线,得到文字区域。

[0081] 本申请上述实施例的字符检测方法,通过采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取,随后基于提取出的待检测图像的特征,采用文字检测模型的字符检测网络预测出待检测图像中每个像素点为字符像素点的概率、以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息;而后基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息;然后将提取出的待检测图像的特征输入文字检测模型的字符映射网络,基于各候选字符的包围盒的位置信息对字符映射网络输出的特征图进行变换,生成用于表征候选字符的特征的字符向量;之后基于候选字符的包围盒的位置信息,在待检测图像中确定各候选字符的近邻候选字符,并根据利用字符向量计算得出的候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组;最后根据字符组中各候选字符的包围盒的位置信息确定待检测图像的文字区域,实现了通用场景的文字检测,提升了不规则文字的检测精度。

[0082] 在一些实施例中,上述文字检测方法还可以包括采用机器学习方法,基于样本图像训练所述文字检测模型的步骤。上述文字检测模型可以包括特征提取网络、字符检测网



络以及字符映射网络。特征提取网络、字符检测网络以及字符映射网络可以是具有不同结构和参数的卷积神经网络。可以将样本图像作为文字检测模型的输入,不断调整文字检测模型的参数,使文字检测模型预测出的文字区域逼近实际的文字区域。

[0083] 在一些可选的实现方式中,可以采用具有较高标注精度的样本图像进行训练。

[0084] 请参考图6,其示出了根据本申请的字符检测方法中训练文字检测模型的一种具体实现方式的流程图。如图6所示,上述训练文字检测模型的流程600可以包括以下步骤:

[0085] 步骤601,获取已采用包围盒方式标记所包含的字符的样本图像。

[0086] 在本实施例中,可以获取具有字符级别标注的样本图像,在该样本图像中,以包围盒的方式标记了所包含的每个字符以及字符组。这里的字符组为距离相近或语义关联的字符的集合。

[0087] 具体地,样本图像可以是随机获取的包含文字字符或不包含文字字符的图像。在获取样本图像时,可以选取标记准确的样本图像形成正样本集合,选取标记错误或标记有偏差的作为负样本集合。在这里,标记错误可以包括将非字符标记为字符和/或将字符标记为非字符等,标记有偏差可以包括标记的字符的包围盒与字符的实际位置之间有偏差等。可选地,可以设定正样本数量与负样本数量的比例为1:3,以保证训练精度。

[0088] 步骤602,将样本图像输入文字检测模型,以对样本图像中的文字区域进行预测,得到样本图像中的像素点是否为字符像素点的预测结果、样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果以及对样本图像中的字符组的预测结果。

[0089] 可以将样本图像输入文字检测模型进行预测,提取经过特征提取网络、字符检测网络处理、并根据字符检测网络输出的各像素点的概率值对像素点将像素点划分为“字符像素点”类和“非字符像素点”类的预测结果,即得出样本图像中的像素点是否为字符像素点的预测结果。可以提取经过特征提取网络、字符检测网络处理、并根据每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息计算得出的候选字符的包围盒的位置信息的预测结果,即得出样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果。还可以提取出经过特征提取网络、字符检测网络和字符映射网络处理、并根据候选字符的包围盒的位置信息以及字符映射网络输出的字符向量得到的样本图像中的字符组的预测结果。

[0090] 步骤603,计算预设损失函数的值,计算文字检测模型中每个参数相对于预设损失函数的梯度,利用反向传播算法更新模型参数,直到预设损失函数的值满足预设的收敛条件。

[0091] 在这里,预设损失函数包括分类损失函数、包围盒回归损失函数以及字符连接损失函数。其中,分类损失函数的值用于表征字符检测网络对样本图像中的像素点是否为字符像素点的预测结果与样本图像中的像素点是否为字符像素点的标记结果之间的差异。包围盒回归损失函数的值用于表征字符检测网络对样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果与样本图像中的对应字符像素点所在字符的包围盒的位置信息的标记结果之间的差异。字符连接损失函数的值用于表征文字检测模型对样本图像中的字符组的预测结果与预测出的字符组中的候选字符的字符向量之间的差异度的一致性程度。

[0092] 分类损失函数用于表征根据字符检测网络对待检测图像中的每个像素点为字符



像素点的概率对像素点进行分类的偏差,包围盒损失函数用于表征根据字符检测网络输出的字符像素点相对于所在字符的包围盒的位置信息计算得出的各字符的位置信息与字符的实际位置信息之间的偏差,字符连接损失函数用于表征在经过候选字符的包围盒位置检测、计算候选字符的字符向量、筛选相关的近邻候选字符之后,文字检测模型对字符组的预测结果的偏差。

[0093] 上述文字检测模型训练过程中所采用的预设损失函数L可以为:

$$[0094] \quad L = L_{\text{class}} + \lambda_1 L_{\text{reg}} + \lambda_2 L_{\text{emb}} \quad (1)$$

[0095] 其中 $L_{\text{class}}$ 为分类损失函数, $L_{\text{reg}}$ 为包围盒回归损失函数, $L_{\text{emb}}$ 为字符连接损失函数, $L_{\text{class}}$ 可以采用Max margin objective(最大边缘目标)函数, $L_{\text{reg}}$ 可以采用交除并损失函数(Intersection-over-Union Loss, IoU loss),可以按照如下方式计算:

$$[0096] \quad L_{\text{emb}} = \overline{J(v_i, v_j, l_{i,j})} \quad (2)$$

$$[0097] \quad J(v_i, v_j, l_{i,j}) = l_{i,j} [D(v_i, v_j)]^2 + (1 - l_{i,j}) \{[\alpha - D(v_i, v_j)]_+\}^2 \quad (3)$$

[0098] 其中, $\lambda_1, \lambda_2$ 为预置的系数, $\overline{J(v_i, v_j, l_{i,j})}$ 表示 $J(v_i, v_j, l_{i,j})$ 的平均值, $v_i, v_j$ 分别为两个字符 $i, j$ 的字符向量, $l_{i,j}$ 为两个字符 $i, j$ 的连接的正负属性,若两个字符 $i, j$ 连接为同一个字符组,则 $l_{i,j} = 1$ ,否则 $l_{i,j} = 0$ , $D$ 为欧式距离计算符, $\alpha$ 为预设的值,表示负连接的边界值, $[\alpha - D(v_i, v_j)]_+$ 表示取0和 $\alpha - D(v_i, v_j)$ 中的最大值。

[0099] 可以基于步骤602得到的样本图像中的像素点是否为字符像素点的预测结果以及样本图像中的像素点是否为字符像素点的标记结果之间的差异计算分类损失函数 $L_{\text{class}}$ 的值,基于步骤602得到的样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果以及对样本图像中的字符组的预测结果计算包围盒回归损失函数 $L_{\text{reg}}$ 的值,基于步骤602得出的字符组的连接属性的预测结果以及字符组内候选字符的字符向量之间的欧式距离计算字符连接损失函数的值,进而得到预设损失函数L的值。

[0100] 可以判断预设损失函数L的值是否满足预设的收敛条件,例如是否收敛到最小值,达到预设的收敛值等,若否,可以计算文字检测模型中的各个参数相对于预设损失函数的梯度,采用梯度下降法更新参数,并利用更新参数后的文字检测模型重新得出预测结果,计算预设损失函数的值,进而判断预设损失函数的值是否满足预设的收敛条件。重复执行上述更新参数、利用更新参数后的文字检测模型重新预测。计算预设损失函数的值并判断预设损失函数的值是否满足预设的收敛条件的步骤,直到预设损失函数的值满足预设的收敛条件,停止更新参数,将当前的文字检测模型的参数确定为训练好的文字模型的参数。

[0101] 在本实施例的另一些可选的实现方式中,可以采用弱监督的方法训练上述文字检测模型,例如可以先检测一些文字框,然后选择一些文字框进行标记,例如可以标记一些词条中的单字,利用标记的词条训练单字检测器,然后采用训练好的单字检测器检测样本图像中的字符,得到已标记所包含的字符的样本图像,之后采用上述步骤602和步骤603,完成对文本检测模型的训练。

[0102] 上述文本检测模型的训练方法,通过不断调整模型的参数,使根据预测结果计算得出的损失函数逐步缩小,从而使得文字检测模型可以有效地学习到字符标记的逻辑,得到较准确的检测结果。同时,该文字检测模型对字符以及字符间的连接没有过多限制,使得该文字检测模型可以适用于扭曲、旋转等通用场景下的文字检测。

[0103] 进一步参考图7,作为对上述各图所示方法的实现,本申请提供了一种文字检测装置的一个实施例,该装置实施例与图2所示的方法实施例相对应,该装置具体可以应用于各种电子设备中。

[0104] 如图7所示,本实施例的文字检测装置700包括:提取单元701、检测单元702、定位单元703、映射单元704、连接单元705以及确定单元706。其中,提取单元701可以用于采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取;检测单元702可以用于基于提取出的待检测图像的特征,采用文字检测模型的字符检测网络预测出待检测图像中每个像素点为字符像素点的概率、以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息;定位单元703可以基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息;映射单元704可以用于将提取出的待检测图像的特征输入文字检测模型的字符映射网络,基于各候选字符的包围盒的位置信息对字符映射网络输出的特征图进行变换,生成用于表征候选字符的特征的字符向量;连接单元705可以用于基于候选字符的包围盒的位置信息,在待检测图像中确定各候选字符的近邻候选字符,并根据利用字符向量计算得出的候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组;确定单元706可以用于根据字符组中各候选字符的包围盒的位置信息确定待检测图像的文字区域。

[0105] 在本实施例中,提取单元701可以首先获取待检测图像,然后将待检测图像输入文字检测模型的特征提取网络中,来提取待检测图像的特征。其中,文字检测模型可以是基于卷积神经网络构建的,包含多个卷积层,多个池化层。上述特征提取网络中的每个卷积层输出的特征图为待检测图像的一个抽象层级的特征,不同的卷积层输出的特征图为不同抽象层级的特征。在本实施例中,提取单元701可以抽取多个不同卷积层输出的特征图,形成多个抽象层级的特征。

[0106] 检测单元702可以基于提取单元701提取出的特征进行候选字符位置的预测。具体可以将提取出的待检测图像的特征输入字符检测网络,输出多个通道的特征图,其中一个通道为概率图,表示待检测图像中每个像素点为字符像素点的概率;其他通道为位置偏移图,表示待检测图像中每个像素点与其所在字符的包围盒的相对位置信息。

[0107] 定位单元703可以根据检测单元702检测得出的各个像素点为字符像素点时该字符像素点与其所在的字符的包围盒的相对位置信息之后,可以确定出各个字符所在的字符的包围盒信息。之后定位单元可以将所在的包围盒的位置信息相同或差异小于预设的差异值的所有像素点聚类,判断同一聚类内的像素点为字符像素点的概率是否满足预设的条件,若是,则可以确定该聚类内的像素点所在的字符为候选字符,该聚类内的像素点所在的字符的包围盒的位置信息即为对应的候选字符的包围盒的位置信息。

[0108] 映射单元704可以利用文字检测模型的字符映射网络将提取单元701提取出的特征映射至高维空间,之后可以将各候选字符的包围盒的位置信息所指示的区域作为感兴趣区域(Region of Interest, ROI),将ROI映射至字符映射网络输出的特征图中,对字符映射网络输出的特征图中的ROI进行池化操作,并利用全连接层对池化操作的结果进行分类,得到对应于各ROI的向量,即为各ROI对应的候选字符的字符向量。

[0109] 连接单元705可以根据候选字符的包围盒的位置信息,计算候选字符之间的距离,从

而确定出每个候选字符的近邻候选字符,然后基于候选字符与近邻候选字符的字符向量计算二者的差异度,将差异度小于预设差异度阈值的两个候选字符确定为相关的候选字符。然后可以连接每个候选字符相关的候选字符,形成字符组。

[0110] 确定单元706可以将字符组中的各候选字符的包围盒连通,形成环绕整个字符组的包围盒,该包围盒即为该字符组对应的文字区域。

[0111] 在一些实施例中,上述提取单元701可以用于采用文字检测模型的特征提取网络,按照如下方式对待检测图像进行多个抽象层级的特征提取:将待检测图像输入特征提取网络中,抽取特征提取网络的多个不同卷积层的输出作为多个抽象层级的特征;拼接多个抽象层级的特征,或利用特征金字塔网络对多个抽象层次的特征进行处理,生成待检测图像的特征。

[0112] 在一些实施例中,上述每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息,包括:每个像素点为字符像素点时该像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点坐标的偏移量。

[0113] 在一些实施例中,上述定位单元703可以用于基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息,按照如下方式确定出各候选字符的包围盒的位置信息:确定概率高于预设概率阈值的像素点为字符像素点;基于确定出的各字符像素点的坐标相对于其所在的字符的矩形包围盒的对角线上的两个顶点的坐标的偏移量,确定由字符像素点定位出的字符的包围盒的坐标;采用非极大值抑制法滤除各字符像素点定位出的字符的包围盒的坐标中被重复定位的字符的包围盒的坐标,得到各候选字符的包围盒的坐标。

[0114] 在一些实施例中,上述连接单元705可以用于基于候选字符的包围盒的位置信息,按照如下方式在待检测图像中确定各候选字符的近邻候选字符:采用K近邻算法,基于各候选字符的包围盒的位置信息对候选字符进行分类,根据分类结果确定各候选字符的近邻候选字符;以及上述连接单元705可以按照如下方式将各候选字符与相关的近邻候选字符连接为字符组:计算候选字符的字符向量与对应的各近邻候选字符的字符向量之间的欧式距离,作为候选字符与对应的各近邻候选字符之间的差异程度;将与候选字符之间的差异程度小于预设差异度阈值的近邻候选字符作为与候选字符相关的近邻候选字符,将各候选字符与相关的近邻候选字符连接为字符组。

[0115] 在一些实施例中,上述确定单元706可以按照如下方式确定待检测图像的文字区域:基于字符组中各候选字符的包围盒的位置信息划定环绕字符组中所有字符的包围线作为待检测图像的文字区域。

[0116] 在一些实施例中,上述装置700还可以包括:训练单元,用于采用机器学习方法,基于样本图像训练文字检测模型。

[0117] 在一些实施例中,上述训练单元可以进一步用于按照如下方式训练文字检测模型:获取已采用包围盒方式标记所包含的字符的样本图像;将样本图像输入文字检测模型,以对样本图像中的文字区域进行预测,得到样本图像中的像素点是否为字符像素点的预测结果、样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果以及对样本图像中的字符组的预测结果;计算预设损失函数的值,计算文字检测模型中每个参数相对于预设损失函数的梯度,利用反向传播算法更新模型参数,直到预设损失函数的值满足预设

的收敛条件。

[0118] 其中,预设损失函数包括分类损失函数、包围盒回归损失函数以及字符连接损失函数。分类损失函数的值用于表征字符检测网络对样本图像中的像素点是否为字符像素点的预测结果与样本图像中的像素点是否为字符像素点的标记结果之间的差异;包围盒回归损失函数的值用于表征字符检测网络对样本图像中的字符像素点所在的字符的包围盒的位置信息的预测结果与样本图像中的对应字符像素点所在字符的包围盒的位置信息的标记结果之间的差异;字符连接损失函数的值用于表征文字检测模型对样本图像中的字符组的预测结果与预测出的字符组中的候选字符的字符向量之间的差异度的一致性程度。

[0119] 装置700中记载的诸单元与参考图2和图6描述的方法中的各个步骤相对应。由此,上文针对方法描述的操作和特征同样适用于装置700及其中包含的单元,在此不再赘述。

[0120] 本申请上述实施例的文字检测装置700,通过提取单元采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取,随后检测单元基于提取出的待检测图像的特征,采用文字检测模型的字符检测网络预测出待检测图像中每个像素点为字符像素点的概率、以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息;而后定位单元基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息;然后映射单元将提取出的待检测图像的特征输入文字检测模型的字符映射网络,基于各候选字符的包围盒的位置信息对字符映射网络输出的特征图进行变换,生成用于表征候选字符的特征的字符向量;之后连接单元基于候选字符的包围盒的位置信息,在待检测图像中确定各候选字符的近邻候选字符,并根据利用字符向量计算得出的候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组;最后确定单元根据字符组中各候选字符的包围盒的位置信息确定待检测图像的文字区域,实现了通用场景的文字检测,提升了不规则文字的检测精度。

[0121] 下面参考图8,其示出了适于用来实现本申请实施例的电子设备的计算机系统800的结构示意图。图8示出的电子设备仅仅是一个示例,不应对本申请实施例的功能和使用范围带来任何限制。

[0122] 如图8所示,计算机系统800包括中央处理单元(CPU)801,其可以根据存储在只读存储器(ROM)802中的程序或者从存储部分808加载到随机访问存储器(RAM)803中的程序而执行各种适当的动作和处理。在RAM 803中,还存储有系统800操作所需的各种程序和数据。CPU 801、ROM 802以及RAM 803通过总线804彼此相连。输入/输出(I/O)接口805也连接至总线804。

[0123] 以下部件连接至I/O接口805:包括键盘、鼠标等的输入部分806;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分807;包括硬盘等的存储部分808;以及包括诸如LAN卡、调制解调器等网络接口卡的通信部分809。通信部分809经由诸如因特网的网络执行通信处理。驱动器810也根据需要连接至I/O接口805。可拆卸介质811,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器810上,以便于从其上读出的计算机程序根据需要被安装入存储部分808。

[0124] 特别地,根据本公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质

上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分809从网络上被下载和安装,和/或从可拆卸介质811被安装。在该计算机程序被中央处理单元(CPU)801执行时,执行本申请的方法中限定的上述功能。需要说明的是,本申请的计算机可读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。在本申请中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本申请中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0125] 可以以一种或多种程序设计语言或其组合来编写用于执行本申请的操作的计算机程序代码,程序设计语言包括面向对象的程序设计语言—诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0126] 附图中的流程图和框图,图示了按照本申请各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0127] 描述于本申请实施例中所涉及到的单元可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中,例如,可以描述为:一种处理器包括提取单元、检测单元、定位单元、映射单元、连接单元以及确定单元。其中,这些单元的名称在某种情况下并不构成对该单元本身的限定,例如,提取单元还可以被描述为“采用文字

检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取的单元”。

[0128] 作为另一方面,本申请还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的装置中所包含的;也可以是单独存在,而未装配入该装置中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被该装置执行时,使得该装置:采用文字检测模型的特征提取网络对待检测图像进行多个抽象层级的特征提取;基于提取出的待检测图像的特征,采用文字检测模型的字符检测网络预测出待检测图像中每个像素点为字符像素点的概率、以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息;基于每个像素点为字符像素点的概率以及每个像素点为字符像素点时该像素点相对于其所在的字符的包围盒的位置信息确定出各候选字符的包围盒的位置信息;将提取出的待检测图像的特征输入文字检测模型的字符映射网络,基于各候选字符的包围盒的位置信息对字符映射网络输出的特征图进行变换,生成用于表征候选字符的特征的字符向量;基于候选字符的包围盒的位置信息,在待检测图像中确定各候选字符的近邻候选字符,并根据利用字符向量计算得出的候选字符与对应的近邻候选字符之间的差异程度将各候选字符与相关的近邻候选字符连接为字符组;根据字符组中各候选字符的包围盒的位置信息确定待检测图像的文字区域。

[0129] 以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解,本申请中所涉及的发明范围,并不限于上述技术特征的特定组合而成的技术方案,同时也应涵盖在不脱离上述发明构思的情况下,由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

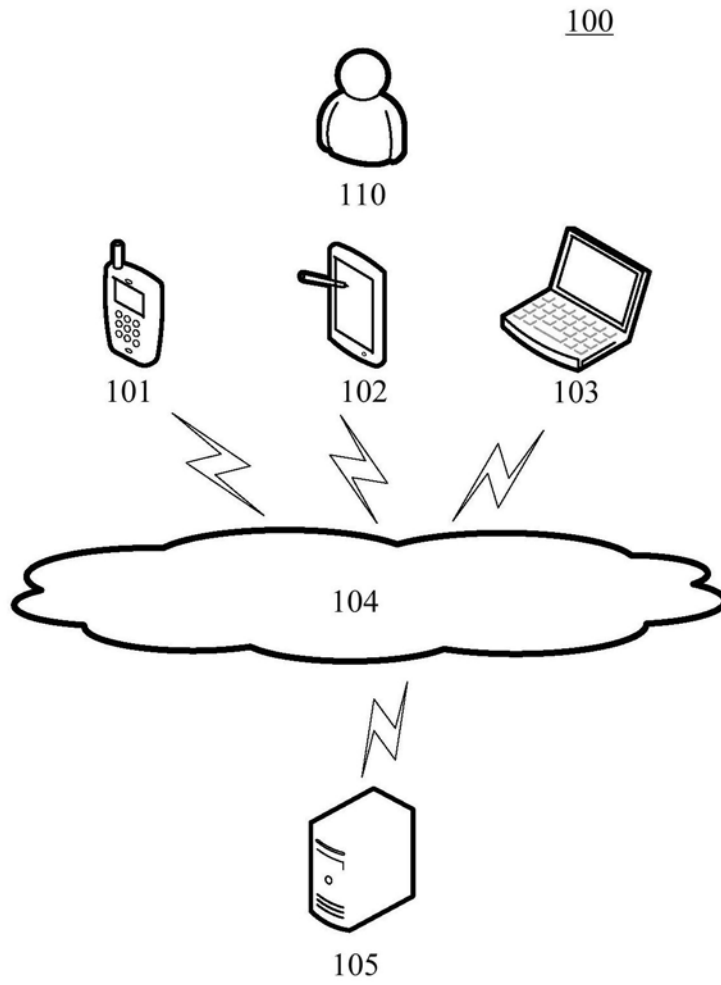


图1

200

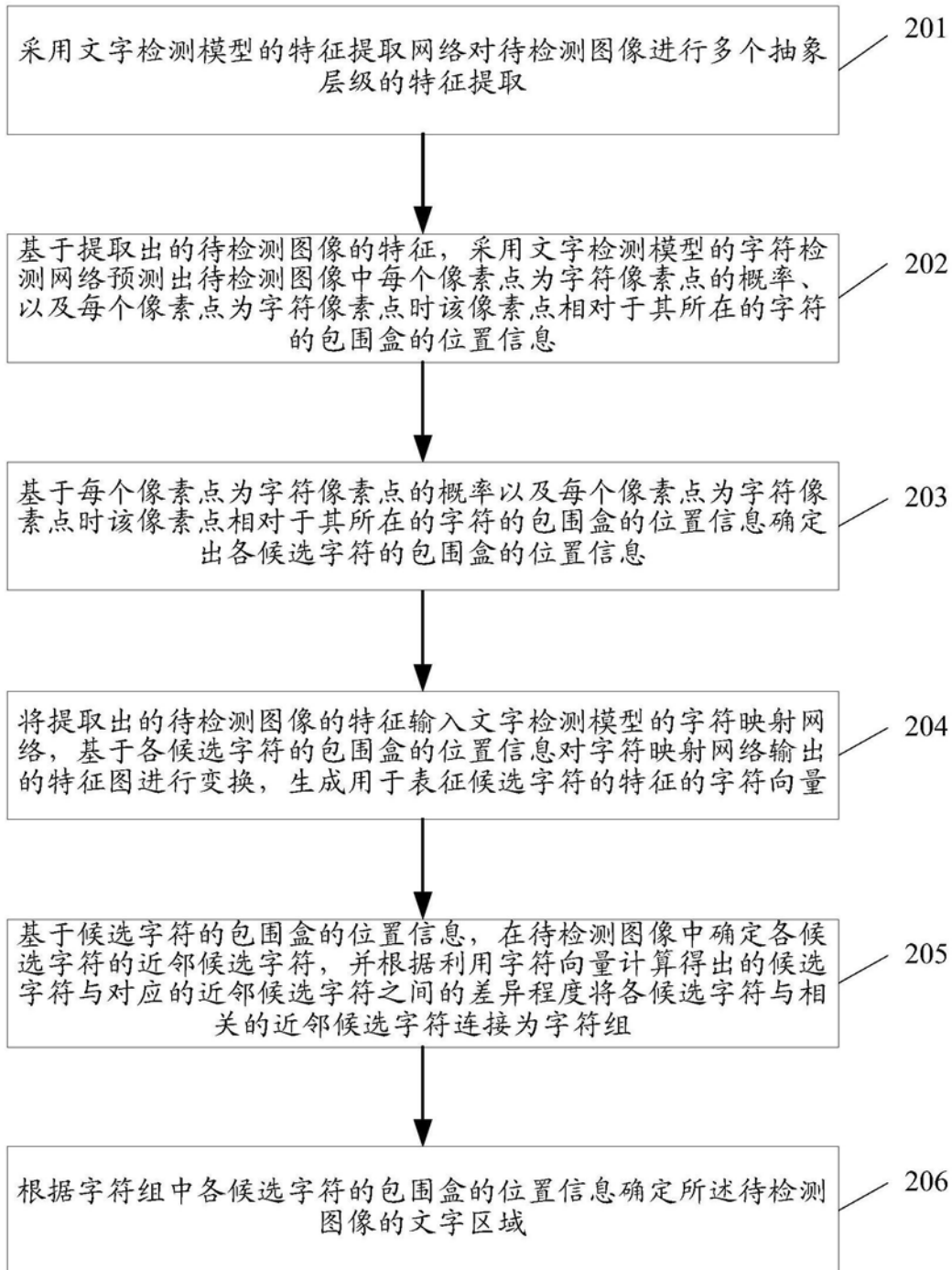


图2



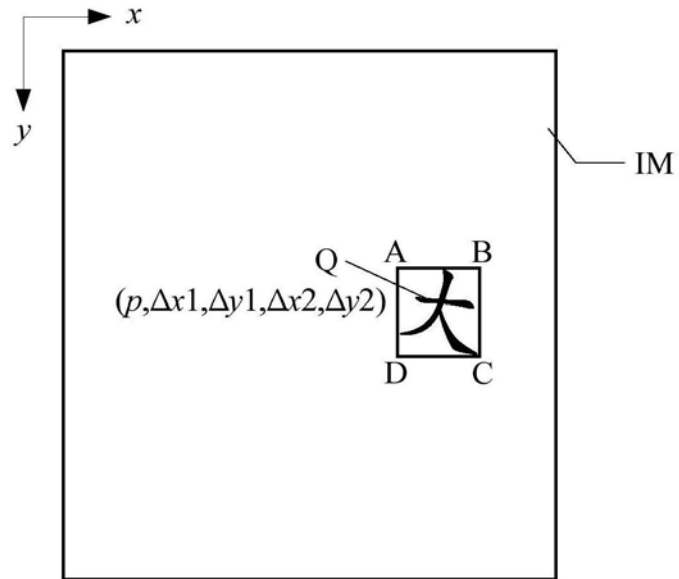


图3

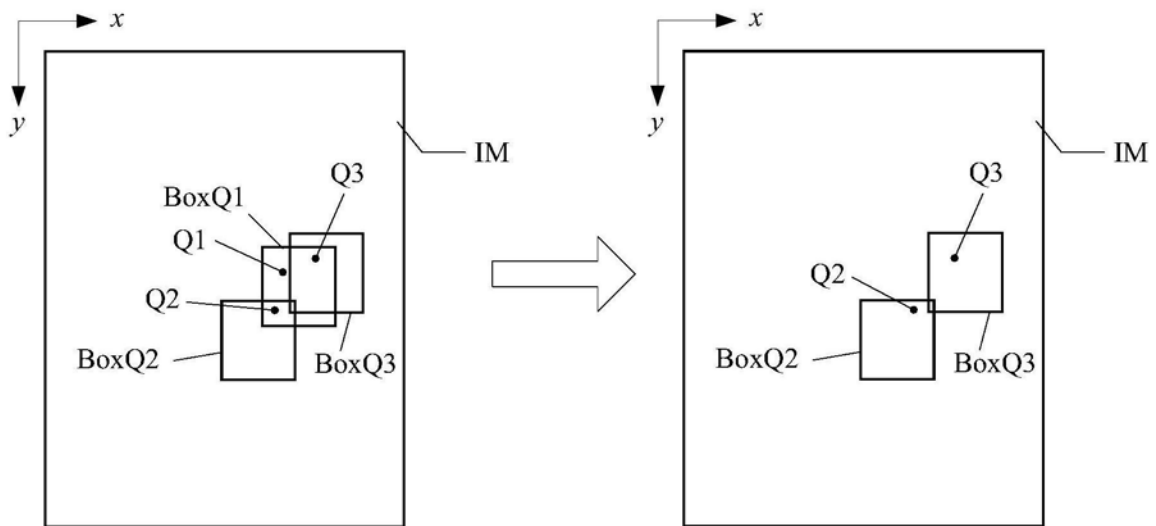


图4

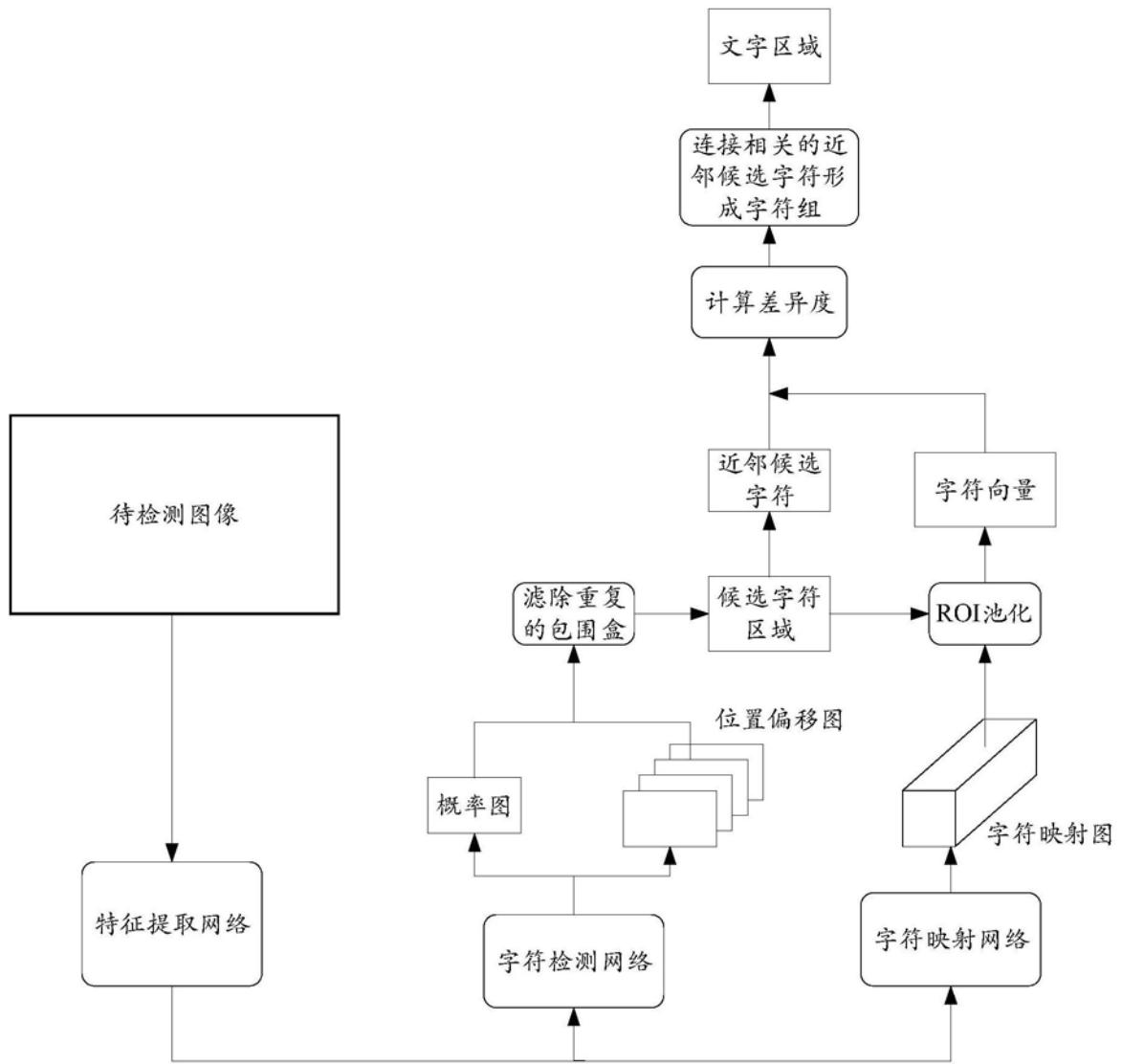


图5

600

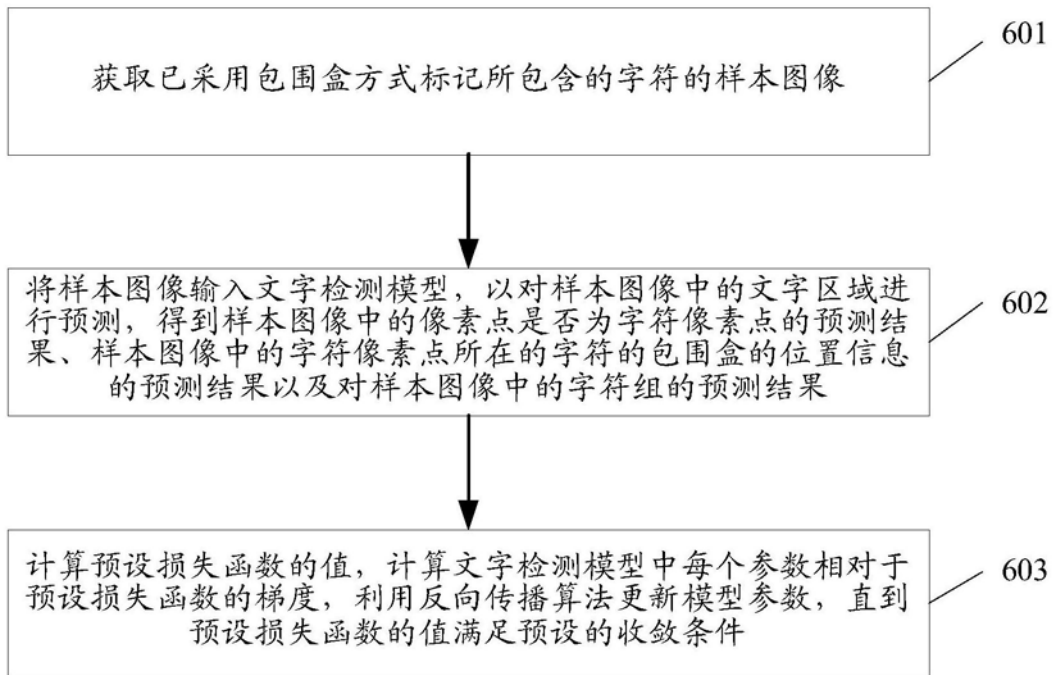


图6

700

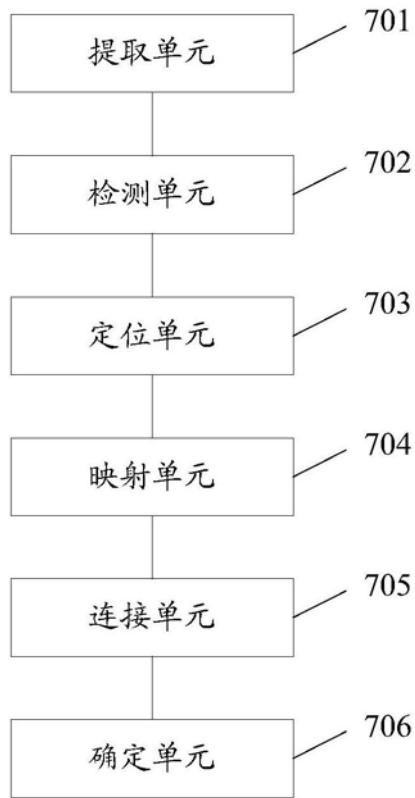


图7

800

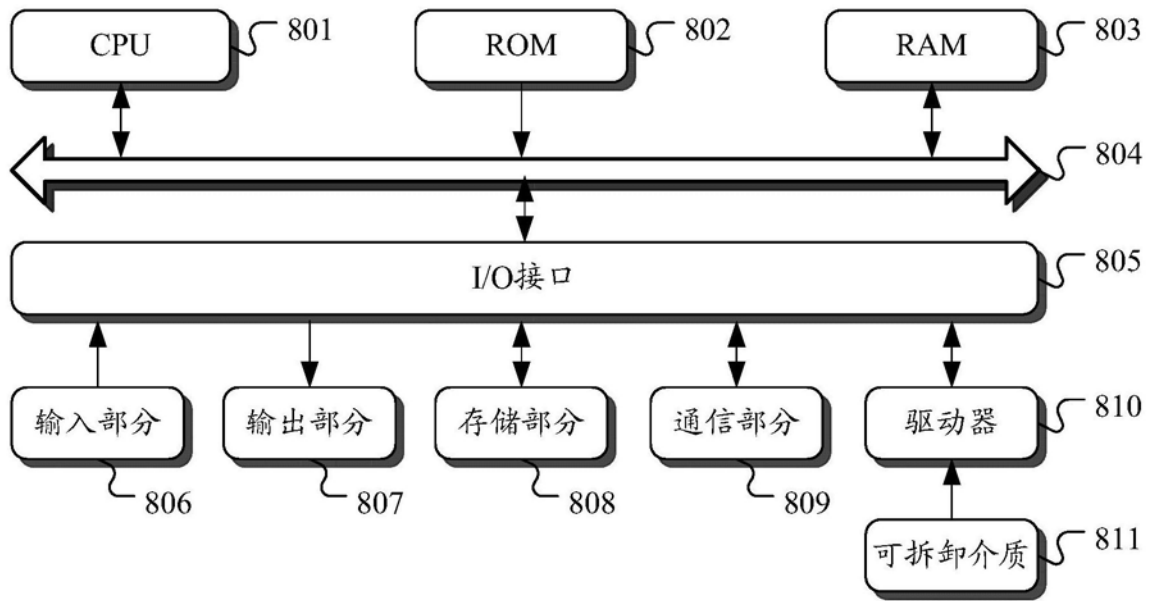


图8