



(12) 发明专利申请

(10) 申请公布号 CN 113435240 A

(43) 申请公布日 2021.09.24

(21) 申请号 202110396302.5 *G06K 9/62* (2006.01)
 (22) 申请日 2021.04.13 *G06T 5/30* (2006.01)
 (71) 申请人 北京易道博识科技有限公司 *G06T 3/40* (2006.01)
 地址 100083 北京市海淀区农大南路1号院 *G06N 3/04* (2006.01)
 2号楼5层办公A-501 *G06N 3/08* (2006.01)

(72) 发明人 周勃宇 王勇 朱军民

(74) 专利代理机构 北京金智普华知识产权代理有限公司 11401

代理人 巴晓艳

(51) Int. Cl.
G06K 9/00 (2006.01)
G06K 9/20 (2006.01)
G06K 9/32 (2006.01)
G06K 9/34 (2006.01)
G06K 9/46 (2006.01)

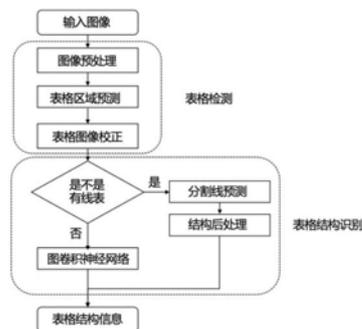
权利要求书2页 说明书8页 附图5页

(54) 发明名称

一种端到端的表格检测和结构识别方法及系统

(57) 摘要

本发明公开了一种端到端的表格检测和结构识别方法及系统,涉及计算机视觉领域。该方法包括:对原始图像进行垂直方向拉伸,并进行长宽比不变的尺寸归一化以及边界补0,形成预处理图像;以编码器-解码器模型为主体结构,确定所述预处理图像中的表格区域,并分类为有线表格图像和无线表格图像;基于所确定的表格区域,从预处理图像中分离出校正后的、仅包含表格区域的表格区域图像;针对所述表格区域图像,根据分类为有线表格图像和无线表格图像分别采用不同方法进行表格结构识别。本发明对不同类别的表格采用不同的结构识别方法,充分结合卷积神经网络图像分割算法、图卷积神经网络算法和传统规则分析方法的优势来提升算法的鲁棒性和通用性。



1. 一种端到端的表格检测和结构识别方法, 输入的原始图像中包含有表格, 其特征在于, 所述方法包括以下步骤:

步骤1: 图像预处理步骤, 对原始图像进行垂直方向拉伸, 并进行长宽比不变的尺寸归一化以及边界补0, 形成预处理图像;

步骤2: 表格区域预测步骤, 以编码器-解码器模型为主体结构, 确定所述预处理图像中的表格区域, 并分类为有线表格图像和无线表格图像;

步骤3: 表格图像校正步骤, 基于所确定的表格区域, 从预处理图像中分离出校正后的、仅包含表格区域的表格区域图像;

步骤4: 表格结构识别步骤, 针对所述表格区域图像, 根据分类为有线表格图像和无线表格图像分别采用不同方式进行表格结构识别。

2. 根据权利要求1所述的方法, 其特征在于, 所述步骤2中, 所述编码器-解码器模型的编码部分利用卷积的方式从第一高分辨表征中下采样出低分辨率表征; 解码部分利用转置卷积或插值的方式从低分辨率表征上采样出第二高分辨率表征。

3. 根据权利要求2所述的方法, 其特征在于, 所述编码部分操作如下:

采用高分辨率网络中多分辨率子网络并行连接的机制生成多分辨率表征, 并引入多分辨率融合模块实现多分辨率表征之间的特征信息交换与融合, 最终输出多种尺度的第一特征图。

4. 根据权利要求3所述的方法, 其特征在于, 所述解码部分操作如下:

首先采用空洞空间卷积池化金字塔模块对所述第一特征图中最小尺寸的特征图以不同采样间隔的空洞卷积并行采样, 后续通过转置卷积的方式分别将其他第一特征图的空间维度扩大两倍, 形成与第一特征图数量相同的多个第二特征图;

将所述第二特征图与来自编码部分同样尺寸的第一特征图拼接在一起, 最后, 卷积生成与预处理图像大小尺寸相同的两张掩膜预测图像;

由此确定表格区域, 并区分为有线表格图像和无线表格图像。

5. 根据权利要求1所述的方法, 其特征在于, 所述步骤3具体包括:

步骤31: 根据掩膜预测图像利用坎尼边缘检测算子计算出表格四周的轮廓;

步骤32: 利用霍夫变换算子检测出轮廓中的所有直线并将部分满足合并条件的直线合并在一起;

步骤33: 根据所有直线的位置计算出准确的表格位置, 由此分离出校正后的、仅包含表格区域的表格区域图像。

6. 根据权利要求1所述的方法, 其特征在于, 所述步骤4中, 针对属于有线表格图像的所述表格区域图像, 具体包括:

根据分割线的掩膜预测图像, 利用坎尼边缘检测算子计算显式分隔线的轮廓图;

利用边界腐蚀方法提取出分隔线的轮廓骨架图;

利用霍夫变换方法从轮廓骨架图中计算出所有直线并将部分满足合并条件的直线融合在一起;

通过计算出所有横线和竖线相交点的位置, 得到表格单元格的位置;

提取表格内文本实例的内容和位置;

根据表格单元格和文本实例的相对位置计算出表格结构信息并输出。

7. 根据权利要求1所述的方法,其特征在于,所述步骤4中,针对属于无线表格图像的所述表格区域图像,具体包括:

以每个文本实例为节点,提取节点特征,所述节点特征由每个文本实例的位置特征、边界框背景特征、行背景特征和列背景特征共同拼接形成;

针对某一节点a,选取其当前所在特征空间内所有节点并计算相似性,选取周边多个最近邻节点;

将节点a的节点特征A及其与所述多个最近邻节点的节点特征的相似度差值分别拼接在一起,输入经训练的图卷积神经网络,输出更新后的节点特征A' ;

重复以上操作,得到所述表格区域图像内所有节点的更新后的节点特征;

通过三个多层感知机网络分别利用更新后的节点特征确定节点与多个最近邻节点的行、列和单元格的结构关系,由此确定表格结构信息并输出。

8. 根据权利要求7所述的方法,其特征在于,所述文本实例的位置特征由边界框的左上角和右下角的坐标组成。

9. 根据权利要求7所述的方法,其特征在于,所述文本实例的边界框背景特征、行背景特征和列背景特征通过感兴趣区域池化图像特征的提取方式从特征图上提取。

10. 一种端到端的表格检测和结构识别装置,其特征在于,所述装置基于根据权利要求1至9中任一项所述方法进行操作,其特征在于,所述装置包括以下部件:

图像预处理单元,用于对原始图像进行垂直方向拉伸,并进行长宽比不变的尺寸归一化以及边界补0,形成预处理图像;

表格区域预测单元,用于以编码器-解码器模型为主体结构,确定所述预处理图像中的表格区域,并分类为有线表格图像和无线表格图像;

表格图像校正单元,用于基于所确定的表格区域,从预处理图像中分离出校正后的、仅包含表格区域的表格区域图像;

表格结构识别单元,用于针对所述表格区域图像,根据分类为有线表格图像和无线表格图像分别采用不同方式进行表格结构识别。

一种端到端的表格检测和结构识别方法及系统

技术领域

[0001] 本发明涉及计算机视觉领域,尤其是一种端到端的表格检测和结构识别方法及系统。

背景技术

[0002] 现实中,表格作为一种关键信息的承载方式广泛存在地存在于PDF、扫描文档、拍照图片等对象中。表格结构识别是很多下游任务的重要前提,如文档分析、信息抽取和可视化。自动表格识别方法通常包含表格检测和表格结构识别两大步骤,其中,表格检测的目的是定位出图片中的表格区域,表格识别则是对于每个区域内的表格内部结构进行识别,以获取最终的结构化数据。人工提取表格内容的方式将耗费大量的人力和时间。相比之下,自动化的方式将会极大的提高工作效率。

[0003] 现实中的表格存在大量不同的风格、版式及内部结构,因此想要采用统一的识别方法通常是十分困难的。传统的表格识别方法通常依赖于手工设计的特征(如行列分隔线、空白区域、单元格数据类型等)以及启发式的规则。表格检测通常采用一种自底向上的策略,如利用表格中显式的文本对齐关系定位出表格中的行列位置,然后再将所有的行列信息融合在一起计算出表格区域。表格样式的多变性和内部结构的复杂性会给行列的检测带来很大的困难,进而影响整体的检测效果。表格结构识别则通常依赖于表格中的显式分隔线,以及分隔线与文本实例的相对位置关系。这种方法在有线表格还可以获得较好的性能,但是却无法应对分隔线部分缺失或完全缺失的无线表格。

[0004] 近些年,深度学习技术推动了计算机视觉的快速发展,在表格识别领域,也得到了较多的应用。概括的讲,相比于传统的方法,深度学习表格识别方法通常具备两大优势。第一,深度学习方法以图像作为输入,原则上可以应用于任何可转换为图像的识别对象,如PDF、扫描文档等。因此,具有统一化方法的优势;第二,得益于强大的自动特征编码能力和统一的端到端可训练方法,相比于手工设计特征和启发式规则为主的传统方法,深度学习具有突出的性能表现。

[0005] 因此,提供一种基于深度学习各项优势的从表格检测到表格结构识别的一体化流程具有良好应用前景。

发明内容

[0006] 为了实现以上目的,本发明提出了一种融合表格检测于一体的结构识别方法,可以高效地从图像中抽取出表格内部结构信息。此方案所使用的图像分割技术不仅可以像素级别的预测方式来准确地计算出表格的边缘,还可以将表格分类为有线表格和无线表格。此方案在后续的步骤中对不同类别的表格采用不同的结构识别方法,充分结合卷积神经网络图像分割算法、图卷积神经网络算法和传统规则分析方法的优势来提升算法的鲁棒性和通用性。

[0007] 具体地讲,该方法首先利用卷积神经网络实现表格区域的检测。对于检测出的有

线表格,采用卷积神经网络完成表格线的检测,并结合后处理规则完成表格结构的识别;对于无线表格,则采用图卷积神经网络实现单元格、行和列关系的预测,进而完成结构的识别。

[0008] 根据本发明的第一方面,提供一种端到端的表格检测和结构识别方法,输入的原始图像中包含有表格,其特征在于,所述方法包括以下步骤:

[0009] 步骤1:图像预处理步骤,对原始图像进行垂直方向拉伸,并进行长宽比不变的尺寸归一化以及边界补0,形成预处理图像;

[0010] 步骤2:表格区域预测步骤,以编码器-解码器(Encoder-Decoder)模型为主体结构,确定所述预处理图像中的表格区域,并分类为有线表格图像和无线表格图像;

[0011] 步骤3:表格图像校正步骤,基于所确定的表格区域,从预处理图像中分离出校正后的、仅包含表格区域的表格区域图像;

[0012] 步骤4:表格结构识别步骤,针对所述表格区域图像,根据分类为有线表格图像和无线表格图像分别采用不同方式进行表格结构识别。

[0013] 进一步地,所述步骤2中,所述编码器-解码器模型的编码部分利用卷积的方式从第一高分辨表征中下采样出低分辨率表征;解码部分利用转置卷积或插值的方式从低分辨率表征上采样出第二高分辨率表征。

[0014] 进一步地,所述编码部分操作如下:

[0015] 采用高分辨率网络(High-Resolution Net,HRNet)中多分辨率子网络并行连接的机制生成多分辨率表征,并引入多分辨率融合模块(Multi-Resolution Fusion Module)实现多分辨率表征之间的特征信息交换与融合,最终输出多种尺度的第一特征图。

[0016] 进一步地,所述解码部分操作如下:

[0017] 首先采用空洞空间卷积池化金字塔(Atrous Spatial Pyramid Pooling,ASPP)模块对所述第一特征图中最小尺寸的特征图以不同采样间隔的空洞卷积并行采样,后续通过转置卷积的方式分别将其他第一特征图的空间维度扩大两倍,形成与第一特征图数量相同的多个第二特征图;

[0018] 将所述第二特征图与来自编码部分同样尺寸的第一特征图拼接在一起,最后,卷积生成与预处理图像大小尺寸相同的两张掩膜(Mask)预测图像;

[0019] 由此确定表格区域,并区分为有线表格图像和无线表格图像。

[0020] 进一步地,所述步骤3具体包括:

[0021] 步骤31:根据掩膜预测图像利用坎尼边缘检测算子(Canny算子)计算出表格四周的轮廓;

[0022] 步骤32:利用霍夫变换算子检测出轮廓中的所有直线并将部分满足合并条件的直线合并在一起;

[0023] 步骤33:根据所有直线的位置计算出准确的表格位置,由此分离出校正后的、仅包含表格区域的表格区域图像。

[0024] 进一步地,步骤32中,合并条件为:

[0025] 首先判断两条线段是否为平行线,如果为平行线,则计算两条线段的垂直距离,当距离大于一定值时,两条线段不可以合并。如果不为平行线,则计算两条线段的斜率差,当差值大于阈值时,两条线段不可以合并。

[0026] 当满足上述条件后,需要继续判断两条线段是否在某一投影方向有重叠,如果有重叠,则分别计算线段的两个端点到另一个线段的垂直距离,当四个距离中的最小值小于阈值时,两条线段可以合并。如果没有重叠,则分别计算两条线段之间端点的距离,当四个距离中的最小值小于阈值时,两条线段可以合并。

[0027] 进一步地,所述步骤4中,针对属于有线表格图像的所述表格区域图像,具体包括:

[0028] 根据分割线的掩膜预测图像,利用坎尼边缘检测算子计算显式分隔线的轮廓图;

[0029] 利用边界腐蚀算法提取出分隔线的轮廓骨架图;

[0030] 利用霍夫变换算法从轮廓骨架图中计算出所有直线并将部分满足合并条件的直线融合在一起;

[0031] 通过计算出所有横线和竖线相交点的位置,得到表格单元格的位置;

[0032] 提取表格内文本实例的内容和位置;

[0033] 根据表格单元格和文本实例的相对位置计算出表格结构信息并输出。

[0034] 进一步地,所述步骤4中,针对属于无线表格图像的所述表格区域图像,具体包括:

[0035] 以每个文本实例为节点,提取节点特征,所述节点特征由每个文本实例的位置特征、边界框背景特征、行背景特征和列背景特征共同拼接形成;

[0036] 针对某一节点a,选取其当前所在特征空间内所有节点并计算相似性,选取周边多个最近邻节点;

[0037] 将节点a的节点特征A及其与所述多个最近邻节点的节点特征的相似度差值分别拼接在一起,输入经训练的图卷积神经网络,输出更新后的节点特征A' ;

[0038] 重复以上操作,得到所述表格区域图像内所有节点的更新后的节点特征;

[0039] 通过三个多层感知机网络分别利用更新后的节点特征确定节点与多个最近邻节点的行、列和单元格的结构关系,由此确定表格结构信息并输出。

[0040] 进一步地,所述最近邻节点的个数优选为10-15个。

[0041] 进一步地,所述文本实例的位置特征由边界框的左上角和右下角的坐标组成。

[0042] 进一步地,所述文本实例的边界框背景特征、行背景特征和列背景特征通过感兴趣区域池化(Region of Interest Pooling,ROI Pooling)图像特征的提取方式从特征图上提取。

[0043] 这里,所述文本实例是指若干相连的字组成的词、句或段。

[0044] 根据本发明的第二方面,提供一种端到端的表格检测和结构识别装置,其特征在于,所述装置基于以上任一方面所述方法进行操作,所述装置包括以下部件:

[0045] 图像预处理单元,用于对原始图像进行垂直方向拉伸,并进行长宽比不变的尺寸归一化以及边界补0,形成预处理图像;

[0046] 表格区域预测单元,用于以编码器-解码器(Encoder-Decoder)模型为主体结构,确定所述预处理图像中的表格区域,并分类为有线表格图像和无线表格图像;

[0047] 表格图像校正单元,用于基于所确定的表格区域,从预处理图像中分离出校正后的、仅包含表格区域的表格区域图像。

[0048] 表格结构识别单元,用于针对所述表格区域图像,根据分类为有线表格图像和无线表格图像分别采用不同方式进行表格结构识别。

[0049] 根据本发明的第三方面,提供一种端到端的表格检测和结构识别系统,所述系统

包括:处理器和用于存储可执行指令的存储器;其中,所述处理器被配置为执行所述可执行指令,以执行如以上任一方面所述的端到端的表格检测和结构识别方法。

[0050] 根据本发明的第四方面,提供一种计算机可读存储介质,其特征在于,其上存储有计算机程序,所述计算机程序被处理器执行时实现如以上任一方面所述的端到端的表格检测和结构识别方法。

[0051] 本发明的有益效果:

[0052] 1,本发明将自动根据表格检测的结果选择不同的结构识别方法。这一选择机制结合了传统规则算法和深度学习算法的优势,提升了算法的鲁棒性和通用性。

[0053] 2,基于图像分割的表格检测方法可以更加准确的计算出表格边缘,尤其是图像中存在倾斜表格的情况下。像素级别的预测方式也可以尽可能排除非表格内容区域对后续表格结构识别工作的干扰。

[0054] 3,结构识别步骤前的投影变换可以帮助获取单元格排列整齐的表格,降低后续结构识别工作的难度。

[0055] 4,图卷积神经网络通过图卷积的方式融合由文本实例的背景特征、文本实例所在行列的图像特征和文本实例的位置特征组成的节点特征,更加高效的提取图结构中的全局特征及局部特征。更新后的节点特征可以帮助更加准确的预测文本实例之间的结构关系,尤其是当表格中存在合并单元格的情况。

附图说明

[0056] 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。

[0057] 图1为根据本发明实施例的有线表格和无线表格示例图;

[0058] 图2为根据本发明实施例的端到端的表格检测和结构识别方法的算法流程图;

[0059] 图3为根据本发明实施例的端到端的表格检测和结构识别方法的算法结构图;

[0060] 图4为根据本发明实施例的图像预处理变换示例图;

[0061] 图5为根据本发明实施例的Decoder中特征图尺寸变化示意图;

[0062] 图6为根据本发明实施例的表格图像分割结果示例图;

[0063] 图7为根据本发明实施例的表格位置优化算法结构图;

[0064] 图8为根据本发明实施例的表格图像校正结果示例图;

[0065] 图9为根据本发明实施例的表格线提取后处理算法结构图;

[0066] 图10为根据本发明实施例的有线表格分隔线分割结果示例图;

[0067] 图11为根据本发明实施例的节点特征提取平面示意图;

[0068] 图12为根据本发明实施例的节点视觉特征编码的方式示意图;

[0069] 图13为根据本发明实施例的图卷积中节点特征更新示意图;

[0070] 图14为根据本发明实施例的表格结构识别算法(图卷积神经网络)结构图;

[0071] 图15为根据本发明实施例的表结构识别结果示例图。

[0072] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

具体实施方式

[0073] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施中所描述的实施方式并不代表与本公开相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本公开的一些方面相一致的装置和方法的例子。

[0074] 本公开的说明书和权利要求书中的术语“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,以便这里描述的本公开的实施例例如能够以除了在这里图示或描述的那些以外的顺序实施。

[0075] 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。

[0076] 多个,包括两个或者两个以上。

[0077] 和/或,应当理解,对于本公开中使用的术语“和/或”,其仅仅是一种描述关联对象的关联关系,表示可以存在三种关系。例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。

[0078] 本发明提供了一种高效的表格结构一体化识别方法。针对倾斜表格及表格结构多样性的问题,本团队创新性地采用图像分割技术实现表格检测及采用规则与深度学习技术相结合的机制实现结构识别,从而大大提升了算法的鲁棒性和通用性。

[0079] 本发明针对表格识别这一难点问题,考虑到表格对象多风格、多板式以及复杂内部结构等特点,立足于端到端统一化的解决思路,充分结合卷积神经网络图像分割、图卷积神经网络以及传统规则分析方法的优点,实现了从表格检测到表格结构识别的一体化流程,具有数据驱动、统一化以及不依赖于特定表格样式等特点,在各类表格上均取得了较好的效果。如图7、9、14所展示的有线表格和无线表格的处理结果。

[0080] 实施例

[0081] 第一步:图像预处理

[0082] 该步骤对输入的图像进行一系列预处理操作。该图像中包含一张或多张表格。针对大部分表格中存在行之间排列紧凑、行间距较小的特点,为了提高行与行之间的区分度,本设计在这一阶段首先完成垂直方向拉伸变换来增大行与行之间的像素距离。接下来的预处理操作还包括对图像做长宽比不变的尺寸归一化以及边界补0,使得图像的尺寸能够支持神经网络的要求,并最大化的保留全局和局部特征信息。训练时,图像预处理阶段还需要完成必要的增强,如图像仿射变换(Rotation, Shear, Scale等)、颜色扭曲等,使得训练样本的分布更接近于潜在的真实样本分布,以缓解可能存在的数据稀缺性问题,从而提升学习模型的鲁棒性和不变性。本发明还引入膨胀变换作为数据增强的方式,其变换首先将输入图像后转化为二值图像,然后使用 2×2 的核算子对所有像素进行膨胀变换,这种变换可以扩大二值图像中黑像素区域。膨胀变换生成的二值图像不仅可以扩大样本集,还可以模仿黑白表格图像模糊不清的情形,提高模型的鲁棒性。预测阶段,算法则只做图像尺寸的归一化处理。

[0083] 第二步:表格区域预测

[0084] 该步骤利用图像分割技术对图像实现像素级别的分类来定位出表格在图像中的实际所在位置,这种方法相比于基于目标检测结果的图像分割方法,如Mask-RCNN等,不需要在目标检测结果的基础上进行图像分割,避免被检测物体的最小外接矩形的影响,可以在边缘精度上获得更好的优势。

[0085] 算法在该步骤中采用以Encoder-Decoder为主体结构的图像分割算法,其中,Encoder部分负责用卷积的方式从高分辨表征中下采样出低分辨率表征,Decoder部分负责用转置卷积或插值的方式从低分辨率表征上采样出高分辨率表征。本设计创新性地采用HRNet模型作为Encoder结构,HRNet网络中的多分辨率子网络并行连接的机制生成多分辨率表征,一直保持语义丰富的高分辨率表征,避免下采样操作所引起的信息丢失。HRNet模型从高分辨率子网络作为第一阶段开始,逐步增加高分辨率到低分辨率的子网络,引入更多的阶段,并将多分辨率子网络并行连接。同时,模型还引入了多分辨率融合模块(Multi-Resolution Fusion Module)实现多分辨率表征之间的特征信息交换与融合,从而得到更加语义丰富和空间位置准确的高分辨率表征。在多分辨率融合模块中,算法采用卷积核为 3×3 以及步长为2的卷积从高分辨率表征中提取出低分辨率表征和双线性插值的方法从低分辨率表征中恢复出高分辨率表征。本设计的Encoder部分最终生成四种尺度的特征图,空间尺寸分别为原图尺寸的 $1/2$ 、 $1/4$ 、 $1/8$ 和 $1/16$ 。

[0086] 模型的Decoder部分首先利用Atrous Spatial Pyramid Pooling (ASPP) 模块对前一阶段最小尺寸的特征图以不同采样间隔的空洞卷积并行采样,帮助模型捕捉更多尺度的特征信息,其中,卷积算子的kernel size分别为1、3和3,dilation rate分别为1、6和12,并且通过padding的方式保持原输入尺寸的输出。接下来,Decoder逐步通过转置卷积的方式将小尺寸的特征图的空间维度扩大两倍,并将其与来自Encoder同样尺寸的特征图拼接在一起。具体过程图如图5所示,其中 s_2 、 s_4 、 s_8 和 s_{16} 分别为Encoder生成的原图尺寸的 $1/2$ 、 $1/4$ 、 $1/8$ 和 $1/16$ 的特征图

[0087] Decoder部分最终利用 1×1 的卷积生成与原图大小尺寸相同、深度为2的Mask图像,实现像素级别的预测。因为像素点一共可能属于有线表格区域、无线表格区域和非表格区域三种情况,所以该分割模型需要输出两张Mask预测图像,在准确计算出表格区域的同时,分类出有线表格和无线表格。Mask图像中的每个像素点位置上的值位于0到1的范围内,两张Mask图像上的像素值分别代表了当前像素点属于有线表格或无线表格的置信度。

[0088] 该步骤输出的类别结果接下来将会帮助模型自动选择所对应的结构识别方法。

[0089] 第三步:表格图像校正

[0090] 该步骤首先的目标是利用上一步得到的Mask预测图像进行边缘拟合计算出完整的表格区域,然后利用投影变换将表格区域从原始图像中分离出来形成一张新的图片。

[0091] 该步骤首先根据Mask图像计算出表格四周的轮廓,接下来利用霍夫变换算子检测出轮廓中的所有直线并将部分满足合并条件的直线融合在一起,最后根据所有直线的位置计算出准确的表格位置。此步骤的算法结构图如图7所示。

[0092] 此步骤中的投影变换可以确保新的图片中在大多数情况下只包含表格内容,并且表格内的大部分单元格是排列整齐的。这样不仅可以排除原始图像中非表格区域内容的干扰,还可以降低识别任务的难度,进一步提升表格识别的精度。

[0093] 第四步:表格结构识别

[0094] 该步骤利用从第二步得到的表格分类结果,对从第三步骤得到表格区域图像采用针对性的结构识别方式。

[0095] 对于有线表格,算法采用显式分隔线检测的方法。算法首先利用图像分割模型预测出显式表格分隔线的位置和OCR引擎提取出表格内文本实例的内容和位置,然后利用后处理算法来计算出文本实例在表格结构中的位置。后处理算法的详细结构图如图9所示。后处理算法首先根据分隔线Mask图像计算出显式分隔线的轮廓图,然后利用边界腐蚀算法提取出分隔线的轮廓骨架图,接下来利用霍夫变换算法从骨架图中计算出所有直线并将满足合并条件的直线融合在一起,通过计算出所有横线和竖线相交点的位置,推理出单元格的位置,最后根据表格单元格和文本实例的相对位置计算出表格结构信息。

[0096] 另一方面,算法采用图卷积神经网络来处理无线表格。在无线表格处理的方法中,ResNet50作为Feature Extractor来提取输入表格的图像特征,OCR引擎负责提取表格内文本实例的内容和位置。图卷积神经网络中的节点以文本实例为单位,节点特征由文本实例的位置特征、边界框背景特征、行背景特征和列背景特征共同拼接形成。文本实例的位置特征由边界框的左上角和右下角的坐标组成。算法引入RoIPooling图像特征的提取方式从特征图上文本实例相对应的位置获取对应的边界框背景特征、行背景特征和列背景特征,如图11所示。

[0097] RoIPooling可以获取文本实例相应位置上固定尺寸大小的feature map,然后算法沿着宽和高两个维度对feature map做全局平均只保留特征图深度维度的特征。

[0098] 图卷积神经网络并不会像普通卷积那样受限于Grid-Structure Data,如图像,中固定的邻接关系,而是会依据特征相似度的度量方式,在图结构中选择和当前节点相似的节点进行特征提取和更新。在多层图卷积网络中,随着节点特征的更新,每个节点的近邻节点在不断变化,一定程度上增大了网络的有效感受野。

[0099] 在本发明中,算法采用欧几里得距离作为图卷积操作中节点特征相似度的度量方式,分别计算出每个节点的多个最近邻节点,然后将节点特征以及节点特征和每个邻居节点特征的差值分别拼接在一起输入到全连接网络中进行特征提取,如图13所示。每一层的图卷积网络层都会将当前节点与所有近邻节点特征的提取结果平均融合在一起作为更新后的节点特征。在本设计算法所采用的图卷积操作中,节点特征代表图结构中的全局特征,节点特征和近邻节点特征的差值代表图结构中的局部特征,多层图卷积操作可以帮助算法充分提取图结构中的全局及局部特征。最终,算法使用的三个多层感知机网络(分类器)分别利用更新后的节点特征判断节点之间的行、列和单元格的结构关系。

[0100] 图14是基于图卷积神经网络的表格结构识别算法的结构图,图15是表格结构识别结果示例图。

[0101] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。

[0102] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0103] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端(可以是手机,计算机,服务器,空调器,或者网络设备等)执行本发明各个实施例所述的方法。

[0104] 上面结合附图对本发明的实施例进行了描述,但是本发明并不局限于上述的具体实施方式,上述的具体实施方式仅仅是示意性的,而不是限制性的,本领域的普通技术人员在本发明的启示下,在不脱离本发明宗旨和权利要求所保护的范围情况下,还可做出很多形式,这些均属于本发明的保护之内。

项目名称	规格	单位	数量	金额	医保类别
复方嗜酸乳杆菌片	0.5g×18片/盒	盒	1	31.01	基本
精度费	1.00次	次	1	-0.02	基本
小儿伪麻美芬滴剂(艾畅)	15ml/瓶	瓶	1	12.03	基本
氯苯那敏片(扑尔敏)	4mg×100片/瓶	片	6	0.46	基本
孟鲁司特钠咀嚼片(顺尔宁)	4mg×5片/盒	盒	1	27.82	自费

项目名称	价格	数量	执行科室	执行医生
换药(大)	40元	1	伤口治疗中心	
水凝胶创伤敷料(吉原 HD-M 25X10CM)	250.8元	5	伤口治疗中心	
液体敷料(3M 3344)	17.6元	7	伤口治疗中心	

项目名称	规格	单位	数量	单价	金额	自费比例	费用等级
120400001	皮下皮内肌肉注射	次	2.00	5.00000	10.00	0.00	甲类
1206c9	换药用医用创面愈合材料	个	1.00	73.50000	73.50	1.00	自费
3314a	特殊脐带夹	个	1.00	102.90000	102.90	1.00	自费

费用小计: 779.49

项目编号	服务项目名称/规格	单价	数量	单位	金额	备注
220301001	彩色多普勒超声常规检查	108	1	次	108	
220302010	脏器声学造影 造影剂另计	108	1	次	108	
220800008	超声计算机图文报告	15	1	次	15	
B超费					231	
实收金额:					231	总金额: 231

有线表格

无线表格

图1

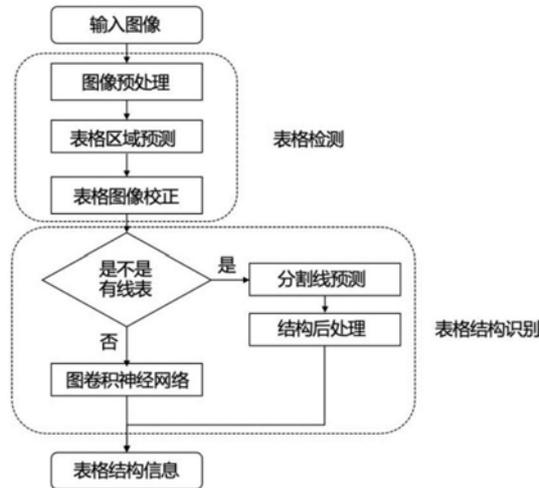


图2

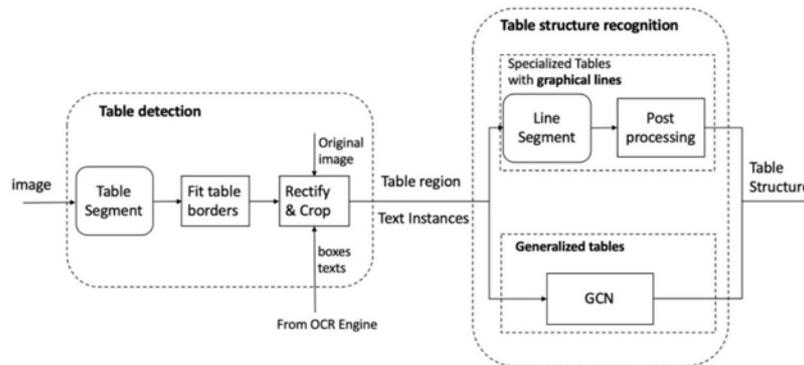
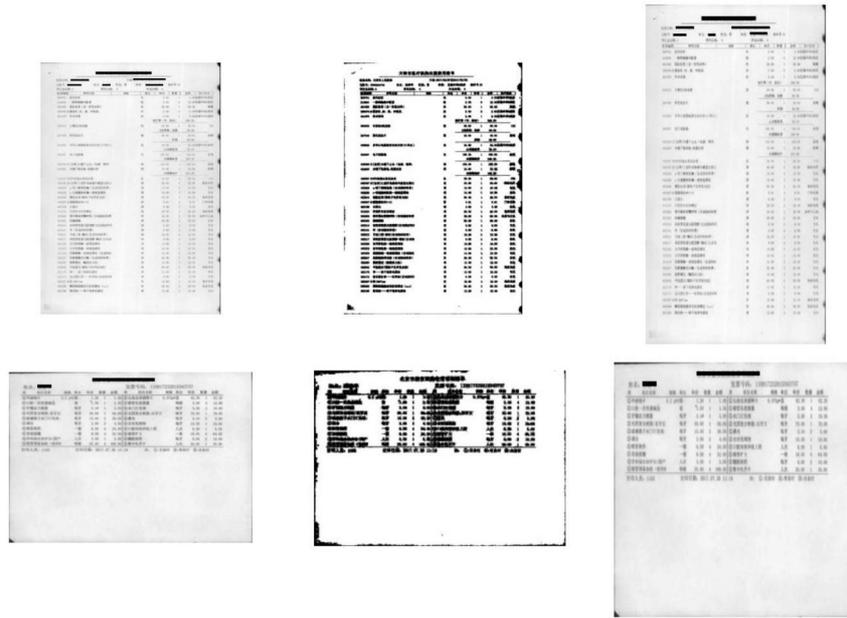


图3



原始图片

膨胀变换

垂直拉伸变换

图4

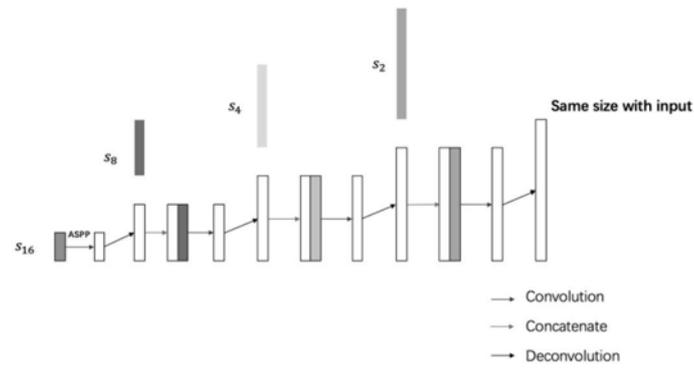


图5

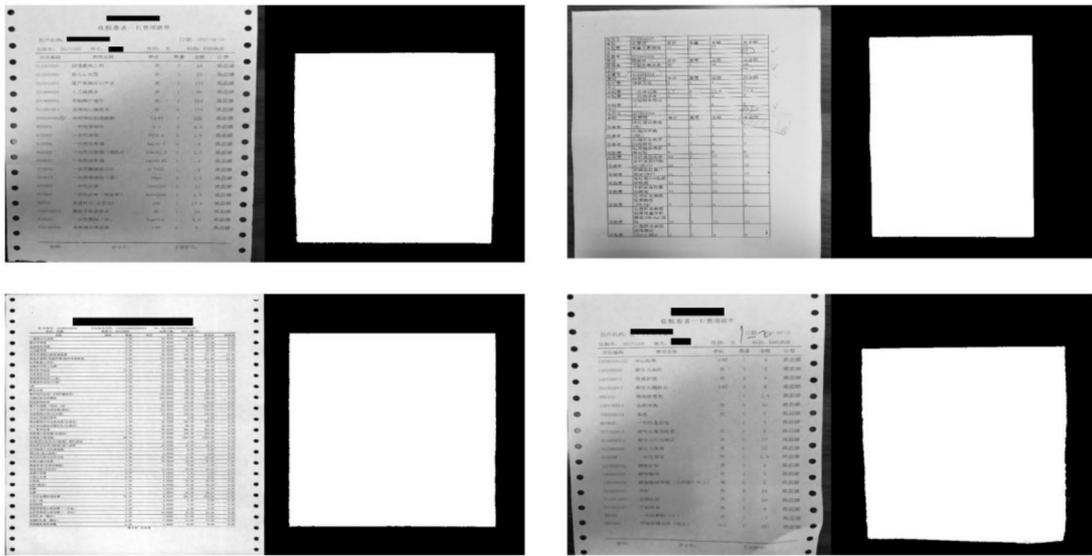


图6



图7

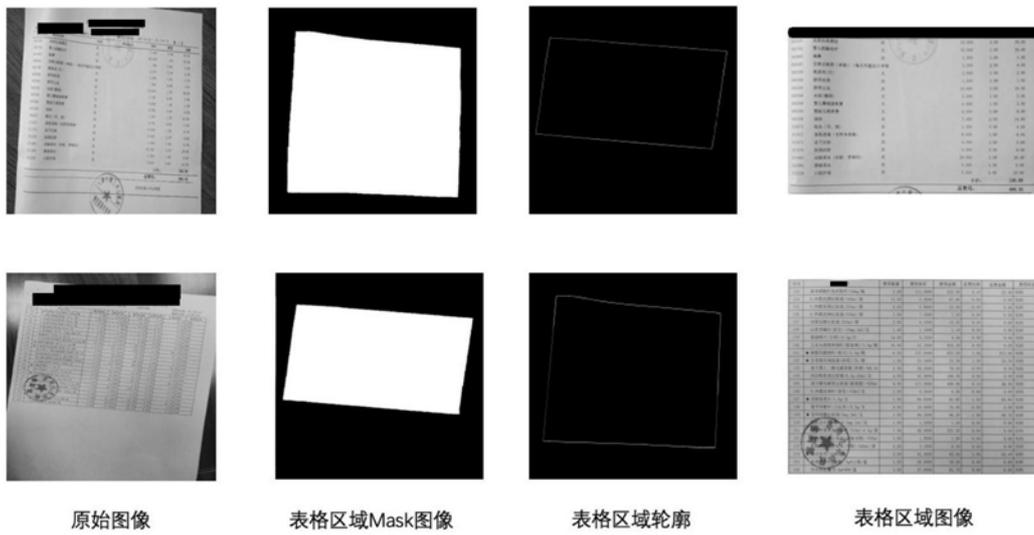


图8



图9

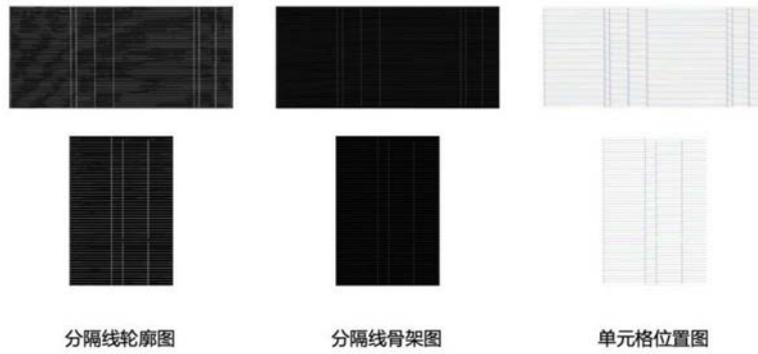


图10

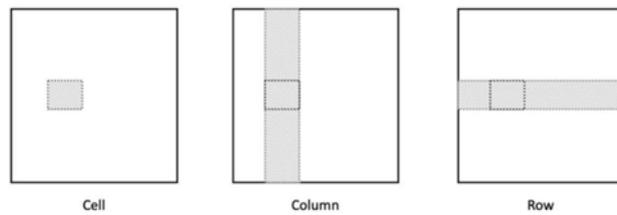


图11

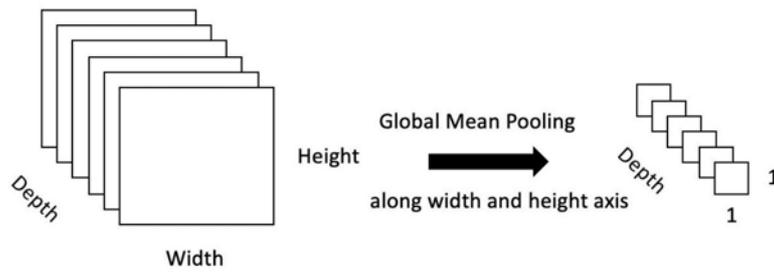


图12

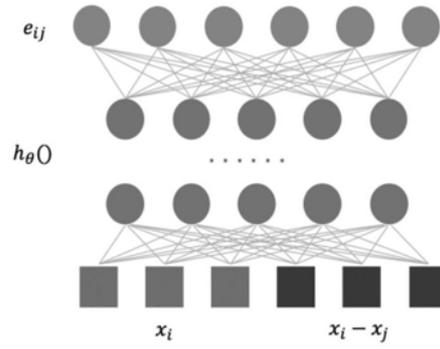


图13

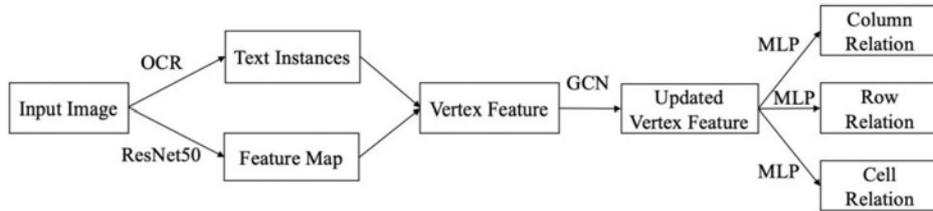


图14

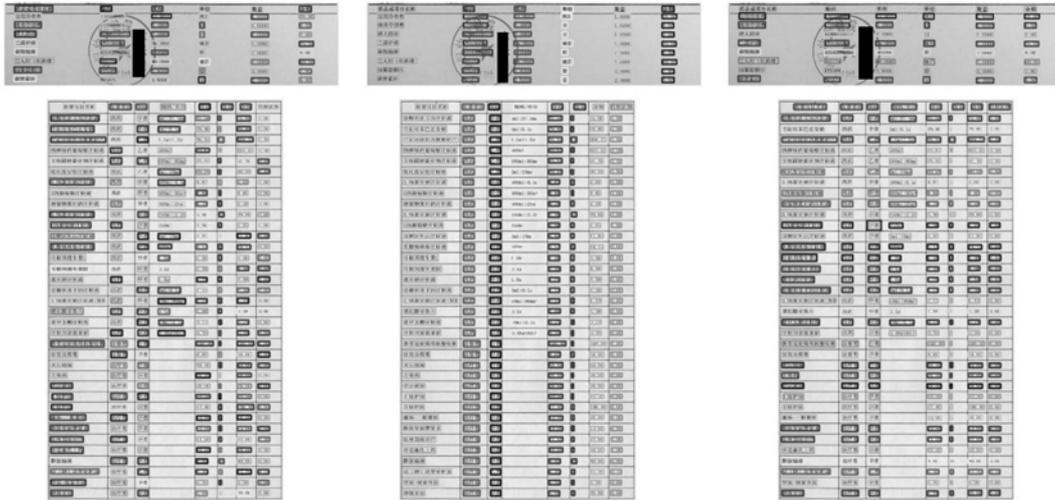


图15