



(12) 发明专利

(10) 授权公告号 CN 114387608 B

(45) 授权公告日 2022. 06. 21

(21) 申请号 202210293274.9

(22) 申请日 2022.03.24

(65) 同一申请的已公布的文献号
申请公布号 CN 114387608 A

(43) 申请公布日 2022.04.22

(73) 专利权人 华南理工大学
地址 510000 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 杨帆

(74) 专利代理机构 东莞卓诚专利代理事务所
(普通合伙) 44754
专利代理师 朱鹏

(51) Int.Cl.

G06V 30/412 (2022.01)
G06V 30/414 (2022.01)
G06V 30/19 (2022.01)
G06V 30/26 (2022.01)
G06K 9/62 (2022.01)
G06N 3/04 (2006.01)
G06N 3/08 (2006.01)
G06F 40/151 (2020.01)
G06F 40/143 (2020.01)

审查员 谭岳峰

权利要求书3页 说明书8页 附图1页

(54) 发明名称

一种联合卷积与图神经网络的表格结构识别方法

(57) 摘要

本发明公开了一种联合卷积与图神经网络的表格结构识别方法,其特征在于,所述方法包括:构建联合卷积图神经网络,所述的联合卷积图神经网络包括深度二维卷积神经网络、二维卷积神经网络和图神经网络;利用不确定度加权的
多任务学习损失函数训练所述的联合卷积图神经网络;使用二维卷积神经网络预测表格各单元格中心点坐标,基于中心点构建表格结构的图,利用所述的图神经网络对图的每条边进行分类;后处理算法处理分好类的表格结构的图,得到各单元格的结构属性,并转换为描述表格结构的标记语言。本发明方法提供一种端到端可训练的基于图神经网络显示建模表格结构的方法来高效识别表格结构。

构建联合卷积图神经网络,所述的联合卷积神经网络包括深度二维卷积神经网络、二维卷积神经网络和图神经网络,所述的深度二维卷积神经网络输出的图像特征,作为所述的二维卷积神经网络和所述的图神经网络的输入

利用不确定度加权的
多任务学习损失函数训练所述的
联合卷积图神经网络

所述的深度二维卷积神经网络用于提取表格图像高维语义特征,使用二维卷积神经网络预测表格各单元格中心点坐标,基于中心点构建表格结构的图,利用所述的图神经网络对图的每条边进行分类

后处理算法处理分好类的表格结构的图,得到各单元格的结构属性,并转换为描述表格结构的标记语言

1. 一种联合卷积与图神经网络的表格结构识别方法,其特征在于,所述方法包括:

构建联合卷积图神经网络,所述的联合卷积图神经网络包括深度二维卷积神经网络、二维卷积神经网络和图神经网络,所述的深度二维卷积神经网络输出的图像特征,作为所述的二维卷积神经网络和所述的图神经网络的输入;

利用不确定度加权的多任务学习损失函数训练所述的联合卷积图神经网络;

所述的深度二维卷积神经网络用于提取表格图像高维语义特征,使用二维卷积神经网络预测表格各单元格中心点坐标,基于中心点构建表格结构的图,利用所述的图神经网络对图的每条边进行分类;

后处理算法处理分好类的表格结构的图,得到各单元格的结构属性,并转换为描述表格结构的标记语言。

2. 如权利要求1所述的一种联合卷积与图神经网络的表格结构识别方法,其特征在于,所述的深度二维卷积神经网络 Ξ_{cnn} 为关键点识别深度卷积神经网络HRNet的前3个阶段:

$$\Xi_{cnn}(x) = stage_3(stage_2(stage_1(x))),$$

其中, x 为输入3通道RGB图像, $stage_i$ 为HRNet网络中第*i*个包含图像不同分辨率特征提取分支的卷积模块;

所述的二维卷积神经网络为 Ξ_{kpt} 为关键点识别深度卷积神经网络HRNet的第4阶段部分:

$$\Xi_{kpt}(\Xi_{cnn}(x)) = stage_4(\Xi_{cnn}(x)).$$

3. 如权利要求2所述的一种联合卷积与图神经网络的表格结构识别方法,其特征在于,所述的图神经网络 Ξ_{gnn} 采用图注意力网络:

$$\Xi_{gnn}(g) = GAT(g),$$

其中, g 为基于中心点构建的表达表格结构的图, $GAT()$ 表示图注意力网络。

4. 如权利要求2所述的一种联合卷积与图神经网络的表格结构识别方法,其特征在于,所述图神经网络 Ξ_{gnn} 采用图卷积神经网络:

$$\Xi_{gnn}(g) = GCN(g),$$

其中, g 为基于中心点构建的表达表格结构的图, $GCN()$ 表示图卷积神经网络。

5. 如权利要求3或4所述的一种联合卷积与图神经网络的表格结构识别方法,其特征在于,基于中心点构建表达表格结构的图,包括如下步骤:

计算CNN特征张量 $h_{feat} = stage'_4(\Xi_{cnn}(x))$,其中 $stage'_4$ 为与 $stage_4$ 结构相同,权重不共享的卷积模块;

对每个顶点采用K-NN算法,构建图 g 的边;

根据各中心点坐标 (h, w) 索引CNN特征张量 h_{kpt} 得到图每个顶点的CNN特征向量

$h_{feat}(h, w, :)$, 其中 h, w 为图像高与宽方向上的坐标;

拼接CNN特征向量 $h_{feat}(h, w, :)$ 与中心点坐标及中心点相对坐标, 得到每个顶点的特征向量 f_n , 其中 n 为顶点编号;

根据边两个顶点的CNN特征向量, 取均值得边的CNN特征向量, 并拼接两顶点的 x 轴绝对与相对距离、 y 轴绝对与相对距离和绝对与相对欧式距离, 得到每个边的特征向量 v_m , 其中 m 为边的编号:

$$v_m = (|w_1 - w_2|, |w_1 - w_2|/W, |h_1 - h_2|, |h_1 - h_2|/H, \sqrt{(w_1 - w_2)^2 + (h_1 - h_2)^2}, \sqrt{(w_1 - w_2)^2 + (h_1 - h_2)^2}/\max\{W, H\})$$

其中 H 和 W 分别为图像的高与宽, w_1 与 w_2 表示两个顶点 x 轴坐标分量, h_1 与 h_2 表示两个顶点 y 轴坐标分量, 由此获得图 $g = (\{f_n\}, \{v_m\}, A)$, 其中 A 为图 g 的邻接矩阵。

6. 如权利要求5所述的一种联合卷积与图神经网络的表格结构识别方法, 其特征在于, 所述的训练联合卷积图神经网络的过程中, 单元格中心点坐标预测网络的损失函数 l_{kpt} 采用均方误差损失:

$$l_{kpt} = \frac{1}{W_{map} \cdot H_{map}} \sum_{i=1}^H \sum_{j=1}^W [h_{kpt}(i, j) - gt_{kpt}(i, j)]^2,$$

其中 $h_{kpt} = \Xi_{kpt}(\Xi_{cnn}(x))$ 为单元格中心点坐标预测网络的输出, 为二阶张量, W_{map}, H_{map} 分别为 h_{kpt} 的宽与高, gt_{kpt} 为各单元格中心点对应的热力图标签;

图分类任务的损失函数 l_{gcls} 采用交叉熵损失函数:

$$l_{gcls} = -\frac{1}{N_e} \sum_{j=1}^{N_e} \sum_{c=1}^3 gt_g(c, j) \log[h_{gnn}(c, j)]$$

其中 $h_{gnn} = \Xi_{gnn}(g)$ 为图神经网络的输出, 为二阶张量, 高等于类别数3, 宽为边的个数 N_e , gt_g 为每条边的类别标签, 3种类别分别是相邻同行、相邻同列和不相邻; $h_{gnn}(c, j)$ 表示 j 边是 c 类别的概率, c 为类别标签, j 为图的边序号;

总损失函数 l_{total} 采用如下以任务不确定度倒数平方 σ_{kpt}^{-2} 和 σ_{gcls}^{-2} 为权重的 l_{kpt} 和 l_{gcls} 的加权求和形式:

$$l_{total} = \frac{1}{\sigma_{kpt}^2} l_{kpt} + \frac{1}{\sigma_{gcls}^2} l_{gcls} + \log \sigma_{kpt} + \log \sigma_{gcls};$$

以最小化 l_{total} 为目标, 利用Adam优化算法数值求解 σ_{kpt} 、 σ_{gcls} 和 Ξ_{kpt} 、 Ξ_{cnn} 、 Ξ_{gnn} 网络参数的最优解。

7. 如权利要求6所述的一种联合卷积与图神经网络的表格结构识别方法, 其特征在于, 所述的预测表格各单元格中心点坐标, 包括以下步骤:

二值化单元格中心点坐标预测网络 Ξ_{kpt} 的输出 h_{kpt} ;

寻找二值图 h_{kpt} 中所有连通区域的轮廓;

利用轮廓中心矩求解每个轮廓的几何中心得到各单元格中心点的预测坐标。

8. 如权利要求7所述的一种联合卷积与图神经网络的表格结构识别方法,其特征在于,所述利用图神经网络 Ξ_{gnn} 对构建的图 g 的每条边进行分类,具体包括,第 j 条边的分类预测结果为 $\arg\max_c h_{gnn}(c, j)$ 。

9. 如权利要求8所述的一种联合卷积与图神经网络的表格结构识别方法,其特征在于,所述各单元格的结构属性,包括起始行、列号和结束行列、列号。

10. 如权利要求9所述的一种联合卷积与图神经网络的表格结构识别方法,其特征在于,所述后处理算法,包括以下步骤:

对分好类的图 g 的每个顶点按行进行排序,对于跨多行的节点,将行号取值为所有左同行相邻顶点中 y 值最小的顶点的行号,得到每个单元格起始行号;

对图 g 每个顶点按行进行排序,对于跨多行的节点,将行号取值为所有左同行相邻顶点中 y 值最大的顶点的行号,得到每个单元格结束行号;

对图 g 每个顶点按列进行排序,对于跨多列的节点,将列号取值为所有下同列相邻顶点中 x 值最小的顶点的列号,得到每个单元格起始列号;

对图 g 每个顶点按列进行排序,对于跨多列的节点,将列号取值为所有下同列相邻顶点中 x 值最大的顶点的列号,得到每个单元格结束列号。

一种联合卷积与图神经网络的表格结构识别方法

技术领域

[0001] 本发明属于图像处理及模式识别技术领域,尤其涉及一种联合卷积与图神经网络的表格结构识别方法。

背景技术

[0002] 表格是易于人理解的,强有力的信息展现工具,常出现在纸张照片或电子扫描图片形式的文档中,用于直观展示信息。但图像中的表格为非结构化数据,无法直接被计算机解析理解。而且表格存在共用行列单元格的情况,具有复杂的结构,加大了表格结构解析算法的开发难度。这些困难阻碍了文档中表格信息自动解析、识别及数字化技术的发展,因此亟需发明具有高精度识别图像中表格结构的计算机算法,以解决上述问题。

[0003] 深度学习技术在表格结构识别任务上取得长足进展,目前基于深度学习的表格结构识别方法可以分为三种:基于分割及后处理的方式、基于图像到序列端到端的方式和基于图神经网络的方式。基于分割的方式得到的是各单元格、行或列的位置信息,还须依赖进一步的复杂后处理才能得到表格结构,这增加了算法的步骤以及引入了复杂后处理出错时新的识别误差。基于图像到序列端到端的方式,虽然省略了后续的处理步骤,一步到位地实现了图像结构的标记语言(如Latex、HTML等)的预测输出,但是现有模型没有显示地利用表格结构天然的逻辑关系信息,限制了模型的精度上限。基于图神经网络的方式,根据表格单元格结构关系,显示地建立图后应用图神经网络对图顶点、边进行分类或回归得到各单元格结构属性,实现表格结构的预测,但是该方法通常需要使用文本检测算法预先得到单元格各文本行的区域,然后再利用这些区域坐标来构建图,而非端到端可训练的方式,这导致一方面降低了模型的识别效率,另一方面没有在训练过程中结合单元格分割任务及图分类任务的标签信息以提高精度。

[0004] 综上所述,现有表格结构识别方法还没有能显示建模表格结构的端到端识别方法,以同时提高模型识别精度及效率。因此,需要提供一种端到端可训练的基于图神经网络显示建模表格结构的方法来识别表格结构。

发明内容

[0005] 有鉴于此,有必要针对上述技术问题,提供一种联合卷积与图神经网络的表格结构识别方法,所述方法提供了一种端到端可训练的基于图神经网络显示建模表格结构的方式,能够高精度、高效地识别表格结构。

[0006] 一种联合卷积与图神经网络的表格结构识别方法,包括以下步骤:

[0007] 步骤1,构建联合卷积图神经网络,所述的联合卷积图神经网络包括深度二维卷积神经网络、二维卷积神经网络和图神经网络,所述的深度二维卷积神经网络输出的图像特征,作为所述的二维卷积神经网络和所述的图神经网络的输入;

[0008] 步骤2,利用不确定度加权的任务学习损失函数训练所述的联合卷积图神经网络;

[0009] 步骤3,所述的深度二维卷积神经网络用于提取表格图像高维语义特征,使用二维卷积神经网络预测表格各单元格中心点坐标,基于中心点构建表格结构的图,利用所述的图神经网络对图的每条边进行分类;

[0010] 步骤4,后处理算法处理分好类的表格结构的图,得到各单元格的结构属性,并转换为描述表格结构的标记语言。

[0011] 具体地,所述的深度二维卷积神经网络 Ξ_{cnn} 为关键点识别深度卷积神经网络HRNet的前3个阶段:

[0012] $\Xi_{cnn}(x) = stage_3(stage_2(stage_1(x)))$,

[0013] 其中, x 为输入3通道RGB图像, $stage_i$ 为HRNet网络中第*i*个包含图像不同分辨率特征提取分支的卷积模块;

[0014] 所述的二维卷积神经网络为 Ξ_{kpt} 为关键点识别深度卷积神经网络HRNet的第4阶段部分:

[0015] $\Xi_{kpt}(\Xi_{cnn}(x)) = stage_4(\Xi_{cnn}(x))$ 。

[0016] 可选地,所述的图神经网络 Ξ_{gnn} 采用图注意力网络:

[0017] $\Xi_{gnn}(g) = GAT(g)$,

[0018] 其中, g 为基于中心点构建的表达表格结构的图, $GAT()$ 表示图注意力网络。

[0019] 可选地,所述图神经网络 Ξ_{gnn} 采用图卷积神经网络:

[0020] $\Xi_{gnn}(g) = GCN(g)$,

[0021] 其中, g 为基于中心点构建的表达表格结构的图, $GCN()$ 表示图卷积神经网络。

[0022] 进一步地,所述的基于中心点构建表格结构的图,包括如下步骤:

[0023] 计算CNN特征张量 $h_{feat} = stage'_4(\Xi_{cnn}(x))$,其中 $stage'_4$ 为与 $stage_4$ 结构相同,权重不共享的卷积模块;

[0024] 对每个顶点采用K-NN算法,构建图 g 的边;

[0025] 根据各中心点坐标 (h, w) 索引CNN特征张量 h_{kpt} 得到图每个顶点的CNN特征向量 $h_{feat}(h, w, :)$,其中 h, w 为图像高与宽方向上的坐标;

[0026] 拼接CNN特征向量 $h_{feat}(h, w, :)$ 与中心点坐标及中心点相对坐标,得到每个顶点的特征向量 f_n ,其中 n 为顶点编号;

[0027] 根据边两个顶点的CNN特征向量,取均值得边的CNN特征向量,并拼接两顶点的 x 轴绝对与相对距离、 y 轴绝对与相对距离和绝对与相对欧式距离,得到每个边的特征向量 v_m ,其中 m 为边的编号:

$$v_m = (|w_1 - w_2|, |w_1 - w_2|/W, |h_1 - h_2|, |h_1 - h_2|/H, \\ [0028] \quad \frac{\sqrt{(w_1 - w_2)^2 + (h_1 - h_2)^2}}{\sqrt{(w_1 - w_2)^2 + (h_1 - h_2)^2 / \max\{W, H\}}})$$

[0029] 其中 H, W 分别为图像的高与宽,由此获得图 $g = (\{f_n\}, \{v_m\}, A)$,其中 A 为图 g 的邻接矩阵。

[0030] 更进一步地,所述的训练联合卷积图神经网络的过程中,单元格中心点坐标预测网络的损失函数 l_{kpt} 采用均方误差损失:

$$[0031] \quad l_{kpt} = \frac{1}{W \cdot H} \sum_{i=1}^H \sum_{j=1}^W [h_{kpt}(i, j) - gt_{kpt}(i, j)]^2,$$

[0032] 其中 $h_{kpt} = \Xi_{kpt}(\Xi_{cnn}(x))$ 为单元格中心点坐标预测网络的输出,为二阶张量, W, H 分别为 h_{kpt} 的宽与高, gt_{kpt} 为各单元格中心点对应的热力图标签;

[0033] 图分类任务的损失函数 l_{gcls} 采用交叉熵损失函数:

$$[0034] \quad l_{gcls} = -\frac{1}{N_e} \sum_{j=1}^{N_e} \sum_{c=1}^3 gt_g(c, j) \log[h_{gmn}(c, j)]$$

[0035] 其中 $h_{gmn} = \Xi_{gmn}(g)$ 为图神经网络的输出,为二阶张量,高等于类别数3,宽为边的个数 N_e , gt_g 为每条边的类别标签,3种类别分别是相邻同行、相邻同列和不相邻; $h_{gmn}(c, j)$ 表示 j 边是 c 类别的概率, c 为类别标签, j 为图的边序号;

[0036] 总损失函数 l_{total} 采用如下以任务不确定度倒数平方 σ^{-2}_{kpt} 和 σ^{-2}_{gcls} 为权重的 l_{kpt} 和 l_{gcls} 的加权求和形式:

$$[0037] \quad l_{total} = \frac{1}{\sigma^2_{kpt}} l_{kpt} + \frac{1}{\sigma^2_{gcls}} l_{gcls} + \log \sigma_{kpt} + \log \sigma_{gcls};$$

[0038] 以最小化 l_{total} 为目标,利用Adam优化算法数值求解 σ_{kpt} 、 σ_{gcls} 和 Ξ_{kpt} 、 Ξ_{cnn} 、 Ξ_{gmn} 网络参数的最优解。

[0039] 优选地,所述的预测表格各单元格中心点坐标,包括以下步骤:

[0040] 二值化单元格中心点坐标预测网络 Ξ_{kpt} 的输出 h_{kpt} ;

[0041] 寻找二值图 h_{kpt} 中所有连通区域的轮廓;

[0042] 利用轮廓中心矩求解每个轮廓的几何中心得到各单元格中心点的预测坐标。

[0043] 具体地,所述利用图神经网络 Ξ_{gmn} 对构建的图 g 的每条边进行分类,具体包括,第 j 条边的分类预测结果为 $\arg\max_c h_{gmn}(c, j)$ 。

[0044] 具体地,所述各单元格的结构属性,包括起始行、列号和结束行列、列号。

[0045] 更进一步地,所述后处理算法,包括以下步骤:

[0046] 对分好类的图 g 的每个顶点按行进行排序,对于跨多行的节点,将行号取值为所有左同行相邻顶点中 y 值最小的顶点的行号,得到每个单元格起始行号;

[0047] 对 g 每个顶点按行进行排序,对于跨多行的节点,将行号取值为所有左同行相邻顶点中 y 值最大的顶点的行号,得到每个单元格结束行号;

[0048] 对 g 每个顶点按列进行排序,对于跨多列的节点,将列号取值为所有下同列相邻顶点中 x 值最小的顶点的列号,得到每个单元格起始列号;

[0049] 对 g 每个顶点按列进行排序,对于跨多列的节点,将列号取值为所有下同列相邻顶点中 x 值最大的顶点的列号,得到每个单元格结束列号。

[0050] 与现有技术相比,本发明的有益效果在于,本发明方法能够同时完成表格各单元格中心点坐标预测任务与表格结构的图分类任务,通过多任务学习方式联合训练中心点预测卷积神经网络与用于表格结构图分类的图神经网络,提高了网络整体预测的性能,能够高精度地对表格各单元格中心点连接构成的图的边关系进行分类,基于边分好类的图通过后处理计算各单元格结构属性,完成表格(全线表、各类省线表)结构的识别。

附图说明

[0051] 图1示出了本发明实施方法的流程示意图。

具体实施方式

[0052] 为了使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明作进一步地详细描述,显然,所描述的实施例仅仅是本发明一部份实施例,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做出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0053] 为了更好地理解本实施例的技术内容,先介绍一下本实施例中涉及到的术语解释。

[0054] 端到端可训练:深度学习提供了一种端到端的学习范式,整个学习的流程并不进行人为的子问题划分,而是完全交给深度学习模型直接学习从原始数据到期望输出的映射。

[0055] 卷积神经网络:卷积神经网络(Convolutional Neural Networks, CNN)是一类包含卷积计算且具有深度结构的前馈神经网络(Feedforward Neural Networks),是深度学习(deep learning)的代表算法之一。

[0056] 图神经网络:图神经网络(GNNs)是一种连接主义模型,它通过图节点之间的消息传递来捕获图的依赖性。

[0057] 图卷积神经网络(GCN, Graph Convolutional Network):模型由图上的卷积操作构成的一种图神经网络。

[0058] 图注意力网络(GAT, Graph Attention Network):模型由自注意力模块构成的一种图神经网络。

[0059] 多任务学习:多任务学习(Multi-task learning)是和单任务学习(single-task learning)相对的一种机器学习方法。在机器学习领域,标准的算法理论是一次学习一个任

务,也就是系统的输出为实数的情况,而多任务学习是一种联合学习,多个任务并行学习,结果相互影响。

[0060] 损失函数:损失函数(loss function)或代价函数(cost function)是将随机事件或其有关随机变量的取值映射为非负实数以表示该随机事件的“风险”或“损失”的函数。在应用中,损失函数通常作为学习准则与优化问题相联系,即通过最小化损失函数求解和评估模型。

[0061] 表格结构:一般用表格各单元格的结构属性来表示。结构属性包括单元格的起始行号、起始列号,结束行号、结束列号。

[0062] 标记语言:是一种将文本以及文本相关的其他信息结合起来,展现出关于文档结构和数据处理细节的电脑文字编码。

[0063] HTML语言:超文本标记语言(HTML,Hyper Text Markup Language)是一种标记语言。它包括一系列标签,通过这些标签可以将网络上的文档格式统一,使分散的Internet资源连接为一个逻辑整体。

[0064] HRNet(High Resolution Net):高分辨率网络模型。

[0065] 图1示出了本发明实施例的流程示意图。一种联合卷积与图神经网络的表格结构识别方法,包括以下步骤:

[0066] 步骤1,构建联合卷积图神经网络,所述的联合卷积图神经网络包括深度二维卷积神经网络、二维卷积神经网络和图神经网络,所述的深度二维卷积神经网络,分别接入所述的二维卷积神经网络和所述的图神经网络;

[0067] 步骤2,利用不确定度加权的多任务学习损失函数训练所述的联合卷积图神经网络;

[0068] 步骤3,所述的深度二维卷积神经网络用于提取表格图像高维语义特征,使用二维卷积神经网络预测表格各单元格中心点坐标,基于中心点构建表格结构的图,利用所述的图神经网络对图的每条边进行分类;

[0069] 步骤4,后处理算法处理分好类的表格结构的图,得到各单元格的结构属性,并转换为描述表格结构的标记语言。

[0070] 下面结合实例具体说明联合卷积与图神经网络的表格结构识别过程。

[0071] 下面结合实例具体说明联合卷积与图神经网络的表格结构识别过程。

[0072] 执行步骤1,构建联合卷积图神经网络,所述的联合卷积图神经网络包括深度二维卷积神经网络、二维卷积神经网络和图神经网络,所述的深度二维卷积神经网络,分别接入所述的二维卷积神经网络和所述的图神经网络。

[0073] 所述的深度二维卷积神经网络 Ξ_{cm} 为关键点识别深度卷积神经网络HRNet的前3个阶段:

[0074] $\Xi_{cm}(x) = stage_3(stage_2(stage_1(x)))$,

[0075] 其中, x 为输入3通道RGB图像,stage_i为HRNet网络中第i个包含图像不同分辨率特征提取分支的卷积模块;

[0076] 所述的二维卷积神经网络为 Ξ_{kpt} 为关键点识别深度卷积神经网络HRNet的第4阶段部分:

[0077] $\Xi_{kpt}(\Xi_{cnn}(x)) = stage_4(\Xi_{cnn}(x))$ 。

[0078] 可选地,所述的图神经网络 Ξ_{gnn} 采用图注意力网络:

[0079] $\Xi_{gnn}(g) = GAT(g)$,

[0080] 其中,g为基于中心点构建的表达表格结构的图, $GAT()$ 表示图注意力网络。

[0081] 可选地,所述图神经网络 Ξ_{gnn} 采用图卷积神经网络:

[0082] $\Xi_{gnn}(g) = GCN(g)$,

[0083] 其中,g为基于中心点构建的表达表格结构的图, $GCN()$ 表示图卷积神经网络。

[0084] 进一步地,所述的基于中心点构建表格结构的图,包括如下步骤:

[0085] 计算CNN特征张量 $h_{feat} = stage'_4(\Xi_{cnn}(x))$,其中 $stage'_4$ 为与 $stage_4$ 结构相同,权重不共享的卷积模块。

[0086] 对每个顶点采用K-NN算法,构建图g的边。首先取K值为20,然后构建以中心点为图顶点的完全图,并计算每条边的欧式距离;下一步,裁剪距离较远的边,以减少图的边数从而提高算法的计算速度,具体地,遍历完全图的每个顶点,遍历到某顶点时,具体的处理方式,为该顶点的每条边按距离从小到大排序,仅保留前20条边(包括20),去除后20条边,当总边数小于20时则保留全部边,遍历地处理完所有顶点后,便完成图g的构建。

[0087] 根据各中心点坐标 (h, w) 索引CNN特征张量 h_{kpt} 得到图每个顶点的CNN特征向量 $h_{feat}(h, w, :)$,其中h,w为图像高与宽方向上的坐标。

[0088] 拼接CNN特征向量 $h_{feat}(h, w, :)$ 与中心点坐标及中心点相对坐标,得到每个顶点的特征向量fn,其中n为顶点编号。

[0089] 根据边两个顶点的CNN特征向量,取均值得边的CNN特征向量,并拼接两顶点的x轴绝对与相对距离、y轴绝对与相对距离和绝对与相对欧式距离,得到每个边的特征向量vm,其中m为边的编号:

$$v_m = (|w_1 - w_2|, |w_1 - w_2|/W, |h_1 - h_2|, |h_1 - h_2|/H, \\ [0090] \quad \frac{\sqrt{(w_1 - w_2)^2 + (h_1 - h_2)^2}}{\sqrt{(w_1 - w_2)^2 + (h_1 - h_2)^2} / \max\{W, H\}})$$

[0091] 其中H,W分别为图像的高与宽,由此获得图 $g = (\{fn\}, \{vm\}, A)$,其中A为图g的邻接矩阵。

[0092] 执行步骤2,利用不确定度加权的任务学习损失函数训练所述的联合卷积图神经网络。

[0093] 所述的训练联合卷积图神经网络的过程中,单元格中心点坐标预测网络的损失函数 l_{kpt} 采用均方误差损失:

$$l_{kpt} = \frac{1}{W \cdot H} \sum_{i=1}^H \sum_{j=1}^W [h_{kpt}(i, j) - gt_{kpt}(i, j)]^2 ,$$

[0095] 其中 $h_{kpt} = \Xi_{kpt}(\Xi_{cnn}(x))$ 为单元格中心点坐标预测网络的输出,为二阶张量,W、H分

别为 $\mathbf{h}_{kpt}$ 的宽与高, $\mathbf{g}t_{kpt}$ 为各单元格中心点对应的热力图标签;

[0096] 图分类任务的损失函数 l_{gcls} 采用交叉熵损失函数:

$$[0097] \quad l_{gcls} = -\frac{1}{N_e} \sum_{j=1}^{N_e} \sum_{c=1}^3 \mathbf{g}t_g(c, j) \log[\mathbf{h}_{gmn}(c, j)]$$

[0098] 其中 $\mathbf{h}_{gmn} = \Xi_{gmn}(\mathbf{g})$ 为图神经网络的输出,为二阶张量,高等于类别数3,宽为边的个数 N_e , $\mathbf{g}t_g$ 为每条边的类别标签,3种类别分别是相邻同行、相邻同列和不相邻;

[0099] 总损失函数 l_{total} 采用如下以任务不确定度倒数平方 σ_{kpt}^{-2} 和 σ_{gcls}^{-2} 为权重的 l_{kpt} 和 l_{gcls} 的加权求和形式:

$$[0100] \quad l_{total} = \frac{1}{\sigma_{kpt}^2} l_{kpt} + \frac{1}{\sigma_{gcls}^2} l_{gcls} + \log \sigma_{kpt} + \log \sigma_{gcls};$$

[0101] 以最小化 l_{total} 为目标,利用Adam优化算法数值求解 σ_{kpt} 、 σ_{gcls} 和 Ξ_{kpt} 、 Ξ_{cm} 、 Ξ_{gmn} 网络参数的最优解。

[0102] 执行步骤3,使用二维卷积神经网络预测表格各单元格中心点坐标,基于中心点构建表格结构的图,利用所述的图神经网络对图的每条边进行分类。

[0103] 所述的预测表格各单元格中心点坐标,包括以下步骤:

[0104] 采用基于自适应阈值的二值化方法大津法,对单元格中心点坐标预测网络 Ξ_{kpt} 的输出 $\mathbf{h}_{kpt}$ 进行二值化,得到二值图;

[0105] 寻找二值图中所有连通区域的轮廓;

[0106] 利用轮廓中心矩求解每个轮廓的几何中心得到各单元格中心点的预测坐标。

[0107] 具体地,所述利用图神经网络 Ξ_{gmn} 对构建的图g的每条边进行分类,具体包括,第j条边的分类预测结果为 $\arg \max_c \mathbf{h}_{gm}(c, j)$ 。

[0108] 具体地,所述各单元格的结构属性,包括起始行、列号和结束行列、列号。

[0109] 更进一步地,所述后处理算法,包括以下步骤:

[0110] 对分好类的图g的每个顶点按行进行排序,对于跨多行的节点,将行号取值为所有左同行相邻顶点中y值最小的顶点的行号,得到每个单元格起始行号;

[0111] 对g每个顶点按行进行排序,对于跨多行的节点,将行号取值为所有左同行相邻顶点中y值最大的顶点的行号,得到每个单元格结束行号;

[0112] 对g每个顶点按列进行排序,对于跨多列的节点,将列号取值为所有下同列相邻顶点中x值最小的顶点的列号,得到每个单元格起始列号;

[0113] 对g每个顶点按列进行排序,对于跨多列的节点,将列号取值为所有下同列相邻顶点中x值最大的顶点的列号,得到每个单元格结束列号。

[0114] 本发明方法的技术原理是:本发明实施例中提出一个联合卷积图神经网络及图神经网络的统一框架,支持端到的训练,从而提高了模型的预测以及训练效率。因为本方法显示地构建了表达表格结构的图,并用图神经网络进行边的分类,因此提升了模型整体识别结构的精度,且同时利用多任务损失函数联合训练了卷积神经网络以及图神经网络,充分

地利用了各任务标签信息,因此进一步提升了识别精度。

[0115] 执行步骤4,后处理算法处理分好类的表格结构的图,得到各单元格的结构属性,并转换为描述表格结构的标记语言。

[0116] 具体的,基于各单元格的结构属性,转换成HTML语法描述的表格(HTML表格)。首先根据所有单元格的起始行号,确定HTML的<tr></tr>标签对个数,得到HTML表格的行数,然后依次根据每行的单元格数量,确定每个<tr></tr>标签对中<td></td>标签对的个数,最后根据结束行/列号与起始行/列号的差值得到各单元格的跨行/列属性值并回填到对应<td></td>标签对的属性赋值区,得到表格结构的完整HTML描述,完成标记语言HTML的转换过程。

[0117] 以上实施例的各技术特征或步骤可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征或步骤所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0118] 本领域技术人员应明白,本申请的实施例可提供为方法、系统或计算机程序产品。因此,本申请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,本申请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。

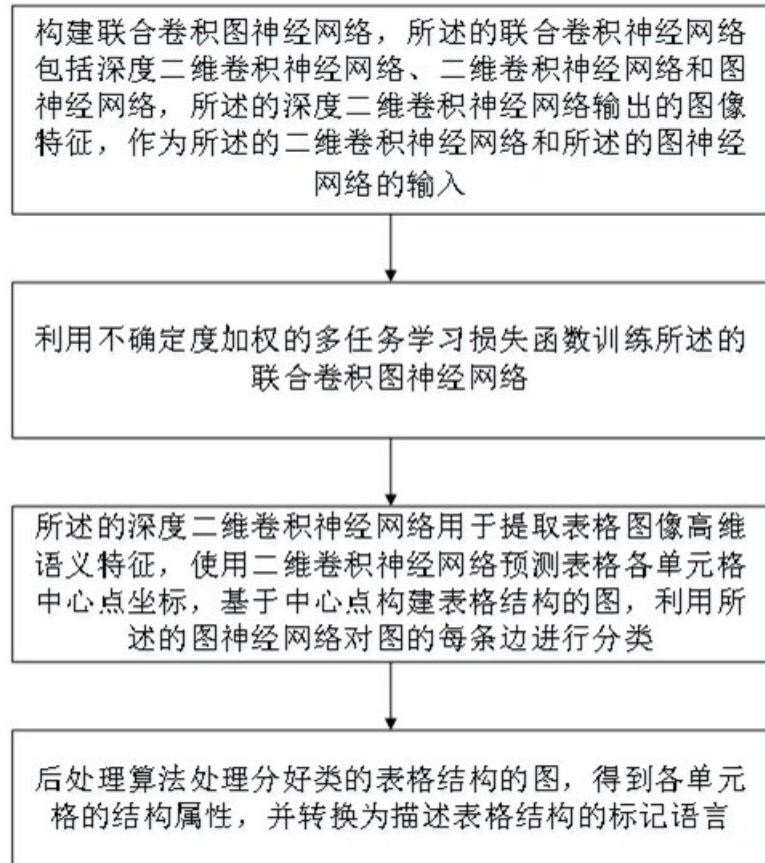


图1