



(12) 发明专利

(10) 授权公告号 CN 113688049 B

(45) 授权公告日 2023. 06. 23

(21) 申请号 202111011124.6

(22) 申请日 2021.08.31

(65) 同一申请的已公布的文献号

申请公布号 CN 113688049 A

(43) 申请公布日 2021.11.23

(73) 专利权人 中国平安财产保险股份有限公司

地址 518000 广东省深圳市福田区益田路

5033号平安金融中心12、13、38、39、40

层

(72) 发明人 郑全威

(74) 专利代理机构 深圳市沃德知识产权代理事

务所(普通合伙) 44347

专利代理师 高杰 于志光

(51) Int. Cl.

G06F 11/36 (2006.01)

G06F 18/22 (2023.01)

G06F 8/20 (2018.01)

G06F 8/38 (2018.01)

G06N 3/0464 (2023.01)

(56) 对比文件

US 2003101336 A1, 2003.05.29

US 2013332387 A1, 2013.12.12

WO 2019144680 A1, 2019.08.01

CN 109446061 A, 2019.03.08

CN 111078552 A, 2020.04.28

US 2018103043 A1, 2018.04.12

审查员 郑银丽

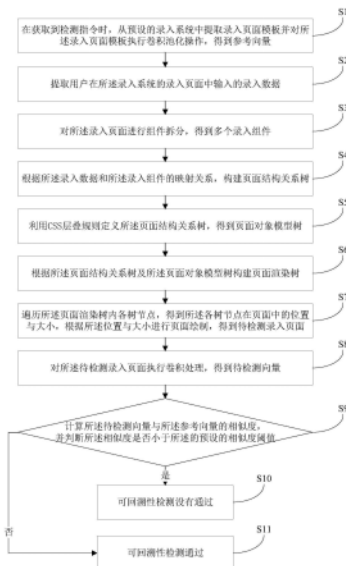
权利要求书2页 说明书12页 附图2页

(54) 发明名称

基于录入信息的可回溯性检测方法、装置、
设备及介质

(57) 摘要

本申请涉及人工智能,提供一种基于录入信息的可回溯性检测方法,包括:获取录入页面模板,计算所述录入页面的参考向量;对录入页面执行组件拆分,并根据用户在录入页面输入的录入数据和所述录入组件的映射关系,构建页面结构关系树;利用CSS层叠规则定义页面结构关系树,并构建页面渲染树;根据页面渲染树进行页面绘制,得到待检测录入页面;计算待检测录入页面的待检测向量;根据参考向量及待检测向量的相似度判断所述录入页面的可回溯性检测是否通过。本发明还涉及区块链技术,所述录入信息模板可以存储在区块链节点中。本发明还提出一种可回溯性检测装置、设备以及介质。本发明可以提高可回溯性检测的效率。



1. 一种基于录入信息的可回溯性检测方法,其特征在于,所述方法包括:

在获取到检测指令时,从预设的录入系统中提取录入页面模板,并对所述录入页面模板执行卷积池化处理,得到参考向量;

提取用户在所述录入系统的录入页面中输入的录入数据;

对所述录入页面进行组件拆分,得到多个录入组件;

根据所述录入数据和所述录入组件的映射关系,构建页面结构关系树;

利用CSS层叠规则定义所述页面结构关系树,得到页面对象模型树;

根据所述页面结构关系树及所述页面对象模型树构建页面渲染树;

遍历所述页面渲染树内各树节点,得到所述各树节点在页面中的位置与大小,根据所述位置与大小进行页面绘制,得到待检测录入页面;

对所述待检测录入页面执行卷积处理,得到待检测向量;

计算所述待检测向量与所述参考向量的相似度,并判断所述相似度是否小于预设相似度阈值;

若所述相似度小于预设相似度阈值,判断所述可回溯性检测没有通过;

若所述相似度大于或等于所述相似度阈值,判断所述可回溯性检测通过;

其中,所述根据所述录入数据和所述录入组件的映射关系,构建页面结构关系树,包括:根据所述录入数据的类别,将所述录入数据拆分成多组页面录入数据;根据多个所述录入组件的属性,将多组所述页面录入数据与多个所述录入组件进行匹配;将成功匹配的多组所述页面录入数据根据所述映射关系嵌入到所对应的多个所述录入组件中,得到所述页面结构关系树;

所述对所述录入页面模板执行卷积池化处理,得到参考向量,包括:根据预构建的卷积神经网络模型的输入格式,调整所述录入页面模板的页面格式;利用所述卷积神经网络模型的卷积核,对调整之后的所述录入页面模板进行卷积处理,得到所述录入页面模板的特征矩阵;将所述特征矩阵进行矩阵特征向量计算,得到所述参考向量。

2. 如权利要求1所述的基于录入信息的可回溯性检测方法,其特征在于,所述根据所述页面结构关系树及所述页面对象模型树构建页面渲染树,包括:

计算所述页面对象模型树中各树节点的样式属性,得到页面样式属性;

去除所述页面结构关系树中各树节点元素中的不可视元素,得到页面可视树节点;

利用所述页面样式属性定义所述页面可视树内各树节点的样式规则,构建所述页面渲染树。

3. 如权利要求1所述的基于录入信息的可回溯性检测方法,其特征在于,所述遍历所述页面渲染树内各树节点,得到所述各树节点在页面中的位置与大小,根据所述位置与大小进行页面绘制,得到待检测录入页面,包括:

从所述页面渲染树根节点进行遍历,得到所述页面渲染树各树节点在页面上的位置与大小;

将所述页面渲染树内各树节点根据所述位置与大小转换成页面像素;

汇集所有所述页面像素,得到所述待检测录入页面。

4. 如权利要求1中所述的基于录入信息的可回溯性检测方法,其特征在于,所述根据预构建的卷积神经网络模型的输入格式,调整所述录入页面模板的页面格式,包括:

根据预构建的卷积神经网络模型的输入格式,对所述录入页面模板执行灰度化处理;

根据所述预构建的卷积神经网络的输入格式确定像素矩阵大小,将所述录入页面模板裁剪成与所述像素矩阵大小相同的图像。

5.如权利要求1至4中任意一项所述的基于录入信息的可回溯性检测方法,其特征在于,所述对所述录入页面进行组件拆分,得到多个录入组件,包括:

获取所述录入页面模板的源代码,并将所述源代码分解成多个录入代码块;

根据所述录入页面模板的组成成分,遍历多个所述录入代码块,提取所述组成成分对应的代码块;

解析所有所述代码块,得到多个所述录入组件。

6.一种基于录入信息的可回溯性检测装置,其特征在于,所述装置包括:

参考页面数据生成模块,用于在获取到检测指令时,从预设的录入系统中提取录入页面模板,并对所述录入页面模板执行卷积池化处理,得到参考向量;

待检测页面数据获取模块,用于提取用户在所述录入系统的录入页面中输入的录入数据,用于对所述录入页面进行组件拆分,得到多个录入组件,根据所述录入数据和所述录入组件的映射关系,构建页面结构关系树,利用CSS层叠规则定义所述页面结构关系树,得到页面对象模型树,根据所述页面结构关系树及所述页面对象模型树构建页面渲染树,遍历所述页面渲染树内各树节点,得到所述各树节点在页面中的位置与大小,根据所述位置与大小进行页面绘制,得到待检测录入页面,对所述待检测录入页面执行卷积处理,得到待检测向量;

可回溯性检测模块,用于计算所述待检测向量与所述参考向量的相似度,并判断所述相似度是否小于预设相似度阈值,若所述相似度小于预设相似度阈值,判断所述可回溯性检测没有通过;若所述相似度大于或等于所述相似度阈值,判断所述可回溯性检测通过;

其中,所述根据所述录入数据和所述录入组件的映射关系,构建页面结构关系树,包括:根据所述录入数据的类别,将所述录入数据拆分成多组页面录入数据;根据多个所述录入组件的属性,将多组所述页面录入数据与多个所述录入组件进行匹配;将成功匹配的多组所述页面录入数据根据所述映射关系嵌入到所对应的多个所述录入组件中,得到所述页面结构关系树;

所述对所述录入页面模板执行卷积池化处理,得到参考向量,包括:根据预构建的卷积神经网络模型的输入格式,调整所述录入页面模板的页面格式;利用所述卷积神经网络模型的卷积核,对调整之后的所述录入页面模板进行卷积处理,得到所述录入页面模板的特征矩阵;将所述特征矩阵进行矩阵特征向量计算,得到所述参考向量。

7.一种电子设备,其特征在于,所述电子设备包括:

至少一个处理器;以及,

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的计算机程序指令,所述计算机程序指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行如权利要求1至5中任一项所述的基于录入信息的可回溯性检测方法。

8.一种计算机可读存储介质,存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至5中任一项所述的基于录入信息的可回溯性检测方法。

基于录入信息的可回溯性检测方法、装置、设备及介质

技术领域

[0001] 本发明涉及人工智能领域,尤其涉及一种基于录入信息的可回溯性检测方法、装置、电子设备及可读存储介质。

背景技术

[0002] 可回溯性检测是对用户在网页页面,如销售页面的操作轨迹予以记录和保存,并基于保存的页面进行用户操作的检测行为。其中,所述操作轨迹包括用户进入和离开所述网页页面的时点、用户在所述网页页面填写或点选的相关内容及相关事件等。目前的可回溯性检测通过在用户操作过程中向服务器上报页面元素,所述服务器将页面元素合成操作过程,并对所述操作过程进行抽样检测。目前的抽样检测都是由人工执行,人工抽样检测存在覆盖范围低、效率低、成本高等资源浪费的问题。

发明内容

[0003] 本发明提供一种基于录入信息的可回溯性检测方法、装置、电子设备及计算机可读存储介质,其主要目的在于提高基于录入信息的可回溯性检测的效率。

[0004] 为实现上述目的,本发明提供的一种基于录入信息的可回溯性检测方法,包括:

[0005] 在获取到检测指令时,从预设的录入系统中提取录入页面模板,并对所述录入页面模板执行卷积池化处理,得到参考向量;

[0006] 提取用户在所述录入系统的录入页面中输入的录入数据;

[0007] 对所述录入页面进行组件拆分,得到多个录入组件;

[0008] 根据所述录入数据和所述录入组件的映射关系,构建页面结构关系树;

[0009] 利用CSS层叠规则定义所述页面结构关系树,得到页面对象模型树;

[0010] 根据所述页面结构关系树及所述页面对象模型树构建页面渲染树;

[0011] 遍历所述页面渲染树内各树节点,得到所述各树节点在页面中的位置与大小,根据所述位置与大小进行页面绘制,得到待检测录入页面;

[0012] 对所述待检测录入页面执行卷积处理,得到待检测向量;

[0013] 计算所述待检测向量与所述参考向量的相似度,并判断所述相似度是否小于预设相似度阈值;

[0014] 若所述相似度小于预设相似度阈值,判断所述可回溯性检测没有通过;

[0015] 若所述相似度大于或等于所述相似度阈值,判断所述可回溯性检测通过。

[0016] 可选地,所述根据所述录入数据和多个所述录入组件的映射关系,构建页面结构关系树,包括:

[0017] 根据所述录入数据的类别,将所述录入数据拆分成多组页面录入数据;

[0018] 根据多个所述录入组件的属性,将多组所述页面录入数据与多个所述录入组件进行匹配;

[0019] 将成功匹配的多组所述页面录入数据根据所述映射关系嵌入到所对应的多个所

述录入组件中,得到所述页面结构关系树。

[0020] 可选地,所述根据所述页面结构关系树以及所述页面对象模型树构建页面渲染树,包括:

[0021] 计算所述页面对象模型树中各树节点的样式属性,得到页面样式属性;

[0022] 去除所述页面结构关系树中各树节点元素中的不可视元素,得到页面可视树节点;

[0023] 利用所述页面样式属性定义所述页面可视树内各树节点的样式规则,构建所述页面渲染树。

[0024] 可选地,所述遍历所述页面渲染树内各树节点,得到所述各树节点在页面中的位置与大小,根据所述位置与大小进行页面绘制,得到待检测录入页面,包括:

[0025] 从所述页面渲染树根节点进行遍历,得到所述页面渲染树各树节点在页面上的位置与大小;

[0026] 将所述页面渲染树内各树节点根据所述位置与大小转换成页面像素;

[0027] 汇集所有所述页面像素,得到所述待检测录入页面。

[0028] 可选地,所述并对所述录入页面模板执行卷积处理,得到参考向量,包括:

[0029] 根据预构建的卷积神经网络模型的输入格式,调整所述录入页面模板的页面格式;

[0030] 利用所述卷积神经网络模型的卷积核,对调整之后的所述录入页面模板进行卷积处理,得到所述录入页面模板的特征矩阵;

[0031] 将所述特征矩阵进行矩阵特征向量计算,得到所述参考向量。

[0032] 可选地,所述根据预构建的卷积神经网络模型的输入格式,调整所述录入页面模板的页面格式,包括:

[0033] 根据预构建的卷积神经网络模型的输入格式,对所述录入页面模板执行灰度化处理;

[0034] 根据所述预构建的卷积神经网络的输入格式确定像素矩阵大小,将所述录入页面模板裁剪成与所述像素矩阵大小相同的图像。

[0035] 可选地,所述对所述录入页面模板进行组件拆分,得到多个录入组件,包括:

[0036] 获取所述录入页面模板的源代码,并将所述源代码分解成多个录入代码块;

[0037] 根据所述录入页面模板的组成成分,遍历多个所述录入代码块,提取所述组成成分对应的代码块;

[0038] 解析所有所述代码块,得到多个所述录入组件。

[0039] 为了解决上述问题,本发明还提供一种基于录入信息的可回溯性检测装置,所述装置包括:

[0040] 参考页面数据生成模块,用于在获取到检测指令时,从预设的录入系统中提取录入页面模板,并对所述录入页面模板执行卷积池化处理,得到参考向量;对所述待检测录入页面执行卷积处理,得到待检测向量;

[0041] 待检测页面数据获取模块,用于提取用户在所述录入系统的录入页面中输入的录入数据,对所述录入页面进行组件拆分,得到多个录入组件,根据所述录入数据和所述录入组件的映射关系,构建页面结构关系树,利用CSS层叠规则定义所述页面结构关系树,得到

页面对象模型树,根据所述页面结构关系树及所述页面对象模型树构建页面渲染树,遍历所述页面渲染树内各树节点,得到所述各树节点在页面中的位置与大小,根据所述位置与大小进行页面绘制,得到待检测录入页面,对所述待检测录入页面执行卷积处理,得到待检测向量;

[0042] 可回溯性检测模块,用于计算所述待检测向量与所述参考向量的相似度,并判断所述相似度是否小于预设相似度阈值,若所述相似度小于预设相似度阈值,判断所述可回溯性检测没有通过;若所述相似度大于或等于所述相似度阈值,判断所述可回溯性检测通过。

[0043] 为了解决上述问题,本发明还提供一种电子设备,所述电子设备包括:

[0044] 存储器,存储至少一个计算机程序;及

[0045] 处理器,执行所述存储器中存储的计算机程序以实现上述所述的基于录入信息的可回溯性检测方法。

[0046] 为了解决上述问题,本发明还提供一种计算机可读存储介质,所述计算机可读存储介质中存储有至少一个计算机程序,所述至少一个计算机程序被电子设备中的处理器执行以实现上述所述的基于录入信息的可回溯性检测方法。

[0047] 本发明实施例提出的基于录入信息的可回溯性检测方法、装置、设备及介质通过组件拆分、构建页面结构关系树构建页面对象模型树、构建页面渲染树、页面绘制等一系列操作,得到待检测录入页面;并通过计算所述待检测录入页面的待检测向量与录入页面模板的参考向量的相似度,进行录入页面的可回溯性检测,可以自动执行对所有录入页面的可回溯性检测,不需要人工抽样,因此可以提高检测效率可回溯性检测的效率。

附图说明

[0048] 图1为本发明一实施例提供的基于录入信息的可回溯性检测方法的流程示意图;

[0049] 图2为本发明一实施例提供的基于录入信息的可回溯性检测装置的模块示意图;

[0050] 图3为本发明一实施例提供的实现基于录入信息的可回溯性检测方法的电子设备的内部结构示意图;

[0051] 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

[0052] 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] 本发明实施例提供一种基于录入信息的可回溯性检测方法。所述基于录入信息的可回溯性检测方法的执行主体包括但不限于服务端、终端等能够被配置为执行本申请实施例提供的该方法的电子设备中的至少一种。换言之,所述基于录入信息的可回溯性检测方法可以由安装在终端设备或服务端设备的软件或硬件来执行,所述软件可以是区块链平台。所述服务端可以是独立的服务器,也可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络(ContentDelivery Network,CDN)、以及大数据和人工智能平台等基础云计算服务的云服务器。

[0054] 参照图1所示的本发明一实施例提供的基于录入信息的可回溯性检测方法的流程

示意图,在本发明实施例中,所述基于录入信息的可回溯性检测方法包括:

[0055] S1、在获取到检测指令时,从预设的录入系统中提取录入页面模板,并对所述录入页面模板执行卷积处理,得到参考向量;

[0056] 本发明实施例中,所述检测指令可以是各平台的后台人员对平台下的某个操作流程进行规范化校验时所输入的操作指令,例如,某电商平台需要对某次网络订单交易的过程进行合法性检测,故在所述电商平台的录入系统中,输入所述检测指令,可提取网络订单交易中各个阶段的页面模板。

[0057] 本发明实施例中,所述预设的录入系统是对网络操作流程中的流程页面模板进行存储的系统,例如,在某电商平台的录入系统中可存储平台用户在产品选择阶段生成的产品选择页面模板、支付阶段生成的支付页面模板。

[0058] 所述录入页面模板可以是所述录入系统中,根据录入页面的要求,将录入信息项使用XXX填充得到的。例如,在某电商平台中得到的校验订单页面有“产品信息:”,则在录入页面模板中“产品信息:”项后面的信息使用XXX作为产品信息填充。

[0059] 详细地,本发明实施例根据预构建的卷积神经网络模型的输入格式,调整所述录入页面模板的页面格式;利用所述卷积神经网络模型的卷积核,对调整之后的所述录入页面模板进行卷积处理,得到所述录入页面模板的特征矩阵;将所述特征矩阵进行矩阵特征向量计算,得到所述参考向量。

[0060] 本发明实施例中,所述预构建的卷积神经网络模型包括输入层、卷积层以及池化层。

[0061] 详细地,本发明实施例根据预构建的卷积神经网络的输入格式,确定像素矩阵大小,并将所述录入页面模板裁剪成与所述像素矩阵大小相同的图像,并对所述图像执行灰度化处理,得到待卷积录入页面模板。

[0062] 进一步地,本发明实施例将所述待卷积录入页面模板输入所述预构建的卷积神经网络模型中;利用所述卷积层对所述待卷积录入页面模板进行卷积操作,得到特征图;对所述特征图进行池化操作,得到特征数据;汇总所有所述特征数据,得到所述录入页面模板的特征矩阵。

[0063] S2、提取用户在所述录入系统的录入页面中输入的录入数据;

[0064] 本发明实施例中,所述录入数据可以是用户在所述录入系统中各录入页面中填写的数据,例如,在电商平台中,录入数据可以用户在选取、购买以及支付等操作环节中勾选和填写的相关数据;在互联网购房领域中,所述录入数据,可以包括但不限于用户打开互联网购房页面、咨询房产人员、个人信息填写以及房源地址信息等。

[0065] S3、对所述录入页面进行组件拆分,得到多个录入组件;

[0066] 详细地,本发明实施例获取所述录入页面的源代码,并将所述源代码分解成多个录入代码块;根据所述录入页面的组成成分,遍历多个所述录入代码块,提取所述组成成分对应的代码块;解析所有所述代码块,得到多个所述录入组件。

[0067] 本发明实施例中,所述代码块可以是根据录入页面的单个功能进行拆分,得到的一段段代码文件。

[0068] 本发明实施例中,所述录入组件可以是能够单独实现页面功能的功能组件,比如,在互联网购房领域中,所述录入组件可以是用户點選进入房源信息进行查看时,出现的房

源信息的“详细地址:”以及“联系电话:”;也可以是在各电商平台下某产品页面中的“详情”按钮,平台用户在产品页面点击“详情”按钮时,会跳转至产品详情页面,单独实现页面的某个特定功能。

[0069] S4、根据所述录入数据和所述录入组件的映射关系,构建页面结构关系树;

[0070] 本发明实施例中,所述映射关系可以表示所述录入数据以及所述录入组件的所属关系,例如,在互联网购房领域中,如录入数据“XX小区XX幢XX号”对应于“详细地址:”,则所述录入数据“XX小区XX幢XX号”与所述“详细地址:”可构建对应关系。

[0071] 本发明实施例中,所述页面结构关系树可以是上述多个所述录入代码块中的主体代码块以及嵌套代码块组成的多层树的关系结构。另外,通俗易懂的来讲,也可以用微机里的文件夹结构来表征,以“我的电脑”为主体,里面嵌套着“C盘”以及“D盘”,每个盘里有着很多的文件夹,文件夹里又嵌套着文件夹,逐层打开文件的过程相当于从根节点遍历到叶子节点的过程,主体与嵌套体构成一棵结构关系树。

[0072] 详细地,所述根据所述录入数据和多个所述录入组件的映射关系,构建页面结构关系树,包括:

[0073] 根据所述录入数据的类别,将所述录入数据拆分成多组页面录入数据;

[0074] 根据多个所述录入组件的属性,将多组所述页面录入数据与多个所述录入组件进行匹配;

[0075] 将成功匹配的多组所述页面录入数据根据所述对应关系嵌入到所对应的多个所述录入组件中,得到所述页面结构关系树。

[0076] 本发明实施例中,所述根据映射关系,将所述录入数据匹配到多组所述录入组件下对应的页面组件中,得到多个重组录入组件可以是将上述录入数据“XX小区XX幢XX号”的信息根据对应关系输入到与所述录入数据对应的录入页面组件“详细地址:”的代码块中,完成匹配的过程。

[0077] S5、利用CSS层叠规则定义所述页面结构关系树,得到CSS对象模型树;

[0078] 本发明实施例中,所述CSS层叠规则可以是HTML中定义样式结构的语言,将网页页面,如将上述互联网购房领域中,将互联网购房页面中的字体、字体大小、字体颜色以及页面组件位置加载进一个网页页面的页面规则文档中,在HTML文件里加入一个超级链接引用所述页面规则文档,则所述网页页面就可按照在所述页面规则文档内定义好的规则进行显示。

[0079] 本发明实施例中,利用所述CSS规则文档定义所述网页页面的页面字体大小、颜色以及页面组件位置等,并将CSS规则文档中的字节内容转换成字符,再依照页面规则组合成属性与对应的属性值的形式,再利用多个所述属性与对应的属性值转换成树节点后链接所述页面结构关系树得到所述CSS对象模型树。

[0080] 本发明实施例中,所述属性与属性值可以是页面规则文档中字体类别、字体大小以及颜色,例如,定义网页的字体种类的属性与属性值为:{font-family:“楷体”,字体大小与颜色为:{font-size:“12px”;color:#666}。

[0081] S6、根据所述页面结构关系树及所述页面对象模型树构建页面渲染树;

[0082] 详细地,本发明实施例计算所述页面对象模型树中各树节点的样式属性,得到页面样式属性;

[0083] 去除所述页面结构关系树中各树节点元素中的不可视元素,得到页面可视树节点;

[0084] 利用所述页面样式属性定义所述页面可视树内各树节点的样式规则,构建所述页面渲染树。

[0085] 本发明实施例中,所述样式属性是所有所述可视属性的属性值,若属性没有被任何匹配的规则定义,则所述样式属性可使用默认值。

[0086] S7、遍历所述页面渲染树内各树节点,得到所述各树节点在页面中的位置与大小,根据所述位置与大小进行页面绘制,得到待检测录入页面;

[0087] 详细地,所述S8,包括:

[0088] 从所述页面渲染树根节点进行遍历,得到所述页面渲染树各树节点在页面上的位置与大小;

[0089] 将所述页面渲染树内各树节点根据所述位置与大小转换成页面像素;

[0090] 汇集所有所述页面像素,得到所述待检测录入页面。

[0091] 本发明实施例中,所述进行页面绘制的应用程序可以是浏览器,包括但不限于是常见的Chrome、Firefox以及IE(Internet Explorer)。

[0092] S8、对所述待检测录入页面执行卷积处理,得到待检测向量;

[0093] 详细地,此处所述对所述待检测录入页面执行卷积处理的过程跟上述S2中对所述录入页面模板执行卷积处理的过程一致,这里不再赘述。

[0094] S9、计算所述待检测向量与所述参考向量的相似度,并判断所述相似度是否小于预设相似度阈值;

[0095] 本发明实施例中,可以采用如下公式计算所述待检测向量以及所述参考向量的相似度;

[0096]
$$Sim(x, y) = \cos(\theta) = \frac{y \cdot x}{|y||x|};$$

[0097] 其中,x为所述参考向量,y为所述待检测向量, θ 为所述参考向量与待检测向量的夹角,Sim(x,y)为所述参考向量以及所述待检测向量的相似度。

[0098] 本发明实施例中,所述预设相似度阈值可以是检测达标指数,若所述相似度小于预设相似度阈值,则所述待检测录入页面录入不合格,比如,比较两张图片的出处是否相同时,若相似度在0.7或以下,则标志着两张图片出处不同。

[0099] 若所述相似度小于预设相似度阈值,则执行S10,判断所述可回溯性检测没有通过;

[0100] 若所述相似度大于或等于所述相似度阈值,则执行S11、判断所述可回溯性检测通过。

[0101] 本发明实施例通过根据用户的录入数据构建待检测录入页面,并计算所述待检测录入页面与对应的录入页面模板之间的相似度,从而执行对所述录入数据的可回溯性检测,提升了录入信息检测的效率,减少了人工抽样检测导致的资源浪费的问题。

[0102] 如图2所示,是本发明基于录入信息的可回溯性检测装置的功能模块图。

[0103] 本发明所述基于录入信息的可回溯性检测装置100可以安装于电子设备中。根据实现的功能,所述基于录入信息的可回溯性检测装置可以包括参考页面数据生成模块101、

待检测页面数据获取模块102、可回溯性检测模块103,本发所述模块也可以称之为单元,是指一种能够被电子设备处理器所执行,并且能够完成固定功能的一系列计算机程序段,其存储在电子设备的存储器中。

[0104] 在本实施例中,关于各模块/单元的功能如下:

[0105] 所述参考页面数据生成模块101用于在获取到检测指令时,从预设的录入系统中提取录入页面模板,并对所述录入页面模板执行卷积池化处理,得到参考向量;

[0106] 本发明实施例中,所述检测指令可以是各平台的后台人员对平台下的某个操作流程进行规范化校验时所输入的操作指令,例如,某电商平台需要对某次网络订单交易的过程进行合法性检测,故在所述电商平台的录入系统中,输入所述检测指令,可在网络订单交易开始以及交易结束期间,提取各个操作阶段的页面模板。

[0107] 本发明实施例中,所述预设的录入系统是对网络操作流程中的操作流程页面模板进行存储的系统,例如,在某电商平台的录入系统中可存储平台用户在选择产品时可参考的选择页面模板、支付阶段可参考的支付页面模板。

[0108] 本发明实施例中,所述录入页面模板可以是所述录入系统中,根据录入页面的要求,将录入信息项使用XXX填充得到的。例如,在某电商平台中得到的校验订单页面有“产品信息:”,则在录入页面模板中“产品信息:”项后面的信息使用XXX作为产品信息填充。

[0109] 本发明实施例中,为了更加精确的得到所述录入页面的全部特征,所述参考页面数据生成模块101需要对所述录入页面利用预构建的卷积神经网络模型执行图像卷积处理。所述卷积神经网络模型通常包括输入层、卷积层以及池化层。

[0110] 详细地,本发明实施例根据预构建的卷积神经网络模型的输入格式,调整所述录入页面模板的页面格式;利用所述卷积神经网络模型的卷积核,对调整之后的所述录入页面模板进行卷积处理,得到所述录入页面模板的特征矩阵;将所述特征矩阵进行矩阵特征向量计算,得到所述参考向量。

[0111] 本发明实施例中,所述预构建的卷积神经网络模型包括输入层、卷积层以及池化层。

[0112] 详细地,本发明实施例根据预构建的卷积神经网络的输入格式,确定像素矩阵大小,并将所述录入页面模板裁剪成与所述像素矩阵大小相同的图像,并对所述图像执行灰度化处理,得到待卷积录入页面模板。

[0113] 进一步地,本发明实施例将所述待卷积录入页面模板输入所述预构建的卷积神经网络中;利用所述卷积层对所述待卷积录入页面模板进行卷积操作,得到特征图;对所述特征图进行池化操作,得到特征数据;汇总所有所述特征数据,得到所述录入页面模板的特征矩阵。

[0114] 所述待检测页面数据获取模块102用于提取用户在所述录入系统的录入页面中输入的录入数据,对所述录入页面进行组件拆分,得到多个录入组件,根据所述录入数据和所述录入组件的映射关系,构建页面结构关系树,利用CSS层叠规则定义所述页面结构关系树,得到页面对象模型树,根据所述页面结构关系树及所述页面对象模型树构建页面渲染树,遍历所述页面渲染树内各树节点,得到所述各树节点在页面中的位置与大小,根据所述位置与大小进行页面绘制,得到待检测录入页面,对所述待检测录入页面执行卷积处理,得到待检测向量;

[0115] 本发明实施例中,所述录入数据可以是用户在所述录入系统中各录入页面中填写的数据,例如,在电商平台中,录入数据可以用户在选取、购买以及支付等操作环节中勾选和填写的相关数据;在互联网购房领域中,所述录入数据,可以包括但不限于用户打开互联网购房页面、咨询房产人员、个人信息填写以及房源地址信息等。

[0116] 详细地,本发明实施例获取所述录入页面的源代码,并将所述源代码分解成多个录入代码块;根据所述录入页面的组成成分,遍历多个所述录入代码块,提取所述组成成分对应的代码块;解析所有所述代码块,得到多个所述录入组件。

[0117] 本发明实施例中,所述代码块可以是根据录入页面的单个功能进行拆分,得到的一段段代码文件。

[0118] 本发明实施例中,所述录入组件可以是能够单独实现页面功能的功能组件,比如,在互联网购房领域中,所述录入组件可以是用户点选进入房源信息进行查看时,出现的房源信息的“详细地址:”以及“联系电话:”;也可以是在各电商平台下某产品页面中的“详情”按钮,平台用户在产品页面点击“详情”按钮时,会跳转至产品详情页面,单独实现页面的某个特定功能。

[0119] 本发明实施例中,所述映射关系可以表示所述录入数据以及所述录入组件的所属关系,例如,在互联网购房领域中,如录入数据“XX小区XX幢XX号”对应于“详细地址:”,则所述录入数据“XX小区XX幢XX号”与所述“详细地址:”可构建对应关系。

[0120] 本发明实施例中,所述页面结构关系树可以是上述多个所述录入代码块中的主体代码块以及嵌套代码块组成的多层树的关系结构。另外,通俗易懂的来讲,也可以用微机里的文件夹结构来表征,以“我的电脑”为主体,里面嵌套着“C盘”以及“D盘”,每个盘里有着很多的文件夹,文件夹里又嵌套着文件夹,逐层打开文件的过程相当于从根节点遍历到叶子节点的过程,主体与嵌套体构成一棵结构关系树。

[0121] 详细地,所述根据所述录入数据和多个所述录入页面组件的映射关系,构建页面结构关系树,包括:根据所述录入数据的类别,将所述录入数据拆分成多组页面录入数据;根据多个所述录入组件的属性,将多组所述页面录入数据与多个所述录入组件进行匹配;将成功匹配的多组所述页面录入数据根据所述对应关系嵌入到所对应的多个所述录入组件中,得到所述页面结构关系树。

[0122] 本发明实施例中,所述根据映射关系,将所述录入数据匹配到多组所述录入组件下对应的页面组件中,得到多个重组录入组件可以是将上述录入数据“XX小区XX幢XX号”的信息根据对应关系输入到与所述录入数据对应的录入组件“详细地址:”的代码块中,完成匹配的过程。

[0123] 本发明实施例中,所述CSS层叠规则可以是HTML中定义样式结构的语言,将网页页面,如将上述互联网购房领域中,将互联网购房页面中的字体、字体大小、字体颜色以及页面组件位置加载进一个网页页面的页面规则文档中,在HTML文件里加入一个超级链接引用所述页面规则文档,则所述网页页面就可按照在所述页面规则文档内定义好的规则进行显示。

[0124] 本发明实施例中,利用所述CSS规则文档定义所述网页页面的页面字体大小、颜色以及页面组件位置等,并将CSS规则文档中的字节内容转换成字符,再依照页面规则组合成属性与对应的属性值的形式,再利用多个所述属性与对应的属性值转换成树节点后链接所

述页面结构关系树得到所述CSS对象模型树。

[0125] 本发明实施例中,所述属性与属性值可以是页面规则文档中字体类别、字体大小以及颜色,例如,定义网页的字体种类的属性与属性值为:{font-family:“楷体”,字体大小与颜色为:{font-size:“12px”;color:#666}}。

[0126] 详细地,本发明实施例计算所述页面对象模型树中各树节点的样式属性,得到页面样式属性;去除所述页面结构关系树中各树节点元素中的不可视元素,得到页面可视树节点;利用所述页面样式属性定义所述页面可视树内各树节点的样式规则,构建所述页面渲染树。

[0127] 本发明实施例中,所述样式属性是所有所述可视属性的属性值,若属性没有被任何匹配的规则定义,则所述样式属性可使用默认值。

[0128] 详细地,所述遍历所述页面渲染树内各树节点,得到所述各树节点在页面中的位置与大小,根据所述位置与大小进行页面绘制,得到待检测录入页面,包括:从所述页面渲染树根节点进行遍历,得到所述页面渲染树各树节点在页面上的位置与大小;将所述页面渲染树内各树节点根据所述位置与大小转换成页面像素;汇集所有所述页面像素,得到所述待检测录入页面。

[0129] 本发明实施例中,所述进行页面绘制的应用程序可以是浏览器,包括但不限于是常见的Chrome、Firefox以及IE(Internet Explorer)。

[0130] 详细地,本发明实施例根据预构建的卷积神经网络模型的输入格式,调整所述待检测页面的页面格式;利用所述卷积神经网络模型的卷积核,对调整之后的所述待检测页面进行卷积处理,得到所述待检测页面的特征矩阵;将所述特征矩阵进行矩阵特征向量计算,得到所述待检测向量。

[0131] 可回溯性检测模块103用于计算所述待检测向量与所述参考向量的相似度,并判断所述相似度是否小于预设相似度阈值,若所述相似度小于预设相似度阈值,判断所述可回溯性检测没有通过;若所述相似度大于或等于所述相似度阈值,判断所述可回溯性检测通过。

[0132] 本发明实施例中,可以采用如下公式计算所述待检测向量以及所述参考向量的相似度;

$$[0133] \quad Sim(x, y) = \cos(\theta) = \frac{y \cdot x}{|y||x|};$$

[0134] 其中,x为所述参考向量,y为所述待检测向量, θ 为所述参考向量与待检测向量的夹角,Sim(x,y)为所述参考向量以及所述待检测向量的相似度。

[0135] 本发明实施例中,所述预设相似度阈值可以是检测达标指数,若所述相似度小于预设相似度阈值,则所述待检测录入页面录入不合格,比如,比较两张图片的出处是否相同时,若相似度在0.7或以下,则标志着两张图片出处不同。

[0136] 若所述相似度小于预设相似度阈值,判断所述可回溯性检测没有通过;

[0137] 若所述相似度大于或等于所述相似度阈值,判断所述可回溯性检测通过。

[0138] 如图3所示,是本发明实现基于录入信息的可回溯性检测方法的电子设备的结构示意图。

[0139] 所述电子设备可以包括处理器10、存储器11、通信总线12和通信接口13,还可以包

括存储在所述存储器11中并可在所述处理器10上运行的计算机程序,如基于录入信息的可回溯性检测程序。

[0140] 其中,所述存储器11至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、移动硬盘、多媒体卡、卡型存储器(例如:SD或DX存储器等)、磁性存储器、磁盘、光盘等。所述存储器11在一些实施例中可以是电子设备的内部存储单元,例如该电子设备的移动硬盘。所述存储器11在另一些实施例中也可以是电子设备的外部存储设备,例如电子设备上配备的插接式移动硬盘、智能存储卡(Smart Media Card,SMC)、安全数字(Secure Digital,SD)卡、闪存卡(Flash Card)等。进一步地,所述存储器11还可以既包括电子设备的内部存储单元也包括外部存储设备。所述存储器11不仅可以用于存储安装于电子设备的应用软件及各类数据,例如基于录入信息的可回溯性检测程序的代码等,还可以用于暂时地存储已经输出或者将要输出的数据。

[0141] 所述处理器10在一些实施例中可以由集成电路组成,例如可以由单个封装的集成电路所组成,也可以是由多个相同功能或不同功能封装的集成电路所组成,包括一个或者多个中央处理器(Central Processing unit,CPU)、微处理器、数字处理芯片、图形处理器及各种控制芯片的组合等。所述处理器10是所述电子设备的控制核心(Control Unit),利用各种接口和线路连接整个电子设备的各个部件,通过运行或执行存储在所述存储器11内的程序或者模块(例如基于录入信息的可回溯性检测程序等),以及调用存储在所述存储器11内的数据,以执行电子设备的各种功能和处理数据。

[0142] 所述通信总线12可以是外设部件互连标准(peripheral component interconnect,简称PCI)总线或扩展工业标准结构(extended industry standard architecture,简称EISA)总线等。该总线可以分为地址总线、数据总线、控制总线等。所述通信总线12总线被设置为实现所述存储器11以及至少一个处理器10等之间的连接通信。为便于表示,图中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0143] 图3仅示出了具有部件的电子设备的,本领域技术人员可以理解的是,图3示出的结构并不构成对所述电子设备的限定,可以包括比图示更少或者更多的部件,或者组合某些部件,或者不同的部件布置。

[0144] 例如,尽管未示出,所述电子设备还可以包括给各个部件供电的电源(比如电池),优选地,电源可以通过电源管理装置与所述至少一个处理器10逻辑相连,从而通过电源管理装置实现充电管理、放电管理、以及功耗管理等功能。电源还可以包括一个或一个以上的直流或交流电源、再充电装置、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。所述电子设备还可以包括多种传感器、蓝牙模块、Wi-Fi模块等,在此不再赘述。

[0145] 可选地,所述通信接口13可以包括有线接口和/或无线接口(如WI-FI接口、蓝牙接口等),通常用于在该电子设备与其他电子设备之间建立通信连接。

[0146] 可选地,所述通信接口13还可以包括用户接口,用户接口可以是显示器(Display)、输入单元(比如键盘(Keyboard)),可选地,用户接口还可以是标准的有线接口、无线接口。可选地,在一些实施例中,显示器可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED(Organic Light-Emitting Diode,有机发光二极管)触摸器等。其中,显示器也可以适当的称为显示屏或显示单元,用于显示在电子设备中处理的信息以及用于显示可

视化的用户界面。

[0147] 应该了解,所述实施例仅为说明之用,在专利申请范围上并不受此结构的限制。

[0148] 所述电子设备中的所述存储器11存储的基于录入信息的可回溯性检测程序是多个计算机程序的组合,在所述处理器10中运行时,可以实现:

[0149] 在获取到检测指令时,从预设的录入系统中提取录入页面模板,并对所述录入页面模板执行卷积池化处理,得到参考向量;

[0150] 提取用户在所述录入系统的录入页面中输入的录入数据;

[0151] 对所述录入页面进行组件拆分,得到多个录入组件;

[0152] 根据所述录入数据和所述录入组件的映射关系,构建页面结构关系树;

[0153] 利用CSS层叠规则定义所述页面结构关系树,得到页面对象模型树;

[0154] 根据所述页面结构关系树及所述页面对象模型树构建页面渲染树;

[0155] 遍历所述页面渲染树内各树节点,得到所述各树节点在页面中的位置与大小,根据所述位置与大小进行页面绘制,得到待检测录入页面;

[0156] 对所述待检测录入页面执行卷积处理,得到待检测向量;

[0157] 计算所述待检测向量与所述参考向量的相似度,并判断所述相似度是否小于预设相似度阈值;

[0158] 若所述相似度小于预设相似度阈值,判断所述可回溯性检测没有通过;

[0159] 若所述相似度大于或等于所述相似度阈值,判断所述可回溯性检测通过。

[0160] 具体地,所述处理器10对上述计算机程序的具体实现方法可参考图1对应实施例中相关步骤的描述,在此不赘述。

[0161] 进一步地,所述电子设备集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。所述计算机可读介质可以是非易失性的,也可以是易失性的。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)。

[0162] 本发明实施例还可以提供一种计算机可读存储介质,所述可读存储介质存储有计算机程序,所述计算机程序在被电子设备的处理器所执行时,可以实现:

[0163] 在获取到检测指令时,从预设的录入系统中提取录入页面模板,并对所述录入页面模板执行卷积池化处理,得到参考向量;

[0164] 提取用户在所述录入系统的录入页面中输入的录入数据;

[0165] 对所述录入页面进行组件拆分,得到多个录入组件;

[0166] 根据所述录入数据和所述录入组件的映射关系,构建页面结构关系树;

[0167] 利用CSS层叠规则定义所述页面结构关系树,得到页面对象模型树;

[0168] 根据所述页面结构关系树及所述页面对象模型树构建页面渲染树;

[0169] 遍历所述页面渲染树内各树节点,得到所述各树节点在页面中的位置与大小,根据所述位置与大小进行页面绘制,得到待检测录入页面;

[0170] 对所述待检测录入页面执行卷积处理,得到待检测向量;

[0171] 计算所述待检测向量与所述参考向量的相似度,并判断所述相似度是否小于预设相似度阈值;

[0172] 若所述相似度小于预设相似度阈值,判断所述可回溯性检测没有通过;

[0173] 若所述相似度大于或等于所述相似度阈值,判断所述可回溯性检测通过。

[0174] 进一步地,所述计算机可用存储介质可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序等;存储数据区可存储根据区块链节点的使用所创建的数据等。

[0175] 在本发明所提供的几个实施例中,应该理解到,所揭露的设备,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

[0176] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0177] 另外,在本发明各个实施例中的各功能模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能模块的形式实现。

[0178] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。

[0179] 因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。

[0180] 本发明所指区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain),本质上是一个去中心化的数据库,是一串使用密码学方法相关联产生的数据块,每一个数据块中包含了一批次网络交易的信息,用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层等。

[0181] 本申请实施例可以基于人工智能技术对相关的数据进行获取和处理。其中,人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

[0182] 此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。系统权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一、第二等词语用来表示名称,而并不表示任何特定的顺序。

[0183] 最后应说明的是,以上实施例仅用以说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或等同替换,而不脱离本发明技术方案的精神和范围。

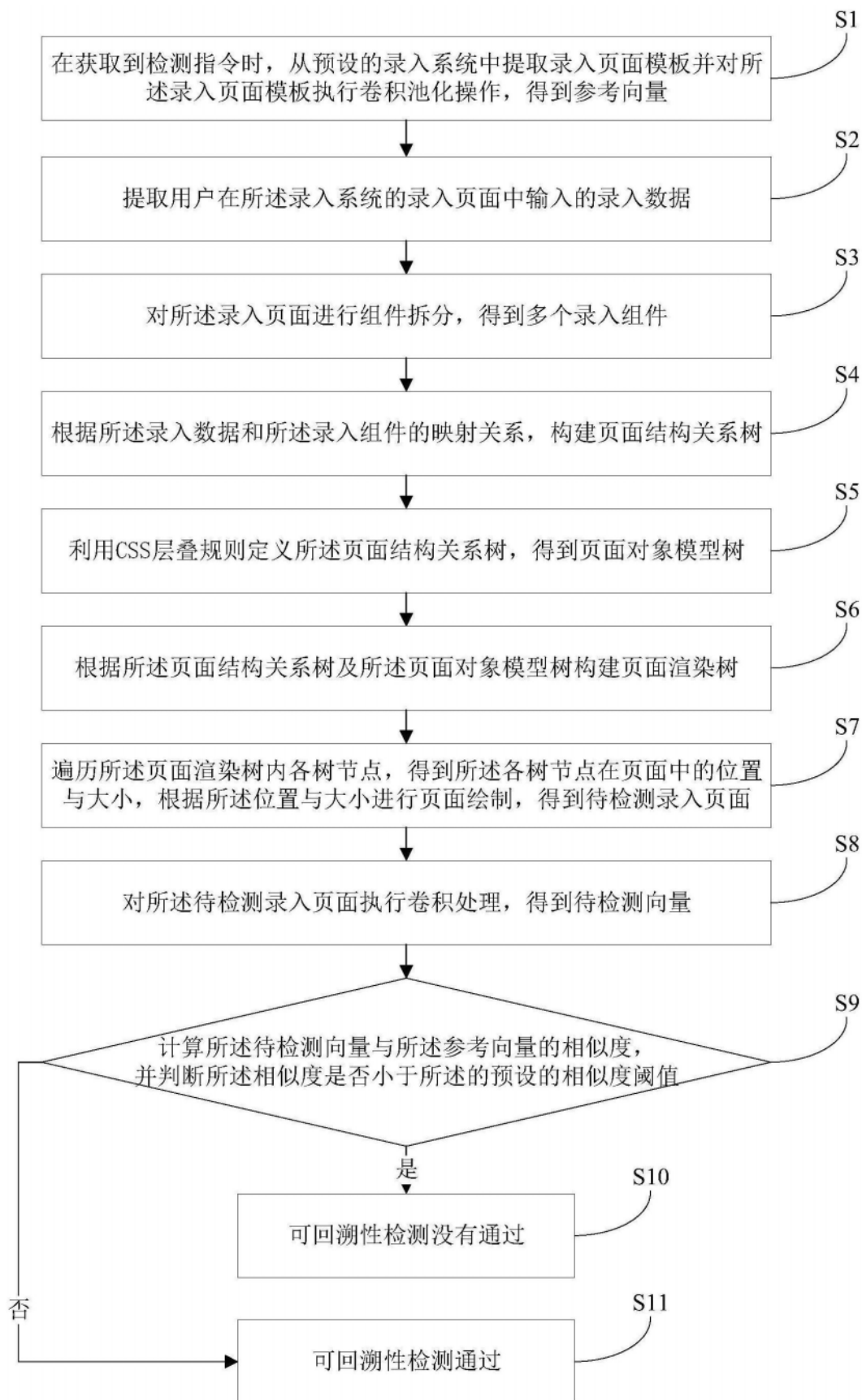


图1

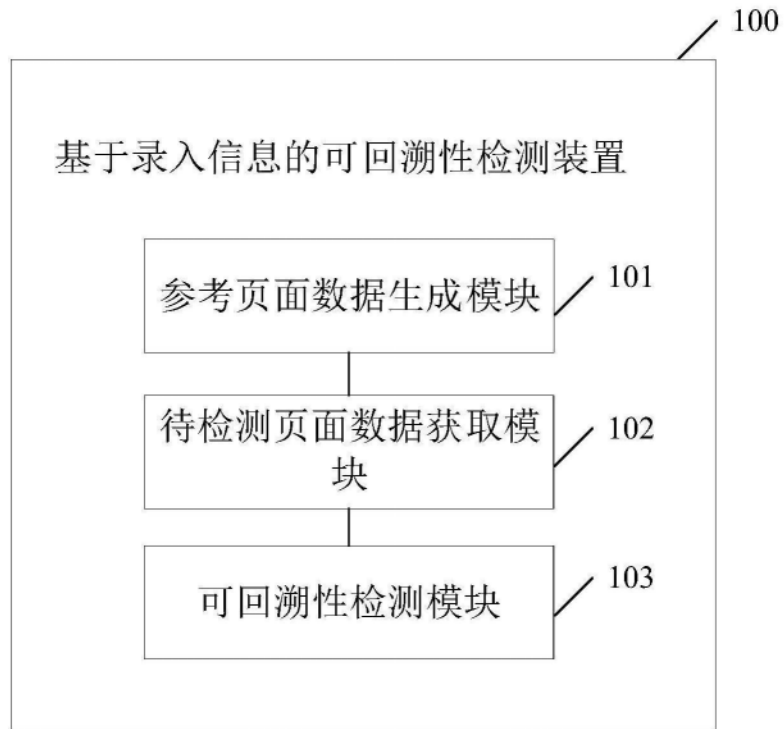


图2

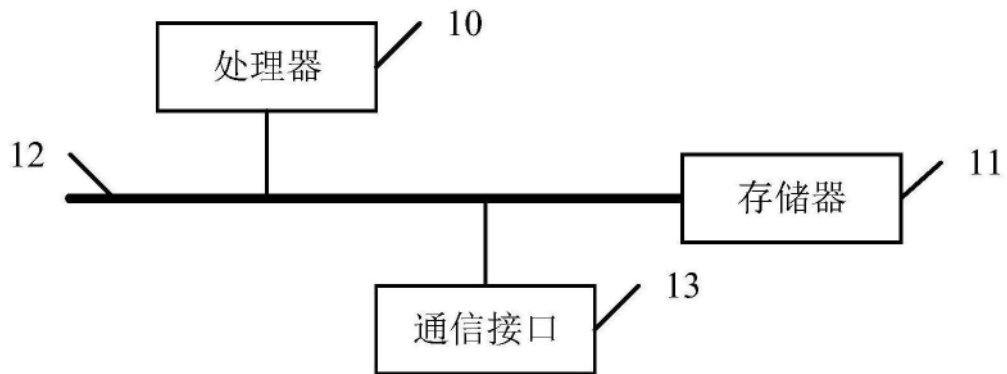


图3