



(12) 发明专利

(10) 授权公告号 CN 113391871 B

(45) 授权公告日 2021. 11. 16

(21) 申请号 202110944521.2

G06F 16/33 (2019.01)

(22) 申请日 2021.08.17

G06F 16/35 (2019.01)

(65) 同一申请的已公布的文献号

G06K 9/00 (2006.01)

申请公布号 CN 113391871 A

G06K 9/20 (2006.01)

(43) 申请公布日 2021.09.14

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

(73) 专利权人 杭州实在智能科技有限公司

审查员 祝子豪

地址 310000 浙江省杭州市余杭区人工智

能小镇六号楼六层

(72) 发明人 欧阳小刚

(74) 专利代理机构 杭州云睿专利代理事务所

(普通合伙) 33254

代理人 张骁敏

(51) Int. Cl.

G06F 9/451 (2018.01)

G06F 16/31 (2019.01)

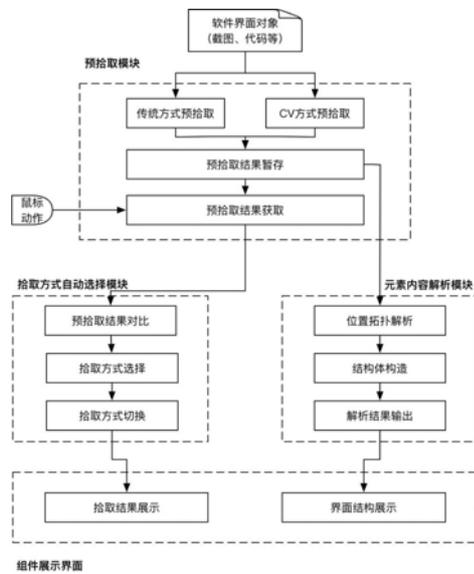
权利要求书2页 说明书9页 附图4页

(54) 发明名称

一种RPA元素智能融合拾取的方法与系统

(57) 摘要

本发明公开了一种RPA元素智能融合拾取的方法与系统,通过将基于深度神经网络的CV元素拾取技术和传统元素拾取技术相结合,并在RPA软件后台根据用户所要操作的软件界面的不同,为用户自动选择更精准、更合适的元素拾取方式(当然也支持用户手动进行拾取方式切换),同时在操作过程中能够实现拾取方式的自动、无感切换。另外,在各个元素实现定位的基础上,提供界面元素的内容解析能力,实现元素的类别、属性、位置、层次等的有序输出,从而支持更多样化的元素操作能力。通过上述功能,可以为用户提供更流畅的使用体验,提高RPA软件可用性和支持范围;并减少用户使用成本,降低用户编辑开发RPA流程的时间。



1. 一种RPA元素智能融合拾取的方法,其特征在于,包括以下步骤:

采用基于深度神经网络的目标检测和OCR识别技术提供CV元素拾取方式,通过分别调用基于操作系统指令、网页结构解析和软件代码解析的传统元素拾取和CV元素拾取两种元素拾取方式对用户所要操作的软件界面进行预先解析,获取到比较全面的界面元素信息,汇总并暂存两种元素拾取方式获得的界面元素信息,再根据用户的鼠标移动或点击动作,匹配到指定位置内的元素;

根据用户指定位置匹配到的元素,获取对应的两种元素拾取方式获得的界面元素信息进行比对,当比对结果出现不一致则停止比对,选择最优的元素拾取方式;当比对结果均一致则随机选择一种元素拾取方式,而后根据比对结果自动切换所选择的元素拾取方式,并将自动切换完成的元素拾取方式获取到的界面元素信息进行输出;

在元素拾取方式的选择和切换的同时,获取暂存的两种元素拾取方式获得的界面元素信息,首先根据界面元素信息中各个元素的位置坐标关系构建元素拓扑结构图,还原各个元素在软件界面上的真实位置关系,再结合构建的元素拓扑结构图进行层次关系分析以还原元素之间的层次关系,并构造出代表软件界面整体结构的数据结构体,同时基于表格识别技术对软件界面中的表格信息以行列坐标的关系抽取,利用NLP技术对表单页面中的键值对关键词进行映射,最终将解析得到的结构体进行输出。

2. 根据权利要求1所述的一种RPA元素智能融合拾取的方法,其特征在于,所述CV元素拾取方式包括图标元素的检测和文字元素的检测,在检测完成后,根据各个元素的交并比关系进行检测框合并,合并后的检测框结合检测框对应元素的类型一并作为界面元素信息进行返回。

3. 根据权利要求2所述的一种RPA元素智能融合拾取的方法,其特征在于,所述交并比简称IOU, IOU的计算公式为:

$$IOU = \text{Intersection}(A, B) / \text{Union}(A, B);$$

其中计算公式中的A和B为两个检测框,上述公式右侧的分子分母分别计算两个检测框的交集面积和并集面积。

4. 根据权利要求1所述的一种RPA元素智能融合拾取的方法,其特征在于,界面元素信息的比对具体通过依次对元素粒度、元素置信度、元素属性方面进行,最终选择最优的元素拾取方式,包括以下步骤:

比对传统元素拾取和CV元素拾取两种元素拾取方式拾取到的元素的粒度,此处所指粒度含义是从元素层次关系看所能操作的更小范围和更精细的区域,选择能够拾取到更细粒度元素的元素拾取方式;

若两种元素拾取方式拾取到的粒度一致,则比对两种元素拾取方式对所定位到元素的置信度,针对CV元素拾取方式表示的是图标或文字检测的置信度,对于传统元素拾取表示其非窗体的概率,选择定位元素置信度更高的元素拾取方式;

若两种元素拾取方式的置信度也一致,则比对其所包含属性的数量,此处属性包含坐标、类型、层次、索引各种元素属性,选择包含属性更多的元素拾取方式;

最后若属性个数也一样,则随机选择一种元素拾取方式作为具体操作该元素的拾取方式。

5. 一种RPA元素智能融合拾取的系统,用于执行如权利要求1所述的RPA元素智能融合拾取的方法,其特征在于,包括:

预拾取模块,结合传统元素拾取和CV元素拾取两种元素拾取方式对用户所要操作的软件界面进行预先解析,获取到比较全面的界面元素信息,汇总并暂存两种元素拾取方式获得的界面元素信息,再根据用户的鼠标移动或点击动作,匹配到指定位置内的元素;

拾取方式自动选择模块,根据传统元素拾取和CV元素拾取获得的元素的详细情况进行比对,从元素的粒度、拾取的精准度角度出发,选择能更精准定位到元素的元素拾取方式,而后根据比对结果自动切换对应选择的元素拾取方式,并将该元素拾取方式获取到的界面元素信息进行输出;

界面元素内容解析模块,在拾取方式自动选择模块进行拾取方式的选择和切换的同时,界面元素内容解析模块根据暂存的预拾取的所有元素信息进行界面内容的重构和解析,构造出代表软件界面整体结构的数据结构体。

6. 根据权利要求5所述的一种RPA元素智能融合拾取的系统,其特征在于,系统还包括组件展示界面,用于展示构造的数据结构体和所要拾取的元素的相关信息。

## 一种RPA元素智能融合拾取的方法与系统

### 技术领域

[0001] 本发明涉及机器人流程自动化技术领域,特别涉及一种RPA元素智能融合拾取的方法与系统。

### 背景技术

[0002] 机器人流程自动化技术(Robotic Process Automation,后简称RPA)是一种用于企业规则化、重复性工作替代的有效技术,目前正广泛应用于各行各业的财务、税务、审批、风控、运营、物流、广告等业务中。RPA技术提供了可视化的流程编辑器,开发人员可以在流程编辑器上设计和配置出能够在企业业务系统中自动执行的流程脚本;在脚本执行过程中,则通过模拟键盘、鼠标操作、调用操作系统接口、分析网页结构等手段,来模拟人类的按钮点击、键盘输入、文件创建、网页跳转、逻辑判断等动作。RPA技术的出现,可以将企业人力从大量简单、重复、规则化的工作中解放出来,投入到更具价值的创造性工作中;同时借助RPA技术,不需要对业务系统进行大规模的底层改造,也不需要开发各类系统间数据对接的复杂接口,因而能够在短时间内实现企业业务流程自动化,投入产出比更高。

[0003] RPA技术的核心是对计算机软件界面上的各类元素,例如按钮、文字、搜索框、输入框、可操作区域等,进行精确地定位(或叫拾取)和操作,进而完成流程执行所需的点击、跳转、输入、复制、粘贴等动作。传统的RPA元素拾取技术是基于操作系统提供的指令接口,或者基于网页的结构解析来完成的。例如对于一般的网页元素拾取,首先会根据其超文本标记语言(Hyper Text Markup Language,后简称HTML)解析为不同的块、标签或文本,再获取这些元素的位置、内容、索引和层次关系,从而实现各类操作,如点击按钮、点击搜索框并输入文字、移动到某张图片上等。然而在RPA的实际场景中,往往有一些应用软件、远程桌面等,无法通过系统指令或代码解析的方式来定位到元素,比如在遇到远程桌面的场景时,RPA软件所能获取到的仅仅是桌面镜像所呈现的一张图片,在这种情况下就需要采用其他技术来实现元素拾取。当前比较流行的是基于计算机视觉(Computer Vision,后简称CV)技术实现的元素拾取,通过将界面上的各类元素作为图片中的一个目标,借助CV领域的目标检测技术和光学字符识别(Optical Character Recognition,后简称OCR)技术定位和识别出每个元素的位置和类型,进而支持一系列的移动、点击、输入等操作。

[0004] 上述两种元素拾取方式相互结合,可以更好地支持不同应用场景下的RPA技术的实施和流程的执行,然而从目前所了解到的情况来看,业界普遍是将两个技术分开使用,需要RPA软件的用户人为选择使用传统拾取技术还是CV拾取技术,或者是在传统拾取技术无法满足要求的情况下,手动切换为CV拾取技术再进行尝试。这种结合和应用的方式,一方面对用户来说不够友好,增加了用户使用RPA软件的难度;另一方面,先尝试传统拾取再使用CV拾取的方式,对于一些传统拾取无法支持的场景来说,每次尝试都增加时间成本,不利于RPA流程的快速开发和部署;并且当前业界对于CV拾取仅仅作为坐标定位使用,限制了其可以支持的范围。

[0005] 基于上述现有技术存在的问题,本申请方案所要解决的技术问题包括:

- [0006] 1、针对传统元素拾取技术支持的场景有限的问题；
- [0007] 2、针对现有的CV元素拾取技术一般仅用于元素定位，而不能支持更多的元素操作的问题；
- [0008] 3、针对现有的多种拾取方式无法自动切换的问题。

### 发明内容

[0009] 针对现有技术存在的不足，本发明的目的在于提供一种RPA元素智能融合拾取的方法与系统，通过将基于深度神经网络的CV元素拾取技术和传统基于操作系统指令、网页结构解析和软件代码解析的传统元素拾取技术相结合，并在RPA软件后台根据用户所要操作的软件界面的不同，为用户自动选择更精准、更合适的元素拾取方式（当然也支持用户手动进行拾取方式切换），同时在操作过程中能够实现拾取方式的自动、无感切换；另外，在各个元素实现定位的基础上，提供界面元素的内容解析能力，实现元素的类别、属性、位置、层次等的有序输出，从而支持更多样化的元素操作能力。

[0010] 通过上述功能，可以为用户提供更流畅的使用体验，提高RPA软件可用性和支持范围；并减少用户使用成本，降低用户编辑开发RPA流程的时间。

[0011] 方案设计中，首先通过基于深度神经网络的、更精准的CV元素拾取技术，实现更细粒度和更通用的软件元素拾取能力，同时为更好地理解软件界面构成和内容提供基础。

[0012] 进而，采用深度目标检测网络、深度文字检测和识别网络、元素位置图分析、表格识别和自然语言处理(Natural Language Processing, 后简称NLP)相结合的方案，对软件界面中各个元素的类别属性、顺序关系、位置关系和层次关系等进行分析 and 排列，实现更智能、更多样化的软件界面元素操作，支持更丰富场景下的RPA流程编辑和运行。

[0013] 进而，通过本发明方案中的预拾取模块、拾取方式自动选择模块对RPA所要操作的软件界面进行智能判断，自动为用户选择和切换最合适的元素拾取方式；例如，当遇到远程桌面等传统元素拾取技术无法拾取的情况下，可以自动切换为高精度的CV元素拾取方式；

[0014] 而当传统元素拾取可以有效支持当前软件时，则选择传统元素拾取方式。这种自动切换方式结合前述的界面内容分析，一方面为用户提供更好、更自如的使用体验；一方面能够支持更复杂场景的RPA流程执行；例如对表单类图片形式的界面通过OCR和表格识别技术抽取出行列信息后，可以完成单元格的各类点击、读取、输入等操作。

[0015] 针对现有技术的不足，本发明方案具体通过以下内容进行解决：

[0016] 1、针对传统元素拾取技术支持的场景有限的问题，采用基于深度神经网络、更精准、更通用的CV元素拾取技术，提供在此类场景下的细粒度元素拾取能力，并且具备各种软件界面的通用拾取能力；

[0017] 例如对于远程桌面等场景，传统元素拾取技术无法支持，借助CV元素拾取则能比较精准地获取其中各个元素的位置、类型、层次关系等信息。

[0018] 2、针对现有的CV元素拾取技术一般仅用于元素定位，而不能支持更多的元素操作的问题，本发明采用深度目标检测网络、OCR、元素位置图分析、表格识别和NLP相结合的方案，分析、排列并结构化输出软件界面上的各个元素的类别属性、位置关系和层次关系等，从而实现对元素加载判断、元素属性获取和表格结构解析等能力，拓展当前CV元素拾取技术所能支持各类元素级别操作和RPA技术所能应用的场景。

[0019] 3、针对现有的多种拾取方式无法自动切换的问题,本发明通过预拾取模块进行软件界面的预先拾取和暂存;通过拾取方式自动选择模块判断在不同的场景下使用哪种拾取方式能够获得更精准、更细粒度的元素操作;结合界面元素内容解析模块将拾取到的元素进行结构化解析、排列和输出,并最终在组件展示界面以统一的元素拾取组件进行呈现。整体对用户而言,无需用户手动切换不同的拾取方式,大大提升用户体验,减少用户使用RPA软件的成本。

[0020] 本发明的上述技术目的是通过以下技术方案得以实现的:

[0021] 一种RPA元素智能融合拾取的方法,包括以下步骤:

[0022] 采用基于深度神经网络的目标检测和OCR识别技术提供CV元素拾取方式,通过分别调用传统元素拾取和CV元素拾取两种元素拾取方式对用户所要操作的软件界面进行预先解析,获取到比较全面的界面元素信息,汇总并暂存两种元素拾取方式获得的界面元素信息,再根据用户的鼠标移动或点击动作,匹配到指定位置内的元素;

[0023] 根据用户指定位置匹配到的元素,获取对应的两种元素拾取方式获得的界面元素信息进行比对,通过依次对元素粒度、元素置信度、元素属性方面进行比对,当比对结果出现不一致则停止比对,选择最优的元素拾取方式;当比对结果均一致则随机选择一种元素拾取方式,而后根据比对结果自动切换所选择的元素拾取方式,并将自动切换完成的元素拾取方式获取到的界面元素信息进行输出;

[0024] 在元素拾取方式的选择和切换的同时,获取暂存的预拾取的所有界面元素信息,首先根据界面元素信息中各个元素的位置坐标关系构建元素拓扑结构图,还原各个元素在软件界面上的真实位置关系,再结合构建的元素拓扑结构图进行层次关系分析以还原元素之间的层次关系,并构造出代表软件界面整体结构的数据结构体,同时基于表格识别技术对软件界面中的表格信息以行列坐标的关系行抽取,利用NLP技术对表单页面中的键值对关键词进行映射,最终将解析得到的结构体进行输出。

[0025] 进一步优选为,所述CV元素拾取方式包括图标元素的检测和文字元素的检测,在检测完成后,根据各个元素的交并比关系进行检测框合并,合并后的检测框结合检测框对应元素的类型一并作为界面元素信息进行返回。

[0026] 进一步优选为,所述交并比简称IOU,IOU的计算公式为:

[0027] 
$$IOU = \text{Intersection}(A, B) / \text{Union}(A, B);$$

[0028] 其中计算公式中的A和B为两个检测框,上述公式左右两侧分别计算两个检测框的交集面积和并集面积。

[0029] 进一步优选为,元素拾取方式选择具体包括以下步骤:

[0030] 比对传统元素拾取和CV元素拾取两种元素拾取方式拾取到的元素的粒度,此处所指粒度含义是从元素层次关系看所能操作的更小范围和更精细的区域,选择能够拾取到更细粒度元素的元素拾取方式;

[0031] 若两种元素拾取方式拾取到的粒度一致,则比对两种元素拾取方式对所定位到元素的置信度,针对CV元素拾取方式表示的是图标或文字检测的置信度,对于传统元素拾取表示其非窗体的概率,选择定位元素置信度更高的元素拾取方式;

[0032] 若两种元素拾取方式的置信度也一致,则比对其所包含属性的数量,此处属性包

含坐标、类型、层次、索引等各种元素属性,选择包含属性更多的元素拾取方式;

[0033] 最后若属性个数也一样,则随机选择一种元素拾取方式作为具体操作该元素的拾取方式。

[0034] 本发明申请的另外一个目的是提供一种RPA元素智能融合拾取的系统,包括预拾取模块、拾取方式自动选择模块、界面元素内容解析模块、组件展示界面;

[0035] 预拾取模块,结合传统元素拾取和CV元素拾取两种元素拾取方式对用户所要操作的软件界面进行预先解析,获取到比较全面的界面元素信息,汇总并暂存两种元素拾取方式获得的界面元素信息,再根据用户的鼠标移动或点击动作,匹配到指定位置内的元素;

[0036] 拾取方式自动选择模块,根据传统元素拾取和CV元素拾取获得的元素的详细情况进行比对,从元素的粒度、拾取的精准度角度出发,选择能更精准定位到元素的元素拾取方式,而后根据比对结果自动切换对应选择的元素拾取方式,并将该元素拾取方式获取到的界面元素信息进行输出;

[0037] 界面元素内容解析模块,在拾取方式自动选择模块进行拾取方式的选择和切换的同时,界面元素内容解析模块根据暂存的预拾取的所有元素信息进行界面内容的重构和解析,构造出代表软件界面整体结构的数据结构体。

[0038] 组件展示界面,用于展示构造的数据结构体和所要拾取的元素的相关信息。

[0039] 综上所述,本发明对比于现有技术的有益效果为:

[0040] 本发明独创性地提出一种RPA元素智能融合拾取的方法与系统,首先底层采用基于深度神经网络的目标检测和OCR识别技术提供CV元素拾取的能力,使得RPA软件在远程桌面等场景下也能精准定位元素,提高了RPA软件的使用范围;其次结合目标检测、OCR、位置图分析、表格识别和NLP技术,对所要操作的界面结构进行全面地解析,解决现有CV技术一般仅用于元素定位的问题,从而除了能够提供坐标信息外,还提供更多的结构、层次和语义信息,进而适应界面分辨率变化的情况,提高了RPA流程的稳定性和可用性,在分辨率变化的情况下,RPA流程不再会失效;最后基于所提出的方法框架和预拾取模块、拾取方式自动选择模块、界面元素内容解析模块和组件展示界面等基础功能模块,能够实现RPA元素拾取方式的自动选择和切换,在后台为用户自动选择最精准、最合适的元素拾取方式,减少用户手动切换的时间和选择难度,提高了用户使用RPA软件开发自动化流程的效率,进一步优化用户体验。

[0041] 整个发明方案可以使得RPA软件适用于更多的场景、实现更多自由度的流程编辑、实现更高效率的流程开发,提供更友好的用户体验,从而构建更智能的新一代RPA软件产品。

## 附图说明

[0042] 通过参考附图阅读下文的详细描述,本发明示例性实施方式的上述以及其他目的、特征和优点将变得易于理解。在附图中,以示例性而非限制性的方式示出了本发明的若干实施方式,其中:

[0043] 图1为实施例的系统功能架构图;

[0044] 图2为实施例预拾取模块中CV元素拾取的模块功能结构图;

[0045] 图3为实施例拾取方式自动选择模块中比对流程图;

[0046] 图4为实施例具体实施情况下的流程图。

### 具体实施方式

[0047] 下面将参考若干示例性实施方式来描述本发明的原理和精神。应当理解,给出这些实施方式仅仅是为了使本领域技术人员能够更好地理解进而实现本发明,而并非以任何方式限制本发明的范围。相反,提供这些实施方式是为了使本公开更加透彻和完整,并且能够将本公开的范围完整地传达给本领域的技术人员。说明书中的“实施例”或“实施方式”既可表示一个实施例或一种实施方式,也可表示一些实施例或一些实施方式的情况。

[0048] 本领域技术人员知道,本发明的实施方式可以实现为一种系统、装置、设备、方法或计算机程序产品。因此,本公开可以具体实现为以下形式,即:完全的硬件、完全的软件(包括固件、驻留软件、微代码等),或者硬件和软件结合的形式。

[0049] 根据本发明的实施方式,提出了一种RPA元素智能融合拾取的方法与系统。

[0050] 需要说明的是,附图中的任何元素数量均用于示例而非限制,以及任何命名都仅用于区分,而不具有任何限制含义。

[0051] 下面对本发明中所涉及的技术术语进行简单描述,以便相关人员更好地理解本方案。

#### [0052] 1、传统元素拾取技术

[0053] 元素拾取实际上是对界面上的元素进行定位、分析和操作的技术,在传统的RPA软件中,上述功能主要通过操作系统指令、网页结构分析、软件代码解析等方式实现。操作系统指令方式是指操作系统自带的指令或接口能够让RPA软件定位并操作某些窗体、程序和桌面图标等;网页结构分析方式如2.1节所述,通过解析网页HTML代码定位、识别到网页上的各类区块、标签、文字和链接,并获取其位置、内容、索引和层次;而软件代码解析的方式与网页结构分析方式基本一致,通过对软件代码进行解析或者反编译,能够获得软件中各类功能按钮、操作区域的坐标、操作接口等,进而支持对该软件各类操作。

#### [0054] 2、目标检测技术

[0055] 大部分的场景,传统元素拾取技术能够胜任对元素的定位和操作;然而在某些情况下,基于指令接口、代码解析这种方式将会失效;比较明显的两个例子分别是远程桌面和某些应用软件。远程桌面对于RPA软件来说,外在的呈现就是一张图片,这种情况下无法通过调用接口或解析代码来获取其镜像图片中的具体信息;同样的,某些应用软件虽然能够获取一些大块区域的窗体信息,但是窗体内部更细节的按钮、文字等无法获取。因此在这些情况下,需要使用CV拾取的方式来拾取更精细的内部元素。

[0056] 目标检测技术是CV拾取中通常采用的技术,基本原理是将软件界面作为一个图片输入,将其中的各类元素作为目标物体进行检测、定位和分类;算法实现上通过标注大量的软件界面样本后进行模型学习,能够得到一个可以定位、分类元素的算法模型;在实际使用中,该算法模型将软件界面图片作为输入,便可以对图片上形似图标、按钮、输入框、搜索框等的区域找到,这里的定位包括了图标中心坐标的获取,和图标大小范围的框定;分类则指的是根据预设的类别将某个图标判断为是按钮、搜索框还是点选框等。当前主流的目标检测算法基于神经网络的方式实现,常用的算法模型包括YOLO、FasterRCNN等。

#### [0057] 3、OCR技术

[0058] RPA所要操作的软件界面中的元素除了各类按钮、图标、搜索框等,还包括了大量的文字信息和文字元素,这些文字元素以按钮提示、网址链接、段落文本、表格内容等方式出现;实际上界面上的文字信息一般要比非文字信息多很多,需要利用OCR技术进行定位和识别。

[0059] OCR是一种通过计算机视觉技术将电子文档、扫描文档、网页图片、拍照图片等图片上的文字字符提取并转化为计算机可读的文本类型数据的技术,基本原理是将图片中的文字或文字区域作为一种特定的目标物体进行检测,将其与其他非文字区域分割开,而后对文字区域中的具体字符进行识别,转录为一个个字符或字符串信息。常见的OCR应用场景包括了各种证件、票据、合同、文档等的识别,并广泛应用在财务报销、材料审核、信息提取、自动录入等业务场景中。OCR技术如今也基本以深度神经网络的方式实现,核心包括文字检测、文字识别两个部分。其中:文字检测主要做文字区域的定位,通过目标检测的思想将图片中的每一块文字区域从图片中检测和定位得到;而文字识别是对区域中的内容进行识别和转化。本发明中,OCR技术主要用于识别网页、软件界面中各类带有文字信息的元素,辅助元素的拾取。

[0060] 一种RPA元素智能融合拾取的方法与系统,设有预拾取模块、拾取方式自动选择模块、界面元素内容解析模块、具体包括以下步骤:

[0061] 预拾取模块采用基于深度神经网络的目标检测和OCR识别技术提供CV元素拾取方式,通过分别调用传统元素拾取和CV元素拾取两种元素拾取方式对用户所要操作的软件界面进行预先解析,获取到比较全面的界面元素信息,汇总并暂存两种元素拾取方式获得的界面元素信息,再根据用户的鼠标移动或点击动作,匹配到指定位置内的元素;

[0062] 拾取方式自动选择模块根据用户指定位置匹配到的元素,获取对应的两种元素拾取方式获得的界面元素信息进行比对,通过依次对元素粒度、元素置信度、元素属性方面进行比对,当比对结果出现不一致则停止比对,选择最优的元素拾取方式;当比对结果均一致则随机选择一种元素拾取方式,而后根据比对结果自动切换所选择的元素拾取方式,并将自动切换完成的元素拾取方式获取到的界面元素信息进行输出;

[0063] 界面元素内容解析模块在元素拾取方式的选择和切换的同时,获取暂存的预拾取的所有界面元素信息,首先根据界面元素信息中各个元素的位置坐标关系构建元素拓扑结构图,还原各个元素在软件界面上的真实位置关系,再结合构建的元素拓扑结构图进行层次关系分析以还原元素之间的层次关系,并构造出代表软件界面整体结构的数据结构体,同时基于表格识别技术对软件界面中的表格信息以行列坐标的关系行抽取,利用NLP技术对表单页面中的键值对关键词进行映射,最终将解析得到的结构体进行输出。

[0064] 本发明所提出的系统功能架构图如图1所示,主要包括:预拾取模块、拾取方式自动选择模块、界面元素内容解析模块、组件展示界面,上述4个模块具体内容如下所述:

[0065] 预拾取模块:以用户所要操作的软件界面对象作为输入,在预拾取模块内部分别调用传统元素拾取和CV元素拾取技术对界面进行预先解析。

[0066] 其中传统元素拾取采用代码解析等方式进行元素拾取,返回元素坐标、类型、索引等信息;同时CV元素拾取则采用目标检测、OCR等技术对元素对象进行检测和识别,返回元素坐标、截图、类别等信息;在预拾取模块中汇总并暂存两种元素拾取方式的结果(界面元素信息),而后根据用户的鼠标移动或点击动作,匹配到指定位置内的元素,并将该元素的

相关信息带入到拾取方式自动选择模块。

[0067] 其中,对于CV元素拾取部分,预拾取模块包括CV拾取模块,参照图2所示,CV拾取模块包括:

[0068] 图标元素检测和文字元素检测,其中图标元素表示不带文字信息的纯图像类元素,如前进、返回、刷新、关闭等按钮;

[0069] 检测完成后,CV拾取模块根据各个元素的交并比(Intersection Over Union,后简称IOU)关系进行检测框合并,例如图标检测和文字检测都定位到同一位置,则需要利用IOU计算检测框的重叠情况,根据重叠情况进行合并,IOU计算公式如下:

[0070]  $IOU = \text{Intersection}(A, B) / \text{Union}(A, B)$ ;

[0071] 其中计算公式中的A和B为两个检测框,上述公式左右两侧分别计算两个检测框的交集面积和并集面积,合并后的检测框结合检测框对应元素的类型(例如是文字还是图标)一并作为元素信息进行返回。

[0072] 拾取方式自动选择模块:元素拾取方式自动选择的原理是根据传统元素拾取和CV元素拾取获得的元素的详细情况进行比对,从元素的粒度、拾取的精准度等角度出发,选择能更精准定位到元素的拾取方式。

[0073] 具体来说,首先获取预拾取模块输出的用户指定的元素信息,并根据其中包含的传统拾取元素信息和CV拾取元素信息进行比对,最终确认所选择的拾取方式。

[0074] 参照图3所示,拾取方式自动选择和比对的逻辑如下:

[0075] 比对传统元素拾取和CV元素拾取两种元素拾取方式拾取到的元素的粒度,此处所指粒度含义是从元素层次关系看所能操作的更小范围和更精细的区域,选择能够拾取到更细粒度元素的元素拾取方式;

[0076] 若两种元素拾取方式拾取到的粒度一致,则比对两种元素拾取方式对所定位到元素的置信度,针对CV元素拾取方式表示的是图标或文字检测的置信度,对于传统元素拾取表示其非窗体的概率,选择定位元素置信度更高的元素拾取方式;

[0077] 若两种元素拾取方式的置信度也一致,则比对其所包含属性的数量,此处属性包含坐标、类型、层次、索引等各种元素属性,选择包含属性更多的元素拾取方式;

[0078] 最后若属性个数也一样,则随机选择一种元素拾取方式作为具体操作该元素的拾取方式。

[0079] 而后根据比对结果自动切换选择的元素拾取方式(传统元素拾取/CV元素拾取),并将该元素拾取方式获取到的元素信息进行输出,同时在RPA软件界面上为用户自动切换拾取方式并进行展示。

[0080] 界面元素内容解析模块:在拾取方式自动选择模块进行元素拾取方式的选择和切换的同时,界面元素内容解析模块根据暂存的预拾取的所有元素信息进行界面内容的重构和解析。

[0081] 具体来说,首先根据各个元素的位置坐标关系,构建元素拓扑结构图,还原各个元素在界面上的真实位置关系,并作为后续匹配和展示的依据。

[0082] 而后根据元素信息构造界面元素的数据结构体。对传统方式拾取的元素来说,除了元素位置外,还包含元素的类型、层次、索引等信息;而对于CV方式拾取的元素,采用基于

深度神经网络的目标检测技术和OCR技术,能够获取到元素的检测框,因此元素的位置坐标和截图区域作为元素属性进行保存,同时输出图标分类和文字的内容,图标分类例如矩形按钮、圆形按钮、单选框、复选框等。结合传统方式和CV拾取方式,能够获取到比较全面的界面元素信息,结合前面构造的元素拓扑结构图可以还原元素之间的层次关系,例如:A是一个窗体、B为窗体中的菜单面板,C为该面板中的一个按钮,则三者结构体中的表达为A包含B包含C的关系。另外结合表格识别技术,能够将界面中包含的表格信息以行列坐标的关系返回;结合NLP技术,可以获取包含表单页面中的键值关系对,例如在某个场景中需要填写表单信息,则可以根据信息的提示比如「手机号」三个字,在对应的输入框中填写相应手机号码,而不会出现输错位置IDE情况。

[0083] 最后,上述构造的元素结构体将和所要拾取的元素一起展示在RPA软件的组件界面。

[0084] 组件展示界面:组件界面即RPA软件的编辑和开发界面,所要拾取的元素的相关信息将展示在界面上供用户确认,用户确认无误则可以完成拾取,而在执行时调用该元素信息进行定位和操作;同时界面结构相关信息也在组件界面展示,提供用户进一步继续拾取元素或操作软件界面的参考,所展示的界面结构信息将以树状结构呈现。

[0085] 参照图4所示,展示了RPA软件中引入智能融合拾取系统的用户操作流程,同时展示本发明的实施方案,及如何将本发明方法和系统与RPA软件结合的细节:

[0086] 1.RPA软件用户首先打开要操作和进行元素拾取的软件界面,该界面的结构信息、代码信息和截图图片会作为输入传递到智能融合拾取系统;

[0087] 2.用户使用RPA中的智能融合拾取系统来对软件界面元素进行拾取,此时在智能融合拾取系统的后台将完成以下一些操作,即该系统的各功能模块实现:

[0088] 1)预拾取模块首先采用传统拾取和CV元素拾取两种元素拾取方式对界面进行预先的解析、拾取和拾取结果暂存,并根据用户鼠标选择动作确定要操作的元素;

[0089] 2)拾取方式自动选择模块根据两种方式获取到的元素信息进行粒度、置信度和属性内容的比对,并自动为用户选择和切换能够更准确、更精细的拾取到元素的方式;

[0090] 3)界面元素内容解析模块根据获取的界面全局元素信息,进行元素拓扑结构图分析、层次关系分析等,并构造出代表界面整体结构的数据结构体,其中的元素层次关系等将被记录,同时基于表格识别技术对界面中的表格内容进行行列关系抽取,利用NLP技术对表单的键值对关键词进行映射,最终将解析得到的结构体进行输出;

[0091] 4)组件展示界面展示拾取结果和界面结构;

[0092] 3.用户在展示界面确认所拾取到的元素就是要操作的元素后确认完成拾取;

[0093] 4.确认拾取完成后RPA流程工程信息中将保存该步所要操作的元素的信息,至此这个操作的RPA流程编辑环节结束,所编辑的RPA流程可用于后续流程运行;

[0094] 5.RPA流程包中包含了整个流程的代码信息和操作信息,其中每一步操作又包含要操作的元素信息或动作信息,当流程运行到该步骤时,RPA软件获取待操作的元素信息;

[0095] 6.根据待操作元素信息和操作动作信息完成指定的操作。

[0096] 本发明的技术关键点:

[0097] 1.创造性地提出一种自动选择和切换元素拾取方式的新方法,区别于传统的人工选择方式,能够为用户提供更合适的拾取方法并提升用户的开发效率,该种思路和方法为

独创；

[0098] 2.创新性地设计了包含预拾取模块和拾取方式自动选择模块在内的元素拾取方式自动切换机制和方法,该方法能保证针对具体界面自动为用户拾取到更精细、更准确的界面元素,此机制和方法为独创；

[0099] 采用目标检测、OCR、位置图分析、表格识别和NLP相结合的技术来分析界面结构,获取到更全面的界面信息,该项融合技术在RPA领域为首创。

[0100] 以上所述仅是本发明的示范性实施方式,而非用于限制本发明的保护范围,本发明的保护范围由所附的权利要求确定。

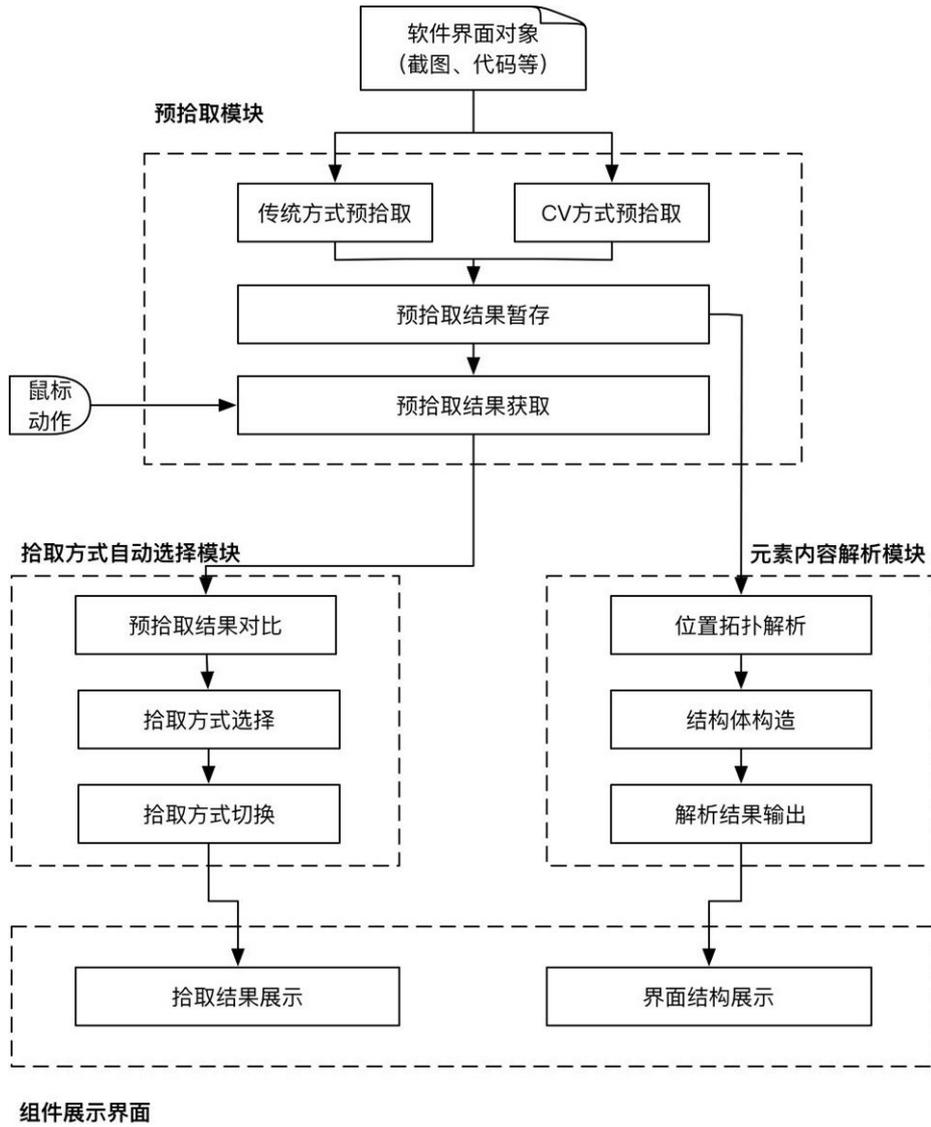


图1

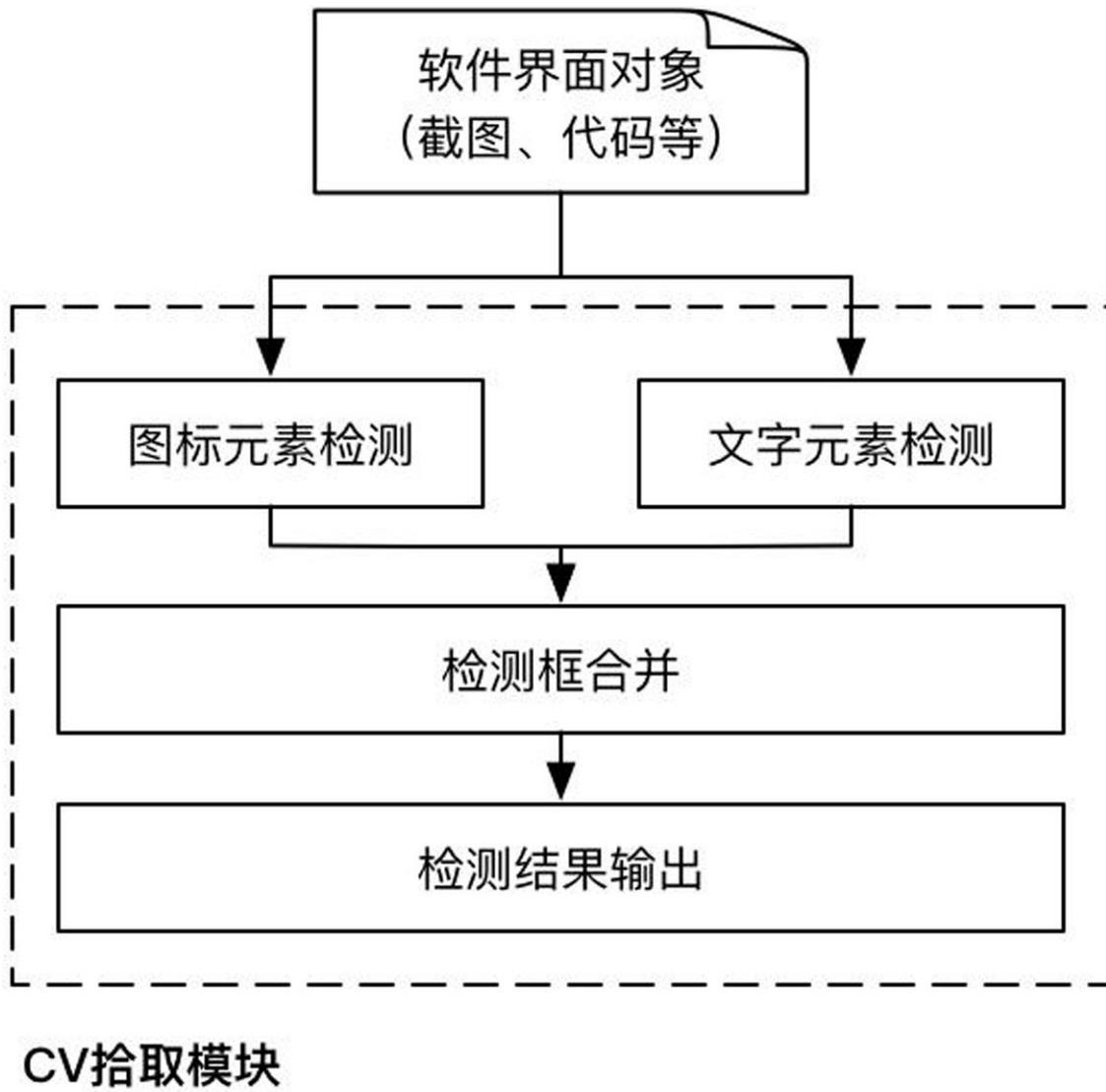


图2

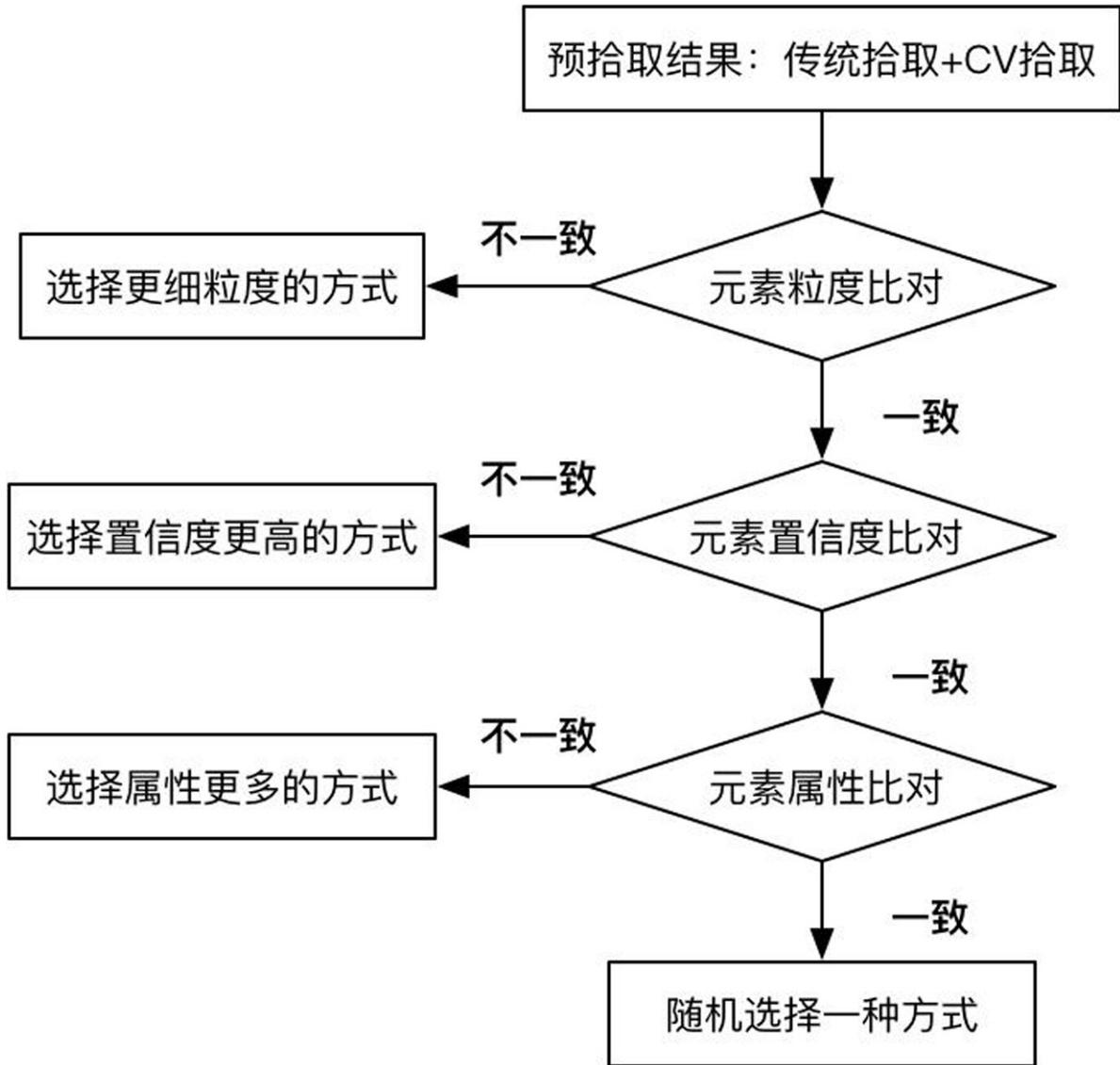


图3

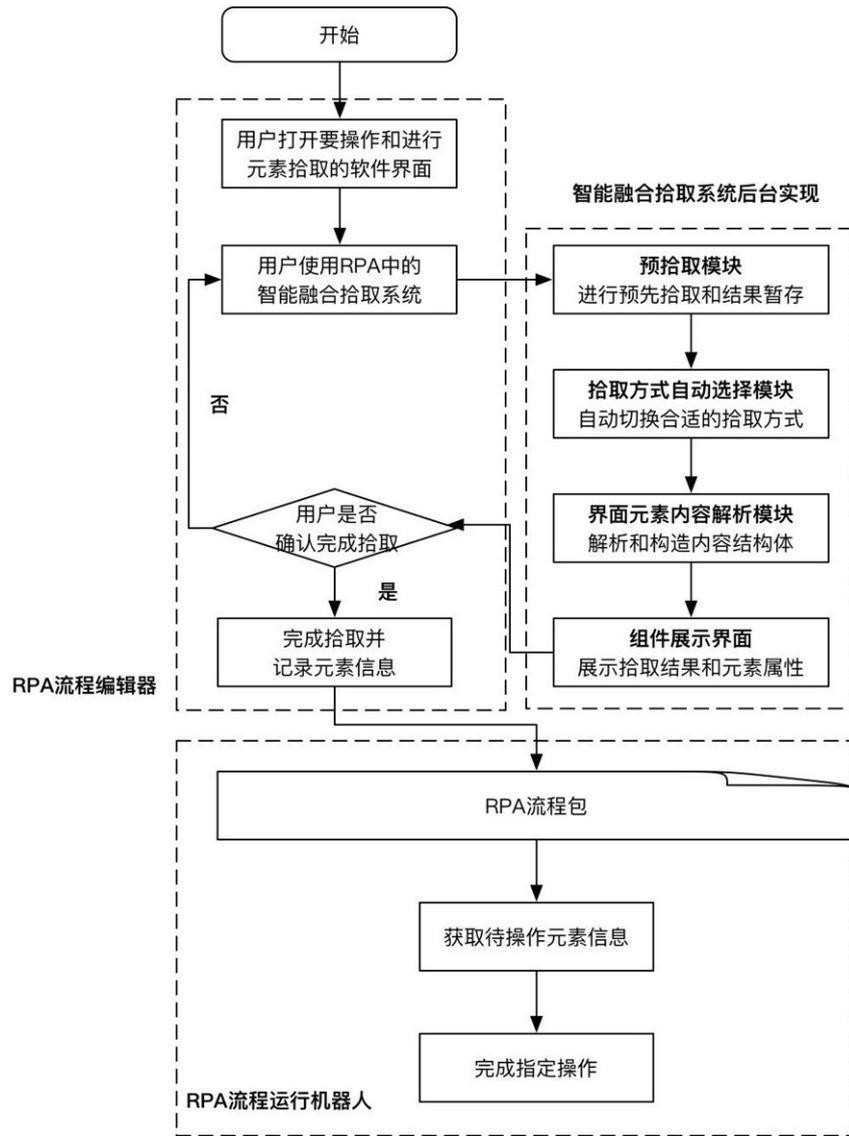


图4