



(12) 发明专利申请

(10) 申请公布号 CN 113971810 A

(43) 申请公布日 2022.01.25

(21) 申请号 202111260590.8

G06V 10/764 (2022.01)

(22) 申请日 2021.10.27

G06K 9/62 (2022.01)

(71) 申请人 北京百度网讯科技有限公司

地址 100085 北京市海淀区上地十街10号  
百度大厦2层

(72) 发明人 谢群义 钦夏孟 恩孟一 张栋栋  
黄聚 徐杨柳 陈毅 姚锟

(74) 专利代理机构 中科专利商标代理有限责任  
公司 11021

代理人 刘丽丽

(51) Int. Cl.

G06V 30/413 (2022.01)

G06V 30/414 (2022.01)

G06V 10/774 (2022.01)

G06V 10/82 (2022.01)

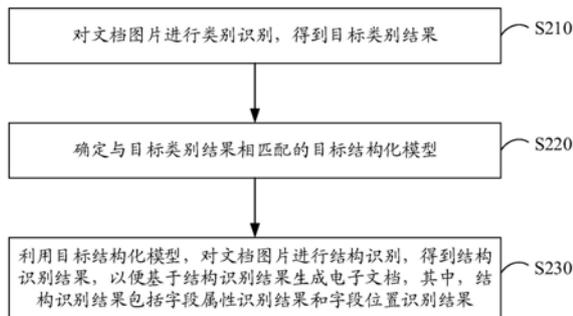
权利要求书2页 说明书11页 附图5页

(54) 发明名称

文档生成方法、装置、平台、电子设备以及存储介质

(57) 摘要

本公开提供了文档生成方法、装置、平台、电子设备、存储介质以及程序产品,涉及人工智能技术领域,尤其涉及计算机视觉和深度学习技术领域,可应用于文字识别等场景。具体实现方案为:对文档图片进行类别识别,得到目标类别结果;确定与目标类别结果相匹配的目标结构化模型;以及利用目标结构化模型,对文档图片进行结构识别,得到结构识别结果,以便基于结构识别结果生成电子文档,结构识别结果包括字段属性识别结果和字段位置识别结果。



1. 一种文档生成方法,包括:
  - 对文档图片进行类别识别,得到目标类别结果;
  - 确定与所述目标类别结果相匹配的目标结构化模型;以及
  - 利用所述目标结构化模型,对所述文档图片进行结构识别,得到结构识别结果,以便基于所述结构识别结果生成电子文档,其中,所述结构识别结果包括字段属性识别结果和字段位置识别结果。
2. 根据权利要求1所述的方法,其中,所述对文档图片进行类别识别,得到目标类别结果包括:
  - 对所述文档图片进行类别识别,得到类别结果;
  - 从至少一个模板类别中识别与所述类别结果相匹配的目标类别结果;以及
  - 在确定检测到所述目标类别结果的情况下,执行所述确定与所述目标类别结果相匹配的目标结构化模型的操作。
3. 根据权利要求1所述的方法,还包括,在所述利用所述目标结构化模型,对所述文档图片进行结构识别,得到结构识别结果之前:
  - 识别所述文档图片的图像方向;以及
  - 在所述图像方向与目标方向不匹配的情况下,对所述文档图片进行旋转校正。
4. 根据权利要求1所述的方法,其中,所述利用所述目标结构化模型,对所述文档图片进行结构识别,得到结构识别结果包括:
  - 利用所述目标结构化模型,对所述文档图片进行结构识别,从多个初始字段结构中确定所述结构识别结果。
5. 根据权利要求1或4所述的方法,还包括,在所述利用所述目标结构化模型,对所述文档图片进行结构识别,得到结构识别结果之后:
  - 基于所述字段位置识别结果,提取目标字段文字;以及
  - 对所述目标字段文字进行文字识别,得到文字识别结果。
6. 根据权利要求1所述的方法,还包括,在所述对文档图片进行类别识别,得到类别结果之前:
  - 对初始图片进行主体检测,得到主体位置结果;以及
  - 基于所述主体位置结果,从所述初始图片中得到所述文档图片。
7. 根据权利要求1-6任一项所述的方法,其中,所述文档图片包括纸质文档或电子文档的图片。
8. 一种文档生成装置,包括:
  - 类别识别模块,用于对文档图片进行类别识别,得到目标类别结果;
  - 模型匹配模块,用于确定与所述目标类别结果相匹配的目标结构化模型;以及
  - 结构识别模块,用于利用所述目标结构化模型,对所述文档图片进行结构识别,得到结构识别结果,以便基于所述结构识别结果生成电子文档,其中,所述结构识别结果包括字段属性识别结果和字段位置识别结果。
9. 根据权利要求8所述的装置,其中,所述类别识别模块包括:
  - 类别识别单元,用于对所述文档图片进行类别识别,得到类别结果;
  - 匹配单元,用于从至少一个模板类别中识别与所述类别结果相匹配的目标类别结果;

以及

执行单元,用于在确定检测到所述目标类别结果的情况下,执行所述确定与所述目标类别结果相匹配的目标结构化模型的操作。

10. 根据权利要求8所述的装置,还包括,在所述结构识别模块之前:

方向识别模块,用于识别所述文档图片的图像方向;以及

旋转校正模块,用于在所述图像方向与目标方向不匹配的情况下,对所述文档图片进行旋转校正。

11. 根据权利要求8所述的装置,其中,所述结构识别模块包括:

结构确定单元,用于利用所述目标结构化模型,对所述文档图片进行结构识别,从多个初始字段结构中确定所述结构识别结果。

12. 根据权利要求8或11所述的装置,还包括,在所述结构识别模块之后:

提取模块,用于基于所述字段位置识别结果,提取目标字段文字;以及

文字识别模块,用于对所述目标字段文字进行文字识别,得到文字识别结果。

13. 根据权利要求8所述的装置,还包括,在所述类别识别模块之前:

主体检测模块,用于对初始图片进行主体检测,得到主体位置结果;以及

主体确定模块,用于基于所述主体位置结果,从所述初始图片中得到所述文档图片。

14. 根据权利要求8-13任一项所述的装置,其中,所述文档图片包括纸质文档或电子文档的图片。

15. 一种文档生成平台,包括:

平台层,用于利用训练数据训练初始文档生成网络,得到文档生成网络,其中,所述文档生成网络用于对待处理文档图片进行类别识别,得到应用目标类别结果,确定与所述应用目标类别结果相匹配的目标结构化模型,利用所述目标结构化模型,对所述待处理文档图片进行结构识别,得到应用结构识别结果,所述应用结构识别结果包括应用字段属性识别结果和应用字段位置识别结果;

原子模型层,用于构建所述初始文档生成网络;以及

业务层,用于接收来自终端设备的所述待处理文档图片。

16. 一种电子设备,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指令,所述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权利要求1-7中任一项所述的方法。

17. 一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1-7中任一项所述的方法。

18. 一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据权利要求1-7中任一项所述的方法。

## 文档生成方法、装置、平台、电子设备以及存储介质

### 技术领域

[0001] 本公开涉及人工智能技术领域,尤其涉及计算机视觉和深度学习技术领域,可应用于文字识别等场景,具体涉及文档生成方法、装置、平台、电子设备、存储介质以及程序产品。

### 背景技术

[0002] 纸质文档电子化的需求广泛存在于日常的生产生活中,例如票据、证件、业务账单等文档的电子录入、存储、管理等。OCR (Optical Character Recognition, 光学字符识别) 技术为纸质文档电子化带来前所未有的变化。提高电子化的处理效率,解放人力。但是随着纸质文档的类型的不断增加,对OCR技术的要求也在不断地提升。

### 发明内容

[0003] 本公开提供了一种文档生成方法、装置、平台、电子设备、存储介质以及程序产品。

[0004] 根据本公开的一方面,提供了一种文档生成方法,包括:对文档图片进行类别识别,得到目标类别结果;确定与所述目标类别结果相匹配的目标结构化模型;以及利用所述目标结构化模型,对所述文档图片进行结构识别,得到结构识别结果,以便基于所述结构识别结果生成电子文档,其中,所述结构识别结果包括字段属性识别结果和字段位置识别结果。

[0005] 根据本公开的另一方面,提供了一种文档生成装置,包括:类别识别模块,用于对文档图片进行类别识别,得到目标类别结果;模型匹配模块,用于确定与所述目标类别结果相匹配的目标结构化模型;以及结构识别模块,用于利用所述目标结构化模型,对所述文档图片进行结构识别,得到结构识别结果,以便基于所述结构识别结果生成电子文档,其中,所述结构识别结果包括字段属性识别结果和字段位置识别结果。

[0006] 根据本公开的另一方面,提供了一种文档生成平台,包括:平台层,用于利用训练数据训练初始文档生成网络,得到文档生成网络,其中,所述文档生成网络用于对待处理文档图片进行类别识别,得到应用目标类别结果,确定与所述应用目标类别结果相匹配的目标结构化模型,利用所述目标结构化模型,对所述待处理文档图片进行结构识别,得到应用结构识别结果,所述应用结构识别结果包括应用字段属性识别结果和应用字段位置识别结果;原子模型层,用于构建所述初始文档生成网络;以及业务层,用于接收来自终端设备的所述待处理文档图片。

[0007] 根据本公开的另一方面,提供了一种电子设备,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指令,所述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上所述的方法。

[0008] 根据本公开的另一方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行如上所述的方法。

[0009] 根据本公开的另一方面,提供了一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现如上所述的方法。

[0010] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其他特征将通过以下的说明书而变得容易理解。

### 附图说明

[0011] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0012] 图1示意性示出了根据本公开实施例的文档生成方法及装置的示例性应用场景;

[0013] 图2示意性示出了根据本公开实施例的文档生成方法的流程图;

[0014] 图3示意性示出了根据本公开实施例的得到目标类别结果的流程图;

[0015] 图4示意性示出了根据本公开实施例的图像校正的流程图;

[0016] 图5示意性示出了根据本公开实施例的得到结构识别结果的流程图;

[0017] 图6示意性示出了根据本公开另一实施例的文档生成方法的流程图;

[0018] 图7示意性示出了根据本公开实施例的文档生成平台的框图;

[0019] 图8示意性示出了根据本公开实施例的文档生成装置的框图;以及

[0020] 图9示意性示出了根据本公开实施例的适于实现文档生成方法的电子设备的框图。

### 具体实施方式

[0021] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0022] 本公开提供了一种文档生成方法、装置、平台、电子设备、存储介质以及程序产品。

[0023] 根据本公开的实施例,文档生成方法可以包括:对文档图片进行类别识别,得到目标类别结果;确定与目标类别结果相匹配的目标结构化模型;以及利用目标结构化模型,对文档图片进行结构识别,得到结构识别结果,以便基于结构识别结果生成电子文档,其中,结构识别结果包括字段属性识别结果和字段位置识别结果。

[0024] 本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,且不违背公序良俗。

[0025] 图1示意性示出了根据本公开实施例的文档生成方法及装置的示例性应用场景。

[0026] 需要注意的是,图1所示仅为可以应用本公开实施例的应用场景的示例,以帮助本领域技术人员理解本公开的技术内容,但并不意味着本公开实施例不可以用于其他设备、系统、环境或场景。

[0027] 如图1所示,纸质文档可以通过电子设备例如扫描仪或者数码相机等转换为文档图片101,文档图片101中的主体图像可以为纸质文档上打印的字符。

[0028] 利用本公开实施例提供的文档生成方法,可以将文档图片101中的主体图像翻译成计算机数据,得到电子文档。

[0029] 本公开实施例所提供的文档生成方法一般可以由终端设备102执行。相应地,本公

开实施例所提供的文档生成装置也可以设置于终端设备102中。根据本公开的实施例,电子设备与终端设备102通信连接,可以经由电子设备将文档图片101传输至终端设备102。

[0030] 或者,本公开实施例所提供的文档生成方法一般也可以由服务器103执行。相应地,本公开实施例所提供的文档生成装置一般可以设置于服务器103中。本公开实施例所提供的文档生成方法也可以由不同于服务器103且能够与终端设备102和/或服务器103通信的服务器或服务器集群执行。相应地,本公开实施例所提供的文档生成装置也可以设置于不同于服务器103且能够与终端设备102和/或服务器103通信的服务器或服务器集群中。根据本公开的实施例,电子设备与终端设备102通信连接,可以经由电子设备将文档图片101传输至终端设备102,经由终端设备102上传至服务器103。

[0031] 根据本公开的实施例,终端设备102可以是具有显示屏并且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、膝上型便携计算机和台式计算机等等。

[0032] 根据本公开的实施例,服务器103可以是提供各种服务的服务器,例如对用户利用终端设备102所传输的文档图片进行文档生成方法处理的后台管理服务器(仅为示例)。

[0033] 应该理解,图1中的电子设备、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终端设备、网络和服务器。

[0034] 图2示意性示出了根据本公开实施例的文档生成方法的流程图。

[0035] 如图2所示,该方法包括操作S210~S230。

[0036] 操作S210,对文档图片进行类别识别,得到目标类别结果。

[0037] 操作S220,确定与目标类别结果相匹配的目标结构化模型。

[0038] 操作S230,利用目标结构化模型,对文档图片进行结构识别,得到结构识别结果,以便基于结构识别结果生成电子文档,其中,结构识别结果包括字段属性识别结果和字段位置识别结果。

[0039] 根据本公开的实施例,文档图片可以包括纸质文档的图片,例如文档图片是通过采集纸质文档中的信息得到的图片,文档图片中的主体图像可以与纸质文档中的文字内容相同。

[0040] 根据本公开的实施例,文档图片的类别可以指纸质文档的类别。例如,可以是身份证、银行卡,但是并不局限于此,还可以是收据、合同等类别。

[0041] 根据本公开的实施例,文档图片的类别不同,文档图片的字段结构例如字段属性、字段位置也相应不同。可以对文档图片进行类别识别来得到目标类别结果。从而确定文档图片的后续处理方式,例如基于目标类别结果来确定目标结构化模型。使得对文档图片的处理更具有针对性。

[0042] 根据本公开的实施例,目标结构化模型可以是与目标类别结果相匹配的结构化模型。可以预先训练得到多个结构化模型,多个结构化模型一一对应多个不同的文档图片的类别。

[0043] 根据本公开的实施例,利用结构化模型可以对文档图片进行结构识别,得到结构识别结果,例如字段属性识别结果和字段位置识别结果。

[0044] 根据本公开的实施例,文档图片中的主体图像一般是由多个字段结构的数据组成,例如,身份证文档图片中的主体图像一般包括名字字段、身份证号字段、地址字段、有效期字段等不同属性字段。通过结构识别可以对不同字段进行区分,基于字段属性识别结果

来确定重要程度,例如为必选字段或者可选字段等。此外,还可以通过结构识别得到字段位置识别结果,利用字段识别结果方便提取对应的字段文字,使得数据的提取更有针对性。

[0045] 根据本公开的实施例,可以基于结构识别,确定字段属性识别结果和字段位置识别结果。利用字段属性识别结果和字段位置识别结果最终生成电子文档,以便得到计算机可以处理的数据,方便保存与应用。

[0046] 根据本公开的实施例,利用目标类别结果来从多个结构化模型中确定目标结构化模型,可以无需人工干预即可明确与文档图片的结构识别方式相匹配的模型,实现解放人力、提高处理效率以及提高识别精度。

[0047] 下面参考图3~图6,结合具体实施例对图2所示的方法做进一步说明。

[0048] 图3示意性示出了根据本公开实施例的得到目标类别结果的流程图。

[0049] 如图3所示,针对操作S210对文档图片进行类别识别,得到目标类别结果,可以通过操作S310、S320、S331、以及S332来执行。

[0050] 操作S310,对文档图片进行类别识别,得到类别结果。

[0051] 操作S320,从至少一个模板类别中识别与类别结果相匹配的目标类别结果。

[0052] 操作S331,在确定检测到目标类别结果的情况下,执行确定与目标类别结果相匹配的目标结构化模型的操作。

[0053] 操作S332,在确定未检测到目标类别结果的情况下,停止操作。

[0054] 根据本公开的实施例,可以预先设置一个模板类别集合,模板类别集合包括至少一个模板类别。至少一个模板类别可以与至少一个文档图片的类别一一对应。

[0055] 根据本公开的实施例,文档图片可以是纸质文档的图片,但是并不局限于此,还可以是电子文档的图片。文档图片的类别可以与对应的纸质文档、或者对应的电子文档的类别相同。例如,可以是合同、收据、发票、证件、档案、卷宗、宣传单等类别中的一种或多种。

[0056] 根据本公开的实施例,文档图片的类别不同,文档图片的字段结构也相对应不同。例如,身份证文档图片中的字段属性、字段位置与合同文档图片中的字段属性、字段位置不同。因此,可以预先设定至少一个模板类别与至少一个结构化模型一一对应。即,每个模板类别配置有对应的结构化模型。也即,针对不同类别的文档图片配置有对应的结构化模型来进行结构识别处理。进而,利用目标类别结果来确定目标结构化模型,利用目标结构化模型来执行结构识别的操作,更具有针对性,避免采用通用模型所造成的识别准确率不高的问题。

[0057] 根据本公开的实施例,在未检测到目标类别结果的情况下,可以执行拒识别操作,即停止执行后续的结构识别操作。提高处理效率。

[0058] 根据本公开的实施例,可以利用主体分类模型来对文档图片进行类别识别,得到目标类别结果。

[0059] 根据本公开的实施例,对主体分类模型的网络架构不做限定,例如可以采用卷积神经网络(CNN)为核心网络,并结合双向长短期记忆网络(BiLSTM)得到主体分类模型。但是并不局限于此。只要是能够以文档图片作为输入数据,以目标类别结果作为输出数据的模型即可。

[0060] 根据本公开的实施例,在执行操作S220确定与目标类别结果相匹配的目标结构化模型之前,可以执行图像校正的预处理操作。例如,识别文档图片的图像方向。在图像方向

与目标方向不匹配的情况下,对文档图片进行旋转校正。在图像方向与目标方向匹配的情况下,可以直接执行后续操作S220。

[0061] 根据本公开的实施例,图像方向可以指文档图片中的主体图像的字符方向。

[0062] 根据本公开的实施例,目标方向可以指预定方向。例如,预定水平方向或者竖直方向为目标方向。

[0063] 根据本公开的示例性实施例,可以预定字符的可读方向例如水平方向为目标方向。

[0064] 图4示意性示出了根据本公开实施例的图像校正的流程图。

[0065] 如图4所示,图像方向可以指文档图片中的主体图像的字符方向。识别并确定文档图片410的图像方向B-B。预定水平方向为目标方向A-A。在图像方向B-B与目标方向A-A不匹配的情况下,例如呈90度夹角,可以对文档图片410进行旋转校正,得到旋转校正后的文档图片420。使得旋转校正后的文档图片420的图像方向与目标方向一致。在图像方向与目标方向匹配的情况下,即说明图像方向与目标方向一致,即可直接执行后续操作。

[0066] 需要说明的是,针对图像校正的预处理操作,可以在执行完类别识别操作之后执行,也可以在执行完类别识别操作之前执行,只要是在执行确定与目标类别结果相匹配的目标结构化模型之前操作即可。

[0067] 根据本公开的示例性实施例,可以利用方向判断模型来识别文档图片的图像方向,例如,利用方向判断模型来识别文档图片的图像方向,得到图像方向相对于目标方向的夹角结果,例如0度、90度、180度、270度等夹角结果。并根据夹角结果来对文档图片进行旋转校正,得到旋转校正后的文档图片。

[0068] 根据本公开的实施例,对方向判断模型的网络架构不做限定,例如可以采用OpenCV算法中的仿射变换函数(Affine Transform)。但是并不局限于此,还可以采用OpenCV算法与Python算法结合。只要是能够以文档图片作为输入数据,以旋转校正后的文档图片作为输出数据的模型即可。

[0069] 利用本公开实施例提供的图像校正的预处理操作,可以避免因文档图片的图像方向不利于字符识别而导致的后续结构识别的精度降低的问题。

[0070] 需要说明的是,在本公开的实施例中,文档生成方法并不局限于旋转校正的预处理操作,还可以包括变换校正,例如对具有不同程度的角度倾斜的图像,可以通过变换校正的预处理操作来处理,由此降低因角度倾斜而造成的干扰。

[0071] 根据本公开的示例性实施例,还可以执行裁切的预处理操作。

[0072] 例如,在执行操作S210在对文档图片进行类别识别,得到类别结果之前,可以对初始图片进行主体检测,得到主体位置结果;以及基于主体位置结果,从初始图片中得到文档图片。

[0073] 根据本公开的实施例,主体检测可以指对初始图片中的主体图像进行检测。例如,利用摄像装置采集身份证的图像而得到的身份证初始图片,会有可能在身份证的图像的周围存在大量的非必要区域。或者,合同初始图片中,在正文两侧和上下部分,也会存在预留的空隙区域。通过主体检测,可以对具有字符的主体图像区域与非必要区域进行区分。

[0074] 根据本公开的实施例,可以将例如具有字符的主体图像区域作为主体位置结果。

[0075] 根据本公开的实施例,基于主体位置结果,可以通过裁切的方式,从初始图片中得

到文档图片,无需人工干预,即可对初始图片进行裁切,去除非必要区域,由此排除背景干扰,进而减少图像处理的复杂度,提高对文档图片的类别的识别精度。

[0076] 根据本公开的示例性实施例,还可以采用主体检测模型来对初始图片进行主体检测,得到主体位置结果,并基于主体位置结果,从初始图片中得到文档图片。

[0077] 根据本公开的实施例,对主体检测模型的网络架构不做限定,例如可以采用OpenCV算法。但是并不局限于此,还可以采用Python、Ruby等通用算法。只要是能够以初始图片作为输入数据,以文档图片作为输出数据的模型即可。

[0078] 根据本公开的实施例,针对操作S230利用目标结构化模型,对文档图片进行结构识别,得到结构识别结果,可以通过如下操作来执行。

[0079] 例如,利用目标结构化模型,对文档图片进行结构识别,从多个初始字段结构中确定结构识别结果。

[0080] 根据本公开的实施例,目标结构化模型可以是与目标类别结果相匹配的结构化模型。可以预先训练得到多个结构化模型,多个结构化模型一一对应多个不同的文档图片的类别。

[0081] 根据本公开的实施例,利用结构化模型可以对文档图片进行结构识别,得到结构识别结果,例如字段属性识别结果和字段位置识别结果。

[0082] 根据本公开的实施例,对结构化模型的网络架构不做限定,例如可以采用CTPN (Connectionist Text Proposal Network,文本检测网络)得到的结构化模型。但是并不局限于此,还可以是卷积神经网络以及RPN网络 (Region Proposal Networks,区域生成网络)结合得到的结构化模型。只要是能够以文档图片或者旋转校正后的文档图片作为输入数据,以结果化识别结果作为输出数据的模型即可。

[0083] 图5示意性示出了根据本公开实施例的得到结构识别结果的流程图。

[0084] 如图5所示,利用目标结构化模型,对证件文档图片510中的主体图像进行结构识别。主体图像中包括多个初始字段结构,初始字段结构包括字段属性以及字段位置。字段属性可以为字段类别,例如证件名称字段、证件号字段、有效期字段、地址字段、使用须知字段等。可以从多个初始字段结构中确定结构识别结果。结构识别结果包括字段属性识别结果和字段位置识别结果。其中,字段属性识别结果可以包括证件名称字段的标签520、证件证号字段的标签530、有效期字段的标签540。字段位置识别结果可以为对应字段属性识别结果的位置标记结果,例如证件名称字段的位置框520'、证件证号字段的位置框530'、有效期字段的位置框540'。

[0085] 根据本公开的实施例,利用目标结构化模型,可以从文档图片中识别并确定出关键内容、目标内容,例如可以通过目标结构化模型来确定生成电子文档需要的字段文字,无需人工干预。此外,字段属性识别结果和字段位置识别结果还可以为后续字段文字提取以及文字识别提供基础,因此,有利于形成连贯的处理流程。

[0086] 根据本公开的实施例,在操作S230利用目标结构化模型,对文档图片进行结构识别,得到结构识别结果之后,还可以执行如下操作。

[0087] 例如,基于字段位置识别结果,提取目标字段文字;以及对目标字段文字进行文字识别,得到文字识别结果。

[0088] 根据本公开的实施例,可以利用文字识别模型来提取目标字段文字,并对目标字

段文字进行文字识别,得到文字识别结果。例如,基于字段位置识别结果获取的对应的图像区域输入至文字识别模型中,得到文字识别结果。根据本公开的实施例,基于文字识别结果生成电子文档。电子文档可以集中存储,使得查找方便快捷,并且有效防止文档的损坏与丢失。

[0089] 根据本公开的实施例,对文字识别模型的网络架构不做限定,例如可以采用卷积神经网络(CNN)为核心网络,并结合激活函数例如softmax得到文字识别模型。但是并不局限于此,还可以是卷积神经网络与循环神经网络(RNN)结合得到的文字识别模型。只要是能够以基于字段位置识别结果获取的对应的图像区域作为输入数据,以文字识别结果作为输出数据的模型即可。

[0090] 图6示意性示出了根据本公开另一实施例的文档生成方法的流程图。

[0091] 如图6所示,文档图片610可以作为输入数据,被主体检测模型620、主体分类模型630、方向判断模型640、目标结构化模型650以及文字识别模型660顺序处理,最终得到能够被计算机识别并处理的数据,生成电子文档670。

[0092] 根据本公开的实施例,由以上5个模型生成端到端的文档生成处理模型,形成流水线模式,实现文档生成的稳健性和高效性。

[0093] 根据本公开的实施例,主体检测模型、主体分类模型、方向判断模型、目标结构化模型、以及文字识别模型,可以分别单独进行模型训练以及模型测试。可以设计一种统一的训练样本,训练样本包括各种类型的文档图片。当模型训练完毕后,使用训练后的模型作为应用模型。可以将各个模型加载于终端设备中,由终端设备完成文档生成方法的操作。

[0094] 图7示意性示出了根据本公开实施例的文档生成平台的框图。

[0095] 如图7所示,文档生成平台700可以包括平台层710、原子模型层720和业务层730。

[0096] 平台层710,用于利用训练数据训练初始文档生成网络,得到文档生成网络,文档生成网络用于对待处理文档图片进行类别识别,得到应用目标类别结果,确定与应用目标类别结果相匹配的目标结构化模型,利用目标结构化模型,对待处理文档图片进行结构识别,得到应用结构识别结果,应用结构识别结果包括应用字段属性识别结果和应用字段位置识别结果。

[0097] 原子模型层720,用于构建初始文档生成网络。

[0098] 业务层730,用于接收来自终端设备的待处理文档图片。

[0099] 根据本公开的实施例,原子模型层可以为文档生成平台的底层,可以构建初始文档生成网络。初始文档生成网络可以包括初始主体检测模型、初始主体分类模型、初始方向判断模型、初始目标结构化模型以及初始文字识别模型。

[0100] 根据本公开的实施例,初始主体检测模型可以用来对初始样本图片进行主体检测,得到样本主体位置结果,并基于样本主体位置结果,从初始样本图片中通过例如裁切的方式得到样本文档图片。

[0101] 根据本公开的实施例,初始主体分类模型可以用来对样本文档图片进行类别识别,得到样本目标类别结果。

[0102] 根据本公开的实施例,初始方向判断模型可以用来识别样本文档图片的图像方向,得到图像方向相对于目标方向的样本夹角结果,并根据样本夹角结果来对样本文档图片进行旋转校正,得到旋转校正后的样本文档图片。

[0103] 根据本公开的实施例,初始目标结构化模型可以用来对样本文档图片进行结构识别,从多个初始样本字段结构中确定样本结构识别结果。

[0104] 根据本公开的实施例,初始文字识别模型可以用来基于样本结构识别结果,提取样本目标字段文字,以及对样本目标字段文字进行文字识别,得到样本文字识别结果。

[0105] 利用本公开实施例提供的原子模型层,可以构建多种不同类别的初始结构化模型,用于处理不同类别的样本文档图片,例如合同、收据、发票、证件、档案、卷宗、宣传单等类别,进而实现对不同待处理文档图片提供更具有针对性的文档生成服务。

[0106] 根据本公开的实施例,平台层可以为文档生成平台的中间层。平台层可以包括模块工具和服务工具等工具。

[0107] 根据本公开的实施例,模块工具可以包括数据解析工具、模型训练工具以及模型测试工具等。数据解析工具可以解析训练数据的数据类别,根据解析结果调用解析模型训练工具启动模型训练。例如,根据数据类别例如样本文档图片的类别的不同,确定与数据类别对应的初始目标结构化模型。模型训练工具用于利用训练数据例如样本文档图片训练初始文档生成网络。模型测试工具用于利用测试集对经训练的文档生成网络进行单独测试,确定性能优良的经训练的文档生成网络。

[0108] 根据本公开的实施例,服务工具可以是一个动态可配的服务工具,能够提供端到端的模型训练、测试服务,也可以提供利用模块工具优化后的文档生成网络进行端到端的文档生成服务。

[0109] 根据本公开的实施例,平台层可以将多项工具例如模块工具和服务工具等进行封装,形成流水线,进而通过流水线提供高效的数据处理能力、模型训练能力以及服务的开发能力。

[0110] 根据本公开的实施例,业务层可以为文档生成平台的顶层。业务层可以用于接收来自终端设备的待处理文档图片,并且对外支持针对多种不同类别的待处理文档图片的文档生成服务。待处理文档图片的类别可以是例如合同、收据、发票、证件、档案、卷宗、宣传单等。

[0111] 利用本公开实施例提供的文档生成平台,可以实现数据管理、端到端的模型训练和服务部署等,提高用户体验。

[0112] 图8示意性示出了根据本公开实施例的文档生成装置的框图。

[0113] 如图8所示,文档生成装置800可以包括类别识别模块810、模型匹配模块820、结构识别模块830。

[0114] 类别识别模块810,用于对文档图片进行类别识别,得到目标类别结果。

[0115] 模型匹配模块820,用于确定与目标类别结果相匹配的目标结构化模型。

[0116] 结构识别模块830,用于利用目标结构化模型,对文档图片进行结构识别,得到结构识别结果,以便基于结构识别结果生成电子文档,其中,结构识别结果包括字段属性识别结果和字段位置识别结果。

[0117] 根据本公开的实施例,类别识别模块可以包括类别识别单元、匹配单元、执行单元。

[0118] 类别识别单元,用于对文档图片进行类别识别,得到类别结果。

[0119] 匹配单元,用于从至少一个模板类别中识别与类别结果相匹配的目标类别结果。

[0120] 执行单元,用于在确定检测到目标类别结果的情况下,执行确定与目标类别结果相匹配的目标结构化模型的操作。

[0121] 根据本公开的实施例,文档生成装置还可以包括,在结构识别模块之前:方向识别模块、旋转校正模块。

[0122] 方向识别模块,用于识别文档图片的图像方向。

[0123] 旋转校正模块,用于在图像方向与目标方向不匹配的情况下,对文档图片进行旋转校正。

[0124] 根据本公开的实施例,结构识别模块可以包括结构识别单元、结构确定单元。

[0125] 结构确定单元,用于利用目标结构化模型,对文档图片进行结构识别,从多个初始字段结构中确定结构识别结果。

[0126] 根据本公开的实施例,文档生成装置还可以包括,在结构识别模块之后:提取模块、文字识别模块。

[0127] 提取模块,用于基于字段位置识别结果,提取目标字段文字。

[0128] 文字识别模块,用于对目标字段文字进行文字识别,得到文字识别结果。

[0129] 根据本公开的实施例,文档生成装置还可以包括,在类别识别模块之前:主体检测模块、主体确定模块。

[0130] 主体检测模块,用于对初始图片进行主体检测,得到主体位置结果。

[0131] 主体确定模块,用于基于主体位置结果,从初始图片中得到文档图片。

[0132] 根据本公开的实施例,文档图片包括纸质文档或电子文档的图片。

[0133] 根据本公开的实施例,本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0134] 根据本公开的实施例,一种电子设备,包括:至少一个处理器;以及与至少一个处理器通信连接的存储器;其中,存储器存储有可被至少一个处理器执行的指令,指令被至少一个处理器执行,以使至少一个处理器能够执行如上所述的方法。

[0135] 根据本公开的实施例,一种存储有计算机指令的非瞬时计算机可读存储介质,其中,计算机指令用于使计算机执行如上所述的方法。

[0136] 根据本公开的实施例,一种计算机程序产品,包括计算机程序,计算机程序在被处理器执行时实现如上所述的方法。

[0137] 图9示出了可以用来实施本公开的实施例的示例电子设备900的示意性框图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0138] 如图9所示,设备900包括计算单元901,其可以根据存储在只读存储器(ROM) 902中的计算机程序或者从存储单元908加载到随机访问存储器(RAM) 903中的计算机程序,来执行各种适当的动作和处理。在RAM 903中,还可存储设备900操作所需的各种程序和数据。计算单元901、ROM 902以及RAM 903通过总线904彼此相连。输入/输出(I/O)接口905也连接至总线904。

[0139] 设备900中的多个部件连接至I/O接口905,包括:输入单元906,例如键盘、鼠标等;输出单元907,例如各种类型的显示器、扬声器等;存储单元908,例如磁盘、光盘等;以及通信单元909,例如网卡、调制解调器、无线通信收发机等。通信单元909允许设备900通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0140] 计算单元901可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元901的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元901执行上文所描述的各个方法和处理,例如文档生成方法。例如,在一些实施例中,文档生成方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元908。在一些实施例中,计算机程序的部分或者全部可以经由ROM 902和/或通信单元909而被载入和/或安装到设备900上。当计算机程序加载到RAM 903并由计算单元901执行时,可以执行上文描述的文档生成方法的一个或多个步骤。备选地,在其他实施例中,计算单元901可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行文档生成方法。

[0141] 本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、复杂可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0142] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0143] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0144] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的

反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0145] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(LAN)、广域网(WAN)和互联网。

[0146] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。服务器可以是云服务器,也可以是分布式系统的服务器,或者是结合了区块链的服务器。

[0147] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0148] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

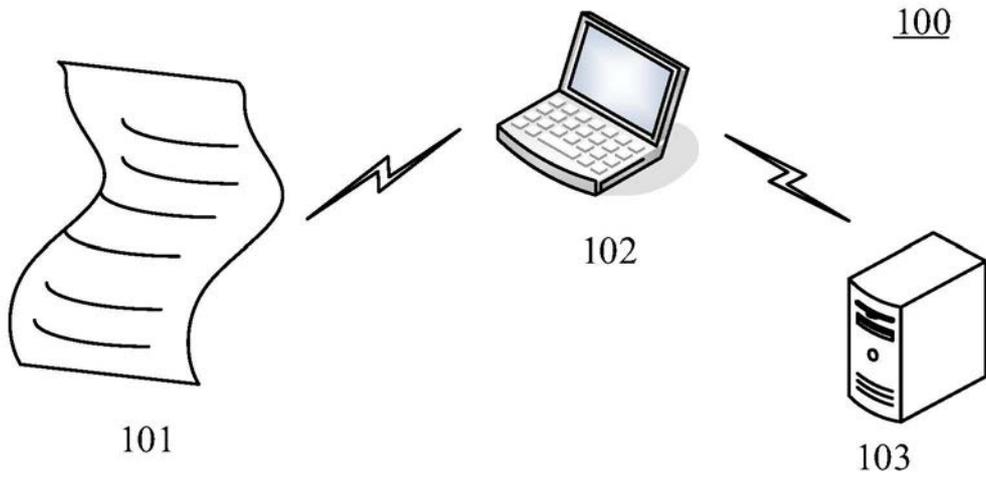


图1

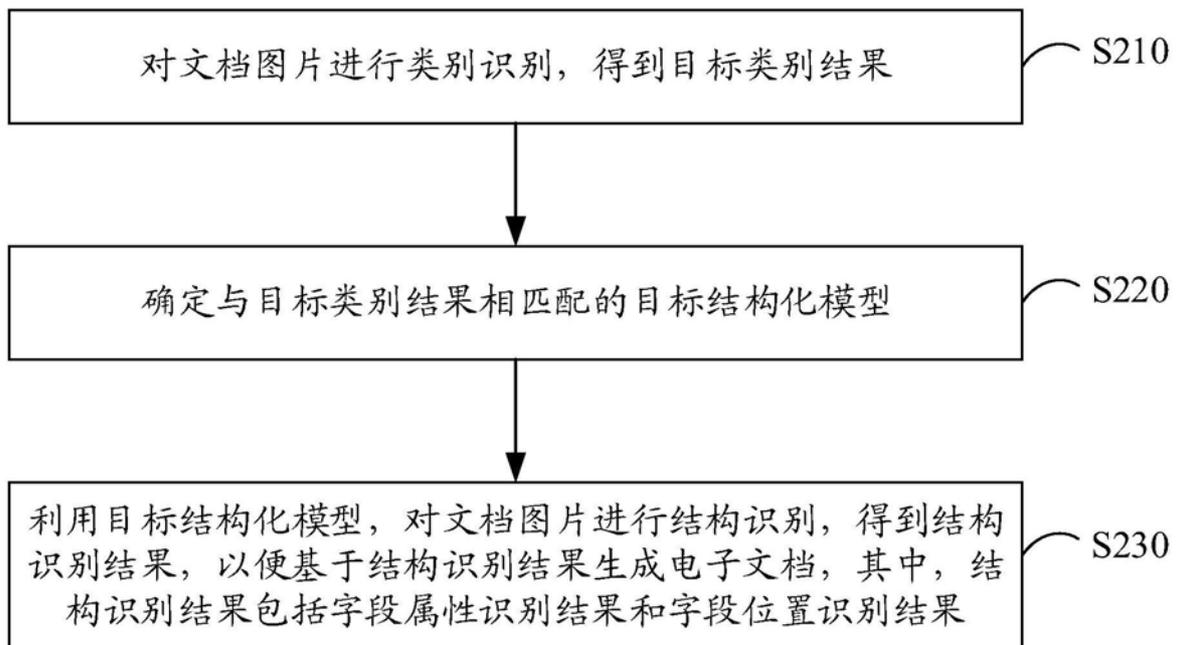


图2

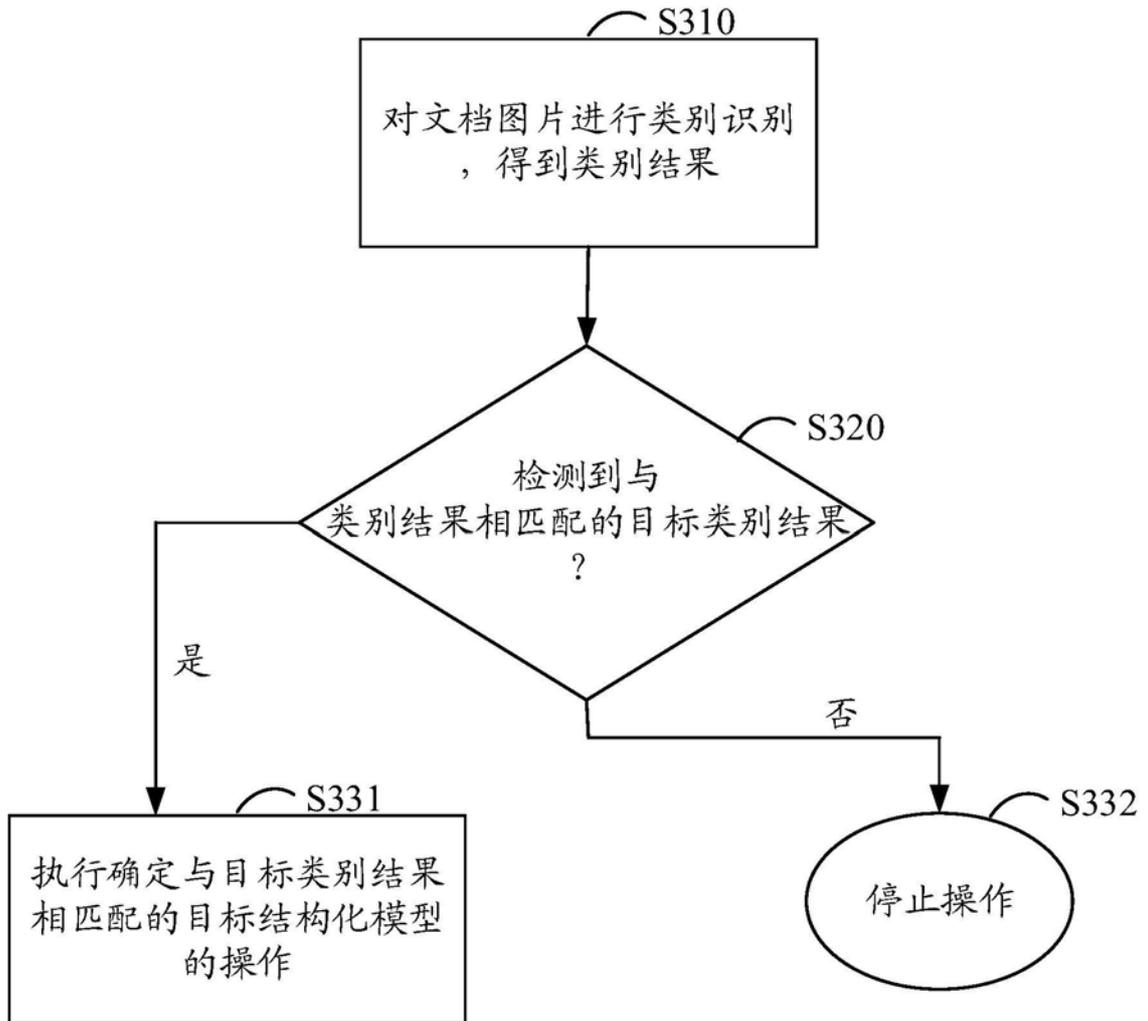


图3

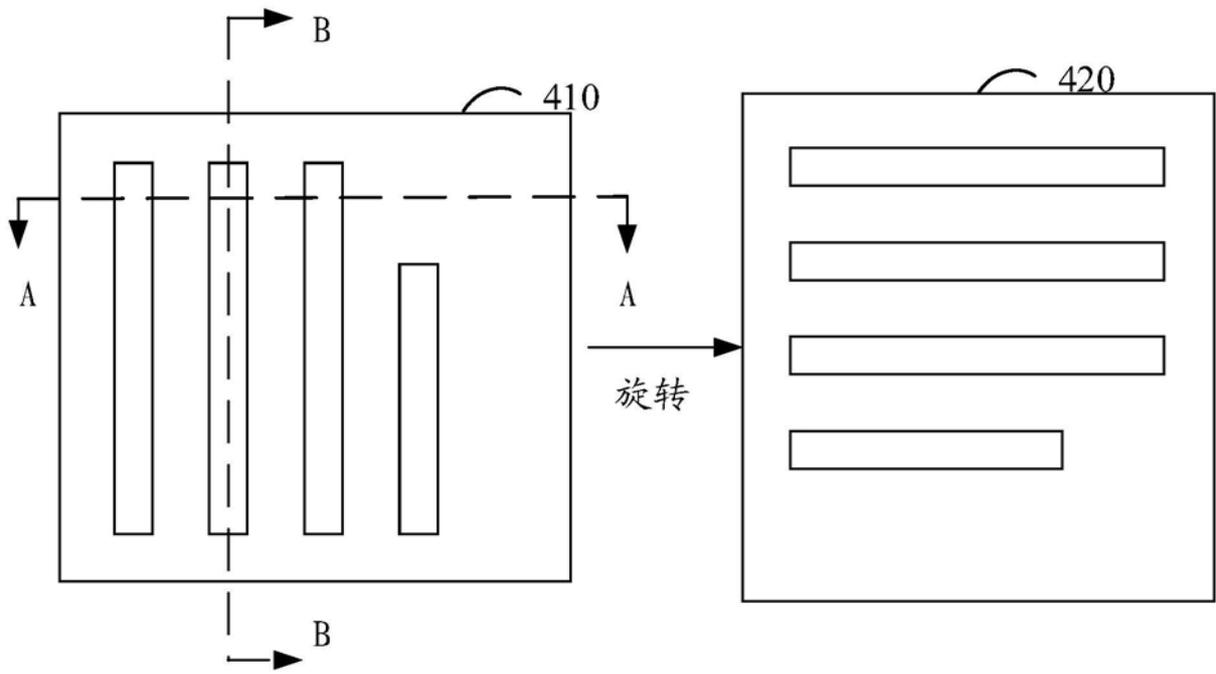


图4

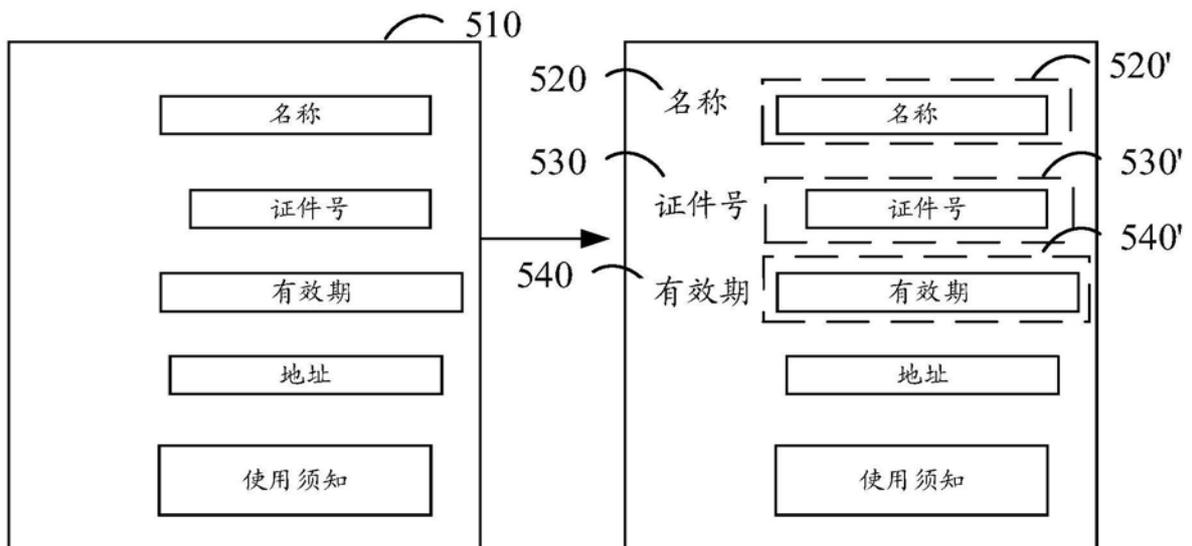


图5

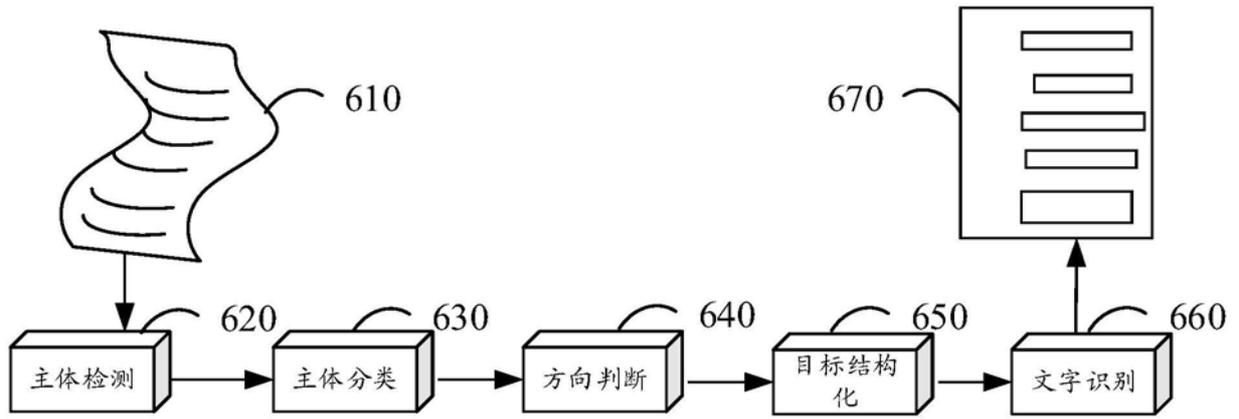


图6



图7

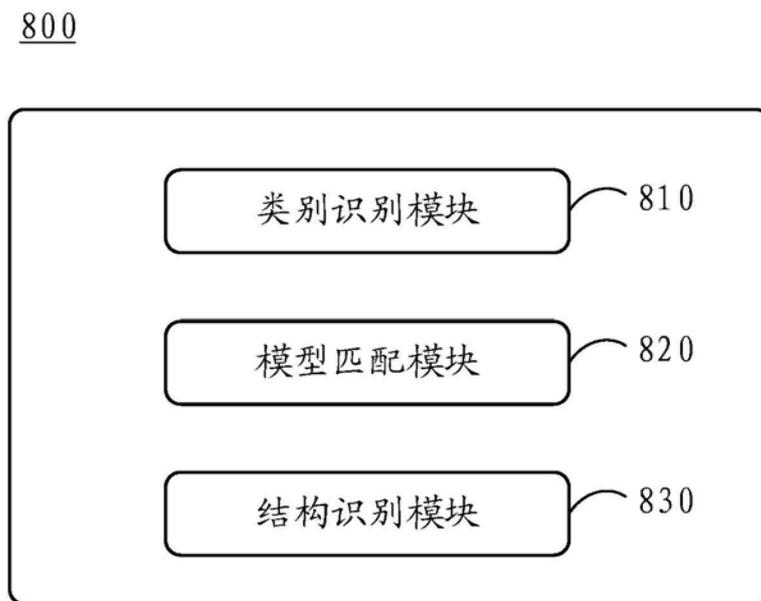


图8

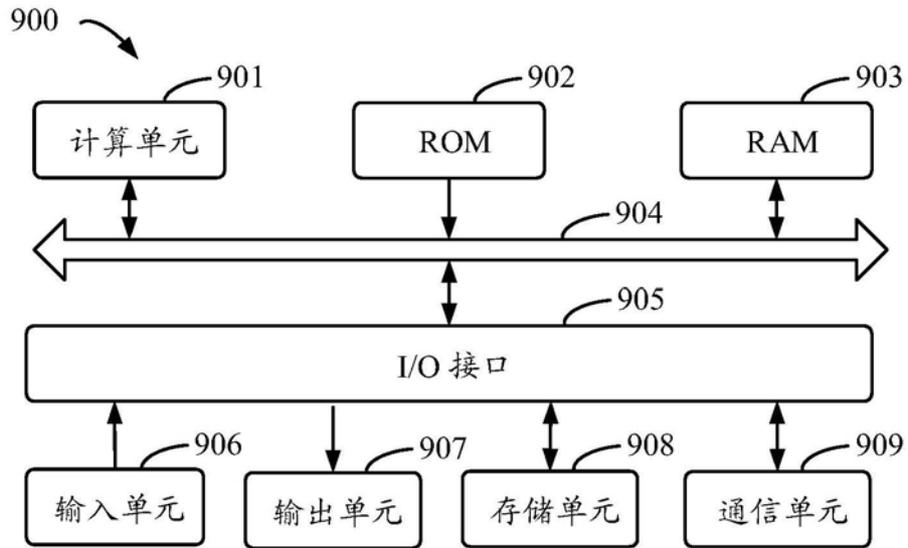


图9