

# (12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织  
国际局

(43) 国际公布日  
2022 年 5 月 19 日 (19.05.2022)



(10) 国际公布号  
**WO 2022/100376 A1**

(51) 国际专利分类号:  
**G06K 9/00** (2022.01) **G06F 40/166** (2020.01)

(21) 国际申请号: PCT/CN2021/124605

(22) 国际申请日: 2021 年 10 月 19 日 (19.10.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:  
202011264865.0 2020 年 11 月 13 日 (13.11.2020) CN

(71) 申请人: 深圳壹账通智能科技有限公司(ONE CONNECT SMART TECHNOLOGY CO., LTD. (SHENZHEN)) [CN/CN]; 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518000 (CN)。

(72) 发明人: 高超(GAO, Chao); 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong

518000 (CN)。徐国强(XU, Guoqiang); 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518000 (CN)。

(74) 代理人: 深圳中一联合知识产权代理有限公司(SHENZHEN ZHONGYI UNION INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深圳市福田区莲花街道紫荆社区深南大道 6008 号深圳特区报业大厦 33 层, Guangdong 518034 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** TEXT PARAGRAPH STRUCTURE RESTORATION METHOD AND APPARATUS, AND DEVICE AND COMPUTER STORAGE MEDIUM

(54) 发明名称: 文本段落结构还原方法、装置、设备及计算机存储介质

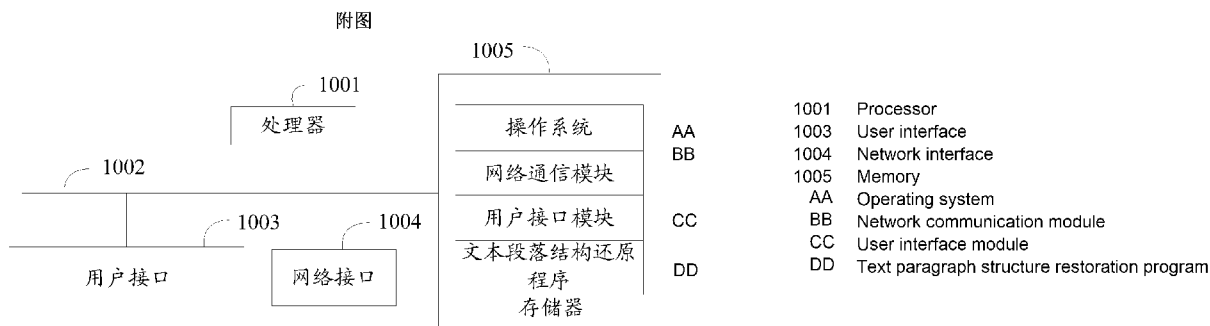


图 1

(57) **Abstract:** The present application relates to the technical field of artificial intelligence. Disclosed are a text paragraph structure restoration method and apparatus, and a device and a computer storage medium. The method comprises: carrying out recognition on a target picture, and on the basis of a recognition result of the recognition, determining all text boxes in the target picture and text box positions of all of the text boxes; sorting the text boxes according to the text box positions, and on the basis of a sorting result of the sorting, inputting text features of all of the text boxes into a preset deep learning model for training; and organizing all of the text boxes on the basis of a result of the training, so as to obtain all text paragraphs corresponding to the target picture. By means of the present application, the accuracy of text paragraph structure restoration is improved.

(57) **摘要:** 本申请涉及人工智能技术领域, 公开了一种文本段落结构还原方法、装置、设备及计算机存储介质, 该方法包括: 对目标图片进行识别, 基于所述识别的识别结果确定所述目标图片中所有文本框和各所述文本框的文本框位置; 根据各所述文本框位置对各所述文本框进行排序, 并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练; 基于所述训练结果对各所述文本框进行合并处理, 以获取所述目标图片对应的所有文本段落。本申请提高了文本段落结构还原的准确性。

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,  
UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国 (除另有指明, 要求每一种可提供的地区  
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,  
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,  
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,  
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,  
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,  
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,  
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

## 发明名称：文本段落结构还原方法、装置、设备及计算机存储介质

[0001] 本申请要求于2020年11月13日提交中国专利局、申请号为202011264865.0、发明名称为“文本段落结构还原方法、装置、设备及计算机存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

### 技术领域

[0002] 本申请涉及人工智能技术领域，尤其涉及一种文本段落结构还原方法、装置、设备以及计算机可读存储介质。

### 背景技术

[0003] 在纸质文档电子化的过程中，需要将文档录入并保留原本格式，发明人意识到，目前基于文本行的检测、识别方法无法直接得到文本段落信息。目前存在两种方法，即自上而下，也就是先进行整个页面的版面分析，分割出段落，再对段落区域中的文本行进行检测识别。这类方法在做版面分析时不能捕捉局部文字细节特征，并且只使用图片信息而没有文字内容信息，准确率不高。或自下而上，即先检测出文本行，再对文本行进行合并，得到段落。这类方法主要是通过文本框位置的相容性，使用一定的规则或启发式算法，对文本框合并得到段落，需要手工提取大量特征，并且难以参考文字内容信息，因此准确率也不高。

### 发明概述

#### 技术问题

[0004] 本申请的主要目的在于提供一种文本段落结构还原方法、装置、设备及计算机存储介质，旨在解决如何提高文本段落结构还原的准确性的技术问题。

#### 问题的解决方案

#### 技术解决方案

[0005] 为实现上述目的，本申请提供一种文本段落结构还原方法，所述文本段落结构还原方法包括：

[0006] 对目标图片进行识别，基于所述识别的识别结果确定所述目标图片中所有文本

框和各所述文本框的文本框位置；

- [0007] 根据各所述文本框位置对各所述文本框进行排序，并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练；
- [0008] 基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落。
- [0009] 可选地，基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落的步骤，包括：
- [0010] 基于所述训练结果确定各所述文本框对应的文本标签，遍历各所述文本标签，检测遍历的文本标签对应的遍历标签信息是否为段落；
- [0011] 若所述遍历标签信息是段落，则确定所述遍历的文本标签对应的文本框为所述目标图片对应的文本段落。
- [0012] 可选地，检测遍历的文本标签对应的遍历标签信息是否为段落的步骤之后，包括：
- [0013] 若否，则检测所述遍历标签信息是否为段落内容；
- [0014] 若所述遍历标签信息是段落内容，则确定遍历的文本标签的前一位文本标签的标签信息是否为段落起始信息；
- [0015] 若所述前一位文本标签的标签信息是段落起始信息，则基于所述遍历的文本标签和前一位文本标签确定所述目标图片对应的文本段落。
- [0016] 可选地，基于所述遍历的文本标签和前一位文本标签确定所述目标图片对应的文本段落的步骤，包括：
- [0017] 检测各所述文本标签中是否存在连续相邻内容标签；
- [0018] 若存在连续相邻内容标签，则确定所述连续相邻内容标签中是否存在遍历的文本标签；
- [0019] 若不存在遍历的文本标签，则将所述遍历的文本标签对应的文本框和所述前一位文本标签对应的文本框进行合并，以获取合并后的文本框，并将所述合并后的文本框作为所述目标图片对应的文本段落。
- [0020] 可选地，确定所述连续相邻内容标签中是否存在遍历的文本标签的步骤之后，包括：

- [0021] 若存在遍历的文本标签，则将具有遍历的文本标签的连续相邻内容标签对应的所有文本框和所述前一位文本标签对应的文本框进行合并，以获取合并文本框，并将合并文本框作为所述图片对应的文本段落。
- [0022] 可选地，基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练的步骤，包括：
- [0023] 依次提取各所述文本框的文本特征，根据所述排序的排序结果将各所述文本特征融合为序列特征，并将所述序列特征输入至预设的深度学习模型进行训练。
- [0024] 可选地，依次提取各所述文本框的文本特征的步骤，包括：
- [0025] 依次遍历各所述文本框，并提取遍历的文本框的位置特征，语言特征和图像特征，将所述位置特征，语言特征和图像特征作为所述遍历的文本框的文本特征。
- [0026] 此外，为实现上述目的，本申请还提供一种文本段落结构还原装置，所述文本段落结构还原装置包括：
- [0027] 确定模块，用于对目标图片进行识别，基于所述识别的识别结果确定所述目标图片中所有文本框和各所述文本框的文本框位置；
- [0028] 输入模块，用于根据各所述文本框位置对各所述文本框进行排序，并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练；
- [0029] 获取模块，用于基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落。
- [0030] 此外，为实现上述目的，本申请还提供一种文本段落结构还原设备；
- [0031] 所述文本段落结构还原设备包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，其中：
- [0032] 所述计算机程序被所述处理器执行时实现如上所述的文本段落结构还原方法的步骤。
- [0033] 此外，为实现上述目的，本申请还提供一种计算机可读存储介质；
- [0034] 所述计算机可读存储介质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上所述的文本段落结构还原方法的步骤。

## 发明的有益效果

### 有益效果

[0035] 本申请通过对目标图片进行识别，基于所述识别的识别结果确定所述目标图片中所有文本框和各所述文本框的文本框位置；根据各所述文本框位置对各所述文本框进行排序，并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练；基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落。通过根据对目标图片的识别结果确定各个文本框和文本框位置，并根据各个文本框位置对各个文本框进行排序，将各个文本框的文本特征输入至预设的深度学习模型进行训练，基于训练结果对各个文本框进行合并处理，以获取文本段落，从而避免了用户手动操作，导致得到的文本段落的准确性较低的现象发生，提高了文本段落结构还原的准确性。

## 对附图的简要说明

### 附图说明

[0036] 图1是本申请实施例方案涉及的硬件运行环境的文本段落结构还原设备结构示意图；

[0037] 图2为本申请文本段落结构还原方法第一实施例的流程示意图；

[0038] 图3为本申请文本段落结构还原装置的功能模块示意图；

[0039] 图4为本申请文本段落结构还原方法的文本框排序示意图。

[0040] 本申请目的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

## 发明实施例

### 本发明的实施方式

[0041] 应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

[0042] 如图1所示，图1是本申请实施例方案涉及的硬件运行环境的文本段落结构还原设备结构示意图。

[0043] 如图1所示，该文本段落结构还原设备可以包括：处理器1001，例如CPU，网络

接口1004，用户接口1003，存储器1005，通信总线1002。其中，通信总线1002用于实现这些组件之间的连接通信。用户接口1003可以包括显示屏（Display）、输入单元比如键盘（Keyboard），可选用户接口1003还可以包括标准的有线接口、无线接口。网络接口1004可选的可以包括标准的有线接口、无线接口（如WI-FI接口）。存储器1005可以是高速RAM存储器，也可以是稳定的存储器（non-volatile memory），例如磁盘存储器。存储器1005可选的还可以是独立于前述处理器1001的存储装置。

[0044] 可选地，文本段落结构还原设备还可以包括摄像头、RF（Radio Frequency，射频）电路，传感器、音频电路、WiFi模块等等。其中，传感器比如光传感器、运动传感器以及其他传感器。具体地，光传感器可包括环境光传感器及接近传感器，其中，环境光传感器可根据环境光线的明暗来调节显示屏的亮度。当然，文本段落结构还原设备还可配置陀螺仪、气压计、湿度计、温度计、红外线传感器等其他传感器，在此不再赘述。

[0045] 本领域技术人员可以理解，图1中示出的文本段落结构还原设备结构并不构成对文本段落结构还原设备的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置。

[0046] 如图1所示，作为一种计算机存储介质的存储器1005中可以包括操作系统、网络通信模块、用户接口模块以及文本段落结构还原程序。

[0047] 在图1所示的文本段落结构还原设备中，网络接口1004主要用于连接后台服务器，与后台服务器进行数据通信；用户接口1003主要用于连接客户端（用户端），与客户端进行数据通信；而处理器1001可以用于调用存储器1005中存储的文本段落结构还原程序，并执行本申请实施例提供的文本段落结构还原方法。

[0048] 参照图2，本申请提供一种文本段落结构还原方法，在文本段落结构还原方法一实施例中，文本段落结构还原方法包括以下步骤：

[0049] 步骤S10，对目标图片进行识别，基于所述识别的识别结果确定所述目标图片中所有文本框和各所述文本框的文本框位置；

[0050] 在本实施例中，当需要对目标图片中的文本段落进行还原时，可以将对目标图片检测识别得到的文本行信息（即文本框），并转换为序列特征输入至预设的

深度学习模型中，通过深度学习模型进行序列标注，从而得到每个文本框的类别，根据各个类别分别进行合并，得到具体的文本段落。

[0051] 因此在本实施例中可以先对目标图片进行识别，而识别的方式可以通过文本识别技术，如OCR (Optical Character Recognition, 光学字符识别) 技术来对目标图片进行识别，以确定目标图片中是否存在文本内容，若存在文本内容，且文本内容分布在不同的位置，则可以根据OCR识别的识别结果来获取目标图片中的各个文本框位置 (如 $x_1$ ,  $y_1$ ,  $x_2$ ,  $y_2$ ) 和文字内容信息，还可以是采用其它文本识别模型进行。并且当通过文本识别技术对目标图片进行识别获取到文本框和文本框的文本框位置后，则检测文本框是否有且只有一个，若是，则可以直接将该文本框作为目标图片中的文本段落，其文本框位置也就是文本段落的位置。但是若检测到的文本框存在多个，则需要根据对各个文本框及其对应的文本框位置来确定目标图片中的文本段落。

[0052] 步骤S20，根据各所述文本框位置对各所述文本框进行排序，并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练；

[0053] 当获取到各个文本框和各个文本框的文本框位置后，可以根据各个文本框位置对各个文本框进行编号，而编号的顺序可以根据用户的需求自行进行设置，如可以设置从目标图片的上方开始进行编号，直至每个文本框都具有各自的编号，并根据各个编号对各个文本框件排序，得到排序后的各个文本框。例如，如图4所示，可以对各个文本框按照文本框的文本框位置，从上到下，从左到右依次进行编号排序，如1-12。

[0054] 当完成各个文本框的排序，得到排序的排序结果后，可以依次获取各个文本框的文本特征，如位置特征，语言特征和图像特征等。其中，位置特征可以是文本框的顶点坐标，中心点，文本框的宽和高等特征。语言特征可以是文本框中文本的语言模型特征，例如：文本的词向量，句向量，文本ngram (n元模型) 得分等。图像特征可以是使用卷积神经网络对图像中的文本区域进行特征提取后的特征等。并将获取到的文本特征作为序列特征输入到提前设置训练好的深度学习模型中进行训练，以获取训练结果。

[0055] 而对深度学习模型的训练过程可以先收集大量文档图片，并使用OCR系统进

行文字的检测识别，得到文档图片中各个文本框的三类特征，即位置特征，语言特征和图像特征。并且还会人工对文档图片中各个文本框的三类特征进行标注，即标注文档图片中各个文本框属于文本段落的哪一部分，确定文档图片中各个文本框的人工标注结果。再将各个文本框的三类特征输入到深度学习模型中进行训练，得到模型输出结果，即文档图片中各个文本框的标签（即具有文本段落信息的标签），并依次遍历文档图片中各个文本框的标签，将遍历的文档图片中文本框的标签和与其对应的人工标注结果进行比对，若存在差异，则需要对深度学习模型进行优化，如使用梯度下降法优化深度学习模型，直至优化完成，得到训练好的深度学习模型，即预设的深度学习模型。其中，深度学习模型可以为各种序列模型，如循环神经网络模型，卷积神经网络模型等。

[0056] 步骤S30，基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落。

[0057] 当获取到训练结果后，可以得到各个文本框的标签，如BIO标签，并各个文本框对应的标签进行合并操作，以得到目标图片对应的所有文本段落。而合并操作可以是根据不同的BIO标签进行。例如，若文本框的标签是O标签，则可以确定该文本框自成一个段落，其文本框位置即文本段落范围。若文本框的标签是B标签，则可以确定该文本段落的覆盖范围是从B标签所在文本框开始的，直至序列中最后一个连续的I标签结束，此时该文本段落的段落范围可以是B标签和此连续的I标签对应的所有文本框的文本框位置。而目标图片中的所有文本段落范围则可以根据各个BIO标签来确定。

[0058] 也就是当获取到深度学习模型的训练结果后，可以根据此训练结果来确定各个文本框对应的文本标签，如B标签，I标签和O标签等。然后再遍历各个文本框对应的文本标签，并需要检测遍历的文本标签对应的标签信息，即遍历标签信息。再确定该遍历标签信息是否为段落，以便根据确定结果来判断遍历的文本标签对应的文本框是否为一个文本段落，如O标签对应的标签信息。当经过判断发现遍历标签信息是段落，则可以确定遍历标签信息对应的文本框为目标图片对应的文本段落。

[0059] 但是若遍历标签信息不是段落，则还需要检查遍历标签信息是否为段落内容，

如I标签对应的标签信息。当经过判断发现遍历标签信息是段落内容时，则需要确定遍历的文本标签的前一位标签的标签信息是否为段落起始信息，如B标签对应的标签信息，若是，则可以直接根据遍历的文本标签对应的文本框和前一位文本标签对应的文本框来确定目标图片的文本段落。而且在确定文本段落时，还需要检测在各个文本框中是否存在连续相邻内容标签，即确定是否存在连续相邻的内容标签，如连续相邻的I标签。若存在连续相邻内容标签，则还需要判断连续相邻内容标签中是否存在本次遍历的文本标签，若不存在遍历的文本标签，则可以直接将遍历的文本标签对应的文本框和前一位文本标签对应的文本框进行合并处理，并将合并后的文本框作为目标图片对应的文本段落。

[0060] 但是若存在遍历的文本标签，则需要确定具有遍历的文本标签的所有相邻文本标签，并将具有遍历的文本标签的连续相邻内容标签对应的所有文本框进行合并处理。例如若遍历的文本标签是I4，连续相邻的文本标签是I1-I5，则连续相邻的文本标签中包含有遍历的文本标签I4，此时就可以将连续相邻的所有文本标签对应的文本框进行合并，即将I1，I2，I3，I4和I5对应的文本框进行合并，得到合并后的文本框，再将其与前一位文本标签对应的文本框进行合并，即和B标签对应的文本框进行合并处理，得到合并文本框，此时就可以直接将合并文本框作为图片对应的文本段落。

[0061] 并且在本实施例中，通过对目标图片进行检测，确定各个文本框及各个文本框的文本特征，根据各个文本特征形成序列特征输入至深度学习模型中进行训练预测，能够避免手工设计规则，支持各种复杂结构，并取得较高的准确率。并且可以将文本结果中分散的文本行信息转化为文本段落，提高文档录入和纸质文本电子化的效率，更好地推动各行业信息化和数字化的发展。

[0062] 在本实施例中，通过对目标图片进行识别，基于所述识别的识别结果确定所述目标图片中所有文本框和各所述文本框的文本框位置；根据各所述文本框位置对各所述文本框进行排序，并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练；基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落。通过根据对目标图片的识别结果确定各个文本框和文本框位置，并根据各个文本框位置对各个文

本框进行排序，将各个文本框的文本特征输入至预设的深度学习模型进行训练，基于训练结果对各个文本框进行合并处理，以获取文本段落，从而避免了用户手动操作，导致得到的文本段落的准确性较低的现象发生，提高了文本段落结构还原的准确性。

[0063] 进一步地，在本申请第一实施例的基础上，提出了本申请文本段落结构还原方法的第二实施例，本实施例是本申请第一实施例的步骤S10，基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落的步骤，包括：

[0064] 步骤a，基于所述训练结果确定各所述文本框对应的文本标签，遍历各所述文本标签，检测遍历的文本标签对应的遍历标签信息是否为段落；

[0065] 在本实施例中，当获取到深度学习模型的训练结果后，可以根据此训练结果来确定各个文本框对应的文本标签，如B标签，I标签和O标签等。然后再遍历各个文本框对应的文本标签，并需要检测遍历的文本标签对应的标签信息，即遍历标签信息。再确定该遍历标签信息是否为段落，以便根据确定结果来判断遍历的文本标签对应的文本框是否为一个文本段落，如O标签对应的标签信息。

[0066] 步骤b，若所述遍历标签信息是段落，则确定所述遍历的文本标签对应的文本框为所述目标图片对应的文本段落。

[0067] 当经过判断发现遍历标签信息是段落，则可以确定遍历标签信息对应的文本框文目标图片对应的文本段落。并且在本实施例中，可以对所有的标签信息均采用相同的方式进行检测，直至确定所有的文本段落。

[0068] 在本实施例中，通过根据训练结果确定各个文本框对应的文本标签，并遍历各个文本标签，在遍历的文本标签的遍历标签信息是段落时，将遍历的文本标签对应的文本框作为目标图片对应的文本段落，从而保障了获取的文本段落的准确性。

[0069] 进一步地，检测遍历的文本标签对应的遍历标签信息是否为段落的步骤之后，包括：

[0070] 步骤c，若否，则检测所述遍历标签信息是否为段落内容；

[0071] 当经过判断发现遍历标签信息不是段落，则还需要检查遍历标签信息是否为段

落内容，如I标签对应的标签信息，并根据不同的检测结果执行不同的操作。

[0072] 步骤d，若所述遍历标签信息是段落内容，则确定遍历的文本标签的前一位文本标签的标签信息是否为段落起始信息；

[0073] 当经过判断发现遍历标签信息是段落内容，则还需要确定遍历的文本标签的前一位标签的标签信息是否为段落起始信息，如B标签对应的标签信息，以便确定遍历标签信息对应的文本框所在段落的起始位置，并根据不同的检测结果执行不同的操作。

[0074] 步骤e，若所述前一位文本标签的标签信息是段落起始信息，则基于所述遍历的文本标签和前一位文本标签确定所述目标图片对应的文本段落。

[0075] 当经过判断发现前一位文本标签的标签信息不是段落起始信息，而是段落内容，则确定所有连续相邻标签信息为段落内容的连续相邻内容文本标签，并从中确定具有遍历的文本标签的目标连续相邻内容文本标签，将目标连续相邻内容文本标签对应的所有文本框和目标连续相邻内容文本标签前一位文本标签对应的文本框进行合并处理，作为目标图片中的文本段落。

[0076] 若前一位文本标签的标签信息是段落起始信息，则需要确定遍历的文本标签的下一位文本标签的标签信息是否为段落内容，若不是段落内容，则可以直接将遍历的文本标签对应的文本框和前一位文本标签对应的文本框合并在一起作为目标图片中的一个文本段落。若下一位文本标签的标签信息是段落内容，则需要继续对下一位文本标签的下一位文本标签进行相同的检测操作，直至标签信息不是段落内容，此时会将标签信息为段落内容的所有相邻的包含有遍历的文本标签的标签对应的文本框进行合并，并将其作为目标图片中的一个文本段落。

[0077] 在本实施例中，通过在确定遍历标签信息是段落内容，且遍历的文本标签的前一位文本标签的标签信息是段落起始信息时，根据遍历的文本标签和前一位文本标签确定文本段落，从而保障了获取的文本段落的准确性。

[0078] 具体地，基于所述遍历的文本标签和前一位文本标签确定所述目标图片对应的文本段落的步骤，包括：

[0079] 步骤f，检测各所述文本标签中是否存在连续相邻内容标签；

- [0080] 在确定文本段落时，还需要检测在各个文本框中是否存在连续相邻内容标签，即确定是否存在连续相邻的内容标签，如连续相邻的I标签。并根据不同检测结果执行不同的操作。其中，若文本标签的标签信息是段落内容，并且和该文本标签相邻的文本标签的标签信息也是段落内容，则将该文本标签和与之相邻的文本标签都作为连续相邻内容标签。
- [0081] 步骤g，若存在连续相邻内容标签，则确定所述连续相邻内容标签中是否存在遍历的文本标签；
- [0082] 当经过判断发现存在连续相邻内容标签，且不同的连续相邻内容标签存在多个，则需要继续确定在各个连续相邻内容标签中是否存在遍历的文本标签，并根据不同的确定结果执行不同的操作。若不存在连续相邻内容标签，则直接将遍历的文本标签对应的文本框和前一位文本标签对应的文本框进行合并，并将合并结果作为目标图片的文本段落。
- [0083] 步骤h，若不存在遍历的文本标签，则将所述遍历的文本标签对应的文本框和所述前一位文本标签对应的文本框进行合并，以获取合并后的文本框，并将所述合并后的文本框作为所述目标图片对应的文本段落。
- [0084] 若不存在遍历的文本标签，则可以直接将遍历的文本标签对应的文本框和前一位文本标签对应的文本框进行合并处理，并将合并后的文本框作为目标图片对应的文本段落。
- [0085] 在本实例中，通过确定各个文本标签中存在连续相邻内容标签，且连续相邻内容标签中不存在遍历的文本标签时，对遍历的文本标签对应的文本框和前一位文本标签对应的文本框进行合并，将合并后的文本框作为文本段落，从而保障了获取的文本段落的准确性。
- [0086] 进一步地，确定所述连续相邻内容标签中是否存在遍历的文本标签的步骤之后，包括：
- [0087] 步骤k，若存在遍历的文本标签，则将具有遍历的文本标签的连续相邻内容标签对应的所有文本框和所述前一位文本标签对应的文本框进行合并，以获取合并文本框，并将合并文本框作为所述图片对应的文本段落。
- [0088] 当经过判断发现遍历的文本标签，则需要确定具有遍历的文本标签的所有相邻

文本标签，并将具有遍历的文本标签的连续相邻内容标签对应的所有文本框进行合并处理。例如若遍历的文本标签是I4，连续相邻的文本标签是I1-I5，则连续相邻的文本标签中包含有遍历的文本标签I4，此时就可以将连续相邻的所有文本标签对应的文本框进行合并，即将I1，I2，I3，I4和I5对应的文本框进行合并，得到合并后的文本框，再将其与前一位文本标签对应的文本框进行合并，即和B标签对应的文本框进行合并处理，得到合并文本框，此时就可以直接将合并文本框作为图片对应的文本段落。

[0089] 在本实施例中，通过在确定连续相邻内容标签中存在遍历的文本标签时，将具有遍历的文本标签的连续相邻内容标签对应的所有文本框和前一位文本标签对应的文本框进行合并，并将合并文本框作为文本段落，从而保障了获取的文本段落的准确性。

[0090] 进一步地，基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练的步骤，包括：

[0091] 步骤m，依次提取各所述文本框的文本特征，根据所述排序的排序结果将各所述文本特征融合为序列特征，并将所述序列特征输入至预设的深度学习模型进行训练。

[0092] 在本实施例中，当获取到各个文本框，并且各个文本框进行排序后，可以依次提取各个文本框中的文本特征，如位置特征，语言特征和图像特征，并在提取到各个文本特征后，可以根据各个文本框的排序结果将各个文本特征如何为序列特征，并将序列特征作为预设的深度学习模型中的输入，输入值预设的深度学习模型中进行训练，以获取输出结果，即训练结果。

[0093] 在本实施例中，通过依次提取各个文本框的文本特征，并对各个文本特征进行融合，将融合后的序列特征输入至预设的深度学习模型进行训练，从而保障了训练的有效进行。

[0094] 进一步地，依次提取各所述文本框的文本特征的步骤，包括：

[0095] 步骤n，依次遍历各所述文本框，并提取遍历的文本框的位置特征，语言特征和图像特征，将所述位置特征，语言特征和图像特征作为所述遍历的文本框的文本特征。

- [0096] 在本实施例中，提取所有文本框的文本特征时，可以依次遍历各个文本框，并提取遍历的文本框的位置特征，语言特征和图像特征，再将位置特征，语言特征和图像特征作为遍历的文本框的文本特征。也就是对所有的文本框均采用相同的提取操作。其中，位置特征可以是文本框的顶点坐标，中心点，文本框的宽和高等特征。语言特征可以是文本框中文本的语言模型特征，例如：文本的词向量，句向量，文本ngram得分等。图像特征可以是使用卷积神经网络对图像中的文本区域进行特征提取后的特征等。
- [0097] 在本实施例中，通过提取遍历的文本框的位置特征，语言特征和图像特征，并将其作为遍历的文本框的文本特征，从而保障了获取到的文本特征的有效性。
- [0098] 此外，参照图3，本申请实施例还提出一种文本段落结构还原装置，所述文本段落结构还原装置包括：
- [0099] 确定模块A10，用于对目标图片进行识别，基于所述识别的识别结果确定所述目标图片中所有文本框和各所述文本框的文本框位置；
- [0100] 输入模块A20，用于根据各所述文本框位置对各所述文本框进行排序，并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练；
- [0101] 获取模块A30，用于基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落。
- [0102] 进一步地，所述获取模块A30，还用于：
- [0103] 基于所述训练结果确定各所述文本框对应的文本标签，遍历各所述文本标签，检测遍历的文本标签对应的遍历标签信息是否为段落；
- [0104] 若所述遍历标签信息是段落，则确定所述遍历的文本标签对应的文本框为所述目标图片对应的文本段落。
- [0105] 进一步地，所述获取模块A30，还用于：
- [0106] 若否，则检测所述遍历标签信息是否为段落内容；
- [0107] 若所述遍历标签信息是段落内容，则确定遍历的文本标签的前一位文本标签的标签信息是否为段落起始信息；
- [0108] 若所述前一位文本标签的标签信息是段落起始信息，则基于所述遍历的文本标

签和前一位文本标签确定所述目标图片对应的文本段落。

[0109] 进一步地，所述获取模块A30，还用于：

[0110] 检测各所述文本标签中是否存在连续相邻内容标签；

[0111] 若存在连续相邻内容标签，则确定所述连续相邻内容标签中是否存在遍历的文本标签；

[0112] 若不存在遍历的文本标签，则将所述遍历的文本标签对应的文本框和所述前一位文本标签对应的文本框进行合并，以获取合并后的文本框，并将所述合并后的文本框作为所述目标图片对应的文本段落。

[0113] 进一步地，所述获取模块A30，还用于：

[0114] 若存在遍历的文本标签，则将具有遍历的文本标签的连续相邻内容标签对应的所有文本框和所述前一位文本标签对应的文本框进行合并，以获取合并文本框，并将合并文本框作为所述图片对应的文本段落。

[0115] 进一步地，所述输入模块A20，还用于：

[0116] 依次提取各所述文本框的文本特征，根据所述排序的排序结果将各所述文本特征融合为序列特征，并将所述序列特征输入至预设的深度学习模型进行训练。

[0117] 进一步地，所述输入模块A20，还用于：

[0118] 依次遍历各所述文本框，并提取遍历的文本框的位置特征，语言特征和图像特征，将所述位置特征，语言特征和图像特征作为所述遍历的文本框的文本特征。

[0119] 其中，文本段落结构还原装置的各个功能模块实现的步骤可参照本申请文本段落结构还原方法的各个实施例，此处不再赘述。

[0120] 本申请还提供一种文本段落结构还原设备，所述文本段落结构还原设备包括：存储器、处理器及存储在所述存储器上的文本段落结构还原程序；所述处理器用于执行所述文本段落结构还原程序，以实现以下步骤：

[0121] 对目标图片进行识别，基于所述识别的识别结果确定所述目标图片中所有文本框和各所述文本框的文本框位置；

[0122] 根据各所述文本框位置对各所述文本框进行排序，并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练；

- [0123] 基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落。
- [0124] 本申请还提供了一种计算机可读存储介质，所述计算机可读存储介质可以是非易失性，也可以是易失性。所述计算机可读存储介质存储有一个或者一个以上程序，所述一个或者一个以上程序还可被一个或者一个以上的处理器执行以用于实现上述文本段落结构还原方法各实施例的步骤。
- [0125] 本申请计算机可读存储介质具体实施方式与上述文本段落结构还原方法各实施例基本相同，在此不再赘述。
- [0126] 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、物品或者系统不仅包括那些要素，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、物品或者系统所固有的要素。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、物品或者系统中还存在另外的相同要素。
- [0127] 上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。
- [0128] 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例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，本申请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如上所述的一个存储介质(如ROM/RAM、磁碟、光盘)中，包括若干指令用以使得一台终端设备(可以是手机，计算机，服务器，空调器，或者网络设备等)执行本申请各个实施例所述的方法。
- [0129] 以上仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

## 权利要求书

- [权利要求 1] 一种文本段落结构还原方法，其中，所述文本段落结构还原方法包括以下步骤：
- 对目标图片进行识别，基于所述识别的识别结果确定所述目标图片中所有文本框和各所述文本框的文本框位置；
- 根据各所述文本框位置对各所述文本框进行排序，并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练；
- 基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落。
- [权利要求 2] 如权利要求1所述的文本段落结构还原方法，其中，所述基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落的步骤，包括：
- 基于所述训练结果确定各所述文本框对应的文本标签，遍历各所述文本标签，检测遍历的文本标签对应的遍历标签信息是否为段落；
- 若所述遍历标签信息是段落，则确定所述遍历的文本标签对应的文本框为所述目标图片对应的文本段落。
- [权利要求 3] 如权利要求2所述的文本段落结构还原方法，其中，所述检测遍历的文本标签对应的遍历标签信息是否为段落的步骤之后，包括：
- 若否，则检测所述遍历标签信息是否为段落内容；
- 若所述遍历标签信息是段落内容，则确定遍历的文本标签的前一位文本标签的标签信息是否为段落起始信息；
- 若所述前一位文本标签的标签信息是段落起始信息，则基于所述遍历的文本标签和前一位文本标签确定所述目标图片对应的文本段落。
- [权利要求 4] 如权利要求3所述的文本段落结构还原方法，其中，所述基于所述遍历的文本标签和前一位文本标签确定所述目标图片对应的文本段落的步骤，包括：
- 检测各所述文本标签中是否存在连续相邻内容标签；

若存在连续相邻内容标签，则确定所述连续相邻内容标签中是否存在遍历的文本标签；

若不存在遍历的文本标签，则将所述遍历的文本标签对应的文本框和所述前一位文本标签对应的文本框进行合并，以获取合并后的文本框，并将所述合并后的文本框作为所述目标图片对应的文本段落。

[权利要求 5] 如权利要求4所述的文本段落结构还原方法，其中，所述确定所述连续相邻内容标签中是否存在遍历的文本标签的步骤之后，包括：

若存在遍历的文本标签，则将具有遍历的文本标签的连续相邻内容标签对应的所有文本框和所述前一位文本标签对应的文本框进行合并，以获取合并文本框，并将合并文本框作为所述图片对应的文本段落。

[权利要求 6] 如权利要求1-5任一项所述的文本段落结构还原方法，其中，所述基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练的步骤，包括：

依次提取各所述文本框的文本特征，根据所述排序的排序结果将各所述文本特征融合为序列特征，并将所述序列特征输入至预设的深度学习模型进行训练。

[权利要求 7] 如权利要求6所述的文本段落结构还原方法，其中，所述依次提取各所述文本框的文本特征的步骤，包括：

依次遍历各所述文本框，并提取遍历的文本框的位置特征，语言特征和图像特征，将所述位置特征，语言特征和图像特征作为所述遍历的文本框的文本特征。

[权利要求 8] 一种文本段落结构还原装置，其中，所述文本段落结构还原装置包括：

确定模块，用于对目标图片进行识别，基于所述识别的识别结果确定所述目标图片中所有文本框和各所述文本框的文本框位置；

输入模块，用于根据各所述文本框位置对各所述文本框进行排序，并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练；

获取模块，用于基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落。

[权利要求 9] 一种文本段落结构还原设备，其中，所述文本段落结构还原设备包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的文本段落结构还原程序，所述文本段落结构还原程序被所述处理器执行时实现如下步骤：

对目标图片进行识别，基于所述识别的识别结果确定所述目标图片中所有文本框和各所述文本框的文本框位置；

根据各所述文本框位置对各所述文本框进行排序，并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练；

基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落。

[权利要求 10] 如权利要求9所述的文本段落结构还原设备，其中，所述基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落的步骤，包括：

基于所述训练结果确定各所述文本框对应的文本标签，遍历各所述文本标签，检测遍历的文本标签对应的遍历标签信息是否为段落；

若所述遍历标签信息是段落，则确定所述遍历的文本标签对应的文本框为所述目标图片对应的文本段落。

[权利要求 11] 如权利要求10所述的文本段落结构还原设备，其中，所述检测遍历的文本标签对应的遍历标签信息是否为段落的步骤之后，包括：

若否，则检测所述遍历标签信息是否为段落内容；

若所述遍历标签信息是段落内容，则确定遍历的文本标签的前一位文本标签的标签信息是否为段落起始信息；

若所述前一位文本标签的标签信息是段落起始信息，则基于所述遍历的文本标签和前一位文本标签确定所述目标图片对应的文本段落。

[权利要求 12] 如权利要求11所述的文本段落结构还原设备，其中，所述基于所述遍

历的文本标签和前一位文本标签确定所述目标图片对应的文本段落的步骤，包括：

检测各所述文本标签中是否存在连续相邻内容标签；

若存在连续相邻内容标签，则确定所述连续相邻内容标签中是否存在遍历的文本标签；

若不存在遍历的文本标签，则将所述遍历的文本标签对应的文本框和所述前一位文本标签对应的文本框进行合并，以获取合并后的文本框，并将所述合并后的文本框作为所述目标图片对应的文本段落。

[权利要求 13] 如权利要求12所述的文本段落结构还原设备，其中，所述确定所述连续相邻内容标签中是否存在遍历的文本标签的步骤之后，包括：

若存在遍历的文本标签，则将具有遍历的文本标签的连续相邻内容标签对应的所有文本框和所述前一位文本标签对应的文本框进行合并，以获取合并文本框，并将合并文本框作为所述图片对应的文本段落。

[权利要求 14] 如权利要求9-13任一项所述的文本段落结构还原设备，其中，所述基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练的步骤，包括：

依次提取各所述文本框的文本特征，根据所述排序的排序结果将各所述文本特征融合为序列特征，并将所述序列特征输入至预设的深度学习模型进行训练。

[权利要求 15] 一种计算机存储介质，其中，所述计算机存储介质上存储有文本段落结构还原程序，所述文本段落结构还原程序被处理器执行时实现如下步骤：

对目标图片进行识别，基于所述识别的识别结果确定所述目标图片中所有文本框和各所述文本框的文本框位置；

根据各所述文本框位置对各所述文本框进行排序，并基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练；

基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图

片对应的所有文本段落。

[权利要求 16] 如权利要求15所述的计算机存储介质，其中，所述基于所述训练结果对各所述文本框进行合并处理，以获取所述目标图片对应的所有文本段落的步骤，包括：

基于所述训练结果确定各所述文本框对应的文本标签，遍历各所述文本标签，检测遍历的文本标签对应的遍历标签信息是否为段落；

若所述遍历标签信息是段落，则确定所述遍历的文本标签对应的文本框为所述目标图片对应的文本段落。

[权利要求 17] 如权利要求16所述的计算机存储介质，其中，所述检测遍历的文本标签对应的遍历标签信息是否为段落的步骤之后，包括：

若否，则检测所述遍历标签信息是否为段落内容；

若所述遍历标签信息是段落内容，则确定遍历的文本标签的前一位文本标签的标签信息是否为段落起始信息；

若所述前一位文本标签的标签信息是段落起始信息，则基于所述遍历的文本标签和前一位文本标签确定所述目标图片对应的文本段落。

[权利要求 18] 如权利要求17所述的计算机存储介质，其中，所述基于所述遍历的文本标签和前一位文本标签确定所述目标图片对应的文本段落的步骤，包括：

检测各所述文本标签中是否存在连续相邻内容标签；

若存在连续相邻内容标签，则确定所述连续相邻内容标签中是否存在遍历的文本标签；

若不存在遍历的文本标签，则将所述遍历的文本标签对应的文本框和所述前一位文本标签对应的文本框进行合并，以获取合并后的文本框，并将所述合并后的文本框作为所述目标图片对应的文本段落。

[权利要求 19] 如权利要求18所述的计算机存储介质，其中，所述确定所述连续相邻内容标签中是否存在遍历的文本标签的步骤之后，包括：

若存在遍历的文本标签，则将具有遍历的文本标签的连续相邻内容标签对应的所有文本框和所述前一位文本标签对应的文本框进行合并，

以获取合并文本框，并将合并文本框作为所述图片对应的文本段落。

[权利要求 20] 如权利要求15-19任一项所述的计算机存储介质，其中，所述基于所述排序的排序结果将各所述文本框的文本特征输入至预设的深度学习模型进行训练的步骤，包括：

依次提取各所述文本框的文本特征，根据所述排序的排序结果将各所述文本特征融合为序列特征，并将所述序列特征输入至预设的深度学习模型进行训练。

附图

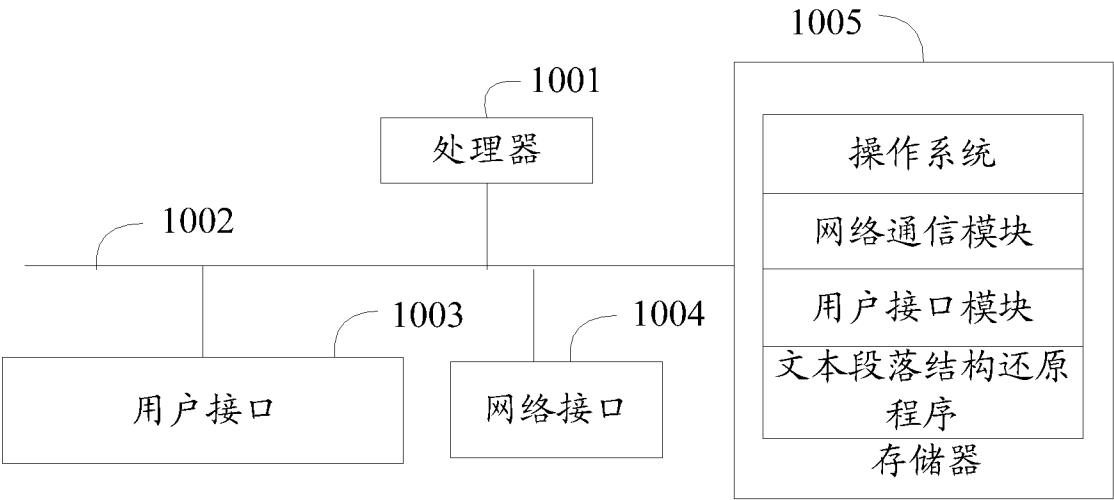


图 1

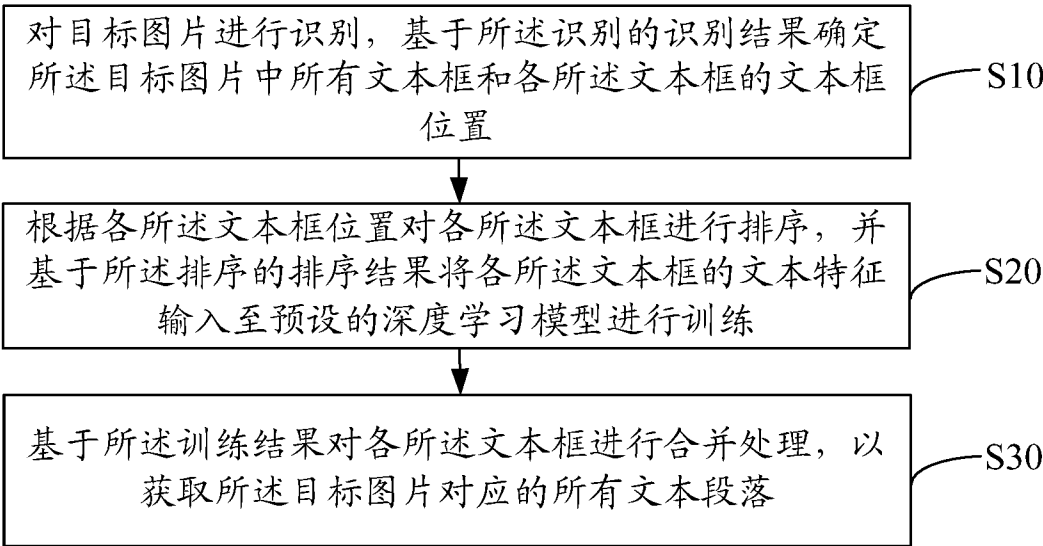


图 2

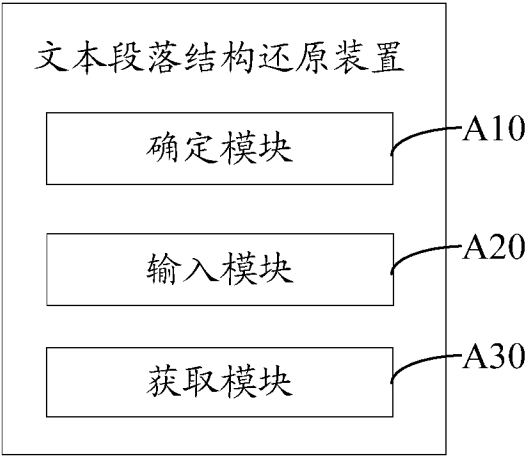


图 3

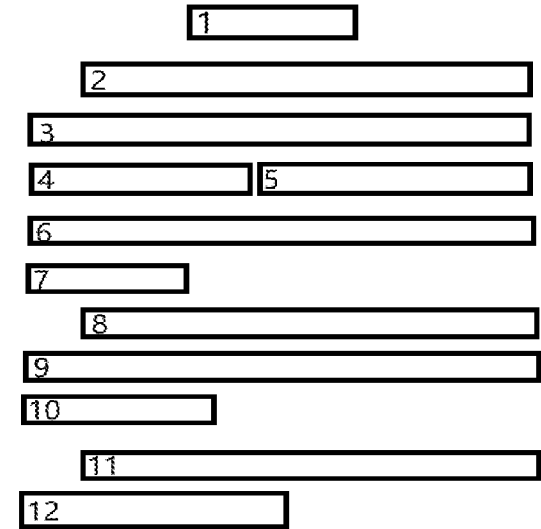


图 4

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/124605

**A. CLASSIFICATION OF SUBJECT MATTER**

G06K 9/00(2022.01)i; G06F 40/166(2020.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G06K; G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DWPI; SIPOABS; CNABS; CNKI; CNTXT; IEEE: 图像, 文本, 文档, 框, 行, 合并, 段落, 分段, 标签, 神经网络, 深度学习, 模型, CNN, IMAGE, TEXT, DOCUMENT, BOX, LINE, PARAGRAPH, NEURAL NETWORK, LABEL, COMBINE

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                 | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| PX        | CN 112070076 A (SHENZHEN ONECONNECT TECHNOLOGY CO., LTD.) 11 December 2020 (2020-12-11)<br>claims 1-10                                                                             | 1-20                  |
| X         | CN 111507112 A (SOGOU (HANGZHOU) INTELLIGENT TECHNOLOGY CO., LTD. et al.) 07 August 2020 (2020-08-07)<br>description paragraphs [0033]-[0102], [0104], [0127], and figures 4, 6, 7 | 1, 6-9, 14, 15, 20    |
| Y         | CN 111507112 A (SOGOU (HANGZHOU) INTELLIGENT TECHNOLOGY CO., LTD. et al.) 07 August 2020 (2020-08-07)<br>description paragraphs [0033]-[0102], [0104], [0127], and figures 4, 6, 7 | 2-5, 10-13, 16-19     |
| Y         | CN 111639175 A (UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA) 08 September 2020 (2020-09-08)<br>description paragraphs [0062]-[0066]                                   | 2-5, 10-13, 16-19     |
| A         | US 10565443 B2 (WIPRO LTD.) 18 February 2020 (2020-02-18)<br>entire document                                                                                                       | 1-20                  |
| A         | US 2014250361 A1 (MACROMEDIA INC) 04 September 2014 (2014-09-04)<br>entire document                                                                                                | 1-20                  |



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

12 January 2022

Date of mailing of the international search report

20 January 2022

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/  
CN)  
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing  
100088, China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2021/124605**

| Patent document<br>cited in search report |            |    | Publication date<br>(day/month/year) | Patent family member(s) |              |    | Publication date<br>(day/month/year) |
|-------------------------------------------|------------|----|--------------------------------------|-------------------------|--------------|----|--------------------------------------|
| CN                                        | 112070076  | A  | 11 December 2020                     | CN                      | 112070076    | B  | 06 April 2021                        |
| CN                                        | 111507112  | A  | 07 August 2020                       | None                    |              |    |                                      |
| CN                                        | 111639175  | A  | 08 September 2020                    | None                    |              |    |                                      |
| US                                        | 10565443   | B2 | 18 February 2020                     | US                      | 2019258854   | A1 | 22 August 2019                       |
|                                           |            |    |                                      | IN                      | 201841006073 | A  | 23 August 2019                       |
| US                                        | 2014250361 | A1 | 04 September 2014                    | None                    |              |    |                                      |

## 国际检索报告

国际申请号

PCT/CN2021/124605

## A. 主题的分类

G06K 9/00(2022.01)i; G06F 40/166(2020.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

## B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06K; G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

DWPI; SIPOABS; CNABS; CNKI; CNTXT; IEEE: 图像, 文本, 文档, 框, 行, 合并, 段落, 分段, 标签, 神经网络, 深度学习, 模型, CNN, IMAGE, TEXT, DOCUMENT, BOX, LINE, PARAGRAPH, NEURAL NETWORK, LABEL, COMBINE

## C. 相关文件

| 类 型* | 引用文件, 必要时, 指明相关段落                                                                                       | 相关的权利要求         |
|------|---------------------------------------------------------------------------------------------------------|-----------------|
| PX   | CN 112070076 A (深圳壹账通智能科技有限公司) 2020年12月11日 (2020 - 12 - 11)<br>权利要求1-10                                 | 1-20            |
| X    | CN 111507112 A (搜狗杭州智能科技有限公司等) 2020年8月7日 (2020 - 08 - 07)<br>说明书第[0033]-[0102]、[0104]、[0127]段, 以及图4、6、7 | 1、6-9、14、15、20  |
| Y    | CN 111507112 A (搜狗杭州智能科技有限公司等) 2020年8月7日 (2020 - 08 - 07)<br>说明书第[0033]-[0102]、[0104]、[0127]段, 以及图4、6、7 | 2-5、10-13、16-19 |
| Y    | CN 111639175 A (电子科技大学) 2020年9月8日 (2020 - 09 - 08)<br>说明书第[0062]- [0066]段                               | 2-5、10-13、16-19 |
| A    | US 10565443 B2 (WIPRO LTD) 2020年2月18日 (2020 - 02 - 18)<br>全文                                            | 1-20            |
| A    | US 2014250361 A1 (MACROMEDIA INC) 2014年9月4日 (2014 - 09 - 04)<br>全文                                      | 1-20            |

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

\* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&amp;” 同族专利的文件

国际检索实际完成的日期

2022年1月12日

国际检索报告邮寄日期

2022年1月20日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

胡燕

电话号码 62089101

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2021/124605

| 检索报告引用的专利文件 |            |    | 公布日<br>(年/月/日) | 同族专利              | 公布日<br>(年/月/日) |
|-------------|------------|----|----------------|-------------------|----------------|
| CN          | 112070076  | A  | 2020年12月11日    | CN 112070076 B    | 2021年4月6日      |
| CN          | 111507112  | A  | 2020年8月7日      | 无                 |                |
| CN          | 111639175  | A  | 2020年9月8日      | 无                 |                |
| US          | 10565443   | B2 | 2020年2月18日     | US 2019258854 A1  | 2019年8月22日     |
|             |            |    |                | IN 201841006073 A | 2019年8月23日     |
| US          | 2014250361 | A1 | 2014年9月4日      | 无                 |                |