

(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
2023 年 2 月 2 日 (02.02.2023)



(10) 国际公布号
WO 2023/005253 A1

(51) 国际专利分类号:
G06K 9/62 (2022.01) *G06K 9/00* (2022.01)

(21) 国际申请号: PCT/CN2022/085149

(22) 国际申请日: 2022 年 4 月 2 日 (02.04.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110858410.X 2021年7月28日 (28.07.2021) CN

(71) 申请人: 北京百度网讯科技有限公司 (BEIJING BAIDU NETCOM SCIENCE TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。

(72) 发明人: 章成全 (ZHANG, Chengquan); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 吕鹏原 (LV, Pengyuan); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 李煜林 (LI, Yulin); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 庾悦晨 (YU, Yuechen); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 姚锟 (YAO, Kun); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 韩钧宇 (HAN, Junyu); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 刘经拓 (LIU, Jingtuo); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 丁二锐 (DING, Errui); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 吴

(54) Title: METHOD, APPARATUS AND SYSTEM FOR TRAINING TEXT RECOGNITION MODEL FRAMEWORK

(54) 发明名称: 文本识别模型框架的训练方法、装置及系统

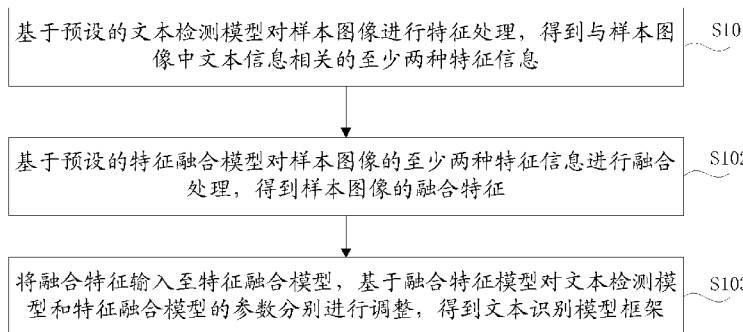


图 1

- S101 Perform feature processing on a sample image on the basis of a preset text detection model, so as to obtain at least two types of feature information related to text information in the sample image
- S102 Perform fusion processing on the at least two types of feature information of the sample image on the basis of a preset feature fusion model, so as to obtain a fused feature of the sample image
- S103 Input the fused feature into the feature fusion model, and respectively adjust parameters of the text detection model and the feature fusion model on the basis of the fused feature, so as to obtain a text recognition model framework

(57) Abstract: The present disclosure relates to the technical field of artificial intelligence, and in particular to the technical fields of computer vision and deep learning, and can be applied in smart city and smart finance scenarios. Provided are a method, apparatus and system for training a text recognition model framework. The method comprises: performing feature processing on a sample image on the basis of a preset text detection model, so as to obtain at least two types of feature information related to text information in the sample image; performing fusion processing on the at least two types of feature information of the sample image on the basis of a

甜(WU, Tian); 中国北京市海淀区上地十街10号百度大厦2层, Beijing 100085 (CN)。王海峰(WANG, Haifeng); 中国北京市海淀区上地十街10号百度大厦2层, Beijing 100085 (CN)。

(74) 代理人: 北京同立钧成知识产权代理有限公司(LEADER PATENT & TRADEMARK FIRM); 中国北京市海淀区西直门北大街32号枫蓝国际A座8F-6, Beijing 100082 (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, JM, 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, 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))。

preset feature fusion model, so as to obtain a fused feature of the sample image; and inputting the fused feature into the feature fusion model, and respectively adjusting parameters of the text detection model and the feature fusion model on the basis of the fused feature, so as to obtain a text recognition model framework. The text detection model and the feature fusion model in the text recognition model framework have a relatively high relevance, so as to realize the integrity and comprehensiveness of a training process, thereby improving the accuracy and reliability of the text recognition model framework.

(57) 摘要: 本公开提供了一种文本识别模型框架的训练方法、装置及系统, 涉及人工智能技术领域, 具体为计算机视觉和深度学习技术领域, 可应用于智慧城市和智慧金融场景。包括: 基于预设的文本检测模型对样本图像进行特征处理, 得到与样本图像中文本信息相关的至少两种特征信息, 基于预设的特征融合模型对样本图像的至少两种特征信息进行融合处理, 得到样本图像的融合特征, 将融合特征输入至特征融合模型, 基于融合特征模型对文本检测模型和特征融合模型的参数分别进行调整, 得到文本识别模型框架, 文本识别模型框架中的文本检测模型与特征融合模型之间具有较高的关联性, 从而实现了训练过程的完整性和全面性, 提高了文本识别模型框架的准确性和可靠性。

文本识别模型框架的训练方法、装置及系统

本公开要求于 2021 年 07 月 28 日提交中国专利局、申请号为 CN202110858410.X、
5 申请名称为“文本识别模型框架的训练方法、装置及系统”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

本公开涉及人工智能技术领域，具体为计算机视觉和深度学习技术领域，尤其涉
10 及一种文本识别模型框架的训练方法、装置及系统，可应用于智慧城市和智慧金融场
景。

背景技术

随着人工智能技术的发展，对图像中的文本信息的识别由人工识别发展为了自动
15 识别，如预先训练用于辅助训练文本识别模型的文本识别模型框架（也可以称为用于
辅助训练文本识别模型的结构化解析框架模型），在该结构化框架模型的基础上，训
练生成用于对待识别的图像中的文本信息进行识别的文本识别模型。

在现有技术中，通常基于文本检测模型和特征融合模型训练得到文本识别模型框
架，其中，文本检测模型与特征融合模型为两个相互独立的模型，特征融合模型具体
20 基于文本检测模型的线下识别结果完成训练。

然而，文本检测模型与特征融合模型二者在训练过程中相互独立，可能导致训练
得到的文本识别模型框架的准确性偏低的技术问题。

发明内容

25 本公开提供了一种用于提高文本识别模型框架的准确性的文本识别模型框架的训
练方法及装置。

根据本公开的第一方面，提供了一种文本识别模型框架的训练方法，所述方法包
括：基于预设的文本检测模型对样本图像进行特征处理，得到与所述样本图像中文本
信息相关的至少两种特征信息；

30 基于预设的特征融合模型对所述样本图像的至少两种特征信息进行融合处理，得
到所述样本图像的融合特征；

将所述融合特征输入至所述特征融合模型，基于所述融合特征模型对所述文本检测模型和所述特征融合模型的参数分别进行调整，得到文本识别模型框架，其中，所述文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型。

根据本公开的第二方面，提供了一种文本识别方法，包括：

5 获取待识别图像；

将所述待识别图像输入至预先训练的文本识别模型，得到所述待识别图像中的文本信息，其中，所述文本识别模型是基于预先训练的文本识别模型框架对待训练图像进行训练生成的，所述文本识别模型框架为由第一方面所述训练方法训练获得，所述待训练图像中包括文本信息。

10 根据本公开的第三方面，提供了一种文本识别模型框架的训练装置，所述装置包括：

处理单元，用于基于预设的文本检测模型对样本图像进行特征处理，得到与所述样本图像中文本信息相关的至少两种特征信息；

15 融合单元，用于基于预设的特征融合模型对所述样本图像的至少两种特征信息进行融合处理，得到所述样本图像的融合特征；

训练单元，用于将所述融合特征输入至所述特征融合模型，基于所述融合特征模型对所述文本检测模型和所述特征融合模型的参数分别进行调整，得到文本识别模型框架，其中，所述文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型。

20 根据本公开的第四方面，提供了一种文本识别装置，包括：

获取单元，用于获取待识别图像；

25 识别单元，用于将所述待识别图像输入至预先训练的文本识别模型，得到所述待识别图像中的文本信息，其中，所述文本识别模型是基于预先训练的文本识别模型框架对待训练图像进行训练生成的，所述文本识别模型框架为由第一方面训练方法训练获得，所述待训练图像中包括文本信息。

根据本公开的第五方面，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，

30 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第一方面所述的方法；或者，以使所述至少一个处理器能够执行第二方面所述的方法。

根据本公开的第六方面，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第一方面所述的方法；或者，所述计算机指令用于使所述计算机执行第二方面所述的方法。

根据本公开的第七方面，提供了一种计算机程序产品，所述计算机程序产品包括：

- 5 计算机程序，所述计算机程序存储在可读存储介质中，电子设备的至少一个处理器可以从所述可读存储介质读取所述计算机程序，所述至少一个处理器执行所述计算机程序使得电子设备执行第一方面所述的方法；或者，所述至少一个处理器执行所述计算机程序使得电子设备执行第二方面所述的方法。

- 10 根据本公开的第八方面，提供了一种文本识别模型框架的训练系统，所述系统包括：

文本检测模型，用于对样本图像进行特征处理，得到与所述样本图像中文本信息相关的至少两种特征信息；

特征融合模型，用于对所述样本图像的至少两种特征信息进行融合处理，得到所述样本图像的融合特征；

- 15 所述特征融合模型还用于，对所述文本检测模型和所述特征融合模型的参数分别进行调整，得到文本识别模型框架，其中，所述文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型。

根据第九方面，本申请实施例提供了一种计算机程序，包括程序代码，当计算机运行所述计算机程序时，所述程序代码执行如上第一方面或者第二方面所述的方法。

- 20 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。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

附图用于更好地理解本方案，不构成对本公开的限定。其中：

- 25 附图用于更好地理解本方案，不构成对本公开的限定。其中：

图 1 是根据本公开第一实施例的示意图；

图 2 是根据本公开第二实施例的示意图；

图 3 是根据本公开实施例的文本识别模型框架的训练方法的场景示意图；

图 4 是根据本公开第三实施例的示意图；

- 30 图 5 是根据本公开第四实施例的示意图；

图 6 是根据本公开第五实施例的示意图；

图 7 是根据本公开第六实施例的示意图；

图 8 是用来实现本公开实施例的文本识别模型框架的训练方法、文本识别方法的电子设备的框图；

图 9 是根据本公开第七实施例的示意图。

5

具体实施方式

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。

文本识别技术是指对图像中的文本信息的识别，且文本识别技术被广泛地应用于各个领域，如教育领域，金融领域，医疗领域，交通领域，以及保险领域等。

例如，当文本识别技术应用于医疗领域中，可以基于文本识别技术对病历本图像中的文本信息进行识别。又如，当文本识别技术应用于保险领域中，可以基于文本识别技术对保险单图像中的文本信息进行识别等，此处不再一一列举。

而随着人工智能技术中深度学习技术的发展，深度学习技术可以与其他技术相结合，例如，可以将深度学习技术应用于文本识别技术，从而提高对文本信息的识别的准确性和可靠性。

例如，可以基于深度学习技术训练用于对文本信息进行识别的文本识别模型。而对文本识别模型的训练通常需要基于文本识别模型框架，即一般而言，先训练得到文本识别模型框架，而后在文本识别模型框架的基础上，训练得到文本识别模型。

在相关技术中，通常由两个相互独立的模型训练得到文本识别模型框架，两个相互独立的模型分别为文本检测模型和特征融合模型，且在训练文本识别模型框架时，特征融合模型是基于文本检测模型的线下识别结果。

具体地，文本检测模型可以为光学字符识别(Optical Character Recognition, OCR)模型，特征融合模型可以为 transfromer 模型，transfromer 模型具体基于光学字符识别模型的线下识别结果完成训练，得到文本识别模型框架。

然而，光学字符识别模型与 transfromer 模型二者在训练过程中相互独立，可能导致训练得到的文本识别模型框架的准确性偏低的技术问题。

为了避免上述技术问题，本公开的发明人经过创造性地劳动，得到了本公开的发明构思：基于文本检测模型和特征融合模型，得到融合特征，由特征融合模型基于融

合特征，对文本识别模型和特征识别模型进行整体训练，以得到文本识别模型框架。

基于上述发明构思，本公开提供一种文本识别模型框架的训练方法、装置及系统，应用于人工智能技术领域中的计算机视觉和深度学习技术领域，可应用于智慧城市和智慧金融场景，以提高文本识别模型框架的准确性。

5 请参阅图 1，图 1 是根据本公开第一实施例的示意图。

如图 1 所示，本公开实施例提供的文本识别模型框架的训练方法，包括：

S101：基于预设的文本检测模型对样本图像进行特征处理，得到与样本图像中文本信息相关的至少两种特征信息。

10 示例性地，本实施例的执行主体可以为文本识别模型框架的训练装置（下文简称训练装置），训练装置可以为服务器（如本地服务器，或者，云端服务器），也可以为终端设备，也可以为处理器，也可以为芯片等，本实施例不做限定。

其中，样本图像中包括文本信息，例如，针对医疗领域，样本图像可以为病历单的图像，样本图像中包括如病人身份的文本信息、以及病例的文本信息等。又如，针对保险领域，样本图像可以为保险单的图像，样本图像中包括如保险人身份的文本信息、以及保险内容的文本信息等。

应该理解地是，用于训练文本识别模型框架的样本图像的数量可以由训练装置基于需求、历史记录、以及试验等方式进行设置，本实施例不做限定。

20 文本检测模型为可以对样本图像中，与文本信息相关的特征进行检测的模型。例如，针对医疗领域，文本检测模型可以对病历单的图像中的病人身份的文本信息进行检测。

具体地，文本检测模型可以为光学字符识别模型。

在本实施例中，特征信息用于表征与样本图像中文本信息相关的特征，至少两种特征信息可以包括：与文本内容相关的信息、与文本视觉相关的信息、以及各文字在空间关系上的信息等，此处不再一一列举。

25 S102：基于预设的特征融合模型对样本图像的至少两种特征信息进行融合处理，得到样本图像的融合特征。

其中，特征融合模型是指，可以对多种特征信息进行融合处理的模型。例如，特征融合模型可以为 transfromer 模型。

30 融合处理可以为对多种特征信息进行拼接，也可以为对多种特征信息进行组合，可以为对多种特征信息进行连接等，本实施例不做限定，融合处理地详细处理过程可以参见相关技术，此处不再赘述。

S103: 将融合特征输入至特征融合模型, 基于融合特征模型对文本检测模型和特征融合模型的参数分别进行调整, 得到文本识别模型框架。

其中, 文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型。

在本实施例中, 将融合特征输入至特征融合模型, 可以基于融合特征调整文本检测模型的参数, 且调整特征融合模型的参数, 从而文本识别模型框架。

可以理解地是, 文本识别模型框架的训练为迭代的过程, 即对文本检测模型的参数和特征融合模型的参数反复进行调整的过程, 当迭代次数达到预设次数阈值, 或者迭代时的损失函数小于预设的损失阈值, 则说明训练已经得到要求, 从而得到文本识别模型框架。

基于上述分析可知, 本公开实施例提供了一种文本识别模型框架的训练方法, 该方法包括: 基于预设的文本检测模型对样本图像进行特征处理, 得到与样本图像中文本信息相关的至少两种特征信息, 基于预设的特征融合模型对样本图像的至少两种特征信息进行融合处理, 得到样本图像的融合特征, 将融合特征输入至特征融合模型, 基于融合特征模型对文本检测模型和特征融合模型的参数分别进行调整, 得到文本识别模型框架, 其中, 文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型, 在本实施例中, 引入了: 基于融合特征对文本检测模型和特征融合模型的参数分别进行调整, 以得到文本识别模型框架的技术特征, 使得文本识别模型框架中的文本检测模型与特征融合模型之间具有较高的关联性, 从而实现了训练过程的完整性和全面性, 避免了相关技术中文本检测模型与特征融合模型相互独立, 使得训练文本识别模型框架时缺乏从整体维度的考虑, 从而导致文本识别模型框架的准确性偏低的弊端, 提高了文本识别模型框架的准确性和可靠性。

请参阅图 2, 图 2 是根据本公开第二实施例的示意图。

如图 2 所示, 本公开实施例提供的文本识别模型框架的训练方法, 包括:

S201: 基于文本检测模型确定样本图像中文本行的位置信息, 并根据位置信息确定至少两种特征信息。

其中, 关于本实施例与第一实施例的相同特征, 本实施例不做赘述。

基于上述分析可知, 本实施例的文本识别模型框架的训练方法可以应用于不同的领域, 如保险领域和医疗领域等, 现结合本实施例的文本识别模型框架的训练方法应用于保险领域为例, 对本实施例进行示范性地描述。

如图 3 所示, 样本图像为保险单图像, 保险单图像中包括文本信息, 如图 3 中所示的“姓名: XXX”、“保险类型: XXXXXX”、以及“保险年限: XXXX”等。

在一些实施例中，可以通过扫描的方式将样本图像传输至训练装置，并由训练装置中的文本检测模型对样本图像中文本行的位置信息进行确定。

在另一些实施例中，如图3所示，也可以由训练装置与外接装置（如存储装置等）连接，并接收由外接装置传输的样本图像，以由训练装置中的文本检测模型对样本图像中5 文本行的位置信息进行确定。

其中，文本行是指，文本信息所在的行。文本行的位置信息是指，文本信息所在的行的位置相关的信息，具体可以为文本信息所在的行、在样本图像中的坐标。

例如，文本检测模型在对样本图像进行识别时，可以基于预设矩形框对样本图像中的文本行进行框选，并确定该矩形框在样本图像中的坐标。

10 在本实施例中，通过确定样本图像中文本行的位置信息，以基于位置信息从样本图像中确定至少两种特征信息，可以实现通过相对较高准确性定位的方式确定至少两种特征信息，进而可以提高至少两种特征信息的准确性和可靠性的技术效果。

在一些实施例中，根据位置信息确定至少两种特征信息，包括：根据位置信息对样本图像进行裁剪操作，得到文本区域，并从文本区域中获取至少两种特征信息。

15 例如，结合上述实施例，在确定出位置信息之后，可以基于位置信息从样本图像中裁剪出矩形框所框选的区域，该区域即为文本区域，并通过对文本区域中的文本信息进行识别的方式，得到至少两种特征信息。

在本实施例中，基于位置信息从样本图像中裁剪得到文本区域，可以使得文本区域中包括几乎全量的文本信息，避免文本信息的遗漏，也使得裁剪操作具有较高的准确性，从而使得文本区域具有较高的准确性和可靠性，进而使得基于文本区域确定的至少两种特征信息具有较高的全面性和可靠性的技术效果。

在一些实施例中，从文本区域中获取至少两种特征信息，包括：从文本区域中提取样本图像的图像特征，并对图像特征进行识别，得到至少两种特征信息。

其中，图像特征可以从两个大的维度理解，两个大的维度分别为内容维度和外观25 维度。如在本实施例中，样本图像为包括文本信息的图像，则内容维度的图像特征是指，图像特征中包括的与文本信息的内容相关的特征，如文本内容；外观维度的图像特征是指，图像特征中包括的与文本信息的颜色和纹理等相关的特征。

因此，在本实施例中，至少两种特征信息可以包括基于两个大的维度（即内容维度和外观维度）分别确定的两种特征信息。当然，结合上述第一实施例中的分析可知，30 也可以将该两个大的维度拆分为更小的维度，并基于更小的维度确定三种及以上的特征信息，本实施例不做限定。

在本实施例中，由于文本区域具有较高的准确性和全面性，因此，从文本区域中提取到的图像特征具有较高的准确性和全面性，且在对图像特征进行识别时，得到特征信息时，可以从多个维度进行分析，从而得到多个维度的特征信息，所以，可以提高特征信息的准确性、全面性、以及可靠性的技术效果。

5 在一些实施例中，至少两种特征信息包括文本特征和视觉特征。

其中，文本特征可以理解为基于内容维度的特征信息，视觉特征可以理解为外观维度的特征信息。

S202：基于预设的特征融合模型对文本特征和视觉特征进行融合处理，得到样本图像的融合特征。

10 其中，关于 S202 的实现原理可以参见第一实施例，此处不再赘述。

S203：构建用于表征文本特征的多个文本特征块，并构建用于表征视觉特征的多个视觉特征块。

例如，构建与文本特征具有映射关系的多个文本特征块，多个文本特征块可以用于表征文本特征。

15 示例性地，可以基于需求、历史记录、以及试验等方式，确定文本特征块的数量，并将文本特征映射至多个文本特征块中，多个文本特征块可以对文本特征进行表征。

具体地，文本特征块可以为 $2*2$ （像素）的特征块，且可以基于文本特征的语义信息，将文本特征拆分并存储至多个 $2*2$ （像素）的特征块中，从而得到多个文本特征块。

20 其中，语义信息可以理解文本信息在字段分类上相关的信息，也可以理解为文本信息在字段间位置相关的信息，也可以理解为文本信息在表征含义上相关的信息。

同理，关于构建用于表征视觉特征的多个视觉特征块的原理和实现，可以参阅关于构建用于表征文本特征的多个文本特征块，此处不再赘述。

25 S204：由特征融合模型根据融合特征和多个文本特征块，对文本检测模型和特征融合模型的参数分别进行调整；和/或，由特征融合模型根据融合特征和多个视觉特征块，对文本检测模型和特征融合模型的参数分别进行调整。

一个示例中，可以结合融合特征和多个文本特征块，对文本检测模型的参数进行调整，且对特征融合模型的参数进行调整。

30 在一些实施例中，结合融合特征和多个文本特征块，对文本检测模型的参数进行调整，且对特征融合模型的参数进行调整，可以包括如下步骤：

第一步骤：由特征融合模型随机遮盖融合特征中的部分文本特征，并根据多个文

本特征块对遮盖的部分文本特征进行预测补齐处理,得到预测补齐后的部分文本特征。

基于上述分析可知,训练文本识别模型框架的过程是迭代的过程,因此,在训练的过程中,当前次迭代随机遮盖融合特征中的部分文本特征与在前迭代随机遮盖融合特征中的部分文本特征不相同。

5 示例性地,每次迭代随机遮盖融合特征的部分文本特征完全不同。

例如,在第一次迭代时,随机遮盖融合特征中的部分文本特征为,文本特征中的前百分之六的文本特征,在第二次迭代时,随机遮盖融合特征中的部分文本特征为,文本特征中的前百分之六至前百分之十二之间的文本特征,以此类推,不再一一列举。

又如,在第一次迭代时,随机遮盖融合特征中的部分文本特征为,文本特征中的百分之六的文本特征,在第二次迭代时,随机遮盖融合特征中的部分文本特征为,文本特征中的除第一次迭代时遮盖的融合特征之外的百分之六的文本特征。

示例性地,每次迭代随机遮盖融合特征的部分文本特征不完全相同。

例如,在第一次迭代时,随机遮盖融合特征中的部分文本特征为,文本特征中的百分之六的文本特征,在第二次迭代时,随机遮盖融合特征中的部分文本特征为,文本特征中的百分之六的文本特征,且第一次迭代中遮盖的百分之六的文本特征、与第二次迭代中遮盖的百分之六的文本特征中存在相同的文本特征。

基于上述分析,多个文本特征块可以用于表征文本特征,因此,当遮挡部分文本特征之后,可以基于多个文本特征块对遮挡的部分文档进行补齐预测,从而得到预测补齐后的部分文本特征。

20 例如,融合特征中的文本特征为 A,文本特征 A 被遮挡的部分文本特征为 a1,其他未被遮挡的部分文档特征为 a2,则训练装置可以基于多个文本块特征、以及部分文档特征 a2 推断出部分文本特征的内容(即预测补齐后的部分文本特征)。

第二步骤:根据预测补齐后的部分文本特征、以及融合特征中除被遮盖的部分文本特征以外的特征,对文本检测模型和特征融合模型的参数分别进行调整。

25 结合上述实施例,该步骤可以理解为:训练装置根据推断出的部分文本特征的内容(即预测补齐后的部分文本特征)、部分文档特征为 a2,对文本检测模型的参数进行调整,且对特征融合模型的参数进行调整。

在本实施例中,通过遮盖融合特征中部分文本特征,并基于多个文本特征块对遮盖的部分文本特征进行预测补齐,以基于得到的预测补齐后的部分文本特征对两个模型(即文本检测模型和特征融合模型)的参数分别进行调整,充分考虑了文本特征中各部分文本特征之间的关联关系(既包括文字内容上的关联关系,也包括位置上的关

联关系)，可以提高两个模型的识别辨别能力，从而提高训练得到的文本识别模型框架的准确性和可靠性的技术效果。

在另一些实施例中，结合融合特征和多个文本特征块，对文本检测模型的参数进行调整，且对特征融合模型的参数进行调整，可以包括如下步骤：

5 第一步骤：由特征融合模型根据多个文本特征块中的至少部分文本特征块，对融合特征中的文本特征进行替换处理，得到替换后的文本特征。

第二步骤：根据融合特征中的视觉特征、以及替换后的文本特征，对文本检测模型和特征融合模型的参数分别进行调整。

10 其中，关于本实施例对融合特征中的文本特征的替换原理可以为全部替换，也可以为部分替换，本实施例不做限定。

且关于对融合特征中的文本特征的替换原理，可以参见上述实施例中，对融合特征中的文本特征中的部分文本特征的遮盖处理的原理，此处不再赘述。

同理，在本实施例中，通过对融合特征中的文本特征进行替换处理，并基于替换后的文本特征、以及融合特征中的视觉特征，对两个模型（即文本检测模型和特征融合模型）的参数分别进行调整，可以提高两个模型的识别辨别能力，从而提高训练得到的文本识别模型框架的准确性和可靠性的技术效果。

另一个示例中，可以结合融合特征和多个视觉特征块，对文本检测模型的参数进行调整，且对特征融合模型的参数进行调整。

20 在一些实施例中，结合融合特征和多个视觉特征块，对文本检测模型的参数进行调整，且对特征融合模型的参数进行调整，可以包括如下步骤：

第一步骤：由特征融合模型随机遮盖融合特征中的部分视觉特征，并根据多个视觉特征块对被遮盖的部分视觉特征进行预测补齐处理，得到预测补齐后的部分视觉特征。

25 第二步骤：根据预测补齐后的部分视觉特征、以及融合特征中除被遮盖的部分视觉特征以外的特征，对文本检测模型和特征融合模型的参数分别进行调整。

其中，关于本实施例的实现原理，可以参见上述实施例中，结合融合特征和多个文本特征块的实现原理，此处不再赘述。

同理，在本实施例中，通过遮盖融合特征中部分视觉特征，并基于多个视觉特征块对遮盖的部分视觉特征进行预测补齐，以基于得到的预测补齐后的部分视觉特征对两个模型（即文本检测模型和特征融合模型）的参数分别进行调整，可以提高两个模型的识别辨别能力，从而提高训练得到的文本识别模型框架的准确性和可靠性的技术

效果。

在另一些实施例中，结合融合特征和多个视觉特征块，对文本检测模型的参数进行调整，且对特征融合模型的参数进行调整，可以包括如下步骤：

5 第一步骤：由特征融合模型根据多个视觉特征块中的至少部分视觉特征块，对融合特征中的视觉特征进行替换处理，得到替换后的视觉特征。

第二步骤：根据融合特征中的文本特征、以及替换后的视觉特征，对文本检测模型和特征融合模型的参数分别进行调整。

其中，关于本实施例的实现原理，可以参见上述实施例中，结合融合特征和多个文本特征块的实现原理，此处不再赘述。

10 同理，在本实施例中，通过替换融合特征中的视觉特征，并基于替换后的视觉特征、以及融合特征中的文本特征，对两个模型（即文本检测模型和特征融合模型）的参数分别进行调整，可以提高两个模型的识别辨别能力，从而提高训练得到的文本识别模型框架的准确性和可靠性的技术效果。

15 再一个示例中，可以结合融合特征、多个文本特征块、以及多个视觉特征块，对文本检测模型的参数进行调整，且对特征融合模型的参数进行调整。

例如，该示例可以包括如下步骤：

第一步骤：由特征融合模型根据融合特征和多个文本特征块，确定用于对文本检测模型和特征融合模型的进行调整的第一调整任务结果。

20 第二步骤：由特征融合模型根据融合特征和多个视觉特征块，确定用于对文本检测模型和特征融合模型的进行调整的第二调整任务结果。

第三步骤：根据第一调整任务结果和第二调整任务结果的加权平均信息，对文本检测模型和特征融合模型的参数分别进行调整。

25 结合上述实施例，在一些实施例中，可以将遮挡融合特征中的部分文本特征作为第一训练任务，得到第一训练结果；将遮挡融合特征中的部分视觉特征作为第二训练任务，得到第二训练结果；将替换融合特征中的文本特征作为第三训练任务，进行训练，得到第三训练结果；对第一训练结果、第二训练结果、以及第三训练结果进行加权平均处理，从而得到最终用于调整文本检测模型的参数和特征融合模型参数，并基于得到的最终用于调整文本检测模型的参数对文本检测模型的参数进行调整，基于得到的最终用于对调整特征融合模型的参数对特征融合模型的参数进行调整。

30 在另一些实施例中，可以将遮挡融合特征中的部分文本特征作为第一训练任务，得到第一训练结果；将遮挡融合特征中的部分视觉特征作为第二训练任务，得到第二

训练结果；将替换融合特征中的视觉特征作为第三训练任务，进行训练，得到第三训练结果；对第一训练结果、第二训练结果、以及第三训练结果进行加权平均处理，从而得到最终用于调整文本检测模型的参数和特征融合模型参数，并基于得到的最终用于调整文本检测模型的参数对文本检测模型的参数进行调整，基于得到的最终用于对

5 调整特征融合模型的参数对特征融合模型的参数进行调整。

关于本实施例中训练任务的组合方式，可以为上述三种训练任务相结合（应该理解地是，上述三种训练任务相结合的方式，只是用于示范性地说明，而不能理解为对相结合的方式的限定，即还可以为除上述组合方式之外的其他组合方式，此处不再一一列举），也可以为两种训练任务相结合，其实现原理可以参见三种训练任务相结合

10 的实现原理，此处不再赘述。

在本实施例中，采用多任务训练的方式训练得到文本识别模型框架，将融合特征和多个文本特征块进行的训练作为其中的一个训练任务，将融合特征和多个视觉特征块进行的训练作为另一个训练任务，并基于两个训练任务的训练结果确定最终用于调整文本检测模型的参数和特征融合模型的参数，从而实现对文本检测模型的参数和特

15 征融合模型的参数的调整，而通过多任务训练的方式调整文本检测模型的参数和特征融合模型的参数，可以实现调整的准确性和可靠性的技术效果。

值得说明地是，在本实施例中，通过基于不同的方式，对文本检测模型和特征融合模型的参数分别进行调整，可以提高参数调整的灵活性和多样性，从而实现训练文本识别模型框架的灵活性和多样性的技术效果。

图 4 是根据本公开第三实施例的示意图，如图 4 所示，本实施例的文本识别方法包括：

20

S401：获取待识别图像。

S402：将待识别图像输入至预先训练的文本识别模型，得到待识别图像中的文本信息。

其中，文本识别模型是基于预先训练的文本识别模型框架对待训练图像进行训练生成的，文本识别模型框架为由如上任意实施例所述训练方法训练获得，待训练图像中包括文本信息。

25

基于上述分析可知，文本识别模型框架包括文本检测模型和特征融合模型，具有较高的准确性和可靠性，因此，当基于文本识别模型框架训练得到文本识别模型时，

30 可以使得文本识别模型具有较高的准确性和可靠性的技术效果，而当基于文本识别模型对待识别图像进行识别时，可以提高识别的有效性和可靠性的技术效果。

图 5 是根据本公开第四实施例的示意图，如图 5 所示，本实施例的文本识别模型框架的训练装置 500 包括：

处理单元 501，用于基于预设的文本检测模型对样本图像进行特征处理，得到与样本图像中文本信息相关的至少两种特征信息。

5 融合单元 502，用于基于预设的特征融合模型对样本图像的至少两种特征信息进行融合处理，得到样本图像的融合特征。

训练单元 503，用于将融合特征输入至特征融合模型，基于融合特征模型对文本检测模型和特征融合模型的参数分别进行调整，得到文本识别模型框架，其中，文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型。

10 图 6 是根据本公开第五实施例的示意图，如图 6 所示，本实施例的文本识别模型框架的训练装置 600 包括：

处理单元 601，用于基于预设的文本检测模型对样本图像进行特征处理，得到与样本图像中文本信息相关的至少两种特征信息。

结合图 6 可知，在一些实施例中，处理单元 601 包括：

15 第一确定子单元 6011，用于基于文本检测模型确定样本图像中文本行的位置信息。

第二确定子单元 6012，用于根据位置信息确定至少两种特征信息。

在一些实施例中，第二确定子单元 6012 包括：

裁剪模块，用于根据位置信息对样本图像进行裁剪操作，得到文本区域。

获取模块，用于从文本区域中获取至少两种特征信息。

20 在一些实施例中，获取模块用于，从文本区域中提取样本图像的图像特征，并对图像特征进行识别，得到至少两种特征信息。

融合单元 602，用于基于预设的特征融合模型对样本图像的至少两种特征信息进行融合处理，得到样本图像的融合特征。

其中，至少两种特征信息包括文本特征和视觉特征。

25 构建单元 603，用于构建用于表征文本特征的多个文本特征块，并构建用于表征视觉特征的多个视觉特征块。

训练单元 604，用于将融合特征输入至特征融合模型，基于融合特征模型对文本检测模型和特征融合模型的参数分别进行调整，得到文本识别模型框架，其中，文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型。

30 结合图 6 可知，在一些实施例中，训练单元 604 包括：

第一遮盖子单元 60411，用于随机遮盖融合特征中的部分文本特征。

第二预测补齐子单元 60412, 用于根据多个文本特征块对遮盖的所述部分文本特征进行预测补齐处理, 得到预测补齐后的部分文本特征。

第一调整子单元 60413, 用于根据预测补齐后的部分文本特征、以及融合特征中除被遮盖的部分文本特征以外的特征, 对文本检测模型和特征融合模型的参数分别进行调整。

在另一些实施例中, 训练单元 604 包括:

第二遮盖子单元 60414, 用于随机遮盖融合特征中的部分视觉特征。

第二预测补齐子单元 60415, 用于根据多个视觉特征块对被遮盖的部分视觉特征进行预测补齐处理, 得到预测补齐后的部分视觉特征。

第二调整子单元 60416, 用于根据预测补齐后的部分视觉特征、以及融合特征中除被遮盖的部分视觉特征以外的特征, 对文本检测模型和特征融合模型的参数分别进行调整。

结合图 6 可知, 在一些实施例中, 训练单元 604 还可以包括:

第一替换子单元 60417, 用于根据多个文本特征块中的至少部分文本特征块, 对融合特征中的文本特征进行替换处理, 得到替换后的文本特征。

第三调整子单元 60418, 用于根据融合特征中的视觉特征、以及替换后的文本特征, 对文本检测模型和特征融合模型的参数分别进行调整。

在另一些实施例中, 训练单元 604 包括:

第二替换子单元 60419, 用于根据多个视觉特征块中的至少部分视觉特征块, 对融合特征中的视觉特征进行替换处理, 得到替换后的视觉特征。

第四调整子单元 60420, 用于根据融合特征中的文本特征、以及替换后的视觉特征, 对文本检测模型和特征融合模型的参数分别进行调整。

结合图 6 可知, 在一些实施例中, 若由特征融合模型根据融合特征、多个文本特征块、以及多个视觉特征块, 对文本检测模型和特征融合模型的参数分别进行调整, 训练单元 604 还可以包括:

第三确定子单元 60421, 用于根据融合特征和多个文本特征块, 确定用于对文本检测模型和特征融合模型的进行调整的第一调整任务结果。

第四确定子单元 60422, 用于根据融合特征和多个视觉特征块, 确定用于对文本检测模型和特征融合模型的进行调整的第二调整任务结果。

第五调整子单元 60423, 用于根据第一调整任务结果和第二调整任务结果的加权平均信息, 对文本检测模型和特征融合模型的参数分别进行调整。

图 7 是根据本公开第六实施例的示意图，如图 7 所示，本实施例的文本识别装置 700，包括：

获取单元 701，用于获取待训练图像，待训练图像中包括文本信息。

5 识别单元 702，用于将待识别图像输入至预先训练的文本识别模型，得到待识别图像中的文本信息，其中，文本识别模型是基于预先训练的文本识别模型框架对待训练图像进行训练生成的，文本识别模型框架为由如上任一实施例所述训练方法训练获得，待训练图像中包括文本信息。

根据本公开的实施例，本公开还提供了一种电子设备和一种可读存储介质。

10 根据本公开的实施例，本公开还提供了一种计算机程序产品，计算机程序产品包括：计算机程序，计算机程序存储在可读存储介质中，电子设备的至少一个处理器可以从可读存储介质读取计算机程序，至少一个处理器执行计算机程序使得电子设备执行上述任一实施例提供的方案。

图 8 示出了可以用来实施本公开的实施例的示例电子设备 800 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、15 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。

20 如图 8 所示，电子设备 800 包括计算单元 801，其可以根据存储在只读存储器 (ROM) 802 中的计算机程序或者从存储单元 808 加载到随机访问存储器 (RAM) 803 中的计算机程序，来执行各种适当的动作和处理。在 RAM 803 中，还可存储设备 800 操作所需的各种程序和数据。计算单元 801、ROM 802 以及 RAM 803 通过总线 804 彼此相连。输入/输出 (I/O) 接口 805 也连接至总线 804。

25 设备 800 中的多个部件连接至 I/O 接口 805，包括：输入单元 806，例如键盘、鼠标等；输出单元 807，例如各种类型的显示器、扬声器等；存储单元 808，例如磁盘、光盘等；以及通信单元 809，例如网卡、调制解调器、无线通信收发机等。通信单元 809 允许设备 800 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。

30 计算单元 801 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 801 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 801 执行上文所描述的各个方法和处理,例如文本识别模型框架的训练方法、文本识别方法。例如,在一些实施例中,文本识别模型框架的训练方法、文本识别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 808。在一些实施例中,5 中,计算机程序的部分或者全部可以经由 ROM 802 和/或通信单元 809 而被载入和/或安装到设备 800 上。当计算机程序加载到 RAM 803 并由计算单元 801 执行时,可以执行上文描述的文本识别模型框架的训练方法、文本识别方法的一个或多个步骤。备选地,在其他实施例中,计算单元 801 可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行文本识别模型框架的训练方法、文本识别方法。

10 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)、负载可编程逻辑设备(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15 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。

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20 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。

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25 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。机器可读储存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)、30 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。

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。

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。

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，又称为云计算服务器或云主机，是云计算服务体系中的一项主机产品，以解决了传统物理主机与 VPS 服务（"Virtual Private Server"，或简称 "VPS"）中，存在的管理难度大，业务扩展性弱的缺陷。服务器也可以为分布式系统的服务器，或者是结合了区块链的服务器。

根据本公开实施例的另一个方面，本公开实施例还提供了一种文本识别模型框架的训练系统，包括：

文本检测模型，用于对样本图像进行特征处理，得到与样本图像中文本信息相关的至少两种特征信息。

特征融合模型，用于对样本图像的至少两种特征信息进行融合处理，得到样本图像的融合特征。

特征融合模型还用于，对文本检测模型和特征融合模型的参数分别进行调整，得到文本识别模型框架，其中，文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型。

结合上述分析可知，在一些实施例中，文本检测模型可以为光学字符识别模型，特征融合模型可以为 transfromer 模型。

结合图 9 可知，在一些实施例中，本公开实施例的文本识别模型框架的训练系统 900，包括：

光学字符识别模型 901，用于对样本图像中的文本信息进行检测，得到样本图像中文本行的位置信息，并将位置信息传输给区域特征提取器 902。

5 区域特征提取器 902，用于根据位置信息对样本图像进行裁剪操作，得到文本区域，并将文本区域分别传输给文字识别器 903 和视觉识别器 904。

文字识别器 903，用于确定文本区域中的文本特征，并将文本特征传输给 transfromer 模型 905。

10 视觉识别器 904，用于确定文本区域中的视觉特征，并将视觉特征传输给 transfromer 模型 905。

transfromer 模型 905 对文本特征和视觉特征进行融合处理，得到融合特征，并基于融合特征对光学字符识别模型 901 的参数和 transfromer 模型 905 的参数进行调整，从而得到文本识别模型框架。

其中，文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型。

15 应该理解地是，上述实施例中的各组件可以一体集成，也可以独立形成，本实施例不做限定。

例如，光学字符识别模型、区域特征提取器、文字识别器、以及视觉识别器，为彼此独立的组件；又如，区域特征提取器为集成与光学字符识别模型中的组件，与文字识别器和视觉识别器分别独立，且文字识别器和视觉识别器为两个独立的组件等，
20 此处不在一一列举。

且关于上述各特征的实现原理可以参见上述方法实施例中地描述，此处不再赘述。

根据本申请实施例的另一个方面，本申请实施例还提供了一种计算机程序，包括程序代码，当计算机运行所述计算机程序时，所述程序代码执行如上任一实施例所述的方法。

25 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，本发申请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提供的技术方案所期望的结果，本文在此不进行限制。

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
30 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。

权 利 要 求 书

1. 一种文本识别模型框架的训练方法，所述方法包括：

基于预设的文本检测模型对样本图像进行特征处理，得到与所述样本图像中文本信息相关的至少两种特征信息；

5 基于预设的特征融合模型对所述样本图像的至少两种特征信息进行融合处理，得到所述样本图像的融合特征；

将所述融合特征输入至所述特征融合模型，基于所述融合特征模型对所述文本检测模型和所述特征融合模型的参数分别进行调整，得到文本识别模型框架，其中，所述文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型。

10 2. 根据权利要求 1 所述的方法，其中，基于预设的文本检测模型对样本图像进行特征处理，得到与所述样本图像中文本相关的至少两种特征信息，包括：

基于所述文本检测模型确定所述样本图像中文本行的位置信息，并根据所述位置信息确定所述至少两种特征信息。

15 3. 根据权利要求 2 所述的方法，其中，根据所述位置信息确定所述至少两种特征信息，包括：

根据所述位置信息对所述样本图像进行裁剪操作，得到文本区域，并从所述文本区域中获取所述至少两种特征信息。

4. 根据权利要求 3 所述的方法，其中，从所述文本区域中获取所述至少两种特征信息，包括：

20 从所述文本区域中提取所述样本图像的图像特征，并对所述图像特征进行识别，得到所述至少两种特征信息。

5. 根据权利要求 2 至 4 中任一项所述的方法，所述至少两种特征信息包括文本特征和视觉特征；在根据所述位置信息确定所述至少两种特征信息之后，还包括：

25 构建用于表征所述文本特征的多个文本特征块，并构建用于表征所述视觉特征的多个视觉特征块；

以及，将所述融合特征输入至所述特征融合模型，基于所述融合特征模型对所述文本检测模型和所述特征融合模型的参数分别进行调整，包括：由所述特征融合模型根据所述融合特征和所述多个文本特征块，对所述文本检测模型和所述特征融合模型的参数分别进行调整；和/或，由所述特征融合模型根据所述融合特征和所述多个视觉特征块，对所述文本检测模型和所述特征融合模型的参数分别进行调整。

30

6. 根据权利要求 5 所述的方法, 其中, 由所述特征融合模型根据所述融合特征和所述多个文本特征块, 对所述文本检测模型和所述特征融合模型的参数分别进行调整, 包括:

5 由所述特征融合模型随机遮盖所述融合特征中的部分文本特征, 并根据所述多个文本特征块对遮盖的所述部分文本特征进行预测补齐处理, 得到预测补齐后的部分文本特征;

根据所述预测补齐后的部分文本特征、以及所述融合特征中除被遮盖的所述部分文本特征以外的特征, 对所述文本检测模型和所述特征融合模型的参数分别进行调整。

7. 根据权利要求 5 所述的方法, 其中, 由所述特征融合模型根据所述融合特征和
10 所述多个视觉特征块, 对所述文本检测模型和所述特征融合模型的参数分别进行调整, 包括:

由所述特征融合模型随机遮盖所述融合特征中的部分视觉特征, 并根据所述多个视觉特征块对被遮盖的所述部分视觉特征进行预测补齐处理, 得到预测补齐后的部分视觉特征;

15 根据所述预测补齐后的部分视觉特征、以及所述融合特征中除被遮盖的所述部分视觉特征以外的特征, 对所述文本检测模型和所述特征融合模型的参数分别进行调整。

8. 根据权利要求 5 所述的方法, 其中, 由所述特征融合模型根据所述融合特征和所述多个文本特征块, 对所述文本检测模型和所述特征融合模型的参数分别进行调整, 包括:

20 由所述特征融合模型根据所述多个文本特征块中的至少部分文本特征块, 对所述融合特征中的文本特征进行替换处理, 得到替换后的文本特征;

根据所述融合特征中的视觉特征、以及所述替换后的文本特征, 对所述文本检测模型和所述特征融合模型的参数分别进行调整。

9. 根据权利要求 5 所述的方法, 其中, 由所述特征融合模型根据所述融合特征和
25 所述多个视觉特征块, 对所述文本检测模型和所述特征融合模型的参数分别进行调整, 包括:

由所述特征融合模型根据所述多个视觉特征块中的至少部分视觉特征块, 对所述融合特征中的视觉特征进行替换处理, 得到替换后的视觉特征;

30 根据所述融合特征中的文本特征、以及所述替换后的视觉特征, 对所述文本检测模型和所述特征融合模型的参数分别进行调整。

10. 根据权利要求 5 至 9 中任一项所述的方法, 其中, 若由所述特征融合模型根据

所述融合特征、所述多个文本特征块、以及所述多个视觉特征块，对所述文本检测模型和所述特征融合模型的参数分别进行调整，则对所述文本检测模型和所述特征融合模型的参数分别进行调整，包括：

5 由所述特征融合模型根据所述融合特征和所述多个文本特征块，确定用于对所述文本检测模型和所述特征融合模型的进行调整的第一调整任务结果；

由所述特征融合模型根据所述融合特征和所述多个视觉特征块，确定用于对所述文本检测模型和所述特征融合模型的进行调整的第二调整任务结果；

根据所述第一调整任务结果和第二调整任务结果的加权平均信息，对所述文本检测模型和所述特征融合模型的参数分别进行调整。

10 11. 一种文本识别方法，包括：

获取待识别图像；

15 将所述待识别图像输入至预先训练的文本识别模型，得到所述待识别图像中的文本信息，其中，所述文本识别模型是基于预先训练的文本识别模型框架对待训练图像进行训练生成的，所述文本识别模型框架为由权利要求 1 至 10 中任一项所述训练方法训练获得，所述待训练图像中包括文本信息。

12. 一种文本识别模型框架的训练装置，所述装置包括：

处理单元，用于基于预设的文本检测模型对样本图像进行特征处理，得到与所述样本图像中文本信息相关的至少两种特征信息；

20 融合单元，用于基于预设的特征融合模型对所述样本图像的至少两种特征信息进行融合处理，得到所述样本图像的融合特征；

训练单元，用于将所述融合特征输入至所述特征融合模型，基于所述融合特征模型对所述文本检测模型和所述特征融合模型的参数分别进行调整，得到文本识别模型框架，其中，所述文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型。

25 13. 根据权利要求 12 所述的装置，其中，所述处理单元，包括：

第一确定子单元，用于基于所述文本检测模型确定所述样本图像中文本行的位置信息；

第二确定子单元，用于根据所述位置信息确定所述至少两种特征信息。

14. 根据权利要求 13 所述的装置，其中，所述第二确定子单元，包括：

30 裁剪模块，用于根据所述位置信息对所述样本图像进行裁剪操作，得到文本区域；
获取模块，用于从所述文本区域中获取所述至少两种特征信息。

15. 根据权利要求 14 所述的装置, 其中, 所述获取模块用于, 从所述文本区域中提取所述样本图像的图像特征, 并对所述图像特征进行识别, 得到所述至少两种特征信息。

16. 根据权利要求 13 至 15 中任一项所述的装置, 所述至少两种特征信息包括文本特征和视觉特征; 还包括:

构建单元, 用于构建用于表征所述文本特征的多个文本特征块, 并构建用于表征所述视觉特征的多个视觉特征块;

以及, 所述训练单元用于, 由所述特征融合模型根据所述融合特征和所述多个文本特征块, 对所述文本检测模型和所述特征融合模型的参数分别进行调整; 和/或, 用于由所述特征融合模型根据所述融合特征和所述多个视觉特征块, 对所述文本检测模型和所述特征融合模型的参数分别进行调整。

17. 根据权利要求 16 所述的装置, 其中, 所述训练单元, 包括:

第一遮盖子单元, 用于随机遮盖所述融合特征中的部分文本特征;

第一预测补齐子单元, 用于根据所述多个文本特征块对遮盖的所述部分文本特征进行预测补齐处理, 得到预测补齐后的部分文本特征;

第一调整子单元, 用于根据所述预测补齐后的部分文本特征、以及所述融合特征中除被遮盖的所述部分文本特征以外的特征, 对所述文本检测模型和所述特征融合模型的参数分别进行调整。

18. 根据权利要求 16 所述的装置, 其中, 所述训练单元, 包括:

第二遮盖子单元, 用于随机遮盖所述融合特征中的部分视觉特征;

第二预测补齐子单元, 用于根据所述多个视觉特征块对被遮盖的所述部分视觉特征进行预测补齐处理, 得到预测补齐后的部分视觉特征;

第二调整子单元, 用于根据所述预测补齐后的部分视觉特征、以及所述融合特征中除被遮盖的所述部分视觉特征以外的特征, 对所述文本检测模型和所述特征融合模型的参数分别进行调整。

19. 根据权利要求 16 所述的装置, 其中, 所述训练单元, 包括:

第一替换子单元, 用于根据所述多个文本特征块中的至少部分文本特征块, 对所述融合特征中的文本特征进行替换处理, 得到替换后的文本特征;

第三调整子单元, 用于根据所述融合特征中的视觉特征、以及所述替换后的文本特征, 对所述文本检测模型和所述特征融合模型的参数分别进行调整。

20. 根据权利要求 16 所述的装置, 其中, 所述训练单元, 包括:

第二替换子单元，用于根据所述多个视觉特征块中的至少部分视觉特征块，对所述融合特征中的视觉特征进行替换处理，得到替换后的视觉特征；

第四调整子单元，用于根据所述融合特征中的文本特征、以及所述替换后的视觉特征，对所述文本检测模型和所述特征融合模型的参数分别进行调整。

- 5 21. 根据权利要求 16 至 20 中任一项所述的装置，其中，若由所述特征融合模型根据所述融合特征、所述多个文本特征块、以及所述多个视觉特征块，对所述文本检测模型和所述特征融合模型的参数分别进行调整，则所述训练单元，包括：

第三确定子单元，用于根据所述融合特征和所述多个文本特征块，确定用于对所述文本检测模型和所述特征融合模型的进行调整的第一调整任务结果；

- 10 第四确定子单元，用于根据所述融合特征和所述多个视觉特征块，确定用于对所述文本检测模型和所述特征融合模型的进行调整的第二调整任务结果；

第五调整子单元，用于根据所述第一调整任务结果和第二调整任务结果的加权平均信息，对所述文本检测模型和所述特征融合模型的参数分别进行调整。

22. 一种文本识别装置，包括：

- 15 获取单元，用于获取待识别图像；

识别单元，用于所述待识别图像输入至预先训练的文本识别模型，得到所述待识别图像中的文本信息，其中，所述文本识别模型是基于预先训练的文本识别模型框架对待训练图像进行训练生成的，所述文本识别模型框架为由权利要求 1 至 10 中任一项所述训练方法训练获得，所述待训练图像中包括文本信息。

- 20 23. 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，

- 25 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 至 10 中任一项所述的方法；或者，以使所述至少一个处理器能够执行权利要求 11 所述的方法。

24. 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权利要求 1 至 10 中任一项所述的方法；或者，所述计算机指令用于使所述计算机执行权利要求 11 所述的方法。

- 30 25. 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 至 10 中任一项所述的方法；或者，所述计算机程序在被处理器执行时实现根据权利要求 11 所述的方法。

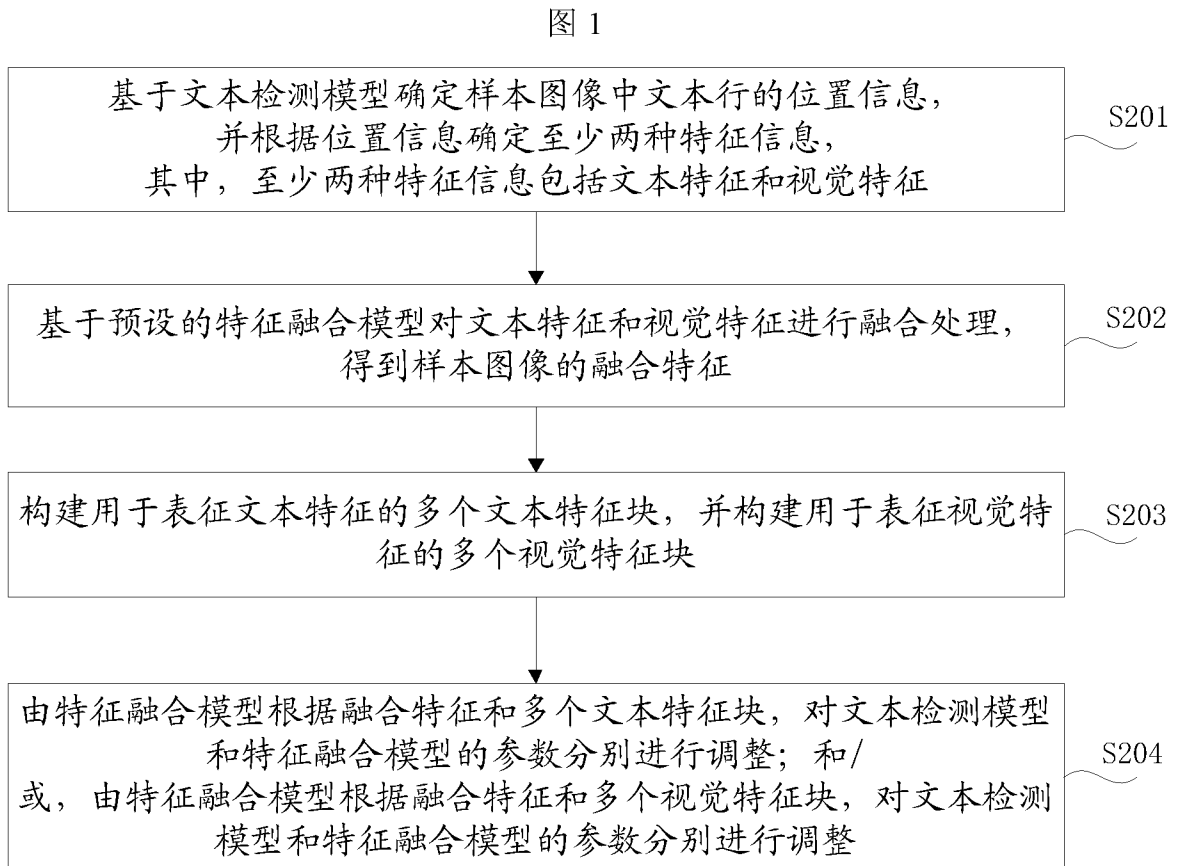
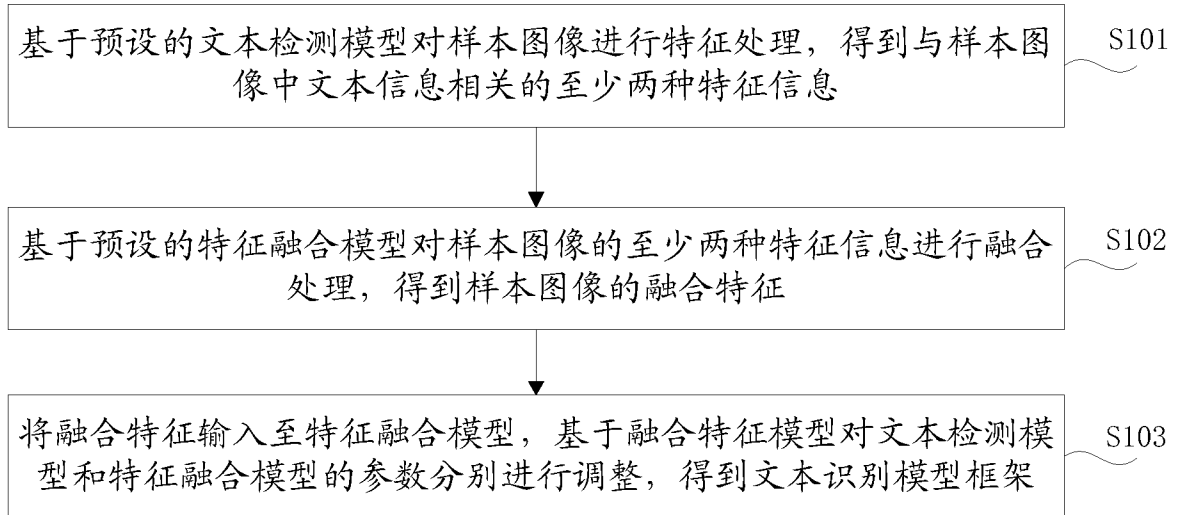
26.一种文本识别模型框架的训练系统，所述系统包括：

文本检测模型，用于对样本图像进行特征处理，得到与所述样本图像中文本信息相关的至少两种特征信息；

5 特征融合模型，用于对所述样本图像的至少两种特征信息进行融合处理，得到所述样本图像的融合特征；

所述特征融合模型还用于，对所述文本检测模型和所述特征融合模型参数分别进行调整，得到文本识别模型框架，其中，所述文本识别模型框架中包括调整后的文本检测模型和调整后的特征融合模型。

27.一种计算机程序，包括程序代码，当计算机运行所述计算机程序时，所述程序
10 代码执行如权利要求 1-10 中任一项所述的方法；或者，所述程序代码执行如权利要求
11 所述的方法。



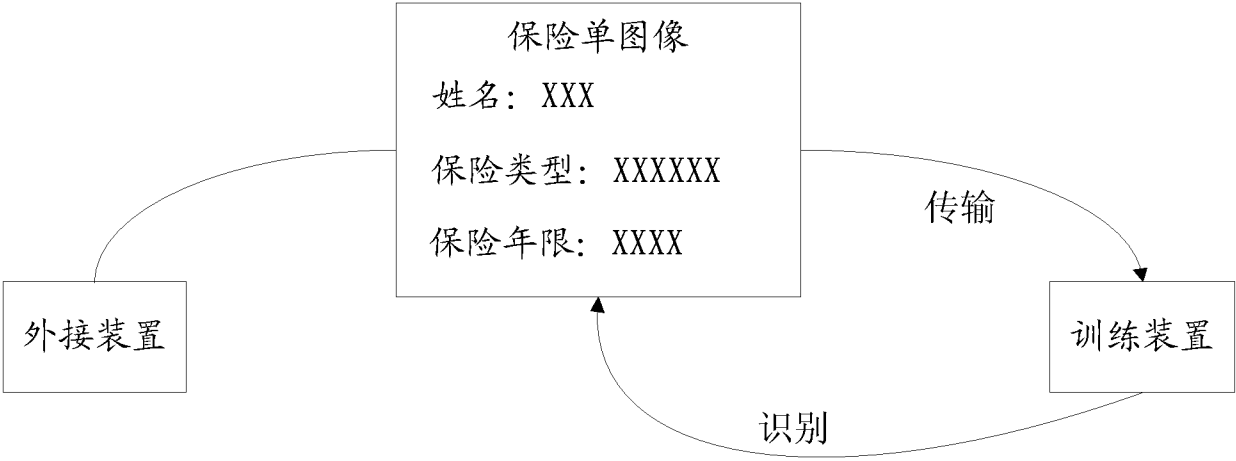


图 3

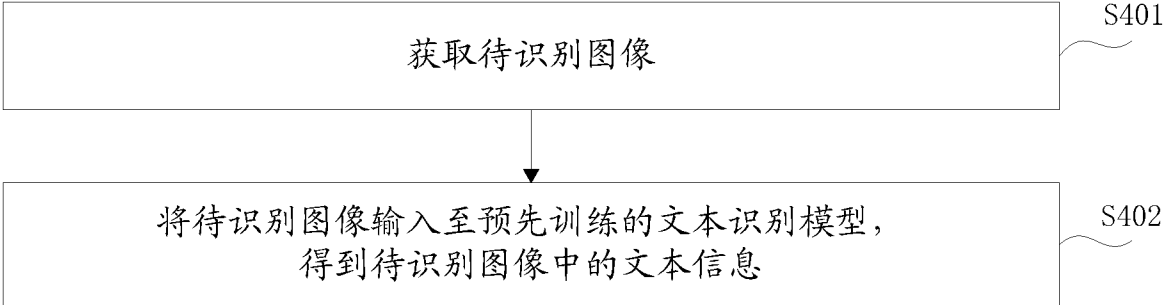


图 4

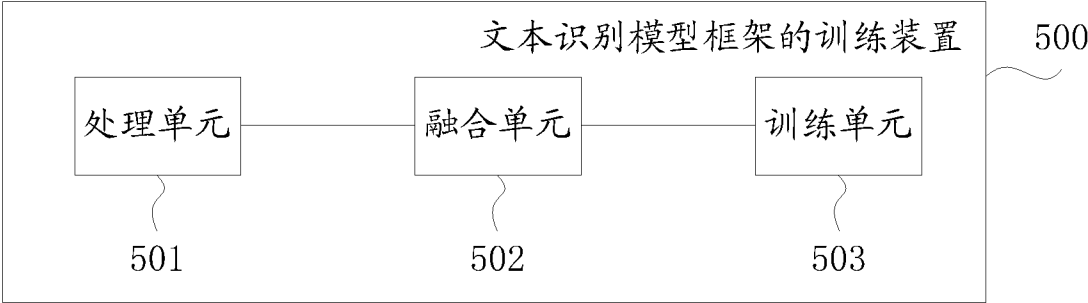


图 5

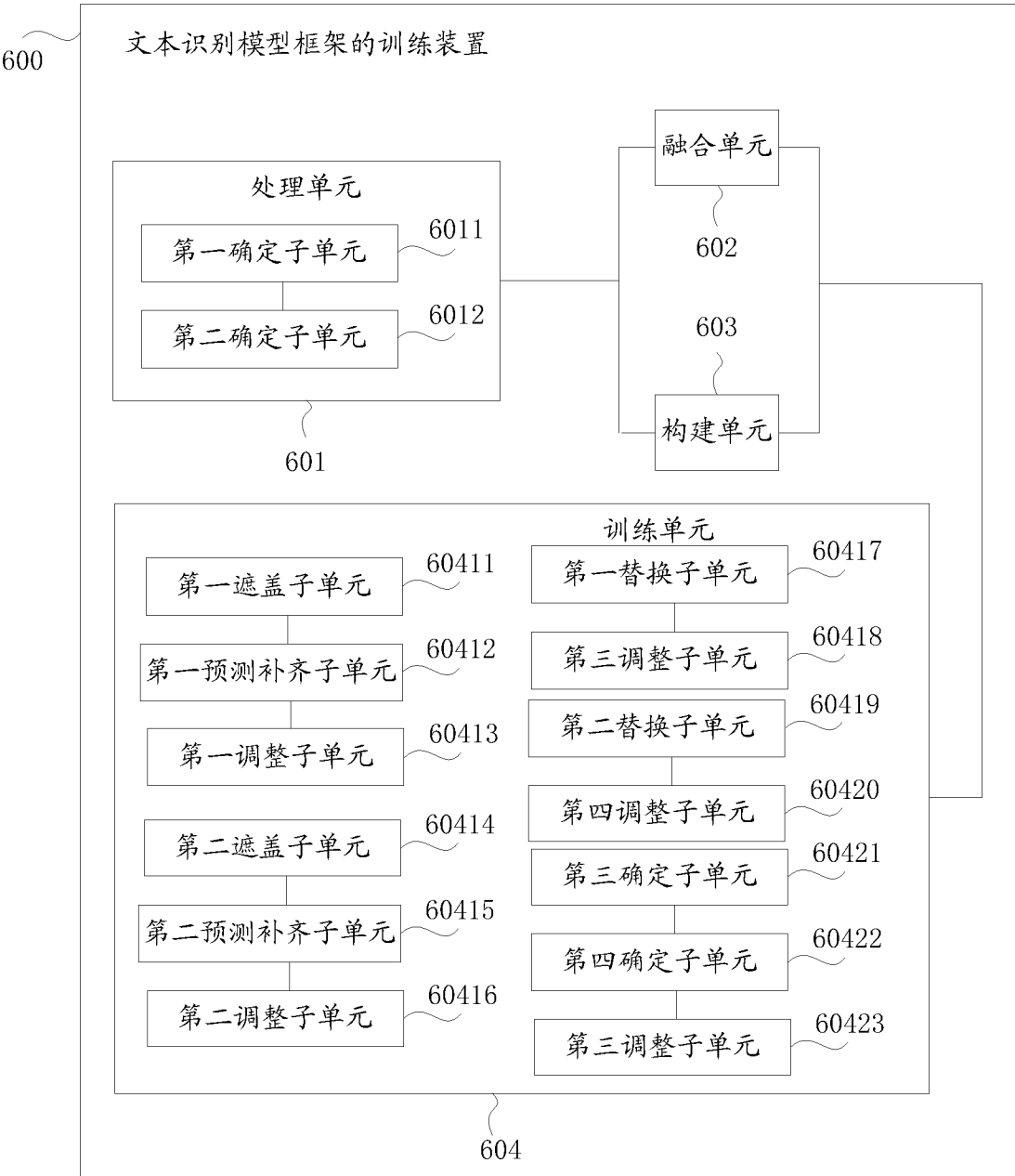


图 6

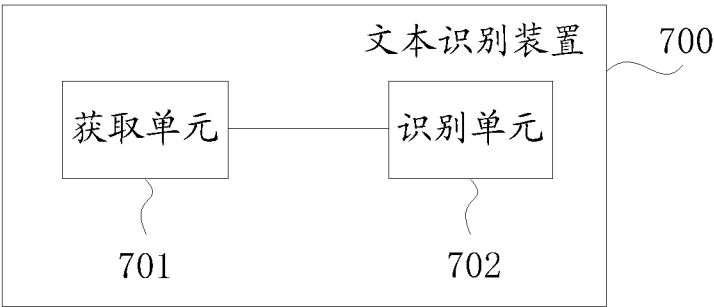


图 7

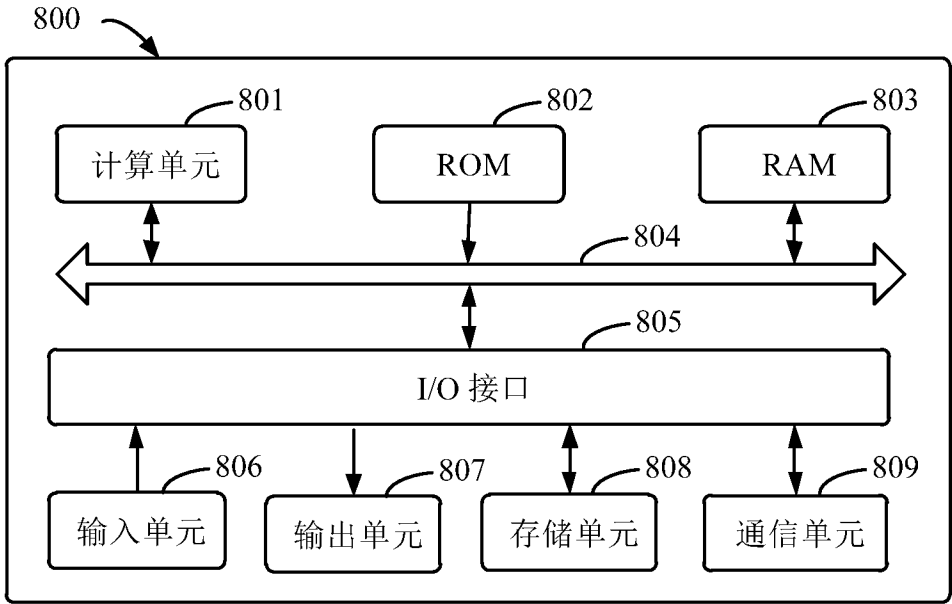


图 8

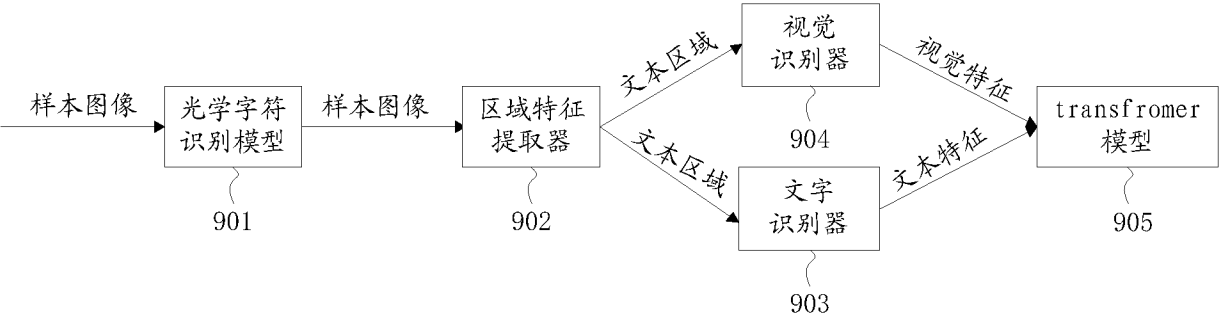


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/085149

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2022.01)i; G06K 9/00(2022.01)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

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)

CNABS; CNTXT; CNKI; WPABS; DWPI; USTXT; WOTXT; EPTXT; IEEE: 北京百度网讯科技有限公司, 章成全, 文本, 识别, 检测, 位置, 特征, 融合, 模型, 参数, 调整, text, recognition, detection, position, feature, fusion, model, parameter, adjust

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113591864 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 02 November 2021 (2021-11-02) claims 1-26, and description, paragraphs [0041]-[0209]	1-27
Y	CN 109271967 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 25 January 2019 (2019-01-25) description, paragraphs [0043]-[0173]	1-27
Y	CN 111507355 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 07 August 2020 (2020-08-07) description, paragraphs [0027]-[0100]	1-27
A	CN 111738251 A (BEIJING ACAD OF ARTIFICIAL INTELLIGENCE) 02 October 2020 (2020-10-02) entire document	1-27
A	CN 110163110 A (CHINA POWERTECH BIG DATA RESEARCH INSTITUTE CO., LTD.) 23 August 2019 (2019-08-23) entire document	1-27



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

05 May 2022

Date of mailing of the international search report

18 May 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/CN2022/085149

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
CN	113591864	A	02 November 2021	None		
CN	109271967	A	25 January 2019	None		
CN	111507355	A	07 August 2020	None		
CN	111738251	A	02 October 2020	CN	111738251	B 04 December 2020
CN	110163110	A	23 August 2019	None		

国际检索报告

国际申请号

PCT/CN2022/085149

A. 主题的分类

G06K 9/62(2022.01)i; G06K 9/00(2022.01)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

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))

CNABS;CNTXT;CNKI;WPABS;DWPI;USTXT;WOTXT;EPTXT;IEEE: 北京百度网讯科技有限公司, 章成全, 文本, 识别, 检测, 位置, 特征, 融合, 模型, 参数, 调整, text, recognition, detection, position, feature, fusion, model, parameter, adjust

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 113591864 A (北京百度网讯科技有限公司) 2021年11月2日 (2021 - 11 - 02) 权利要求第1-26项, 说明书第[0041]-[0209]段	1-27
Y	CN 109271967 A (腾讯科技深圳有限公司) 2019年1月25日 (2019 - 01 - 25) 说明书第[0043]-[0173]段	1-27
Y	CN 111507355 A (北京百度网讯科技有限公司) 2020年8月7日 (2020 - 08 - 07) 说明书第[0027]-[0100]段	1-27
A	CN 111738251 A (北京智源人工智能研究院) 2020年10月2日 (2020 - 10 - 02) 全文	1-27
A	CN 110163110 A (中电科大数据研究院有限公司) 2019年8月23日 (2019 - 08 - 23) 全文	1-27

☐ 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

“&” 同族专利的文件

国际检索实际完成的日期

2022年5月5日

国际检索报告邮寄日期

2022年5月18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

受权官员

顾明海

电话号码 (86-512) 88995878

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/085149

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	113591864	A	2021年11月2日	无	
CN	109271967	A	2019年1月25日	无	
CN	111507355	A	2020年8月7日	无	
CN	111738251	A	2020年10月2日	CN 111738251	B 2020年12月4日
CN	110163110	A	2019年8月23日	无	