(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 114118075 B
(45)授权公告日 2022.04.22

(21)申请号 202210103724.3
(22)申请日 2022.01.28
(65)同一申请的已公布的文献号
申请公布号 CN 114118075 A
(43)申请公布日 2022.03.01
(73)专利权人 北京易真学思教育科技有限公司
地址 102200 北京市昌平区未来科学城英才北三街16号院16号楼401室
(72)发明人 秦勇
(74)专利代理机构 北京北汇律师事务所 11711
代理人 张臻贤
(51)Int.Cl.
G06F 40/279(2020.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

审查员 李咏梅

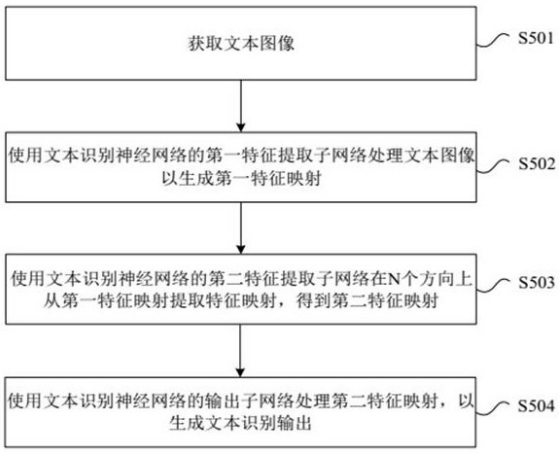
权利要求书2页 说明书12页 附图6页

(54)发明名称

文本识别方法、装置、电子设备和存储介质

(57)摘要

本公开提供一种文本识别方法、装置、电子设备和存储介质，其中，文本识别方法包括：获取文本图像；使用文本识别神经网络的第一特征提取子网络处理文本图像以生成第一特征映射；使用文本识别神经网络的第二特征提取子网络在多个(N个)方向上从第一特征映射提取特征映射，得到第二特征映射；使用文本识别神经网络的输出子网络处理第二特征映射，以生成文本识别输出。根据本公开，提取文本图像的特征映射，进一步在多个方向上从该特征映射提取特征映射，基于多个方向对应的特征映射生成文本识别输出，由于从多个方向上提取的特征映射表征了文本图像上文本行中字符之间的位置关系，可以降低多识别字符或漏识别字符，提高文本识别准确率。



CN 114118075 B

1.一种文本识别方法,其特征在于,包括:

获取文本图像,其中,所述文本图像记录有单行文本;

使用文本识别神经网络的第一特征提取子网络处理所述文本图像以生成第一特征映射;

使用所述文本识别神经网络的第二特征提取子网络在N个方向上从所述第一特征映射提取特征映射,得到第二特征映射,其中,N为大于等于2的自然数;

使用所述文本识别神经网络的输出子网络处理所述第二特征映射,以生成文本识别输出;

其中,所述第二特征提取子网络包括N个特征提取分支和整合单元,通过N-1个特征提取分支中的每个特征提取分支,按照对应的方向旋转所述第一特征映射,得到对应的第三特征映射;通过所述第二特征提取子网络,处理所述第一特征映射和所述N-1个第三特征映射以生成所述第二特征映射。

2.如权利要求1所述的文本识别方法,其特征在于,所述文本识别神经网络还包括:编码器子网络和融合单元,其中,

所述文本识别方法,还包括:使用所述编码器子网络处理所述第一特征映射以生成特征向量;

使用所述融合单元按照融合规则将所述特征向量和所述第二特征映射进行融合,以生成融合输出;

所述使用所述文本识别神经网络的输出子网络处理所述第二特征映射,以生成文本识别输出,包括:使用所述文本识别神经网络的输出子网络处理所述融合输出,以生成文本识别输出。

3.如权利要求1所述的文本识别方法,其特征在于,其中,所述通过所述第二特征提取子网络处理所述第一特征映射和所述N-1个第三特征映射以生成第二特征映射,包括:

通过所述N-1个特征提取分支中相应的特征提取分支处理相应的第三特征映射,以及通过剩余的一个特征提取分支处理所述第一特征映射;

通过所述整合单元按照预设整合规则处理所述N个特征提取分支的输出,以生成所述第二特征映射。

4.如权利要求1所述的文本识别方法,其特征在于,所述N-1个方向包括:一个或多个预设方向,以及一个或多个随机方向。

5.如权利要求3所述的文本识别方法,其特征在于,通过所述整合单元按照预设整合规则处理所述N个特征提取分支的输出,以生成所述第二特征映射,包括:通过所述整合单元串联拼接所述N个特征提取分支的输出,以生成所述第二特征映射。

6.如权利要求2所述的文本识别方法,其特征在于,所述使用所述融合单元按照融合规则将所述特征向量与所述第二特征映射进行融合,以生成融合输出,包括:

通过所述融合单元将所述特征向量与所述第二特征映射逐点逐通道相乘,以生成所述融合输出。

7.如权利要求2所述的文本识别方法,其特征在于,所述编码器子网络包括:串联的多个基于注意力的编码器,其中,所述使用所述编码器子网络处理所述第一特征映射以生成特征向量,包括:

对于所述第一特征映射中的每个像素点，使用正余弦函数生成一个向量，由所述第一特征映射中每个像素点对应的所述向量形成位置编码；

通过所述串联的多个基于注意力的编码器根据所述位置编码处理所述第一特征映射，以生成所述特征向量。

8.如权利要求2所述的文本识别方法，其特征在于，所述输出子网络包括一个1*1卷积层，其中，

所述使用所述文本识别神经网络的输出子网络处理所述融合输出，以生成所述文本识别输出，包括：通过所述一个1*1卷积层对所述融合输出进行降维，以生成所述文本识别输出。

9.如权利要求1或2所述的文本识别方法，其特征在于，所述第一特征映射包括多尺度的特征映射。

10.一种文本识别装置，其特征在于，包括：

获取模块，用于获取文本图像，其中，所述文本图像记录有单行文本；

识别模块，用于：

使用文本识别神经网络的第一特征提取子网络处理所述文本图像以生成第一特征映射；

使用所述文本识别神经网络的第二特征提取子网络在N个方向上从所述第一特征映射提取特征映射，得到第二特征映射，其中，N为大于等于2的自然数；

使用所述文本识别神经网络的输出子网络处理所述第二特征映射，以生成文本识别输出；

其中，所述第二特征提取子网络包括N个特征提取分支和整合单元，通过N-1个特征提取分支中的每个特征提取分支，按照对应的方向旋转所述第一特征映射，得到对应的第三特征映射；通过所述第二特征提取子网络，处理所述第一特征映射和所述N-1个第三特征映射以生成所述第二特征映射。

11.如权利要求10所述的文本识别装置，其特征在于，所述识别模块，还用于：

使用所述文本识别神经网络的编码器子网络处理所述第一特征映射以生成特征向量；

使用所述文本识别神经网络的融合单元按照融合规则将所述特征向量和所述第二特征映射进行融合，以生成融合输出；

其中，使用所述文本识别神经网络的输出子网络处理所述第二特征映射，以生成文本识别输出，包括：使用所述文本识别神经网络的输出子网络处理所述融合输出，以生成文本识别输出。

12.一种电子设备，其特征在于，包括：

处理器；以及

存储程序的存储器，

其中，所述程序包括指令，所述指令在由所述处理器执行时使所述处理器执行根据权利要求1-9中任一项所述的方法。

13.一种存储有计算机指令的非瞬时计算机可读存储介质，其特征在于，所述计算机指令用于使所述计算机执行根据权利要求1-9中任一项所述的方法。

文本识别方法、装置、电子设备和存储介质

技术领域

[0001] 本发明涉及图像处理技术领域，尤其涉及文本识别方法、装置、电子设备和存储介质。

背景技术

[0002] 自然场景文字识别是带文字的图片中识别出字符序列的过程，示例性的，对于中文，一个字符可为一个汉字，对于英文，一个字符可为一个字母。它是一项具有极大挑战性的课题，除了图片背景复杂，光照变化等因素外，识别输出空间的复杂性也是一大困难，由于文字由数量不固定的字符组成，因此，自然场景文字识别需要从图片中识别长度不固定的序列。相关技术中使用序列到序列的方法，先将图像编码，然后进行序列解码得出整个字符串。然而该方法存在识别结果多识别字符或漏识别字符的问题。

发明内容

[0003] 本公开提供了一种文本识别方法、装置、电子设备和存储介质，以至少解决相关技术中文本识别存在识别结果多识别字符或漏识别字符的问题。

[0004] 根据本公开的一方面，提供了一种文本识别方法，包括：

[0005] 获取文本图像，其中，文本图像记录有单行文本；

[0006] 使用文本识别神经网络的第一特征提取子网络处理文本图像以生成第一特征映射；

[0007] 使用文本识别神经网络的第二特征提取子网络在N个方向上从第一特征映射提取特征映射，得到第二特征映射，其中，N为大于等于2的自然数；

[0008] 使用文本识别神经网络的输出子网络处理第二特征映射，以生成文本识别输出。

[0009] 根据本公开的另一方面，提供了一种文本识别装置，包括：

[0010] 获取模块，用于获取文本图像，其中，文本图像记录有单行文本；

[0011] 识别模块，用于：

[0012] 使用文本识别神经网络的第一特征提取子网络处理文本图像以生成第一特征映射；

[0013] 使用文本识别神经网络的第二特征提取子网络在N个方向上从第一特征映射提取特征映射，得到第二特征映射，其中，N为大于等于2的自然数；

[0014] 使用文本识别神经网络的输出子网络处理第二特征映射，以生成文本识别输出。

[0015] 根据本公开的又一方面，提供了一种电子设备，包括：处理器；以及存储程序的存储器，其中，该程序包括指令，该指令在由处理器执行时使处理器执行本公开的文本识别方法。

[0016] 根据本公开的再一方面，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，该计算机指令用于使计算机执行本公开的文本识别方法。

[0017] 本申请实施例中提供的一个或多个技术方案，获取记录有单行文本的文本图像，

提取文本图像的特征映射，进一步在多个方向上从该特征映射提取特征映射，基于多个方向对应的特征映射生成文本识别输出，由于从多个方向上提取的特征映射表征了文本图像上文本行中字符之间的位置关系，可以降低多识别字符或漏识别字符的问题，提高文本识别准确率。

附图说明

[0018] 在下面结合附图对于示例性实施例的描述中，本公开的更多细节、特征和优点被公开，在附图中：

[0019] 图1示出了根据本公开示例性实施例的文本识别系统的示意性框图；

[0020] 图2示出了根据本公开示例性实施例的第二特征提取子网络230的示意性框图；

[0021] 图3示出了根据本公开示例性实施例的编码器子网络220的示意性框图；

[0022] 图4示出了根据本公开示例性实施例的第一特征提取子网络210的示意性框图；

[0023] 图5示出了根据本公开示例性实施例的文本识别方法的流程图；

[0024] 图6示出了根据本公开示例性实施例的文本识别方法的另一流程图；

[0025] 图7示出了根据本公开示例性实施例的多尺度特征提取和处理的示意图；

[0026] 图8示出了根据本公开示例性实施例的文本识别装置的结构框图；以及

[0027] 图9示出了能够用于实现本公开的实施例的示例性电子设备的结构框图。

具体实施方式

[0028] 下面将参照附图更详细地描述本公开的实施例。虽然附图中显示了本公开的某些实施例，然而应当理解的是，本公开可以通过各种形式来实现，而且不应该被解释为限于这里阐述的实施例，相反提供这些实施例是为了更加透彻和完整地理解本公开。应当理解的是，本公开的附图及实施例仅用于示例性作用，并非用于限制本公开的保护范围。

[0029] 应当理解，本公开的方法实施方式中记载的各个步骤可以按照不同的顺序执行，和/或并行执行。此外，方法实施方式可以包括附加的步骤和/或省略执行示出的步骤。本公开的范围在此方面不受限制。

[0030] 本文使用的术语“包括”及其变形是开放性包括，即“包括但不限于”。术语“基于”是“至少部分地基于”。术语“一个实施例”表示“至少一个实施例”；术语“另一实施例”表示“至少一个另外的实施例”；术语“一些实施例”表示“至少一些实施例”。其他术语的相关定义将在下文描述中给出。需要注意，本公开中提及的“第一”、“第二”等概念仅用于对不同的装置、模块或单元进行区分，并非用于限定这些装置、模块或单元所执行的功能的顺序或者相互依存关系。

[0031] 需要注意，本公开中提及的“一个”、“多个”的修饰是示意性而非限制性的，本领域技术人员应当理解，除非在上下文另有明确指出，否则应该理解为“一个或多个”。

[0032] 本公开实施方式中的多个装置之间所交互的消息或者信息的名称仅用于说明性的目的，而并不是用于对这些消息或信息的范围进行限制。

[0033] 文本图像记录的文本行中多个字符的中心点可能不在同一直线上，其一个可能的原因是书写难以保证字符在同一直线上，另一个可能的原因是拍摄纸张上的文字时纸张形变引起字符位置变化，再一个可能的原因是扫描纸张上的文字时引起的字符位置变化。示

例性的，可以根据字符书写方向及走势将文本大致分为三种，正常文本(也称为直文本)、带角度的倾斜文本，以及弯曲文本。以从左向右为例，正常文本的字符大致在一条直线上，这条直线和水平方向几乎重合；带角度的倾斜文本的字符大致在一条直线上，且这条直线和水平方向有一定的夹角；弯曲文本的字符至少部分不在同一直线上，字符的中心点连起来之后大致呈一条曲线。

[0034] 使用神经网络进行文本识别时，可能因字符分布在图像上(相对于文本行字符序列的走向)不同位置，导致识别结果可能会有偏差，由此可能出现多识别字符或漏识别字符的问题。

[0035] 基于此，本公开示例性实施例提供了一种使用文本识别神经网络识别文本的方法、装置及电子设备，以降低文本行中字符之间位置不同导致的多识别字符或漏识别字符的问题，至少可提高文本识别的精度。下面对本公开示例性实施例的文本识别神经网络的系统及文本识别方法进行描述。应当理解，术语“文本行”不是对文本方向的限定，在水平书写中，文本行可为从左到右的若干字符，在竖向书写中，文本行可为从上到下的若干字符。

[0036] 图1示出了根据本公开示例性实施例的文本识别系统的示意性框图。文本识别系统100是被实现为一个或多个位置中一个或多个计算机上的计算机程序的系统的示例。该一个或多个位置中的一个或多个计算机可包括终端和/或服务端及其他具备计算能力的设备。

[0037] 文本识别系统100被配置为处理文本图像101以生成文本识别输出102。在本公开示例性实施例中，文本图像101记录有单行文本。单行文本也就是文本图像上记录一个文本行，或称为一行文本。照片、扫描文档、屏幕截图等图像可经剪裁、预处理等形成承载有单行文本的文本图像101，图像的剪裁可使用本领域公知的技术，本公开对此不作限定。在一些可能的实施方式中，可获取照片、扫描文档、屏幕截图等图像，对获取的图像进行剪裁或其他预处理，得到承载有单行文本的文本图像101。在一些可能的实施方式中，获取用户在图像上选定的区域，从图像中剪裁出该区域得到文本图像101。示例性的，该图像可为通过相机拍摄的照片，也可为扫描文档(例如，扫描得到的PDF文档)，也可以包括屏幕截图(例如，对照片、扫描文档的截图)等。

[0038] 文本图像101记录的文本包括一个或多个字符，字符可包括汉字、英文字母等。在一些可能的情形中，字符可包括手写字符和/或打印字符，例如，文本图像可来自试卷或作业，其内容可包括题干对应的打印字符和解答对应的手写字符。

[0039] 在一些可能的实施方式中，不同的文本图像101可能记录的字符数目不同，文本图像101上字符的数量对于文本识别系统100而言是未知的，也就是文本识别系统101可进行字符长度非固定的文本图像101的识别。

[0040] 在一些可能的实施方式中，文本图像101为预设大小，不同的文本图像101具有相同的大小。

[0041] 参考图1所示，文本识别系统100包括：文本识别神经网络200。文本识别神经网络200被配置为处理文本图像101以生成文本识别输出102(例如识别文本的概率矩阵)。

[0042] 在一些实施例中，参考图1所示，文本识别神经网络200可包括：第一特征提取子网络210、第二特征提取子网络230和输出子网络250。

[0043] 第一特征提取子网络210，被配置为处理文本图像101，以生成第一特征映射1011。

第二特征提取子网络230，被配置为在多个(表示为N个)方向上从第一特征映射1011提取特征映射，得到第二特征映射1012。该多个方向可包括第一特征映射1011的初始方向，以及其他一个或多个方向。输出子网络250，被配置为处理第二特征映射1012，以生成文本识别输出102。

[0044] 第二特征提取子网络230在N个方向上从第一特征映射1011提取特征映射，可以获取到文本图像101上记录的单行文本中字符的位置的信息，对于带角度的倾斜文本、弯曲文本，可以降低因单行文本中字符的位置不同带来的漏识别字符或多识别字符的情况。直文本、倾斜文本和弯曲文本，其特征可被全方位提取。例如，第一个字符位于水平线上，第二个字符向上偏离，第三个字符向下偏离，第四个字符向下偏离(并可与第三个字符的偏离程度不同)，通过在多个方向上从第一特征映射中提取特征映射，可以提取到前述的字符位置变化相关的信息。

[0045] 在一些可能的实施方式中，第二特征提取子网络230被配置为在预设的一个或多个方向以及随机生成一个或多个方向从第一特征映射1011提取特征映射，以适应文本图像101上单行文本中字符位置的不确定性。在一些示例中，第二特征提取子网络230被配置为在预先设置角度范围(例如30°到60°)内随机生成一个或多个方向。在一些示例中，第二特征提取子网络230被配置为在预先设置多个方向(例如30°方向、45°方向、60°方向等)中随机选择一个或一个以上的方向。

[0046] 在一些可能的实施方式中，第二特征提取子网络230被配置为在第一特征映射1011的初始方向，以及相对于该初始方向的90°方向、180°方向，从第一特征映射1011提取特征映射，这些方向便于进行特征映射的操作。可选地，还包括在30°方向、45°方向和60°方向中随机选择的一个方向，由此可适度适应单行文本中字符位置的不确定性。

[0047] 在一些可能的实施方式中，第二特征提取子网络230被配置为：按照N-1个方向(例如，90°方向、180°方向和60°方向)旋转第一特征映射1011，得到相应的N-1个第三特征映射(分别与90°方向、180°方向和60°方向对应)；处理第一特征映射(对应于初始方向，或称为0°方向)和N-1个第三特征映射以生成第二特征映射1012。

[0048] 图2示出了根据本公开示例性实施例的第二特征提取子网络230的示意性框图，参考图2所示，第二特征提取子网络230可包括N个特征提取分支，图2中示出为特征提取分支231-1至231-N。

[0049] 参考图2所示，N个特征提取分支231-1至231-N中的每个特征提取分支，被配置为在其对应的方向上从第一特征映射1011提取特征映射。特征提取分支231-1至231-N可采用任意的特征提取结构，示例性的，特征提取分支231-1至231-N可包括预设数目的卷积层。可选地，每个特征提取分支可权重共享。

[0050] 在一些实施例中，参考图2所示，第二特征提取子网络230还包括整合单元232。整合单元232被配置为按照预设整合规则处理N个特征提取分支231-1至231-N的输出，以生成第二特征映射1012。第二特征映射1012整合了N个方向上提取的特征映射，可包括N个方向上的信息，由此可表征文本行中字符之间位置相关的信息。在一些可能的实施方式中，整合单元232被配置为串联拼接N个特征提取分支231-1至231-N的输出，以生成第二特征映射1012。作为一种示例，N个特征提取分支231-1至231-N分别输出m通道p*q的特征映射，整合单元232将N个m通道p*q的输出串联拼接，得到N*m通道p*q的特征映射，其中，m、p和q为正整

数。

[0051] 在一些可能的实施方式中，输出子网络250采用卷积循环神经网络(Convolutional Recurrent Neural Network,CRNN)中的序列化建模层(例如BiLSTM)以及解码层(例如,联接时间分类(Connectionist temporal classification,CTC))。

[0052] 考虑到CRNN参数量过大，训练周期过长。在一些可能的实施方式中，输出子网络250包括1*1卷积层，通过该1*1卷积层对融合输出1014进行降维，以生成文本识别输出102。由此降低文本识别神经网络200的训练周期，并提高识别速度。应当理解，输出子网络250还可以包括一个或多个网络层，例如卷积层、池化层等，以对融合输出1014作预处理，1*1卷积层以预处理后的融合输出1014为输入。

[0053] 在一些实施例中，参考图1所示，文本识别神经网络200还可包括编码器子网络220和融合单元240。编码器子网络220，被配置为处理第一特征映射1011以生成特征向量1013。融合单元240，被配置为按照融合规则将特征向量1013和第二特征映射1012进行融合，以生成融合输出1014。输出子网络250，被配置为处理融合输出1014，以生成文本识别输出102。通过编码器子网络220进行特征筛选，并通过融合方向相关的特征和筛选得到的特征，可提高识别精度。

[0054] 使用编码器子网络220对第一特征映射1011的局域低阶像素值进行归类与分析，从而获得高阶信息。编码器子网络220可使用各种类型的编码器。在一些可能的实施方式中，使用基于注意力的编码器，通过使用注意力机制，引导神经网络关注输入图像上要识别的文字所在的区域。下面对包括基于注意力的编码器的编码器子网络220进行描述。

[0055] 图3示出了根据本公开示例性实施例的编码器子网络220的示意性框图。编码器子网络220可包括位置编码模块221和串联的多个基于注意力的编码器，图3中示出为编码器222-1至222-6。位置编码模块221被配置为对于第一特征映射1011中的每个像素点，使用正余弦函数生成一个向量，由第一特征映射1011中每个像素点对应的向量形成位置编码。串联的多个基于注意力的编码器，被配置为根据位置编码处理第一特征映射1011，以生成特征向量1013。

[0056] 融合单元240被配置为按照融合规则将特征向量1013与第二特征映射1012进行融合，以生成融合输出1014。在一些可能的实施方式中，融合单元240被配置为将特征向量1013与第二特征映射1012逐点逐通道相乘，以生成融合输出1014。通过逐点逐通道相乘，可以减少相邻特征之间的差距。在一些示例中，在进行融合之前，还可将特征向量1013与第二特征映射1012进行预处理，以使两者的维度适配。

[0057] 输出子网络250被配置为处理融合输出1014，以生成文本识别输出102(例如识别文本的概率矩阵)。在一些可能的实施方式中，输出子网络250采用CRNN中的序列化建模层(例如BiLSTM)以及解码层(例如CTC)。

[0058] 考虑到CRNN参数量过大，训练周期过长。在一些可能的实施方式中，输出子网络250包括1*1卷积层，通过该1*1卷积层对融合输出1014进行降维，以生成文本识别输出102。由此降低文本识别神经网络200的训练周期。

[0059] 第一特征提取子网络210可采用各种类型的神经网络。在一些可能的实施方式中，第一特征提取子网210可采用残差神经网络(ResNet)。下面对采用ResNet的第一特征提取子网络210进行描述。

[0060] 图4示出了根据本公开示例性实施例的第一特征提取子网络210的示意性框图。第一特征提取子网络210包括串联的多个残差块，图4中示出为残差块211-1至211-4。在一些可能的实施方式中，由残差块211-4输出第一特征映射1011。

[0061] 在一些可能的实施方式中，第一特征提取子网络210被配置为提取多尺度的特征映射，由多尺度的特征映射组成第一特征映射1011。残差块211-1至211-4中和至少部分残差块的输出尺度依序递减，以输出多个尺度的特征映射。参考图4所示，经过第一个残差块211-1时，文本图像101的高宽不变，经过后面的残差块211-2至211-4时，高宽每次均减半，即大小分别为原始高宽大小的1/2、1/4和1/8的3组特征映射，分别表示为特征映射1011-1、1011-2和1011-3。下面对编码器子网络220、第二特征提取子网络230和融合单元240对多个尺度的特征映射的处理进行说明。

[0062] 在一些可能的实施方式中，编码器子网络220，被配置为：将多尺度的特征映射缩放到同一尺度，将缩放后的特征映射串联拼接作为编码器子网络220的输入。编码器子网络220的输入融合了多个尺度的信息，可以提高特征的丰富程度，进而提高识别精度。

[0063] 在一些可能的实施方式中，结合图2所示，第二特征提取子网络230的每个特征提取分支，被配置为：将多尺度的特征映射缩放为同一尺度，分别处理缩放后的每组特征映射，得到相应的特征映射，将各组相应的特征映射整合(例如串联拼接)得到该特征提取分支对应的特征映射输出。

[0064] 参考图1所示，文本识别系统100还可包括训练装置300，训练装置300被配置为使用训练数据301训练文本识别神经网络200，以生成文本识别神经网络200的参数。训练数据301包括待识别的文本图像，识别的文本图像包括直文本、倾斜文本和弯曲的文本图像。训练数据301还包括待识别的文本图像的标注信息，也就是文本图像上的文本字符信息，标注信息可为文本图像上的整个字符序列。可选地，待识别的文本图像可设置为同一大小。训练装置300被配置为获取训练数据301，使用练数据301和CTC损失函数对文本识别神经网络200进行训练，以生成文本识别神经网络200的参数。

[0065] 在一些可能的实施方式中，训练装置300可在服务端实施，以训练文本识别神经网络200。训练好的文本识别神经网络200可设置于终端，终端使用文本识别神经网络200进行文本识别。

[0066] 本公开示例性实施例提供了一种文本识别方法，本公开示例性实施例的文本识别方法可应用于客户端和/或服务端等计算设备。该文本识别方法可使用本公开前述的文本识别系统100实现，但不限于此。下面对本公开示例性实施例的文本识别方法进行描述。

[0067] 图5示出了根据本公开示例性实施例的文本识别方法的流程图，参考图5所示，本公开示例性实施例的文本识别方法包括步骤S501至步骤S504。

[0068] 步骤S501，获取文本图像。

[0069] 在本公开示例性实施例中，文本图像记录有单行文本，也就是文本图像上包含一行文本(也称为一个文本行)。照片、扫描文档、屏幕截图等图像可经剪裁、预处理等形成承载有单行文本的文本图像，图像的剪裁可使用本领域公知的技术，本公开对此不作限定。

[0070] 在一些可能的实施方式中，在步骤S501中，获取照片、扫描文档、屏幕截图等图像，对获取的图像进行剪裁或其他预处理，得到承载有单行文本的文本图像。

[0071] 在一些可能的实施方式中，在步骤S501之前，获取照片、扫描文档、屏幕截图等图

像,对获取的图像进行剪裁,得到记录有单行文本的文本图像。

[0072] 在一些可能的实施方式中,获取用户在图像上选定的区域,从图像中剪裁出该区域得到文本图像。示例性的,该图像可为通过相机拍摄的照片,也可为扫描文档(例如,扫描得到的PDF文档),也可以包括屏幕截图(例如,对照片、扫描文档的截图)。

[0073] 在一些可能的实施方式中,文本图像记录的文本包括一个或多个字符,字符可包括汉字、英文字母等。在一些可能的实施方式中,字符可包括手写字符和/或打印字符,例如,文本图像可来自试卷,其内容可包括题干对应的打印字符和解答对应的手写字符。

[0074] 在一些示例中,文本图像包含预设数目的字符。在另一些示例中,文本图像具有预设大小,不同的文本图像可能记录的字符数目不同。在一些可能的实施方式中,在步骤S501中,文本图像上字符的数量对于识别过程而言是未知的,也就是该识别过程可进行字符长度非固定的字符序列的识别。

[0075] 在一些可能的实施方式中,文本图像为预设大小,不同的文本图像具有相同的大小。

[0076] 步骤S502,使用文本识别神经网络的第一特征提取子网络处理文本图像以生成第一特征映射。

[0077] 在步骤S502中,第一特征提取子网络可采用各种类型的神经网络。在一些可能的实施方式中,第一特征提取子网可采用残差神经网络(ResNet)。第一特征提取子网络可包括串联的多个残差块。

[0078] 在步骤S502中,示例性的,第一特征映射可包括多个通道的特征映射,第一特征映射可表示为H*W*C,其中,H为特征映射的高度,W为特征映射的宽度,C为特征映射的通道数。

[0079] 步骤S503,使用文本识别神经网络的第二特征提取子网络在N个方向上从第一特征映射提取特征映射,得到第二特征映射。

[0080] 在步骤S503中,在多个(N个)方向上从第一特征映射提取特征映射,可以获取到文本图像上记录的单行文本中字符的位置的信息,对于带角度的倾斜文本、弯曲文本,可以降低因单行文本中字符的位置不同带来的漏识别字符或多识别字符的情况。直文本、倾斜文本和弯曲文本,其特征可被全方位提取。例如,第一个字符位于水平线上,第二个字符向上偏离,第三个字符向下偏离,第四个字符向下偏离(并可与第三个字符的偏离程度不同),通过在多个方向上从第一特征映射中提取特征映射,可以提取到前述的字符位置变化相关的信息。

[0081] 在一些可能的实施方式中,在步骤S503中,可在预设的一个或多个方向以及随机生成一个或多个方向从第一特征映射提取特征映射,以适应文本图像上单行文本中字符位置的不确定性。在一些示例中,步骤S503中,可在预先设置角度范围(例如30°到60°)内随机生成一个或多个方向。在一些示例中,步骤S503中,可在预先设置多个方向(例如30°方向、45°方向、60°方向等)中随机选择一个或一个以上的方向。

[0082] 在一些可能的实施方式中,在步骤S503中,在第一特征映射的初始方向,以及相对于该初始方向的90°方向、180°方向,从第一特征映射提取特征映射,这些方向便于进行特征映射的操作。可选地,还可在30°方向、45°方向和60°方向中随机选择的一个方向,由此可适度适应单行文本中字符位置的不确定性。

[0083] 在一些可能的实施方式中,第二特征提取子网络包括N个特征提取分支,在上述步

骤S503中，通过N个特征提取分支中的每个特征提取分支，在其对应的方向上从第一特征映射提取特征映射。作为一种示例，每个特征提取分支对应一个方向，但不限于此。

[0084] 在一些可能的实施方式中，第二特征提取子网络还包括整合单元，通过整合单元按照预设整合规则处理N个特征提取分支的输出，以生成第二特征映射。作为一种实施方式，通过整合单元串联拼接N个特征提取分支的输出，以生成第二特征映射。

[0085] 在一些可能的实施方式中，在步骤S503中，按照N-1个方向(例如，90°方向、180°方向和60°方向)旋转第一特征映射，得到相应的N-1个第三特征映射(分别与90°方向、180°方向和60°方向对应)；处理第一特征映射(对应于初始方向)和N-1个第三特征映射以生成第二特征映射。进一步的，通过N-1个特征提取分支中相应的特征提取分支处理相应的第三特征映射，以及通过剩余的一个特征提取分支处理第一特征映射；通过整合单元按照预设整合规则处理N个特征提取分支的输出，以生成第二特征映射。

[0086] 步骤S504，使用文本识别神经网络的输出子网络处理第二特征映射，以生成文本识别输出。

[0087] 在一些可能的实施方式中，步骤S504中，文本识别输出为识别文本的概率矩阵。进一步的，可使用贪心算法或者集束搜索(beam search)算法等，根据识别文本的概率矩阵解码得到识别文本，本公开对此不作限定。

[0088] 在一些可能的实施方式中，输出子网络采用CRNN中的序列化建模层(例如BiLSTM)以及解码层(例如CTC)。在步骤S504中，通过序列化建模层和解码层处理步骤S503得到的第二特征映射，以生成文本识别输出。

[0089] 考虑到CRNN参数量过大，训练周期过长。在一些可能的实施方式中，输出子网络可包括1*1卷积层。在步骤S504中，通过该1*1卷积层对第二特征映射进行降维，以生成文本识别输出。由此降低文本识别神经网络的训练周期，并提高识别速度。

[0090] 在一些实施例中，还可对文本图像进行编码，并将编码得到的信息与上述第二特征映射进行特征融合生成融合输出，处理融合输出以生成文本识别输出。下面结合图6对该实施例进行描述。

[0091] 图6示出了根据本公开示例性实施例的文本识别方法的另一流程图。参考图6所示，该方法包括步骤S601至步骤S606。

[0092] 步骤S601，获取文本图像。

[0093] 步骤S602，使用文本识别神经网络的第一特征提取子网络处理文本图像以生成第一特征映射。

[0094] 步骤S603，使用文本识别神经网络的编码器子网络处理第一特征映射以生成特征向量。

[0095] 在步骤S603中，处理第一特征映射以生成特征向量。对第一特征映射的局域低阶像素值进行归类与分析，从而获得高阶信息。编码器子网络可使用各种类型的编码器。在一些可能的实施方式中，使用基于注意力的编码器，通过使用注意力机制，引导神经网络关注输入图像上要识别的文字所在的区域。

[0096] 在一些可能的实施方式中，编码器子网络包括位置编码模块和串联的多个基于注意力的编码器。在步骤S603中，通过位置编码模块对于第一特征映射中的每个像素点，使用正余弦函数生成一个向量，由第一特征映射中每个像素点对应的向量形成位置编码；通过

串联的多个基于注意力的编码器，根据位置编码处理第一特征映射，以生成特征向量。

[0097] 步骤S604，使用文本识别神经网络的第二特征提取子网络在N个方向上从第一特征映射提取特征映射，得到第二特征映射。

[0098] 步骤S605，使用文本识别神经网络的融合单元按照融合规则将特征向量和第二特征映射进行融合，以生成融合输出。

[0099] 在一些可能的实施方式中，可预先设置融合规则，在步骤605中按照融合规则将步骤S603生成的特征向量和步骤S604提取的第二特征映射进行融合，以生成融合输出。融合输出具有第二特征映射中与多个方向相关的信息，以及特征向量中字符序列的编码相关的信息。

[0100] 在一些可能的实施方式中，融合规则可配置为逐点逐通道相乘。在步骤S605中，按照该融合规则，将特征向量与第二特征映射逐点逐通道相乘，以生成融合输出。通过逐点逐通道相乘，可以减少相邻特征之间的差距。应当理解，本公开并不限于此，基于本公开示例性实施例的其他融合规则也是可以被构想的。

[0101] 在一些示例中，在进行融合之前，还可将特征向量与第二特征映射进行预处理，以使两者的维度适配。在另一些示例中，步骤S603输出的特征向量与步骤S604输出的第二特征映射的维度适配。

[0102] 步骤S606，使用文本识别神经网络的输出子网络处理融合输出，以生成文本识别输出。

[0103] 在一些可能的实施方式中，输出子网络采用CRNN中的序列化建模层以及解码层。在步骤S606中，通过序列化建模层和解码层处理步骤S605得到的融合输出，以生成文本识别输出。

[0104] 考虑到CRNN参数量过大，训练周期过长。在一些可能的实施方式中，输出子网络可包括1*1卷积层。在步骤S606中，通过该1*1卷积层对融合输出进行降维，以生成文本识别输出。由此降低文本识别神经网络的训练周期，并提高识别速度。

[0105] 在一些实施例中，还提取多尺度的图像特征，下面结合图7对该实施方式进行描述。

[0106] 图7示出了根据本公开示例性实施例的多尺度特征提取和处理的示意图，参考图7所示，在步骤S602(或步骤S502)中通过第一特征提取子网络提取多尺度的特征映射，形成包括多尺度的特征映射的第一特征映射。在步骤S603处理包括多个尺度的特征映射的第一特征映射，以生成特征向量。在步骤S604(或步骤S503)中处理包括多个尺度的特征映射的第一特征映射，以生成第二特征映射。

[0107] 参考图7所示，在步骤S602(或步骤S502)中，通过第一特征提取子网络提取多尺度的特性映射，图7中示出为文本图像的宽高1/2、1/4和1/8的特征映射。示例性的，特征映射的通道数可设置为128，也就是每个尺度的特征映射包括128个通道。

[0108] 在步骤S604(或步骤S503)中，将第一特征映射中各个尺度的特征映射缩放为同一尺度，图7中示出为缩放为文本图像的宽高的1/8的特征映射。第二特征提取子网络包括N个(图7中示出为4个)特征提取分支，将缩放后的特征映射作为每个特征提取分支的输入，该输入包括与尺度对应的3组特征映射，每组特征映射包括128个通道。通过第一个特征提取分支在初始方向上进行特征提取，通过第二个特征提取分支在90°方向上进行特征提取，通

过第三个特征提取分支在180°方向上进行特征提取，通过第四个特征提取分支在30°、45°和60°中随机选择一个方向进行特征提取。每个特征提取分支可包括与尺度对应的3个特征提取子分支，每个特征提取子分支处理一个尺度对应的特征映射，第二特征提取网络包括12个特征提取子分支。4个特征提取分支可输出12组特征映射，每组特征映射可包括128个通道。通过整合单元串联拼接12组特征映射，得到包括12组特征映射的第二特征映射。

[0109] 在步骤S603中，将第一特征映射中各个尺度的特征映射缩放为同一尺度，图7中示出为缩放为文本图像的宽高的1/8的特征映射。将缩放后的特征映射串联拼接，形成编码器子网络的输入，该输入包括与尺度对应的3组特征映射，每组特征映射包括128个通道。通过编码器子网络处理该输入以生成特征向量。可选地，特征向量的维度与第二特征映射适配，其一个维度为12(对应于12组)，另一个维度为128(对应于128个通道)，另外两个维度与特征映射的宽高一致(例如，32*256)。

[0110] 在步骤S605中，可将第二特征映射与特征向量逐点逐通道相乘，生成融合输出。参考图7所示，融合输出包括12组特征映射，每组特征映射包括128个通道。在步骤S606中，通过1*1卷积层处理融合输出，以生成识别文本的概率矩阵。

[0111] 本公开示例性实施例还提供了一种文本识别装置。图8示出了根据本公开示例性实施例的文本识别装置的结构框图，如图8所示，文本识别装置包括：获取模块810和识别模块820。获取模块810，用于获取文本图像。识别模块820，与获取模块810相连，用于使用文本识别神经网络处理文本图像，以生成文本识别输出。

[0112] 在一些实施例中，识别模块820用于：使用文本识别神经网络的第一特征提取子网络处理文本图像以生成第一特征映射；使用文本识别神经网络的第二特征提取子网络在多个(表示为N个)方向上从第一特征映射提取特征映射，得到第二特征映射；使用文本识别神经网络的输出子网络处理第二特征映射，以生成文本识别输出。

[0113] 在一些实施例中，识别模块820用于：使用文本识别神经网络的第一特征提取子网络处理文本图像以生成第一特征映射；使用文本识别神经网络的编码器子网络处理第一特征映射以生成特征向量；使用文本识别神经网络的第二特征提取子网络在多个(表示为N个)方向上从第一特征映射提取特征映射，得到第二特征映射；使用文本识别神经网络的融合单元按照融合规则将特征向量和第二特征映射进行融合，以生成融合输出；使用文本识别神经网络的输出子网络处理融合输出，以生成文本识别输出。

[0114] 本公开示例性实施例还提供一种电子设备，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。所述存储器存储有能够被所述至少一个处理器执行的计算机程序，所述计算机程序在被所述至少一个处理器执行时用于使所述电子设备执行根据本公开实施例的方法。

[0115] 本公开示例性实施例还提供一种存储有计算机程序的非瞬时计算机可读存储介质，其中，所述计算机程序在被计算机的处理器执行时用于使所述计算机执行根据本公开实施例的方法。

[0116] 本公开示例性实施例还提供一种计算机程序产品，包括计算机程序，其中，所述计算机程序在被计算机的处理器执行时用于使所述计算机执行根据本公开实施例的方法。

[0117] 参考图9，现将描述可以作为本公开的服务器或客户端的电子设备900的结构框图，其是可以应用于本公开的各方面的硬件设备的示例。电子设备旨在表示各种形式的数

字电子的计算机设备，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。

[0118] 如图9所示，电子设备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。

[0119] 电子设备900中的多个部件连接至I/O接口905，包括：输入单元906、输出单元907、存储单元908以及通信单元909。输入单元906可以是能向电子设备900输入信息的任何类型的设备，输入单元906可以接收输入的数字或字符信息，以及产生与电子设备的用户设置和/或功能控制有关的键信号输入。输出单元907可以是能呈现信息的任何类型的设备，并且可以包括但不限于显示器、扬声器、视频/音频输出终端、振动器和/或打印机。存储单元904可以包括但不限于磁盘、光盘。通信单元909允许电子设备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，并且可以包括但不限于调制解调器、网卡、红外通信设备、无线通信收发机和/或芯片组，例如蓝牙TM设备、WiFi设备、WiMax设备、蜂窝通信设备和/或类似物。

[0120] 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，在一些实施例中，文本识别方法及文本识别装置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/或安装到电子设备900上。在一些实施例中，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)而被配置为执行文本识别方法。

[0121] 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。

[0122] 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。

[0123] 如本公开使用的,术语“机器可读介质”和“计算机可读介质”指的是用于将机器指令和/或数据提供给可编程处理器的任何计算机程序产品、设备、和/或装置(例如,磁盘、光盘、存储器、可编程逻辑装置(PLD)),包括,接收作为机器可读信号的机器指令的机器可读介质。术语“机器可读信号”指的是用于将机器指令和/或数据提供给可编程处理器的任何信号。

[0124] 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。

[0125] 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的实施方式交互)、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。

[0126] 计算机系统可以包括客户端和服务器。客户端和服务器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器的关系。

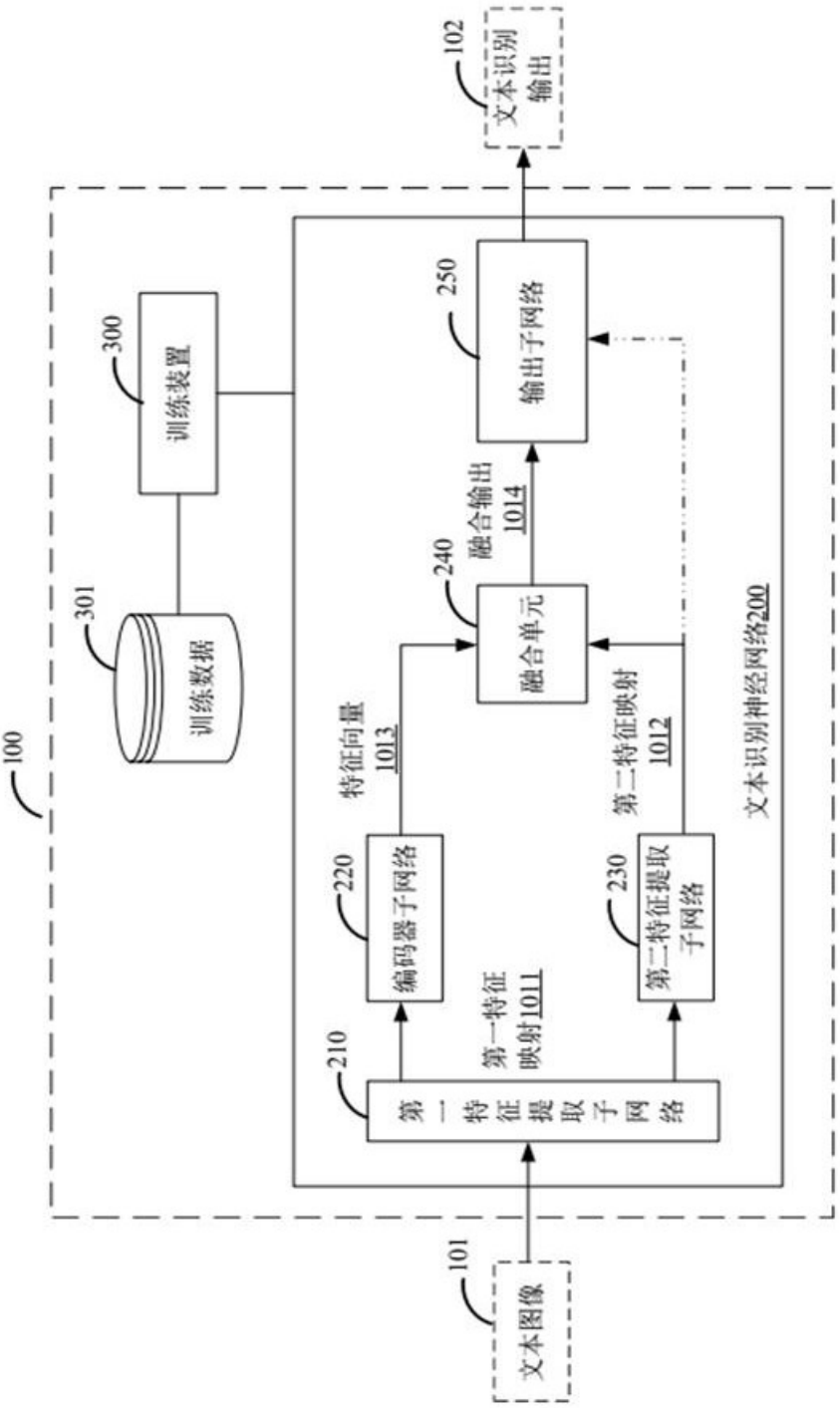

102
文本识别输出
250
输出子网络
300
训练装置
240
融合输出1014
融合单元
301
训练数据
特征向量1013
第二特征映射1012
100
文本识别神经网络200
220
编码器子网络
230
第二特征提取子网络
第一特征映射1011
210
第一特征提取子网络
101
文本图像

图1

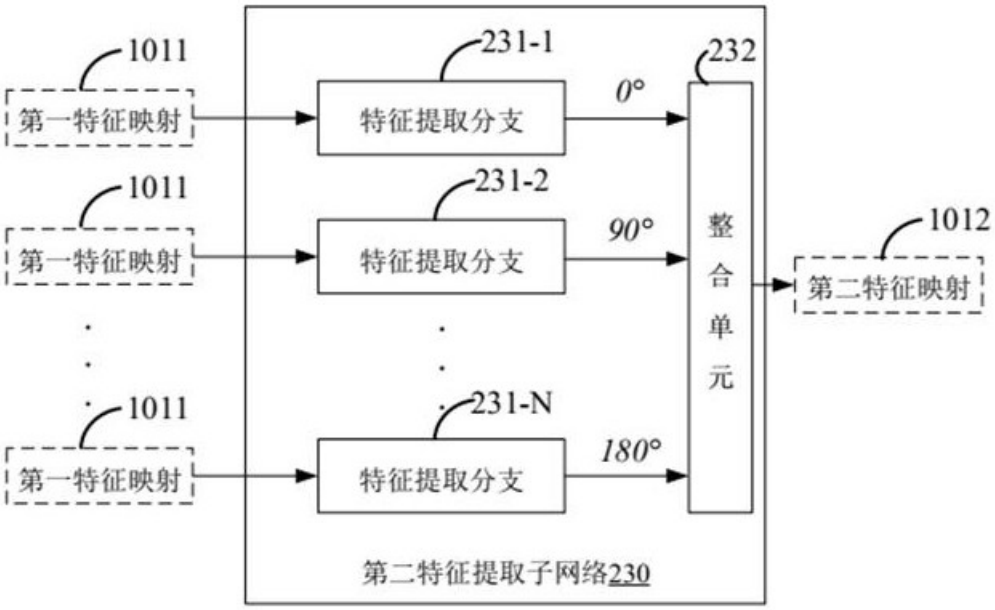

1011
第一特征映射
1011
第一特征映射
1011
第一特征映射
231-1
特征提取分支
231-2
特征提取分支
231-N
特征提取分支
0°
90°
180°
232
整合单元
1012
第二特征映射
第二特征提取子网络230

图2

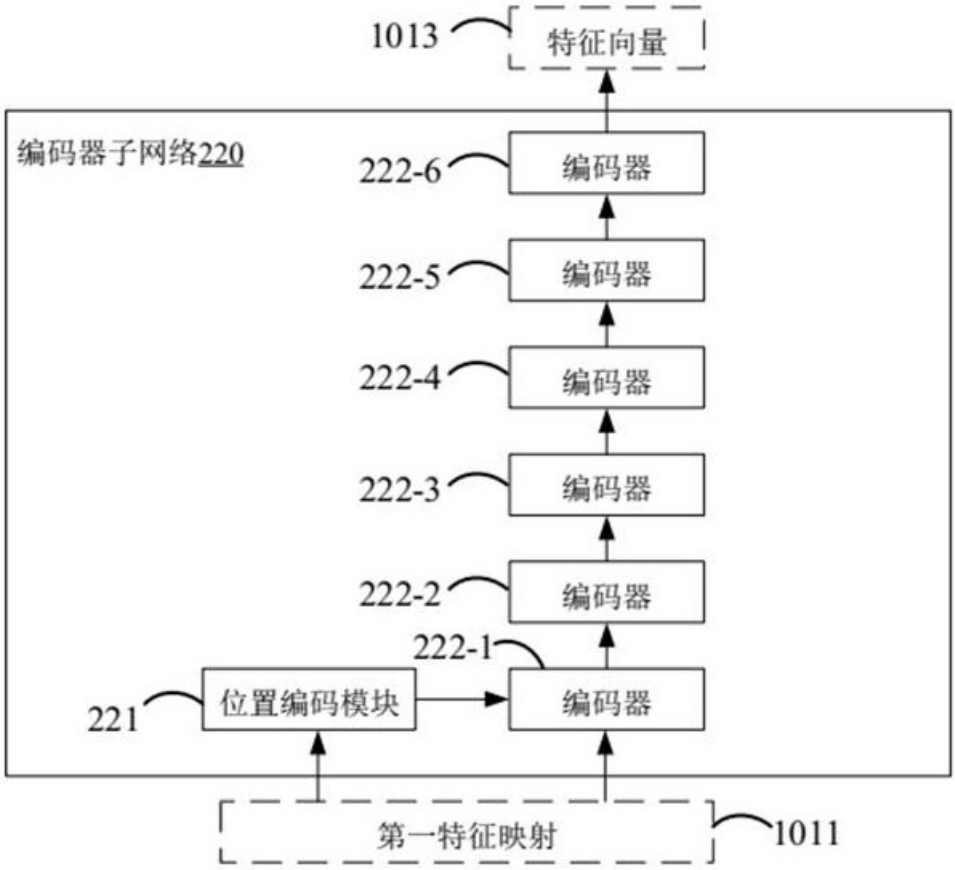

1013
特征向量
编码器子网络220
222-6
编码器
222-5
编码器
222-4
编码器
222-3
编码器
222-2
编码器
222-1
编码器
221
位置编码模块
第一特征映射
1011

图3

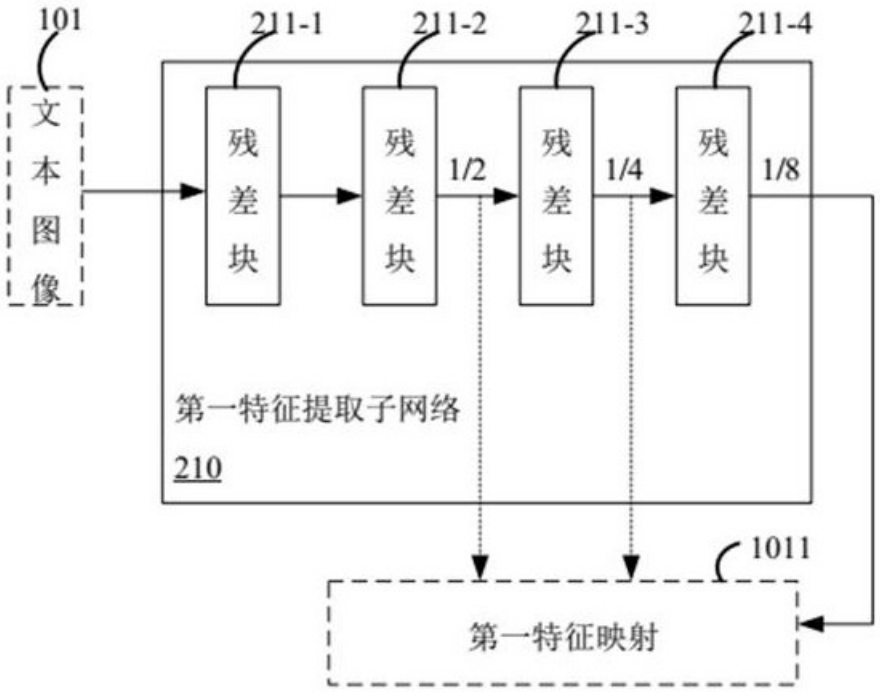

101
文本图像
211-1
211-2
211-3
211-4
残差块
残差块
残差块
残差块
1/2
1/4
1/8
第一特征提取子网络
210
1011
第一特征映射

图4

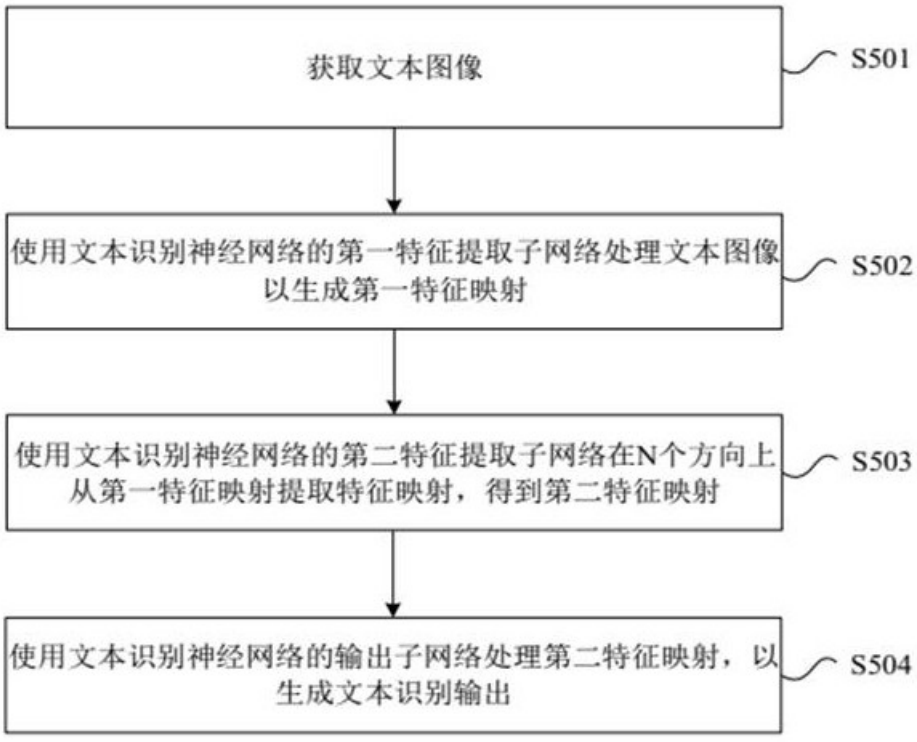

获取文本图像
S501
使用文本识别神经网络的第一特征提取子网络处理文本图像以生成第一特征映射
S502
使用文本识别神经网络的第二特征提取子网络在N个方向上从第一特征映射提取特征映射，得到第二特征映射
S503
使用文本识别神经网络的输出子网络处理第二特征映射，以生成文本识别输出
S504

图5

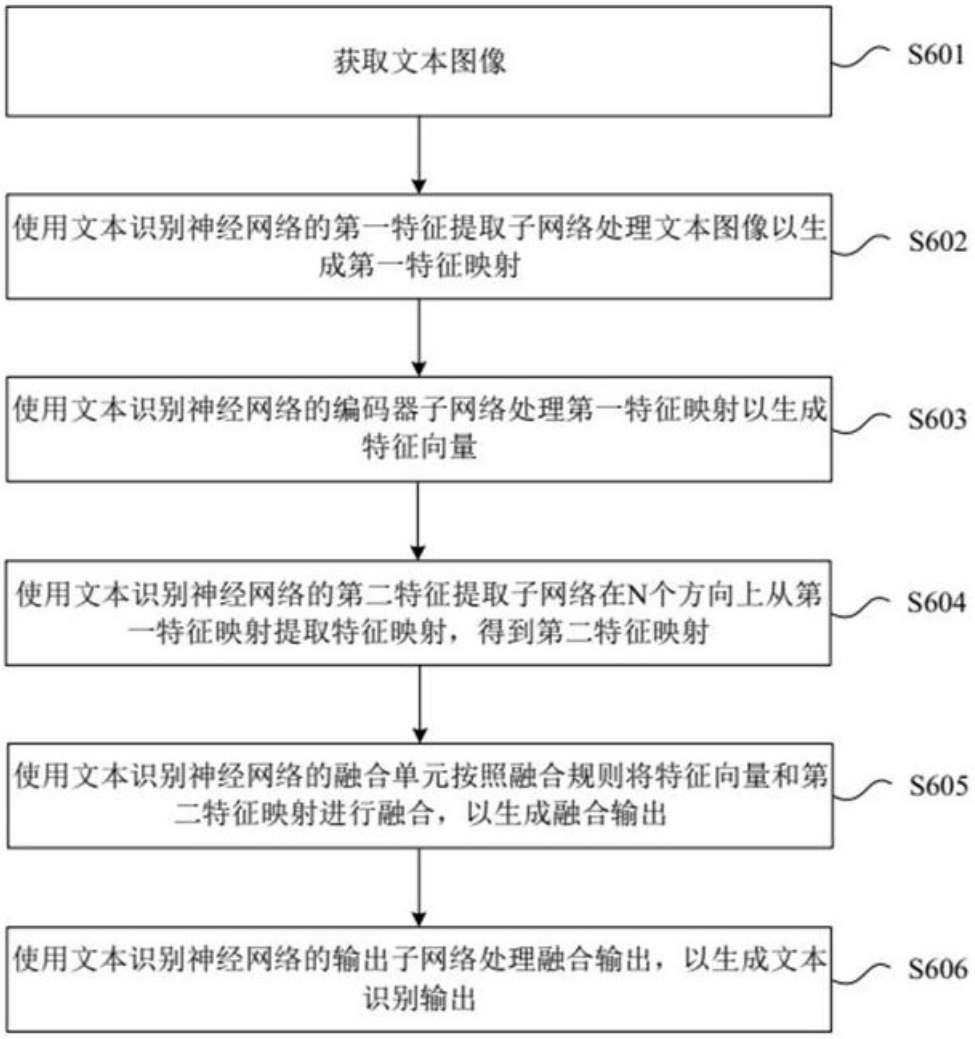

获取文本图像
S601
使用文本识别神经网络的第一特征提取子网络处理文本图像以生成第一特征映射
S602
使用文本识别神经网络的编码器子网络处理第一特征映射以生成特征向量
S603
使用文本识别神经网络的第二特征提取子网络在N个方向上从第一特征映射提取特征映射，得到第二特征映射
S604
使用文本识别神经网络的融合单元按照融合规则将特征向量和第二特征映射进行融合，以生成融合输出
S605
使用文本识别神经网络的输出子网络处理融合输出，以生成文本识别输出
S606

图6

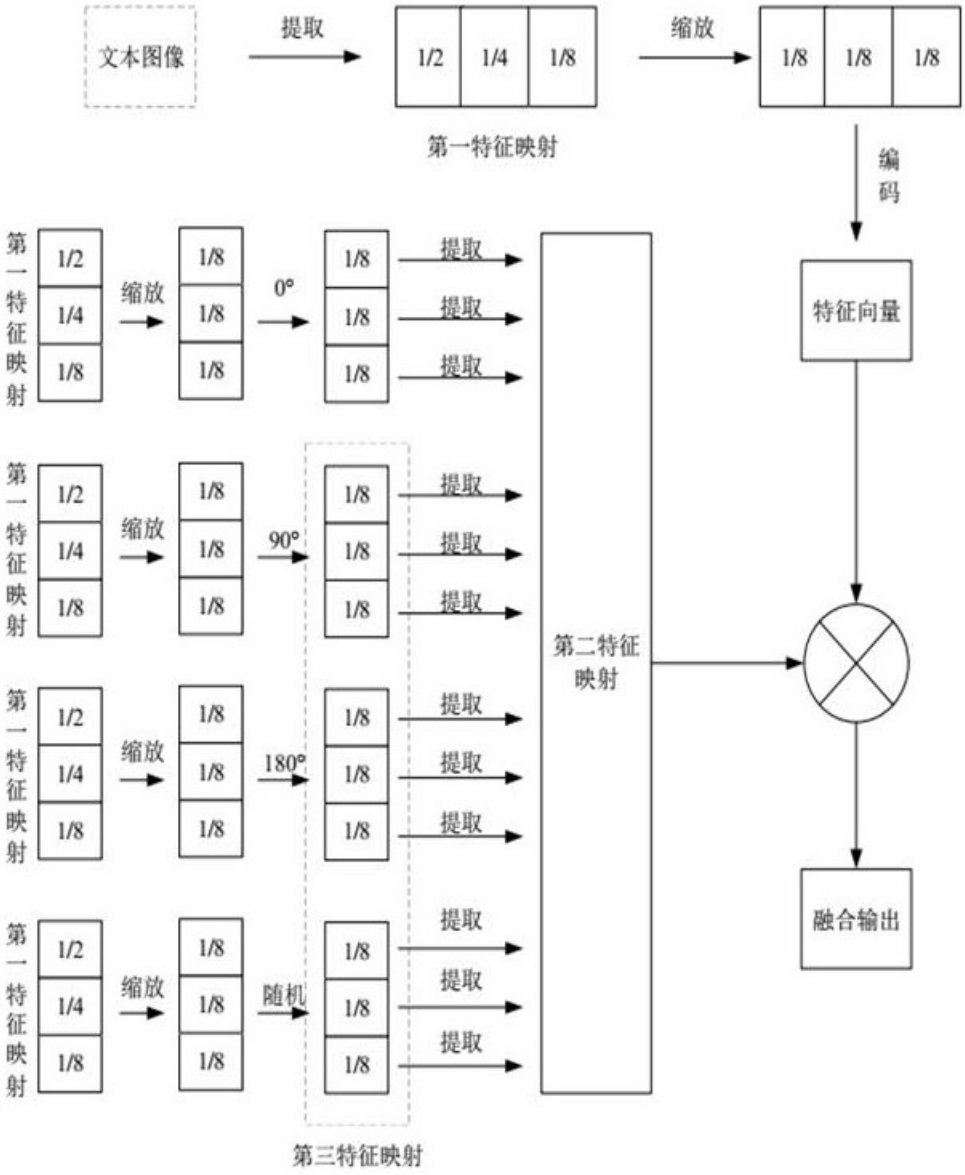

文本图像
提取
1/2
1/4
1/8
缩放
1/8
1/8
1/8
第一特征映射
编码
特征向量
第一特征映射
缩放
0°
90°
180°
随机
提取
第二特征映射
融合输出
第三特征映射

图7


获取模块810
识别模块820

图8

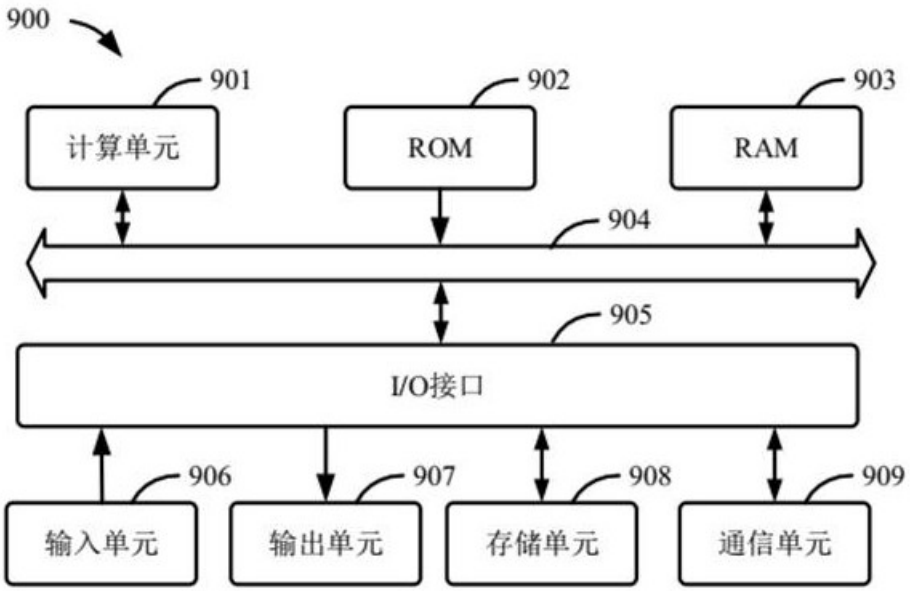

900
901
计算单元
902
ROM
903
RAM
904
905
I/O接口
906
输入单元
907
输出单元
908
存储单元
909
通信单元

图9