

文字图像处理方法、装置、设备及存储介质

相关申请的交叉引用

本申请基于申请号为 201711407535.0、申请日为 2017 年 12 月 22 日的中国专利申请提出，并要求该中国专利申请的优先权，该中国专利申请的
5 全部内容在此引入本申请作为参考。

技术领域

本申请涉及计算机视觉技术，尤其是涉及一种文字图像处理方法、神经网络的训练方法、文字图像处理装置、神经网络的训练装置、电子设备、计算机可读存储介质以及计算机程序。

10 背景技术

文字图像是包含有文字（如汉字或者英文或者其他语种的字符以及数字等）的图像。在有些情况下，图像中的而文字识别率并不能达到预期的准确率。如何提高文字识别的准确率是一个值得关注的技术问题。

15 发明内容

本申请实施方式提供一种文字图像处理技术方案。

根据本申请实施方式的一方面，提供了一种文字图像处理方法，该方法包括：获取包含有待处理文字图像中的文字的至少一个图像块；基于神经网络，获取所述图像块的图像块形态变换信息，所述图像块形态变换信息
20 息用于使所述图像块中的文字方向转变为预定方向；所述神经网络是利用具有形态变换标注信息的图像块样本训练获得的；根据所述图像块形态变换信息对所述待处理文字图像进行形态变换处理；对形态变换之后的待处

理文字图像进行文字识别。

根据本申请实施方式的另一方面，提供了一种神经网络的训练方法，该方法包括：从训练数据集中获取具有形态变换标注信息的图像块样本；基于待训练的神经网络，获取所述图像块样本的图像块样本形态变换信息，
5 所述图像块样本形态变换信息用于使所述图像块样本中的文字方向转变为预定方向；以所述图像块样本的形态变换标注信息为指导信息，对所述待训练的神经网络进行监督学习，以减小所述待训练的神经网络输出的图像块样本形态变换信息与所述形态变换标注信息之间的误差。

根据本申请实施方式的再一方面，提供了一种文字图像处理装置，该
10 装置包括：获取图像块模块，配置为获取包含有待处理文字图像中的文字的至少一个图像块；获取变换信息模块，配置为基于神经网络，获取所述图像块的图像块形态变换信息，所述图像块形态变换信息用于使所述图像块中的文字方向转变为预定方向；所述神经网络是利用具有形态变换标注信息的图像块样本训练获得的；变换处理模块，配置为根据所述图像块形
15 态变换信息对所述待处理文字图像进行形态变换处理；文字识别模块，配置为对形态变换之后的待处理文字图像进行文字识别。

根据本申请实施方式的再一方面，提供了一种神经网络的训练装置，该装置包括：获取样本模块，配置为从训练数据集中获取具有形态变换标注信息的图像块样本；获取样本变换信息模块，配置为基于待训练的神经
20 网络，获取所述图像块样本的图像块样本形态变换信息，所述图像块样本形态变换信息用于使所述图像块样本中的文字方向转变为预定方向；监督学习模块，配置为以所述图像块样本的形态变换标注信息为指导信息，对所述待训练的神经网络进行监督学习，以减小所述待训练的神经网络输出的图像块样本形态变换信息与所述形态变换标注信息之间的误差。

25 根据本申请实施方式的再一方面，提供了一种电子设备，包括：存储

器，用于存储计算机程序；处理器，用于执行所述存储器中存储的计算机程序，且所述计算机程序被执行时，实现本申请任一文字图像处理方法实施方式中的步骤，或者实现本申请任一神经网络的训练方法实施方式中的步骤。

- 5 根据本申请实施方式的再一个方面，提供一种计算机可读存储介质，其上存储有计算机程序，该计算机程序被处理器执行时，实现本申请任一文字图像处理方法实施方式中的步骤，或者，实现本申请任一神经网络的训练方法实施方式中的步骤。

- 根据本申请实施方式的再一个方面，提供一种计算机程序，包括计算机指令，当所述计算机指令在设备的处理器中运行时，实现本申请任一文字图像处理方法实施方式中的步骤，或者，实现本申请任一神经网络的训练方法实施方式中的步骤。

- 基于本申请提供的文字图像处理方法、文字图像处理装置、设备、存储介质以及计算机程序，本申请通过利用神经网络来获得用于使图像块中的文字方向转变为预定方向的图像块形态变换信息，可以方便快捷的根据神经网络输出的图像块形态变换信息对待处理文字图像进行形态变换处理，从而使形态变换处理后的待处理文字图像中的文字方向呈现为预定方向（如水平方向），由此可知，本申请可以有效避免由于文字方向发生倾斜而影响文字图像的文字识别的准确率的现象，最终，本申请提供的文字图像处理的技术方案有利于提高文字识别的准确率。

附图说明

构成说明书的一部分的附图描述了本申请的实施方式，并且连同描述一起用于解释本申请的原理。

- 参照附图，根据下面的详细描述，可以更加清楚地理解本申请，其中：
25 图 1 为本申请实施例提供的文字图像处理方法的流程图；

图 2 为本申请实施例提供的神经网络的一些实施方式的示意图；

图 3 为本申请实施例提供的神经网络的另一些实施方式的示意图；

图 4 为本申请实施例提供的神经网络的训练方法的流程图；

图 5 为本申请实施例提供的形成训练数据的流程图；

5 图 6 为本申请实施例提供的文字图像处理装置的结构示意图；

图 7 为本申请实施例提供的神经网络的训练装置的结构示意图；

图 8 为实现本申请的一示例性设备的框图。

具体实施方式

现在将参照附图来详细描述本申请的各种示例性实施例。应注意到：

10 除非另外具体说明，否则在这些实现方式中阐述的部件和步骤的相对布置、数字表达式和数值不限制本申请的范围。

同时，应当明白，为了便于描述，附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

以下对至少一个示例性实施例的描述实际上仅仅是说明性的，决不作
15 为对本申请及其应用或使用的任何限制。

对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论，但在适当情况下，所述技术、方法和设备应当被视为说明书的一部分。

应当注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一
20 步讨论。

本申请实施例可以应用于终端设备、计算机系统以及服务器等电子设备，其可与众多其它通用或者专用的计算系统环境或者配置一起操作。适于与终端设备、计算机系统以及服务器等电子设备一起使用的众所周知的终端设备、计算系统、环境和/或配置的例子包括但不限于：个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微
25 统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微

处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境，等等。

终端设备、计算机系统以及服务器等电子设备可以在由计算机系统执行的计算机系统可执行指令（诸如程序模块）的一般语境下描述。通常，程序模块可以包括例程、程序、目标程序、组件、逻辑以及数据结构等等，它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施，分布式云计算环境中，任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中，程序模块可以位于包括存储设备的本地或远程计算系统存储介质上。

图 1 为本申请实施例提供的文字图像处理方法的示意性流程图。如图 1 所示，该方法包括：步骤 S100、步骤 S110、步骤 S120 以及步骤 S130。下面对图 1 中的各步骤进行详细说明。

S100、获取包含有待处理文字图像中的文字的至少一个图像块。

在一些实现方式中，本申请中的待处理文字图像可以为基于 RGB（红、绿、蓝）的图像或者 YUV（明亮度、色度、浓度）的图像。本申请中的待处理文字图像可以为呈现静态的图片或者照片等图像，也可以为呈现动态的视频中的视频帧等。该待处理文字图像可以包括由多个文字形成的一个或者多个文字段落，也可以包括不以段落的形式出现的文字，如以表格的形式出现的文字等。本申请中的文字可以包括用于表达特定语意的字符（如汉字或者英文或者其他语种的字符）以及数字等。本申请不限制待处理文字图像中的文字的具体表现形态。

在一些实现方式中，所述待处理文字图像中的文字包括但不限于以下至少之一：

各种类型的字符或数字，例如，汉语言文字、英文单词、拉丁字母等；

各种类型的分隔符、标点符号或数学运算符，例如，逗号、省略号、括号、加号等。

在 S100 中，获取包含文字的至少一个图像块，其中，图像块为长方形或者正方形或其他形状，此外，如果该至少一个图像块为多个图像块，多个图像块中的不同图像块的大小相同或不同，例如，所有图像块具有相同的大小，其中，该大小可以为预设尺寸，或者依赖于待处理文字图像的至少一部分的尺寸。再例如，多个图像块中位于边缘的图像块的尺寸不同于中间的图像块（即不位于边缘的图像块），等等。在一些实施方式中，图像块为具有预定大小的正方形图像块，该图像块的形状以及大小与神经网络对图像块的要求相关。本申请并不限制图像块的形状以及大小等。

可选地，图像块包含整个待处理文字图像或者待处理文字图像中的一部分图像。在一些实现方式中，对待处理文字图像进行文字检测，得到待处理文字图像中的文字区域，并将待处理文字图像中的文字区域划分成至少一个图像块，其中，可以通过多种方式确定待处理文字图像中的文字区域，例如但不限于，将待处理文字图像输入到文字检测神经网络进行处理，得到待处理文字图像中的文字区域的信息。此时，将文字区域划分得到的至少一个图像块称为包含文字的至少一个图像块。

在另一些实现方式中，将待处理文字图像划分成多个图像块，并对多个图像块中的每个图像块进行文字检测处理，得到包含有文字的至少一个图像块，但本公开实施例对此不做限定。

在另一些实施方式中，对待处理文字图像进行切割处理，以获得包含有待处理文字图像中的文字的图像块。其中，该切割处理获得的至少一个图像块的数量为一个或多个，并且切割处理的规则可以预先设定，例如，以最大面积或者最小长宽比或者最大长边或短边等等，本公开实施例对此不做限定。

在一个例子中，至少一个图像块为一个图像块。在这种情况下，可选地，以待处理文字图像的中心位置为中心，以待处理文字图像的短边为边长，从待处理文字图像中切割出一正方形图像块，该正方形图像块为能够从待处理文字图像中切割出的面积最大的图像块。

5 可选地，在上述切割过程中，也可以以待处理文字图像的中心位置为中心，以小于待处理文字图像的短边的长度为边长，从待处理文字图像中切割出一正方形图像块，或者，以介于待处理文字图像的短边与长边的长度之间的长度为长边，以待处理文字图像的短边为短边，从待处理文字图像中切割出一个长方形图像块。

10 在另一个例子中，至少一个图像块为两个图像块。在这种情况下，可选地，以待处理文字图像的两个短边分别为正方形的边长，从待处理文字图像中切割出两个正方形图像块，这两个正方形图像块均为能够从待处理文字图像中切割出的面积最大的图像块。

应理解，上述仅为举例描述，本申请不限制从待处理文字图像中切割
15 出图像块的数量以及切割位置等具体切割方式。

在一些实现方式中，本申请中的图像块包含有整个待处理文字图像。其中，图像块可以为待处理文字图像本身或者是通过对待处理文字图像进行一种或多种处理得到的。例如，对待处理文字图像进行填补处理，以获得包含有整个待处理文字图像的至少一个正方形图像块。

20 在一个例子中，在待处理文字图像的形状为长方形的情况下，可以在该待处理文字图像的短边方向，填补空白内容，从而形成以待处理文字图像的长边为正方形的边长的图像块。或者，在上述填补处理过程中，也可以在该待处理文字图像的短边方向和长边方向分别填补空白内容，从而使形成的正方形图像块的边长比待处理文字图像的长边略长。

25 S110、基于神经网络，获取图像块的图像块形态变换信息。

在一些实现方式中，将图像块直接提供给神经网络，例如形态变换神经网络，并根据该神经网络针对输入的图像块而输出的信息，获得图像块形态变换信息。

在另一些实现方式中，对图像块进行一种或多种处理后提供给神经网络。例如，如果在上述步骤 S100 中，基于待处理文字图像而产生的图像块的大小或形状等并不符合神经网络对输入图像块的要求，则在本步骤 S110 中，先对该图像块的大小或形状等进行调整，使调整后的图像块符合神经网络对输入图像块的要求，然后，再将调整后的图像块提供给神经网络。

图像块的形态变换信息用于使图像块中的文字方向转变为预定方向。

10 在一些实现方式中，神经网络针对输入的图像块所输出的图像块形态变换信息包括：图像块的透视变换信息以及表示 90 度整数倍数的图像块旋转信息中的至少一个。其中，图像块的透视变换信息用于使图像块发生基于透视变换的旋转，其中，可选地，该基于透视变换的旋转角度小于 90 度。图像块的透视变换信息包括但不限于图像块顶点的透视坐标变换信息，例如
15 图像块四个顶点的坐标透视变换信息。表示 90 度整数倍数的图像块旋转信息用于对图像块进行 90 度整数倍数的旋转处理，包括但不限于用于指示进行 90 度整数倍数的旋转处理的旋转角度的指示信息，例如，包括 0、1、2 或者 3 等，其中，0 表示不旋转，1 表示顺时针/逆时针旋转 90 度，2 表示顺时针/逆时针旋转 180 度，3 表示顺时针/逆时针旋转 270 度，等等，本公
20 开实施例对此不做限定。

在一些应用场景中，由于图像的拍摄角度等问题，使得原本标准尺寸的文本，出现了一端大一端小的现象，这种现象会导致识别率低的问题。在本公开实施例中，对图像块进行透视坐标变化，矫正由于拍摄角度造成文字图像中的文字一端大一端小的问题，从而提升文字图像的识别准确率。

25 在还有一些情况下，会出现文字的朝向的改变，例如待处理文字图像

或图像块在输入到神经网络之前进行过旋转处理，在本公开实施例中，通过对图像块进行旋转处理或反向旋转处理，使得文字朝向预定方向，例如将文字行方向旋转到水平方向。例如，旋转角度除了上述 90 度的整数倍数以外，还可以是 60 度、30 度或者任意角度。

5 在一些实现方式中，在神经网络针对图像块输出的图像块形态变换信息包括图像块的透视变换信息以及表示 90 度整数倍数的图像块旋转信息的情况下，可选地，神经网络包括：用于提取图像块的图像特征的共享神经网络、用于输出图像块的透视变换信息的第一分支以及用于输出表示 90 度整数倍数的图像块旋转信息的第二分支；其中的共享神经网络从图像块中
10 提取的图像特征被分别提供给第一分支和第二分支。

在一些实现方式中，图像块形态变换信息还用于确定是否对待处理文字图像或至少一个图像块进行形态变换处理。在一些例子中，图像块形态变换信息还包括用于确定是否对待处理文字图像进行形态变换处理的指示信息，其中，可选地，该指示信息指示图像块中的文字数量是否达到预设
15 要求，或者指示图像块中的文字所占用的面积是否达到预定要求，等等。可选地，该预设要求包括：图像块中的文字的面积或数量或面积所占的比例达到预定比值，其中，该预定比值可以为 0.3 左右或其他数值，其具体数值依赖于实际应用场景，本公开实施例对此不做限定。例如，神经网络针对输入的图像块，还输出指示图像块中的文字的数量或文字所占用的面积
20 是否达到预定要求的指示信息。例如，神经网络输出指示图像块的文字在图像块中占用的面积比例或者文字数量是否达到预设要求的指示符，其中，指示符包括例如但不限于 0 和 1，0 指示未达到预设要求，1 指示达到预设要求，等等。再例如，神经网络针对输入的图像块输出图像块中的文字所占用的面积或数量，或者输出图像块中的文字所占用的面积与该图像块的
25 面积的比值。再例如，神经网络针对输入的图像块输出图像块中的文字所

占用的面积与该图像块的面积的比值达到预定比值的概率值，相应地，在概率值低于预定概率值的情况下，认为图像块中的文字所占用的面积与该图像块的面积的比值未达到预定比值，而在概率值不低于预定概率值的情况下，认为图像块中的文字所占用的面积与该图像块的面积的比值达到预定比值。或者，指示信息也可以以其他方式实现，本公开实施例对此不做限定。

在一些实现方式中，在包含文字的至少一个图像块为多个图像块的情况下，确定每个图像块中的文字在该图像块占用的面积是否达到预设要求，进而确定多个图像块中文字在图像块中占用的面积达到预设要求的图像块的比例。若该比例大于预设比例，则生成指示文字在图像块中所占用的面积达到预定要求的指示信息，并对多个图像块进行旋转处理。在另一些实施方式中，若所述比例不大于所述预设比例，生成指示文字在图像块中所占用的面积达不到预定要求的指示信息。

由于不同类型的文本占用的面积不同，例如，标点符号占用的面积一般小于字符占用的面积，若仅基于单个图像块中文字所占用的面积是否达到预定要求可能不能进行准确的判断，通过统计出待处理文字图像对应的所有图像块中文字占用对应图像块的面积达到预定要求的情况，确定是否对图像块进行形态变换处理，能够进一步提高准确率。

在另一些实现方式中，文字在对应图像块中所占用的面积是否达到预定要求包括：文字在对应图像块中占用的面积达到预设面积阈值，等等，本公开实施例对其具体实现不做限定。

在一些实现方式中，在神经网络针对输入的图像块输出图像块的透视变换信息、表示 90 度整数倍数的图像块旋转信息以及指示图像块中的文字所占用的面积与该图像块的面积的比值是否达到预定比值的指示信息的情况下，神经网络的一个示例如图 2 所示，图 2 中的神经网络包括：配置为

提取输入的图像块的图像特征的共享神经网络 200、用于输出图像块的透视变换信息的第一分支 210、配置为输出表示 90 度整数倍数的图像块旋转信息的第二分支 220 以及用于输出指示图像块中的文字所占用的面积与该图像块的面积的比值是否达到预定比值的指示信息的第三分支 230; 其中的共享神经网络 200 从图像块中提取出的图像特征被分别提供给第一分支 210、
5 第二分支 220 以及第三分支 230。

本公开实施例中的神经网络可以通过多种方式实现。在一些实现方式中, 上述神经网络可为卷积神经网络, 本申请并不限制该卷积神经网络的具体网络结构。例如, 该卷积神经网络包括但不限于: 卷积层、非线性 Relu
10 层、池化层以及全连接层等, 该卷积神经网络所包含的层数越多, 则网络越深。再例如, 该卷积神经网络的网络结构可以采用但不限于 ALEXNet、深度残差网络(Deep Residual Network, ResNet)或 VGGnet(Visual Geometry Group Network, 视觉几何组网络)等神经网络所采用的网络结构。

卷积神经网络的一个示例如图 3 所示, 图 3 中的神经网络包括: 属于
15 共享神经网络的第一卷积层(即图 3 中的 conv1)、第一卷积层的 Relu 层(即图 3 中的 relu1)、第一池化层(即图 3 中的 pool1)、第二卷积层(即图 3 中的 conv2)、第二卷积层的 Relu 层(即图 3 中的 relu2)、第二池化层(即图 3 中的 pool2)、第三卷积层(即图 3 中的 conv3)、第三卷积层的 Relu 层(即图 3 中的 relu3)、第四卷积层(即图 3 中的 conv4)和第四卷积层的
20 Relu 层(即图 3 中的 relu4); 属于第一分支的第五卷积层(即图 3 中的 conv5)、第五卷积层的 Relu 层(即图 3 中的 relu5)、第六全连接层(即图 3 中的 fc6)、第六全连接层的 Relu 层(即图 3 中的 relu6)、第七全连接层(即图 3 中的 fc7)和第七全连接层的 Relu 层(即图 3 中的 relu7); 属于第二分支的第五卷积层(即图 3 中的 conv5_rot)、第五卷积层的 Relu 层(即图 3 中的
25 relu_conv5_rot)、第六全连接层(即图 3 中的 fc6_rot)、第六全连接层的 Relu

层（即图 3 中的 relu_fc6_rot）、第七全连接层（即图 3 中的 fc7_rot）和第七全连接层的 Relu 层（即图 3 中的 relu_fc7_rot）；以及属于第三分支的第五卷积层（即图 3 中的 conv5_cls）、第五卷积层的 Relu 层（即图 3 中的 relu_conv5_cls）、第六全连接层（即图 3 中的 fc6_cls）、第六全连接层的 Relu 层（即图 3 中的 relu_fc6_cls）、第七全连接层（即图 3 中的 fc7_cls）和第七全连接层的 Relu 层（即图 3 中的 relu_fc7_cls）。

本申请实施例中的神经网络是利用包含有多条训练数据的训练数据集成功训练出的神经网络。训练数据集中的训练数据包含有图像块样本以及图像块样本的形态变换标注信息，在一些实现方式中，训练数据还包含指示图像块样本中的文字所占用的面积是否达到预定要求的标注信息。例如，训练数据包含有图像块样本的透视变换标注信息、表示 90 度整数倍数的图像块样本旋转标注信息以及指示图像块样本中的文字所占用的面积与该图像块样本的面积比值是否达到预定比值的标注信息。利用训练数据集对神经网络进行训练的过程可以参见下述针对图 4 的描述。在此不再详细说明。

S120、根据图像块形态变换信息对待处理文字图像进行形态变换处理。

在一些实现方式中，在图像块形态变换信息包括图像块的透视变换信息以及表示 90 度整数倍数的图像块旋转信息，且基于待处理文字图像而产生的包含文字的至少一个图像块为一个图像块的情况下，根据图像块形态变换信息对待处理文字图像进行形态变换处理的一种实现方式包括：

首先，根据表示 90 度整数倍数的图像块旋转信息对待处理文字图像进行 90 度整数倍数的旋转处理（如反向旋转处理），得到旋转处理后的待处理文字图像；

然后，根据图像块的透视变换信息确定出透视变换矩阵。示例性的，图像块的透视变换信息可以包括顶点坐标的位移量，透视变换矩阵是根据

该位移量得到的，透视变换矩阵可以使图像中的各点的位置发生变化。

最后，利用该透视变换矩阵对旋转处理后的待处理文字图像进行透视变换，得到形态变换处理后的文字图像。例如，透视变换后的文字图像即为待处理文字图像对应的形态变换处理后的文字图像。

5 在另一些实现方式中，在图像块形态变换信息包括图像块的透视变换信息以及表示 90 度整数倍数的图像块旋转信息，且基于待处理文字图像而产生的包含文字的至少一个图像块为多个图像块的情况下，根据图像块形态变换信息对待处理文字图像进行形态变换处理的一种实现方式包括：

首先，图像块形态变换信息包括多个图像块中的部分或每一个图像块
10 对应的表示 90 度整数倍数的图像块旋转信息包括概率值，其中，该概率值是神经网络输出的，或者是其他方式得到的，本公开实施例对此不做限定。可选地，该概率值表示图像块进行了相应角度的旋转处理的概率，或者表示图像块进行相应角度的旋转处理的置信度，等等。从多个图像块的图像块旋转信息中确定出概率值最大的表示 90 度整数倍数的图像块旋转信息，
15 例如，第一个图像块的表示顺时针/逆时针旋转 180 度的概率值为 0.6，第二个图像块的表示顺时针/逆时针旋转 90 度的概率值为 0.65，则选取第二个图像块的表示 90 度整数倍数的图像块旋转信息；

然后，根据上述确定出的概率值最大的表示 90 度整数倍数的图像块旋转信息对待处理文字图像进行 90 度整数倍数的旋转处理（如进行相应角度的正向或反向旋转处理）；
20

其次，图像块形态变换信息还包括多个图像块中的部分或每一个图像块对应的图像块的透视变换信息。针对每一个图像块的透视变换信息分别计算出透视变换矩阵，并基于多个图像块对应的透视变换矩阵计算透视变换平均矩阵，其中，该透视变换平均矩阵用于使图像块发生基于透视变换
25 的旋转，且该基于透视变换的旋转角度通常小于 90 度；

最后，利用计算出的透视变换平均矩阵对上述旋转处理后的待处理文字图像进行透视变换，得到待处理文字图像对应的形态变换处理后的文字图像。

在另一些实现方式中，在图像块形态变换信息包括：图像块的透视变换信息、表示 90 度整数倍数的图像块旋转信息以及指示图像块中的文字所占用的面积是否达到预定要求的指示信息，且基于待处理文字图像而产生的包含文字的至少一个图像块为一个图像块的情况下，根据图像块形态变换信息对待处理文字图像进行形态变换处理的一种实现方式包括：

首先，基于指示图像块中的文字所占用的面积是否达到预定要求的指示信息，确定图像块中的文字所占用的面积是否达到预定要求，其中，在一些实现方式中，该指示信息是神经网络输出的或以其他方式得到的，但本公开实施例不限于此。例如，假设指示信息指示图像块中的文字所占用的面积与该图像块的面积的比值是否达到预定比值，判断指示信息是否指示图像块中的文字所占用的面积与该图像块的面积的比值超过预定比值，如果判断结果为图像块中的文字所占用的面积达到预定要求，则根据表示 90 度整数倍数的图像块旋转信息对待处理文字图像进行 90 度整数倍数的旋转处理，然后，根据图像块的透视变换信息确定透视变换矩阵，并利用该透视变换矩阵对旋转处理后的待处理文字图像进行透视变换。可选地，如果判断结果为图像块中的文字所占用的面积未达到预定要求，则不执行根据图像块的透视变换信息和表示 90 度整数倍数的图像块旋转信息对待处理文字图像进行形态变换处理的步骤。

在另一些实现方式中，在图像块形态变换信息包括：图像块的透视变换信息、表示 90 度整数倍数的图像块旋转信息以及指示图像块中的文字所占用的面积是否达到预定要求的指示信息，且基于待处理文字图像而产生的包含文字的至少一个图像块为多个图像块的情况下，图像块形态变换信

息包括多个图像块中所有图像块或部分图像块对应的指示信息。假设图像块形态变换信息包括多个图像块中每个图像块对应的指示信息，根据图像块形态变换信息对待处理文字图像进行形态变换处理的一种实现方式包括：

- 5 首先，判断所有指示图像块中的文字所占用的面积是否达到预定要求的指示信息中，即多个图像块对应的指示信息中，是否存在：指示图像块中的文字所占用的面积达到预定要求的指示信息，即是否存在指示对应的图像块中的文字在所述对应的图像块中所占用的面积达到预定要求的第一指示信息，例如，指示信息指示图像块中的文字所占用的面积与该图像块
10 的面积比值是否达到预定比值，并且该预定比值等于 0.3，则判断神经网络输出的所有指示信息中是否存在：指示图像块中的文字所占用的面积与该图像块的面积的比值超过 0.3 的指示信息；

- 可选地，如果判断结果为存在指示图像块中的文字在该图像块中所占用的面积达到预定要求的指示信息（即第一指示信息），则从神经网络针对
15 每一个图像块输出的表示 90 度整数倍数的图像块旋转信息所对应的概率值中，确定概率值最大的表示 90 度整数倍数的图像块旋转信息。例如，第一个图像块的表示顺时针/逆时针旋转 180 度的概率值为 0.6，第二个图像块的表示顺时针/逆时针旋转 90 度的概率值为 0.65，则确定第二个图像块的表示 90 度整数倍数的图像块旋转信息的概率值最大。此外，基于神经网络针对
20 每一个图像块输出的图像块的透视变换信息计算透视变换矩阵，并基于所有的透视变换矩阵计算透视变换平均矩阵。根据概率值最大的表示 90 度整数倍数的图像块旋转信息对待处理文字图像进行 90 度整数倍数的旋转处理，其中，可选地，在概率值表示进行过相应角度的旋转处理的情况下，可以对待处理文字图像进行相应角度的反向旋转处理，但本公开实施例不
25 限于此；并利用计算出的透视变换平均矩阵对上述旋转处理后的待处理文

字图像进行透视变换，得到待处理文字图像对应的形态变换处理后的文字图像。

可选地，如果判断结果为不存在指示图像块中的文字在该图像块中所占用的面积达到预定要求的第一指示信息，则不执行根据图像块的透视变换信息和表示 90 度整数倍数的图像块旋转信息对待处理文字图像进行形态变换处理的步骤。

S130、对形态变换之后的待处理文字图像进行文字识别。

在一些实现方式中，可以利用 OCR（Optical Character Recognition，光学字符识别）技术等，来对上述形态变换之后的待处理文字图像进行文字识别处理，本申请实施例不限制文字识别的具体实现方式。

可选地，S130 中的文字识别可以包括确定待处理文字图像中的文字内容，或者包括对待处理文字图像进行文本检测并基于文本检测结果确定待处理文字图像中的文字内容，本公开实施例对此不做限定。

在本公开实施例中，通过将图像块的文字方向朝预定方向的调整，可以使得颠倒或倾斜的文字被摆正。针对传真、手持设备拍摄的图像或影印图像中的文字，可能出现倾斜或颠倒的现象。若直接识别倾斜或颠倒的文字，可能会导致错误率低的问题。在本公开实施例中，通过将图像块中的文字的方向进行矫正，使得文本朝向规定的预设方向，从而减少因为文字的朝向问题导致的识别率低的问题，提升图像中文字的识别率。

本申请通过利用神经网络来获得用于使图像块中的文字方向转变为预定方向的图像块形态变换信息，可以方便快捷的根据神经网络输出的图像块形态变换信息对待处理文字图像进行形态变换处理，从而使形态变换处理后的待处理文字图像中的文字方向呈现为预定方向（如水平方向），由此，本申请可以有效避免由于文字方向发生倾斜而影响文字图像的文字识别的准确率的现象。

本申请的文字图像处理的技术方案有利于提高文字识别的准确率。从待处理图像识别出的文字可以用于信息检索、图像分类以及辅助阅读等应用，本申请不限制具体的应用场景。

图 4 为本申请实施例中神经网络进行训练的方法的流程图。如图 4 所示，该方法包括：步骤 S400、步骤 S410 以及步骤 S420。下面对图 4 中的各步骤进行详细说明。

S400、从训练数据集中获取具有形态变换标注信息的图像块样本。

在一些实现方式中，训练数据集中包括多条用于训练神经网络的训练数据，其中，可选地，训练数据包括：图像块样本以及图像块样本的形态变换标注信息等。在本申请实施例中，按照随机读取方式或者按照训练数据排列次序顺序读取方式，每次从训练数据集中读取一条或者多条训练数据。训练数据集中的训练数据的生成方式可以如但不限于下述针对图 5 的描述，在此不再详细说明。

在一些实现方式中，训练数据中的图像块样本的形态变化标注信息包括：图像块样本的透视变换标注信息和/或表示 90 度整数倍数的图像块样本旋转标注信息，例如，包括图像块样本至少一个顶点（如四个顶点）的透视坐标变换标注信息以及表示 90 度整数倍数的图像块样本旋转标注信息。

S410、基于待训练的神经网络，获取图像块样本的图像块样本形态变换信息。

在一些实现方式中，将上述获取到的图像块样本提供给神经网络，从而根据神经网络输出的信息，获得图像块样本形态变换信息。该图像块样本形态变换信息用于使图像块样本中的文字方向转变为预定方向（如水平方向等）。可选地，神经网络包括：共享神经网络、第一分支以及第二分支，或者，神经网络包括：共享神经网络、第一分支、第二分支以及第三分支。或者，该神经网络包括：共享神经网络和第一分支，或者包括：共享神经

网络和第二分支。

S420、以图像块样本的形态变换标注信息为指导信息，对待训练的神经网络进行监督学习，以减小待训练的神经网络输出的图像块样本形态变换信息与形态变换标注信息之间的误差。

5 在一些实现方式中，在训练数据中的图像块样本的形态变化标注信息包括图像块样本的透视变换标注信息和表示 90 度整数倍数的图像块样本旋转标注信息（即神经网络至少包括共享神经网络、第一分支和第二分支）的情况下，可选地，先以图像块样本的透视变换标注信息为指导信息，对神经网络的共享神经网络和第一分支进行监督学习；在第一分支训练完成
10 后，固定共享神经网络和第一分支的网络参数（如权值等），并以表示 90 度整数倍数的图像块样本旋转标注信息为指导信息，对神经网络的第二分支进行监督学习；在第二分支训练完成后，该待训练的神经网络训练完成。训练完成的神经网络可以用于上述待处理文字图像进行形态变换处理的应用中。

15 在另一些实现方式中，在训练数据中的图像块样本的形态变化标注信息包括图像块样本的透视变换标注信息、表示 90 度整数倍数的图像块样本旋转标注信息以及指示图像块样本中的文字所占用的面积是否达到预定要求的标注信息（即神经网络包括共享神经网络、第一分支、第二分支和第三分支）的情况下，可选地，先以图像块样本的透视变换标注信息为指导
20 信息，对神经网络的共享神经网络和第一分支进行监督学习；在第一分支训练完成后，固定共享神经网络和第一分支的网络参数（如权值等），并以表示 90 度整数倍数的图像块样本旋转标注信息为指导信息，对神经网络的第二分支进行监督学习；在第二分支训练完成后，以指示图像块样本中的文字所占用的面积是否达到预定要求的标注信息为指导信息，对神经网络
25 的第三分支进行监督学习；在第三分支训练完成后，该神经网络训练完成。

训练得到的神经网络可以用于上述对待处理文字图像进行形态变换处理的应用中。

需要特别说明的是，在上述可选示例中，对第二分支和第三分支的训练过程的先后顺序没有特殊限定，即可以在第一分支训练完成后，先训练第二分支，再训练第三分支；也可以在第一分支训练完成后，先训练第三分支，再训练第二分支。另外，本申请实施例在对第一分支和第二分支进行训练时，所采用的训练数据中的指示图像块样本中的文字所占用的面积是否达到预定要求的标注信息为第一标注信息，其中，第一标注信息指示图像块样本中的文字所占用的面积达到预定要求，而在对第三分支进行训练时，采用包含有指示图像块样本中的文字所占用的面积达到预定要求的标注信息的训练数据以及包含有指示图像块样本中的文字所占用的面积未达到预定要求的标注信息的训练数据。

可选地，本申请实施例中的用于对神经网络的共享神经网络、第一分支以及第二分支进行训练的训练数据的获得包括：对文字方向为预定方向的文字图像样本中的至少部分区域进行形态变换处理，并根据形态变换处理确定形态变换标注信息，从而利用形态变换标注信息形成训练数据，添加在训练数据集中。图 5 为本申请实施例中形成训练数据集中的训练数据的流程图。如图 5 所示，该方法包括：步骤 S500、步骤 S510 以及步骤 S520。下面对图 5 中的各步骤进行详细说明。

S500、对文字方向为预定方向的文字图像样本中的至少部分区域进行透视变换处理。

在一些实现方式中，文字图像样本中的文字方向为水平方向，例如，文字图像样本包括多个基本呈水平方向的文字段落。本申请实施例中，可选地，可以根据随机产生的透视变换矩阵对整个文字图像样本进行透视变换处理，也可以根据随机产生的透视变换矩阵对文字图像样本的部分区域

进行透视变换处理，例如，进行透视变换处理的区域的边长不超过文字图像样本的长边的五分之一。这样，通过对该区域进行透视变换处理，使透视变换处理后的区域中的文字呈现与水平方向的夹角小于预定度数（如 30 度等）的非水平方向。

- 5 S510、从透视变换处理后的区域中截取文字所占用的面积达到预定要求的图像样本块（例如正方形图像块样本），并对图像样本块进行随机 90 度整数倍数的旋转。

在一些实现方式中，从透视变换处理后的区域中截取的图像样本块中的文字所占用的面积与图像样本块的面积的比值应达到预定比值，例如，
10 比值应超过 0.3。在一个例子中，以透视变换处理后的区域的中心为图像样本块的中心，截取边长小于该区域的边长的正方形图像样本块。在本公开实施例中，按照一定的旋转方向对图像样本块进行随机 90 度整数倍数的旋转，例如，按照顺/逆时针旋转方向对图像样本块进行随机 90 度整数倍数的旋转。另外，可选地，可以根据文字图像样本的二值化矩阵来确定截取的
15 图像样本块中的文字所占用的面积与图像样本块的面积的比值是否达到预定比值，如果没有达到预定比值，可以调整截取位置，以便于使截取的图像样本块中的文字所占用的面积与图像样本块的面积的比值达到预定比值。上述二值化矩阵用于区分文字图像样本中的文字区域和非文字区域。

- 20 S520、根据透视变换处理所采用的透视变换矩阵确定图像样本块顶点的透视坐标变换标注信息，并根据随机 90 度整数倍数的旋转确定表示 90 度整数倍数的图像块旋转标注信息。

在一些实现方式中，根据透视变换处理所采用的透视变换矩阵确定出图像样本块四个顶点的透视坐标变换标注信息，可选地，可以将图像样本块、图像样本块四个顶点的透视坐标变换标注信息以及表示 90 度整数倍数的
25 图像块旋转标注信息作为一条训练数据添加在训练数据集中。或者，也

可以将图像样本块中的文字所占用的面积与图像样本块的面积的比值作为训练数据的一部分。

在一些实现方式中，在对第三分支进行训练的过程中，所使用的所有训练数据包含有图像样本块中的文字所占用的面积与图像样本块的面积的比值达到预定比值的训练数据，也包含有图像样本块中的文字所占用的面积与图像样本块的面积的比值未达到预定比值的训练数据。

可选地，上述训练方法训练得到的神经网络用于对文字图像处理方法中的图像块进行处理，得到图像块的图像块形态变换信息。

图 6 为本申请实施例提供的文字图像处理装置的结构示意图。如图 6 所示，该装置包括：获取图像块模块 600、获取变换信息模块 610、变换处理模块 620 以及文字识别模块 630。在一些实现方式中，该装置还包括：获取占用信息模块 640、获取样本模块 650、获取样本变换信息模块 660、监督学习模块 670 以及标注模块 680。

获取图像块模块 600，配置为获取包含有待处理文字图像中的文字的至少一个图像块。获取图像块模块 600 所执行的具体操作可以参见上述针对 S100 的描述，在此不再重复说明。

获取变换信息模块 610，配置为基于神经网络，获取图像块的图像块形态变换信息。其中，图像块形态变换信息用于使图像块中的文字方向转变为预定方向（如水平方向）；且本申请中的神经网络是下述神经网络的训练装置利用具有形态变换标注信息的图像块样本训练获得的。获取变换信息模块 610 所执行的具体操作可以参见上述针对 S110 的描述，在此不再重复说明。

变换处理模块 620，配置为根据图像块形态变换信息对待处理文字图像进行形态变换处理。在一些实现方式中，在本申请的装置包括获取占用信息模块 640 的情况下，该变换处理模块 620，配置为在指示图像块中的文字

所占用的面积是否达到预定要求的指示信息为文字在对应图像块中所占用的面积达到预定要求的情况下，根据图像块形态变换信息对所述待处理文字图像进行形态变换处理，否则，变换处理模块 620 不执行根据图像块形态变换信息对待处理文字图像进行形态变换处理的步骤。变换处理模块 620 所执行的具体操作可以参见上述 S120 中的相关描述，在此不再重复说明。

文字识别模块 630 配置为对形态变换之后的待处理文字图像进行文字识别。文字识别模块 630 可以利用 OCR 技术等，来对上述形态变换之后的待处理文字图像进行文字识别处理。

获取占用信息模块 640 配置为基于神经网络，获取指示文字在对应图像块中所占用的面积是否达到预定要求的指示信息。获取占用信息模块 640 所执行的具体操作可以参见上述 S120 中的相关描述，在此不再重复说明。

获取样本模块 650 配置为从训练数据集中获取具有形态变换标注信息的图像块样本。获取样本模块 650 所执行的具体操作可以参见上述 S400 中的相关描述，在此不再重复说明。

获取样本变换信息模块 660 用于基于待训练的神经网络，获取图像块样本的图像块样本形态变换信息，其中的图像块样本形态变换信息用于使图像块样本中的文字方向转变为预定方向（如水平方向）。获取样本变换信息模块 660 所执行的具体操作可以参见上述 S410 中的相关描述，在此不再重复说明。

监督学习模块 670 主要用于以图像块样本的形态变换标注信息为指导信息，对待训练的神经网络进行监督学习，以便于减小待训练的神经网络所输出的图像块样本形态变换信息与形态变换标注信息之间的误差。监督学习模块 670 所执行的具体操作可以参见上述 S420 中的相关描述，在此不再重复说明。

标注模块 680 用于对文字方向为预定方向的文字图像样本中的至少部

分区域进行形态变换处理，并根据形态变换处理确定形态变换标注信息。标注模块 680 所执行的具体操作可以参见上述针对图 5 的描述，在此不再重复说明。

5 图 7 为本申请实施例提供的神经网络的训练装置的结构示意图。如图 7 所示，该装置主要包括：获取样本模块 650、获取样本变换信息模块 660 和监督学习模块 670，在一些实现方式中，该装置还包括：标注模块 680。该装置中的各模块所执行的操作可以参见上述文字图像处理装置以及方法实施方式中的描述，在此不再重复说明。

10 图 8 示出了适于实现本申请技术方案的示例性设备 800，设备 800 可以是汽车中配置的控制系统/电子系统、移动终端（例如，智能移动电话等）、个人计算机（PC，例如，台式计算机或者笔记型计算机等）、平板电脑以及服务器等。图 8 中，设备 800 包括一个或者多个处理器、通信部等，所述一个或者多个处理器可以为：一个或者多个中央处理单元（CPU）801，和
15 /或，一个或者多个图像处理器（GPU）813 等，处理器可以根据存储在只读存储器（ROM）802 中的可执行指令或者从存储部分 808 加载到随机访问存储器（RAM）803 中的可执行指令而执行各种适当的动作和处理。通信部 812 可以包括但不限于网卡，所述网卡可以包括但不限于 IB（Infiniband）网卡。处理器可与只读存储器 802 和/或随机访问存储器 803
20 中通信以执行可执行指令，通过总线 804 与通信部 812 相连、并经通信部 812 与其他目标设备通信，从而完成本申请方法实施方式中的相应步骤。

此外，在 RAM 803 中，还可以存储有装置操作所需的各种程序以及数据。CPU801、ROM802 以及 RAM803 通过总线 804 彼此相连。在有 RAM803 的情况下，ROM802 为可选模块。RAM803 存储可执行指令，或在运行时
25 向 ROM802 中写入可执行指令，可执行指令使中央处理单元 801 执行上述物体分割方法所包括的步骤。输入/输出（I/O）接口 805 也连接至总线 804。

通信部 812 可以集成设置,也可以设置为具有多个子模块(例如,多个 IB 网卡),并分别与总线连接。

以下部件连接至 I/O 接口 805: 包括键盘、鼠标等的输入部分 806; 包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分 807; 包括硬盘等的存储部分 808; 以及包括诸如 LAN 卡、调制解调器等
5 等的网络接口卡的通信部分 809。通信部分 809 经由诸如因特网的网络执行通信处理。驱动器 810 也根据需要连接至 I/O 接口 805。可拆卸介质 811, 诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器 810 上,以便于从其上读出的计算机程序根据需要被安装在存储部分 808 中。

10 需要特别说明的是,如图 8 所示的架构仅为一种可选实现方式,在具体实践过程中,可根据实际需要对上述图 8 的部件数量和类型进行选择、删减、增加或替换;在不同功能部件设置上,也可采用分离设置或集成设置等实现方式,例如, GPU 和 CPU 可分离设置,再如理,可将 GPU 集成在 CPU 上,通信部可分离设置,也可集成设置在 CPU 或 GPU 上等。这些
15 可替换的实施方式均落入本申请的保护范围。

特别地,根据本申请的实施方式,下文参考流程图描述的过程可以被实现为计算机软件程序,例如,本申请实施方式包括一种计算机程序产品,其包括有形地包含在机器可读介质上的计算机程序,计算机程序包含用于
20 执行流程图所示的步骤的程序代码,程序代码可包括对应执行本申请提供的步骤对应的指令。

在这样的实施方式中,该计算机程序可以通过通信部分 809 从网络上被下载及安装,和/或从可拆卸介质 811 被安装。在该计算机程序被中央处理单元(CPU) 801 执行时,执行本申请中记载的上述指令。

可能以许多方式来实现本申请的方法和装置、电子设备以及计算机可
25 读存储介质。例如,可通过软件、硬件、固件或者软件、硬件、固件的任

何组合来实现本申请的方法和装置、电子设备以及计算机可读存储介质。
用于方法的步骤的上述顺序仅是为了进行说明，本申请的方法的步骤不限于以上具体描述的顺序，除非以其它方式特别说明。此外，在一些实施方式中，还可将本申请实施为记录在记录介质中的程序，这些程序包括用于
5 实现根据本申请的方法的机器可读指令。因而，本申请还覆盖存储用于执行根据本申请的方法的程序的记录介质。

本申请的描述是为了示例和描述起见而给出的，而并不是无遗漏的或者将本申请限于所公开的形式。很多修改和变化对于本领域的普通技术人员而言是显然的。选择和描述实施方式是为了更好说明本申请的原理和实
10 际应用，并且使本领域的普通技术人员能够理解本申请从而设计适于特定用途的带有各种修改的各种实施方式。