(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111259846 A
(43)申请公布日 2020.06.09

(21)申请号 202010071898.7

(22)申请日 2020.01.21

(71)申请人 第四范式(北京)技术有限公司
地址 100085 北京市海淀区上地东路35号颐泉汇大厦写字楼A座610室

(72)发明人 顾立新 韩景涛 韩锋

(74)专利代理机构 北京铭硕知识产权代理有限公司 11286
代理人 朱志玲 田方

(51)Int.Cl.
G06K 9/00(2006.01)
G06K 9/20(2006.01)
G06K 9/32(2006.01)

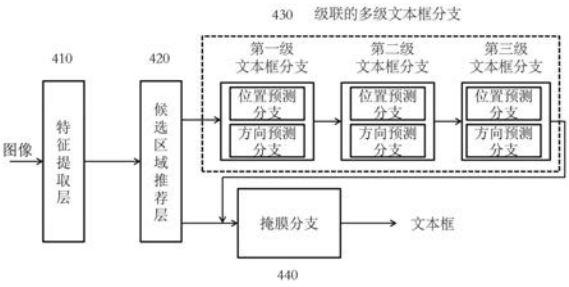
权利要求书2页 说明书21页 附图4页

(54)发明名称

文本定位方法和系统以及文本定位模型训练方法和系统

(57)摘要

提供了一种文本定位方法和系统以及文本定位模型训练方法和系统,其中,所述文本定位方法包括:获取预测图像样本;基于预测图像样本的特征,利用预先训练的文本定位模型,确定用于在预测图像样本中定位文本的文本框的位置并确定所述文本框中的文本的方向,其中,所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。



CN 111259846 A

1.一种在图像中定位文本的方法,包括:

获取预测图像样本;

基于预测图像样本的特征,利用预先训练的文本定位模型,确定用于在预测图像样本中定位文本的文本框的位置并确定所述文本框中的文本的方向,

其中,所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

2.如权利要求1所述的方法,其中,确定所述文本框中的文本的方向包括:

根据预测出的所述文本框的位置以及利用所述文本方向预测分支预测出的与所述文本框的角度有关的值确定所述文本框中的文本的方向。

3.如权利要求2所述的方法,其中,根据预测出的所述文本框的位置以及利用所述文本方向预测分支预测出的与所述文本框的角度有关的值确定所述文本框中的文本的方向包括:

根据预测出的所述文本框的位置以及利用所述文本方向预测分支预测出的与所述文本框的角度有关的值确定所述文本框的四个拐点中的哪个点为起始点,其中,所述起始点能够决定文本的方向,

其中,与所述文本框的角度有关的值包括所述文本框的角度的正弦值和余弦值。

4.一种存储指令的计算机可读存储介质,其中,当所述指令被至少一个计算装置运行时,促使所述至少一个计算装置执行如权利要求1至3中的任一权利要求所述的方法。

5.一种包括至少一个计算装置和存储指令的至少一个存储装置的系统,其中,所述指令在被所述至少一个计算装置运行时,促使所述至少一个计算装置执行如权利要求1至3中的任一权利要求所述的方法。

6.一种在图像中定位文本的系统,包括:

预测图像样本获取装置,被配置为获取预测图像样本;

文本定位装置,被配置为基于预测图像样本的特征,利用预先训练的文本定位模型,确定用于在预测图像样本中定位文本的文本框的位置并确定所述文本框中的文本的方向,

其中,所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

7.一种训练文本定位模型的方法,包括:

获取训练图像样本集,其中,训练图像样本中对文本进行了文本框标记,其中,所述文本框标记包括文本框位置标记和文本框方向标记两者;

基于训练图像样本集训练所述文本定位模型,

其中,所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

8.一种存储指令的计算机可读存储介质,其中,当所述指令被至少一个计算装置运行时,促使所述至少一个计算装置执行如权利要求7所述的方法。

9.一种包括至少一个计算装置和存储指令的至少一个存储装置的系统,其中,所述指令在被所述至少一个计算装置运行时,促使所述至少一个计算装置执行如权利要求7所述的方法。

10.一种训练文本定位模型的系统,包括:

训练图像样本集获取装置，被配置为获取训练图像样本集，其中，训练图像样本中对文本进行了文本框标记，其中，所述文本框标记包括文本框位置标记和文本框方向标记两者；

模型训练装置，被配置为基于训练图像样本集训练所述文本定位模型，

其中，所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

文本定位方法和系统以及文本定位模型训练方法和系统

技术领域

[0001] 本公开总体说来涉及人工智能领域，更具体地，涉及一种在图像中定位文本位置的方法和系统、以及训练文本定位模型的方法和系统。

背景技术

[0002] 图像中的文本蕴含着丰富的信息，提取这些信息(即，文本识别)对图像所处场景的理解等具有重要意义。文本识别分为两个步骤：文本的检测(即，文本的定位)和文本的识别(即，识别文本的内容)，两者缺一不可，而文本检测作为文本识别的前提条件，尤为关键。文本检测需要把图片中文字定位出来，定位出的文本框一般为水平矩形框或者旋转矩形框。定位出文本框后可以根据文本框的位置信息从图像中把对应的文本区域裁剪出来，并把裁剪后的文本区域送给识别网络进行文字识别。但是，根据文本框的位置信息从图像中将对应的文本区域裁剪出来的过程中，非常依赖于文本区域的方向(也可称为文本方向)，然而，现在的文本定位模型通常只会给出文本框位置信息，并不能给出文本方向信息，而且在实际场景中常常会出现因为拍照角度的问题导致图片旋转90、180、270度等，文本定位模型虽然能检测出文本框位置信息，但是因为无法确定文本方向，会导致文本区域裁剪之后送给后续的识别模型后不能识别出文本内容。

发明内容

[0003] 本发明在于至少解决现有文本定位中存在的以上难点，以便在文本定位中既能定位文本位置又能定位文本方向。

[0004] 根据本申请示例性实施例，提供了一种在图像中定位文本的方法，所述方法可包括：获取预测图像样本；基于预测图像样本的特征，利用预先训练的文本定位模型，确定用于在预测图像样本中定位文本的文本框的位置并确定所述文本框中的文本的方向，其中，所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

[0005] 可选地，确定所述文本框中的文本的方向包括：根据预测出的所述文本框的位置以及利用所述文本方向预测分支预测出的与所述文本框的角度有关的值确定所述文本框中的文本的方向。

[0006] 可选地，根据预测出的所述文本框的位置以及利用所述文本方向预测分支预测出的与所述文本框的角度有关的值确定所述文本框中的文本的方向包括：根据预测出的所述文本框的位置以及利用所述文本方向预测分支预测出的与所述文本框的角度有关的值确定所述文本框的四个拐点中的哪个点为起始点，其中，所述起始点能够决定文本的方向，其中，与所述文本框的角度有关的值包括所述文本框的角度的正弦值和余弦值。

[0007] 可选地，根据预测出的文本框的位置以及利用所述文本方向预测分支预测出的与所述文本框的角度有关的值确定所述文本框的四个拐点中的哪个点为起始点包括：根据指示预测出的文本框的位置的四个拐点的坐标值分别计算所述文本框的四条边的正弦值和

余弦值;根据计算出的所述文本框的四条边的正弦值和余弦值以及预测出的与所述文本框的角度有关的值来确定所述文本框的四个拐点中的哪个点为起始点。

[0008] 可选地,根据计算出的所述文本框的四条边的正弦值和余弦值以及预测出的与所述文本框的角度有关的值来确定所述文本框的四个拐点中的哪个点为起始点包括:对预测出的所述文本框的角度的正弦值和余弦值进行处理以使其取值范围与计算出的所述文本框的四条边的正弦值和余弦值的取值范围相同;分别计算处理后的所述文本框的角度的正弦值与四条边中的每条边的正弦值之间的差值以及处理后的所述文本框的角度的余弦值与四条边中的每条边的余弦值之间的差值,将计算出的差值求和,并将最小求和值所对应的边的第一个点作为所述起始点。

[0009] 可选地,所述文本定位模型包括特征提取层、候选区域推荐层、级联的多级文本框分支以及掩膜分支,其中,特征提取层用于提取预测图像样本的特征以生成特征图,候选区域推荐层用于基于生成的特征图在预测图像样本中确定预定数量个候选文本区域,级联的多级文本框分支中的每一级文本框分支包括文本框位置预测分支和文本方向预测分支,并且所述级联的多级文本框分支用于基于特征图中的与每个候选文本区域对应的特征来预测候选水平文本框的位置和与候选水平文本框的角度有关的值,掩膜分支用于基于特征图中与候选水平文本框对应的特征来预测候选水平文本框中的文本的掩膜信息,根据预测出的掩膜信息确定用于在预测图像样本中定位文本的最终文本框。

[0010] 可选地,基于预测图像样本的特征,利用预先训练的文本定位模型,确定用于在预测图像样本中定位文本的文本框的位置并确定所述文本框中的文本的方向的步骤包括:利用特征提取层提取预测图像样本的特征以生成特征图;利用候选区域推荐层基于生成的特征图在预测图像样本中确定预定数量个的候选文本区域;利用级联的多级文本框分支基于特征图中的与每个候选文本区域对应的特征预测初始候选水平文本框,并且通过第一非极大值抑制操作从初始候选水平文本框中筛选出文本框重合度小于第一重合度阈值的水平文本框作为候选水平文本框;利用掩膜分支,基于特征图中与候选水平文本框对应的特征来预测候选水平文本框中的文本的掩膜信息,根据预测出的文本的掩膜信息确定初选文本框,并且通过第二非极大值抑制操作从确定的初选文本框中筛选出文本框重合度小于第二重合度阈值的文本框作为所述最终文本框,其中,第一重合度阈值大于第二重合度阈值;将所述最终文本框的位置确定为用于在预测图像样本中定位文本的文本框的位置,并根据所述最终文本框的位置以及预测出的与候选水平文本框的角度有关的值确定文本的方向。

[0011] 可选地,获取预测图像样本的步骤包括:获取图像,并且对获取的图像进行多尺度缩放来获取与所述图像对应的不同尺寸的多个预测图像样本,其中,所述方法还包括:针对第一尺寸的预测图像样本,在利用所述文本定位模型确定了用于在第一尺寸的预测图像样本中定位文本的的文本框之后从该文本框中选择尺寸大于第一阈值的第一文本框,并且针对第二尺寸的预测图像样本,在利用所述文本定位模型确定了用于在第二尺寸的预测图像样本中定位文本的文本框之后从该文本框中选择尺寸小于第二阈值的第二文本框,其中,第一尺寸小于第二尺寸;利用第三非极大值抑制操作对选择的第一文本框和第二文本框进行筛选,以得到用于在所述图像中定位文本的最终文本框。

[0012] 可选地,所述级联的多级文本框分支是三级文本框分支,其中,利用级联的多级文本框分支基于特征图中的与每个候选文本区域对应的特征预测初始候选水平文本框包括:

利用第一级文本框分支，从特征图中提取与每个候选文本区域对应的特征并预测每个候选文本区域与真实文本区域的位置偏差、每个候选文本区域包括文本的置信度和不包括文本的置信度、以及与每个候选文本区域的角度有关的值，并且根据第一级文本框分支的预测结果确定第一级水平文本框；利用第二级文本框分支，从特征图中提取与第一级水平文本框对应的特征并预测第一级水平文本框与真实文本区域的位置偏差、第一级水平文本框包括文本的置信度和不包括文本的置信度、以及与第一级水平文本框的角度有关的值，并根据第二级文本框分支的预测结果确定第二级水平文本框；利用第三级文本框分支，从特征图中提取与第二级水平文本框对应的特征并预测第二级水平文本框与真实文本区域的位置偏差、第二级水平文本框包括文本的置信度和不包括文本的置信度、以及与第二级水平文本框的角度有关的值，并根据第三级文本框分支的预测结果确定初始候选水平文本框。

[0013] 可选地，利用候选区域推荐层基于生成的特征图在预测图像样本中确定预定数量个的候选文本区域的步骤包括：利用候选区域推荐层基于生成的特征图预测候选文本区域与预先设置的锚点框之间的差异，根据该差异和锚点框确定初始候选文本区域，并利用第四非极大值抑制操作从初始候选文本区域中筛选出所述预定数量个候选文本区域，其中，所述锚点框的宽高比是通过在所述文本定位模型的训练阶段对训练图像样本集中所标记的文本框的宽高比进行统计而确定的。

[0014] 可选地，根据预测出的文本的掩膜信息确定初选文本框包括：根据预测出的文本的掩膜信息确定包含文本的最小外接矩形，并将确定的最小外接矩形作为初选文本框。

[0015] 可选地，所述方法还包括：在所述图像上显示用于在所述图像中定位文本的最终文本框，其中，所述最终文本框包括水平文本框和/或旋转文本框。

[0016] 可选地，所述文本定位模型基于Mask-RCNN框架，特征提取层对应于Mask-RCNN框架中的深度残差网络，候选区域推荐层对应于Mask-RCNN框架中的区域推荐网络RPN层，级联的多级文本框分支中的每一级文本框分支包括Mask-RCNN框架中的RolAlign层和全连接层，掩膜分支包括一系列卷积层。

[0017] 可选地，预测图像样本的特征包括预测图像样本中像素的相关度。

[0018] 根据本申请另一示例性实施例，提供了一种存储指令的计算机可读存储介质，其中，当所述指令被至少一个计算装置运行时，促使所述至少一个计算装置执行如上所述的在图像中定位文本的方法。

[0019] 根据本申请另一示例性实施，提供了一种包括至少一个计算装置和存储指令的至少一个存储装置的系统，其中，所述指令在被所述至少一个计算装置运行时，促使所述至少一个计算装置执行如上所述的在图像中定位文本的方法。

[0020] 根据本申请另一示例性实施例，提供了一种在图像中定位文本的系统，所述系统可包括：预测图像样本获取装置，被配置为获取预测图像样本；文本定位装置，被配置为基于预测图像样本的特征，利用预先训练的文本定位模型，确定用于在预测图像样本中定位文本的文本框的位置并确定所述文本框中的文本的方向，其中，所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

[0021] 根据本申请另一示例性实施例，提供了一种训练文本定位模型的方法，所述方法可包括：获取训练图像样本集，其中，训练图像样本中对文本进行了文本框标记，其中，所述

文本框标记包括文本框位置标记和文本框方向标记两者;基于训练图像样本集训练所述文本定位模型,其中,所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

[0022] 可选地,文本框位置标记通过包含文本的文本框指示,文本框方向标记通过标记包含文本的文本框的起始点来指示,其中,根据所述起始点和文本框位置标记能够确定与文本框有关的角度值。

[0023] 可选地,基于训练图像样本集训练所述文本定位模型包括:将训练图像样本输入所述文本定位模型;将训练图像样本中的文本位置标记作为文本位置预测分支的标记,根据所述起始点和文本框位置标记确定与文本框有关的角度值并将与确定的角度值对应的角度值作为文本方向预测分支的标记;针对每个训练图像样本,通过将文本位置预测分支的输出与文本位置预测分支的标记进行比较以及将文本方向预测分支的输出与文本方向预测分支的标记进行比较来计算文本位置预测分支的预测损失以及文本方向预测分支的预测损失,不断更新文本定位模型的参数来降低预测损失,直至预测损失最小时确定文本定位模型的参数。

[0024] 可选地,与文本框的角度有关的值包括文本框的角度的正弦值和余弦值,其中,根据所述起始点和文本框位置标记确定与文本框有关的角度值并将与确定的角度值对应的角度值作为文本方向预测分支的标记包括:根据所述起始点的坐标值以及在顺时针方向上与所述起始点相邻的文本框拐点的坐标值,确定所述文本框的角度的正弦值和余弦值;将确定的所述文本框的角度的正弦值和余弦值进行处理以使其取值范围满足预定条件,并且将处理后的正弦值和余弦值作为文本方向预测分支的标记。

[0025] 可选地,所述文本定位模型包括特征提取层、候选区域推荐层、级联的多级文本框分支以及掩膜分支,其中,特征提取层用于提取图像的特征以生成特征图,候选区域推荐层用于基于生成的特征图在图像中确定预定数量个候选文本区域,级联的多级文本框分支中的每一级文本框分支包括文本框位置预测分支和文本方向预测分支,并且所述级联的多级文本框分支用于基于特征图中的与每个候选文本区域对应的特征来预测候选水平文本框的位置和与候选水平文本框的角度有关的值,掩膜分支用于基于特征图中与候选水平文本框对应的特征来预测候选水平文本框中的文本的掩膜信息,并根据预测出的掩膜信息确定用于在图像中定位文本的最终文本框。

[0026] 可选地,所述方法还包括:在基于训练图像样本集训练所述文本定位模型之前,对训练图像样本集中的训练图像样本进行尺寸变换和/或透射变换以获得变换后的训练图像样本集,其中,对训练图像样本进行尺寸变换包括:在不保持训练图像样本的原始宽高比的情况下,对训练图像样本进行随机的尺寸变换使得训练图像样本的宽和高在预定范围内;对训练图像样本进行透射变换包括:使训练图像样本中像素的坐标分别绕x轴、y轴和z轴进行随机旋转。

[0027] 可选地,基于训练图像样本集训练所述文本定位模型的步骤包括:将经过变换的训练图像样本输入所述文本定位模型;利用特征提取层提取输入的训练图像样本的特征以生成特征图;利用候选区域推荐层基于生成的特征图在输入的训练图像样本中确定预定数量个的候选文本区域;利用级联的多级文本框分支基于特征图中的与每个候选文本区域对应的特征预测每个候选文本区域的位置与文本框位置标记之间的位置偏差、每个候选文本

区域包括文本的置信度和不包括文本的置信度、以及与每个候选文本区域的角度有关的值，并根据预测的位置偏差、置信度以及与每个候选文本区域的角度有关的值计算与每个候选文本区域对应的文本框预测损失；将所述预定数量个候选文本区域按照其对应的文本框预测损失进行排序，并根据排序结果筛选出文本框预测损失最大的前特定数量个的候选文本区域；利用掩膜分支基于特征图中与筛选出的候选文本区域对应的特征来预测筛选出的候选文本区域中的掩膜信息，并通过比较预测出的掩膜信息与文本的真实掩膜信息来计算掩膜预测损失；通过使文本框预测损失和掩膜预测损失的总和最小来训练文本定位模型。

[0028] 可选地，利用候选区域推荐层基于生成的特征图在输入的训练图像样本中确定预定数量个的候选文本区域包括：利用候选区域推荐层基于生成的特征图预测候选文本区域与预先设置的锚点框之间的差异，根据该差异和锚点框确定初始候选文本区域，并利用非极大值抑制操作从初始候选文本区域中筛选出所述预定数量个候选文本区域。

[0029] 可选地，所述方法还包括：在训练所述文本定位模型之前，统计变换后的训练图像样本集中标记的所有文本框的宽高比，并且根据统计的所有文本框的宽高比设置所述锚点框的宽高比集合。

[0030] 可选地，根据统计的所有文本框的宽高比设置所述锚点框的宽高比集合包括：将统计的所有文本框的宽高比进行排序；根据排序后的宽高比确定所述锚点框的宽高比的上限值和下限值，在上限值和下限值之间等比例地进行插值，并将由上限值和下限值以及通过插值得到的值构成的集合作为所述锚点框的宽高比集合。

[0031] 可选地，根据预测的位置偏差、置信度以及与每个候选文本区域的角度有关的值计算与每个候选文本区域对应的文本框预测损失包括：针对每个候选文本区域，分别根据每一级文本框分支的预测结果和文本框标记来计算每一级文本框分支的文本框预测损失，并通过将各级文本框分支的文本框预测损失求和来确定与每个候选文本区域对应的文本框预测损失，其中，文本框预测损失包括与每个候选文本区域对应的置信度预测损失、位置偏差预测损失以及角度值预测损失，其中，针对每一级文本框分支设置的用于计算每一级文本框分支的文本框预测损失的重叠度阈值彼此不同，并且针对前一级文本框分支设置的重叠度阈值小于针对后一级文本框分支设置的重叠度阈值，其中，重叠度阈值是每一级文本框分支预测出的水平文本框与文本框位置标记之间的重叠度阈值。

[0032] 可选地，所述最终文本框包括水平文本框和/或旋转文本框。

[0033] 可选地，所述文本定位模型基于Mask-RCNN框架，特征提取层对应于Mask-RCNN框架中的深度残差网络，候选区域推荐层对应于Mask-RCNN框架中的区域推荐网络RPN层，级联的多级文本框分支中的每一级文本框分支包括Mask-RCNN框架中的RolAlign层和全连接层，掩膜分支包括一系列卷积层。

[0034] 可选地，图像的特征包括图像中像素的相关度。

[0035] 根据本申请另一示例性实施例，提供了一种存储指令的计算机可读存储介质，其中，当所述指令被至少一个计算装置运行时，促使所述至少一个计算装置执行如上所述的训练文本定位模型的方法。

[0036] 根据本申请另一示例性实施例，提供了一种包括至少一个计算装置和存储指令的至少一个存储装置的系统，其中，所述指令在被所述至少一个计算装置运行时，促使所述至

少一个计算装置执行如上所述的训练文本定位模型的方法。

[0037] 根据本申请另一示例性实施例，提供了一种训练文本定位模型的系统，所述系统可包括：训练图像样本集获取装置，被配置为获取训练图像样本集，其中，训练图像样本中对文本进行了文本框标记，其中，所述文本框标记包括文本框位置标记和文本框方向标记两者；模型训练装置，被配置为基于训练图像样本集训练所述文本定位模型，其中，所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

[0038] 根据本申请示例性实施例，通过使文本定位模型既包括文本框位置预测分支又包括文本方向预测分支，使得在文本定位过程中，既可以定位文本位置，又可以定位文本方向，从而可提供更好的文本定位效果。

附图说明

[0039] 从下面结合附图对本公开实施例的详细描述中，本公开的这些和/或其他方面和优点将变得更加清楚并更容易理解，其中：

[0040] 图1是示出根据本申请示例性实施例的训练文本定位模型的系统的框图；

[0041] 图2是示出根据起始点从图像中裁剪文本区域的过程的示意图；

[0042] 图3是示出根据本申请示例性实施例的文本框标记的示意图；

[0043] 图4是根据本申请示例性实施例的文本定位模型的示意图；

[0044] 图5是示出根据本申请示例性实施例的训练文本定位模型的方法的流程图；

[0045] 图6是示出根据本申请示例性实施例的在图像中定位文本的系统的框图；

[0046] 图7是示出根据本申请示例性实施例的文本框的方向向量的示意图；

[0047] 图8是示出根据本申请示例性实施例的在图像中定位文本的方法的流程图。

具体实施方式

[0048] 为了使本领域技术人员更好地理解本公开，下面结合附图和具体实施方式对本公开的示例性实施例作进一步详细说明。

[0049] 图1是示出根据本申请示例性实施例的训练文本定位模型的系统(在下文中，为描述方便，将其简称为“模型训练系统”)100的框图。

[0050] 如图1所示，模型训练系统100可包括训练图像样本集获取装置110和模型训练装置120。

[0051] 具体地，训练图像样本集获取装置110可获取训练图像样本集。作为示例，训练图像样本集获取装置110可直接从外部获取由其他装置产生的训练图像样本集，或者，训练图像样本集获取装置110可本身执行操作来构建训练图像样本集。例如，训练图像样本集获取装置110可通过手动、半自动或全自动的方式来获取训练图像样本集，并将获取的训练图像样本处理为适当的格式或形式。这里，训练图像样本集获取装置110可通过输入装置(例如，工作站)接收用户手动导入的训练图像样本集，或者训练图像样本集获取装置110可通过全自动的方式从数据源获取训练图像样本集，例如，通过以软件、固件、硬件或其组合实现的定时器机制来系统地请求数据源将训练图像样本集发送给训练图像样本集获取装置110，或者，也可在有人工干预的情况下自动进行训练图像样本集的获取，例如，在接收到特定的

用户输入的情况下请求获取训练图像样本集。当获取到训练图像样本集时，优选地，训练图像样本集获取装置110可将获取的样本集存储在非易失性存储器(例如，数据仓库)中。

[0052] 这里，训练图像样本中对文本进行了文本框标记，具体地，文本框标记可包括文本框位置标记和文本框方向标记两者。例如，文本框位置标记可通过包含文本的文本框指示，换言之，在图像中用文本框标记出了文本位置。根据示例性实施例，文本框方向标记可通过标记包含文本的文本框的起始点来指示。这是因为，在根据文本框的位置信息从图片中把对应的文字区域裁剪出过程中，非常依赖于文本框的起始点，起始点能够决定裁剪后的文字区域方向。

[0053] 下面参照图2对此进行解释。图2是示出根据起始点从图像中裁剪文本区域的过程的示意图。如图2所示，假设左边为定位出来的文本框以及起始点，右边为对应裁剪之后的文本区域。四种情况下文本框位置信息都是一样，但是描述文本框的起始点不一致，而这会导致裁剪后的文本区域相差非常大。在图2的示例中，将裁剪后的文本区域送给识别网络时，只有第一种情况的文本区域内容会被正确识别，而其他三种情况识别网络均无法识别。因此，根据本发明示例性实施例，可通过标注文本框的起始点作为文本方向标记。而根据所述起始点和文本框位置标记能够确定与文本框有关的角度值，而确定的与文本框有关的角度值可用于确定文本方向。可选地，也可以直接标注与文本框有关的角度值作为文本方向标记，但是角度的标注一般较为困难。因此，在本发明中，例如，在训练过程中，为了不增加标注的复杂度，可只标注文本框的起始点，然后，可根据标注的文本框的起始点(文本框的第一个拐点)以及按照顺时针方向找到的文本框的第二个拐点计算出与文本框有关的角度值，计算出的角度值可以作为文本框方向的真实标记。可选地，在数据标注时，也可以默认从文字方向的左上角的拐点开始标注，并将该左上角的拐点作为文本框的起始点，然后可进一步顺时针标注其他三个拐点。随后，根据标注的文本框第一个拐点和第二个拐点计算出与文本框有关的角度值。作为示例，与文本框的角度有关的值包括文本框的角度的正弦值和余弦值，但不限于此。

[0054] 图3是示出根据本申请示例性实施例的文本框标记的示意图。参照图3，例如，可将文字方向的左上角的拐点标注为文本框的起始点(在图3中，灰色标记的点即为起始点)，此外，还可顺时针标注其他拐点。模型训练装置120可基于训练图像样本集训练文本定位模型。这里，文本定位模型可包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。此外，作为示例，文本定位模型可以是基于深度神经网络的文本定位模型，而深度神经网络可以是卷积神经网络，但不限于此。

[0055] 具体地，模型训练装置120可首先将训练图像样本输入所述文本定位模型，然后，将训练图像样本中的文本位置标记作为文本位置预测分支的标记，根据所述起始点和文本框位置标记确定与文本框有关的角度值并将与确定的角度值对应的角度值作为文本方向预测分支的标记。这里，根据文本框位置标记可以获得文本框的包括起始点在内的每个拐点的坐标值信息。如上所述，与文本框有关的角度值可以是文本框的角度的正弦值和余弦值。在起始点已经被标记出的情况下，例如，可以根据所述起始点的坐标值以及在顺时针方向上与所述起始点相邻的文本框拐点的坐标值，确定所述文本框的角度的正弦值和余弦值。参照图3，假设标注的起始点的坐标是(x1，y1)，顺时针方向上与该起始点相邻的文本框

拐点的坐标值是(x2,y2)，则可计算得到文本框的方向向量为(x2-x1,y2-y1)，并进一步可计算出文本框的角度的余弦值(即，cos值)为$(x2-x1)/\sqrt{(x2-x1)^2+(y2-y1)^2}$，正弦值(即，sin值)为$(y2-y1)/\sqrt{(x2-x1)^2+(y2-y1)^2}$。随后，模型训练装置120可将确定的所述文本框的角度的正弦值和余弦值进行处理以使其取值范围满足预定条件，并且将处理后的正弦值和余弦值作为文本方向预测分支的标记。例如，如果根据本发明示例性实施例的文本定位模型是基于深度神经网络的文本定位模型，则因为上面计算出的cos和sin值为-1到1之间，而深度神经网络模型一般会把值压缩到0到1之间，所以需要对cos值和sin值进行处理以使其取值范围可以在0到1之间。例如，可对cos值进行处理，使得处理后的cos值为$\left(\left((x2-x1)/\sqrt{(x2-x1)^2+(y2-y1)^2}\right)+1\right)/2$，并且可对sin值进行处理，使得处理后是sin值为$\left(\left((y2-y1)/\sqrt{(x2-x1)^2+(y2-y1)^2}\right)+1\right)/2$。由此，便得到了文本方向预测分支的标记。

[0056] 在得到了文本位置预测分支的标记以及文本方向预测分支的标记之后，模型训练装置120可针对每个训练图像样本，通过将文本位置预测分支的输出与文本位置预测分支的标记进行比较以及将文本方向预测分支的输出与文本方向预测分支的标记进行比较来计算文本位置预测分支的预测损失以及文本方向预测分支的预测损失，不断更新文本定位模型的参数来降低预测损失，直至预测损失最小时确定文本定位模型的参数。

[0057] 图4是根据本申请示例性实施例的文本定位模型的示意图。如图4所示，文本定位模型可包括特征提取层410、候选区域推荐层420、级联的多级文本框分支430(为方便示意，图4中将多级文本框分支示意为包括三级文本框分支，但这仅是示例，级联的多级文本框分支不限于仅包括三级文本框分支)以及掩膜分支440。具体地，特征提取层用于提取图像的特征以生成特征图，候选区域推荐层用于基于生成的特征图在图像中确定预定数量个候选文本区域，级联的多级文本框分支中的每一级文本框分支包括文本框位置预测分支和文本方向预测分支，并且所述级联的多级文本框分支用于基于特征图中的与每个候选文本区域对应的特征来预测候选水平文本框的位置和与候选水平文本框的角度有关的值，掩膜分支用于基于特征图中与候选水平文本框对应的特征来预测候选水平文本框中的文本的掩膜信息，并根据预测出的掩膜信息确定用于在图像中定位文本的最终文本框。这里，所述最终文本框可包括水平文本框和/或旋转文本框。也就是说，本申请的文本定位模型既可以定位水平文本，也可定位旋转文本。

[0058] 作为示例，图4的文本定位模型可基于Mask-RCNN框架，此时，特征提取层可对应于Mask-RCNN框架中的深度残差网络(例如，resnet 101)，候选区域推荐层可对应于Mask-RCNN框架中的区域推荐网络RPN层，级联的多级文本框分支中的每一级文本框分支可包括Mask-RCNN框架中的RolAlign层和全连接层，掩膜分支包括一系列卷积层。本领域技术人员均清楚Mask-RCNN框架中的深度残差网络、RPN层、RolAlign层和全连接层的功能和操作，因此，这里不对其进行详细介绍。此外，需要说明的是，本文的文本定位模型不限于基于Mask-RCNN框架的文本定位模型，并且本文提到的文本方向预测分支同样适用于其他的文本定位模型，比如east(Efficient and Accurate Scene Text Detector)算法，fots(Fast Oriented Text Spotting with a Unified Network)算法等。

[0059] 本领域技术人员均了解，传统的Mask-RCNN框架只包括一个文本框分支，而且在RPN层确定了预定数量个候选区域(例如，2000个)之后，从这些候选区域中随机抽样一些候选区域(例如，512个)，并将抽样的候选区域分别送给文本框分支和掩膜分支。然而，这样的结构以及随机抽样候选区域分别送给文本框分支和掩膜分支的操作导致传统Mask-RCNN框架的文本定位效果较差。这是因为，一级文本框分支仅能检测与真实文本框标记的重叠度在一定范围内的候选区域，而随机抽样不利于模型对难样本的学习，比如，如果2000个候选区域存在大量简单样本，较少难样本，则随机抽样会较大概率把一些简单样本送给文本框分支和掩膜分支，从而导致模型学习效果较差。针对此，本发明包括多级文本框分支并且将多级文本框分支点的输出作为掩膜分支的输入，可有效地提高文本定位效果。

[0060] 下面，将对本发明的文本定位模型的训练进行详细描述。自然场景中由于图像拍摄角度不一，会存在文本变形的可能，并且可能存在平面旋转和三维立体旋转，因此，模型训练装置120可在基于训练图像样本集训练所述文本定位模型之前，对训练图像样本集中的训练图像样本进行尺寸变换和/或透射变换以获得变换后的训练图像样本集，从而使得训练图像样本更切近真实场景。具体而言，模型训练装置120可在不保持训练图像样本的原始宽高比的情况下，对训练图像样本进行随机的尺寸变换使得训练图像样本的宽和高在预定范围内。这里，之所以不保持训练图像样本的原始宽高比就是为了模拟真实场景中的压缩和拉伸。例如，可将训练图像样本的宽和高随机变换到640至2560个像素之间，但是预定范围不限于此。此外，对训练图像样本进行透射变换可以包括使训练图像样本中像素的坐标分别绕x轴、y轴和z轴进行随机旋转。例如，可以将训练图像样本中的每个像素绕x轴随机旋转(-45,45)，绕y轴随机旋转(-45,45)，绕z轴随机旋转(-30,30)，增强后的训练图像样本将更加符合真实场景。例如，可通过下面的等式对文本框坐标进行变换：

[0061] $$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

[0062] 其中，

[0063] $$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ \cos\theta_x & -\sin\theta_x & 0 \\ \sin\theta_x & \cos\theta_x & 0 \end{bmatrix} \begin{bmatrix} \sin\theta_y & \cos\theta_y & 0 \\ 0 & 0 & 1 \\ \cos\theta_y & -\sin\theta_y & 0 \end{bmatrix}$$ 为透射变换矩阵，θ_x为绕x轴随机旋转(-45,45)，θ_y为绕y轴随机旋转(-45,45)，θ_z为绕z轴随机旋转(-30,30)得到，$\begin{bmatrix} x \\ y \\ z \end{bmatrix}$为变换前的坐标，通常z的取值为1，$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}$为变换后的坐标，变换后的文本框坐标可表示为$x=x'/z'$，$y=y'/z'$。

[0064] 在对训练图像样本集进行变换之后，模型训练装置120可基于变换后的训练图像样本集训练上述文本定位模型。具体地，模型训练装置120可以进行以下操作来训练上述文本定位模型：将经过变换的训练图像样本输入所述文本定位模型；利用特征提取层提取输

入的训练图像样本的特征以生成特征图;利用候选区域推荐层基于生成的特征图在输入的训练图像样本中确定预定数量个的候选文本区域;利用级联的多级文本框分支基于特征图中的与每个候选文本区域对应的特征预测每个候选文本区域的位置与文本框位置标记之间的位置偏差、每个候选文本区域包括文本的置信度和不包括文本的置信度、以及与每个候选文本区域的角度有关的值,并根据预测的位置偏差、置信度以及与每个候选文本区域的角度有关的值计算与每个候选文本区域对应的文本框预测损失;将所述预定数量个候选文本区域按照其对应的文本框预测损失进行排序,并根据排序结果筛选出文本框预测损失最大的前特定数量个的候选文本区域;利用掩膜分支基于特征图中与筛选出的候选文本区域对应的特征来预测筛选出的候选文本区域中的掩膜信息,并通过比较预测出的掩膜信息与文本的真实掩膜信息来计算掩膜预测损失;通过使文本框预测损失和掩膜预测损失的总和最小来训练文本定位模型。

[0065] 作为示例,图像的特征可以包括图像中像素的相关度,但不限于此。模型训练装置120可利用特征提取层提取训练图像样本中像素的相关度来生成特征图。随后,模型训练装置120可利用候选区域推荐层基于生成的特征图预测候选文本区域与预先设置的锚点框之间的差异,根据该差异和锚点框确定初始候选文本区域,并利用非极大值抑制操作从初始候选文本区域中筛选出所述预定数量个候选文本区域。这里,由于预测出的初始候选文本区域可能会存在彼此重叠的现象,因此,本申请利用非极大值抑制操作来对初始候选文本区域进行筛选。下面,简要地对非极大值抑制操作进行描述。具体地,可从与锚点框的差异最小的初始候选文本区域开始,分别判断其他初始候选文本框与该初始候选文本区域的重叠度是否大于某个设定的阈值,如果存在大于该阈值的初始候选文本区域则将其去除,也就是说,保留重叠度小于该阈值的初始候选文本区域。然后,再在所有保留下来的初始候选文本区域之中再选择一个与锚点框的差异最小的初始候选文本区域,并继续判断该初始候选文本区域与其他初始候选文本区域的重叠度,如果重叠度大于阈值则删除,否则保留,直至筛选出预定数量个候选文本区域。

[0066] 这里,预先设置的锚点框是预先设置的图像中每个可能的文本框,以用于与真实文本框进行匹配。传统的基于Mask-RCNN框架的模型的锚点的宽高比集合是固定的,该集合为[0.5,1,2],也就是说,锚点的宽高比仅有0.5、1和2这三种。利用这三种宽高比的锚点在一些通用的目标检测数据集(例如,coco数据集)上基本能够覆盖目标,但是,在文本场景中确远远不足以覆盖文本。这是因为,文本场景中宽高比范围很大,1:5,5:1的文本很常见,如果用传统Mask-RCNN的仅具有三种固定宽高比的锚点框会导致锚点框和真实的文本框匹配不上,从而导致文本漏检。因此,根据本申请示例性实施例,模型训练装置120还可在训练所述文本定位模型之前,统计变换后的训练图像样本集中标记的所有文本框的宽高比,并且根据统计的所有文本框的宽高比设置所述锚点框的宽高比集合。也就是说,本发明可对锚点框的宽高比进行重新设计。具体地,例如,在统计了变换后的训练图像样本集中标记的所有文本框的宽高比之后,可将统计的所有文本框的宽高比进行排序,根据排序后的宽高比确定锚点框的宽高比的上限值和下限值,在上限值和下限值之间等比例地进行插值,并将由上限值和下限值以及通过插值得到的值构成的集合作为所述锚点框的宽高比集合。例如,可以将所有文本框的宽高比由小到大排序后处于第5%的宽高比和处于第95%的宽高比分别确定为锚点框的宽高比的下限值和上限值,然后在上限值和下限值之间等比例地进

行三次插值来得到另外三个宽高比，并将由上限值和下限值以及通过插值得到的三个值构成的集合作为锚点框的宽高比集合。然而，以上确定锚点框的宽高比集合的方式仅是示例，上限值和下限值的选取方式以及插值的方式和次数均不限于以上示例。通过根据以上方式设计锚点框的宽高比集合，可以有效地减少文本框的漏检。

[0067] 如上所述，在确定了预定数量个候选文本区域之后，模型训练装置120可利用级联的多级文本框分支基于特征图中的与每个候选文本区域对应的特征预测每个候选文本区域的位置与文本框位置标记之间的位置偏差、每个候选文本区域包括文本的置信度和不包括文本的置信度、以及与每个候选文本区域的角度有关的值，并根据预测的位置偏差、置信度以及与每个候选文本区域的角度有关的值计算与每个候选文本区域对应的文本框预测损失。作为示例，如图4所示，所述级联的多级文本框分支可以是三级文本框分支，但不限于此。在每一级文本框分支中均包括文本框位置预测分支以及文本方向预测分支，其中，文本框位置预测分支的输出包括位置偏差，而文本方向预测分支的输出包括角度值。

[0068] 另外，如上所述，本发明提出了难样本学习机制，也就是说，将所述预定数量个候选文本区域按照其对应的文本框预测损失进行排序，根据排序结果筛选出文本框预测损失最大的前特定数量个的候选文本区域，并将筛选出的候选文本区域输入掩膜分支进行掩膜信息预测。例如，可根据文本框预测损失从2000个候选区域中选出文本框预测损失较大的512个候选文本区域。为此，模型训练装置120可根据预测的位置偏差、置信度以及与每个候选文本区域的角度有关的值计算与每个候选文本区域对应的文本框预测损失。具体而言，例如，针对每个候选文本区域，模型训练装置120可分别根据每一级文本框分支的预测结果和文本框标记来计算每一级文本框分支的文本框预测损失，并通过将各级文本框分支的文本框预测损失求和来确定与每个候选文本区域对应的文本框预测损失。这里，文本框预测损失包括与每个候选文本区域对应的置信度预测损失、位置偏差预测损失以及角度值预测损失。此外，针对每一级文本框分支设置的用于计算每一级文本框分支的文本框预测损失的重叠度阈值彼此不同，并且针对前一级文本框分支设置的重叠度阈值小于针对后一级文本框分支设置的重叠度阈值。这里，重叠度阈值是每一级文本框分支预测出的水平文本框与文本框标记之间的重叠度阈值。重叠度(IOU)可以是两个文本框之间的交集除以两个文本框的并集所获得的值。例如，在所述多级文本框分支是三级文本框分支的情况下，针对第一级文本框分支至第三级文本框分支设置的重叠度阈值可以分别是0.5、0.6和0.7。具体地，例如，在计算第一级文本框预测损失时，如果针对候选文本区域预测出的水平文本框与训练图像样本中的文本框标记之间的重叠度阈值大于0.5，则该候选文本区域被确定为是针对第一级文本框分支的正样本，小于0.5则被确定为是负样本。但是当阈值取0.5时会有较多的误检，因为0.5的阈值会使得正样本中有较多的背景，这是较多文本位置误检的原因。如果用0.7的重叠度阈值，则可以减少误检，但检测效果不一定最好，主要原因在于重叠度阈值越高，正样本的数量就越少，因此过拟合的风险就越大。然而，根据本发明示例性实施例，由于采取级联的多级文本框分支，并且针对每一级文本框分支设置的用于计算每一级文本框分支的文本框预测损失的重叠度阈值彼此不同，而且针对前一级文本框分支设置的重叠度阈值小于针对后一级文本框分支设置的重叠度阈值，因此能够让每一级文本框分支都专注于定位与真实文本框标记重叠度在某一范围内的候选文本区域，因此文本定位效果会越来越好。

[0069] 在筛选出文本框预测损失较大的候选文本区域之后，模型训练装置120可利用掩膜分支基于特征图中与筛选出的候选文本区域对应的特征来预测筛选出的候选文本区域中的掩膜信息(具体地，可将预测为文本的像素的掩膜设置为1，不是文本的像素的掩膜设置为0)，并通过比较预测出的掩膜信息与文本的真实掩膜信息来计算掩膜预测损失。具体地，例如，模型训练装置120可利用筛选出的候选文本区域内的像素之间的相关度来预测掩膜信息。这里，可以默认认为文本框标记中的像素的掩膜值均为1，并且将其作为真实掩膜信息。模型训练装置120可通过不断利用训练图像样本对文本定位模型进行训练，直至使文本框预测损失和掩膜预测损失的总和最小来训练文本定位模型的总和最小，从而完成文本定位模型的训练。

[0070] 以上，已经参照图1至图4对根据本申请示例性实施例的模型训练系统和文本定位模型进行了描述。利用上述模型训练系统训练出的文本定位模型不仅可以定位文本位置，而且可以定位文本方向，因而能够提供较佳的文本定位效果。此外，由于本申请的文本定位模型包括级联的多级文本框分支，并且在训练前对训练样本集进行了尺寸和/或旋转变化，重新设计了锚点框，并且在训练过程中加入了难样本学习机制，因此，训练出的文本定位模型可提供更好的文本定位效果。

[0071] 需要说明的是，尽管以上在描述模型训练系统100时将其划分为用于分别执行相应处理的装置(例如，训练图像样本集获取装置110和模型训练装置120)，然而，本领域技术人员清楚的是，上述各装置执行的处理也可以在模型训练系统100不进行任何具体装置划分或者各装置之间并无明确划界的情况下执行。此外，以上参照图1所描述的模型训练系统100并不限于包括以上描述的装置，而是还可以根据需要增加一些其他装置(例如，存储装置、数据处理装置等)，或者以上装置也可被组合。

[0072] 图5是示出根据本申请示例性实施例的训练文本定位模型的方法(以下，为描述方便，将其简称为“模型训练方法”)的流程图。

[0073] 这里，作为示例，图5所示的模型训练方法可由图1所示的模型训练系统100来执行，也可完全通过计算机程序或指令以软件方式实现，还可通过特定配置的计算系统或计算装置来执行，例如，可通过包括至少一个计算装置和至少一个存储指令的存储装置的系统来执行，其中，所述指令在被所述至少一个计算装置运行时，促使所述至少一个计算装置执行上述模型训练方法。为了描述方便，假设图5所示的模型训练方法由图1所示的模型训练系统100来执行，并假设模型训练系统100可具有图1所示的配置。

[0074] 参照图5，在步骤S510，训练图像样本集获取装置110可获取训练图像样本集，其中，训练图像样本中对文本进行了文本框标记，并且所述文本框标记包括文本框位置标记和文本框方向标记两者。接下来，在步骤S520，模型训练装置120可基于训练图像样本集训练所述文本定位模型。这里，所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。由于以上已经参照图2至图4对文本框的标记以及文本定位模型等内容进行了介绍，因此，这里为简洁起见不再赘述。另外，已经在图1的描述中对基于训练图像样本集训练所述文本定位模型的细节进行了描述，因此，这里也不再对步骤S520涉及的具体操作进行赘述，相关内容可参见以上关于图1的相关描述。事实上，由于图5所示的模型训练方法由图1所述的模型训练系统100执行，因此，以上参照图1在描述模型训练系统中包括的各个装置时所提及的内

容均适用于这里，故关于以上步骤中所涉及的相关细节，均可参见图1的相应描述，这里都不再赘述。

[0075] 在下文中，将参照图6至图8对利用上述训练出的文本定位模型在图像中定位文本的过程进行描述。

[0076] 图6是示出根据本申请示例性实施例的在图像中定位文本的系统(以下，为描述方便，将其简称为“文本定位系统”)600的框图。

[0077] 参照图6，文本定位系统600可包括预测图像样本获取装置610和文本定位装置620。具体地，预测图像样本获取装置610可被配置为获取预测图像样本，文本定位装置620可被配置为基于预测图像样本的特征，利用预先训练的文本定位模型，确定用于在预测图像样本中定位文本的文本框的位置并确定所述文本框中的文本的方向。这里，文本定位模型可包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

[0078] 作为示例，文本定位装置620可根据预测出的文本框的位置以及利用文本方向预测分支预测出的与文本框的角度有关的值确定文本框中的文本的方向。如以上在关于模型训练的描述中所提及的，文本框的起始点能够决定文本的方向，因此，作为示例，文本定位装置620可根据预测出的文本框的位置以及利用所述文本方向预测分支预测出的与文本框的角度有关的值确定文本框的四个拐点中的哪个点为起始点，进而确定文本的方向。这里，与文本框的角度有关的值包括文本框的角度的正弦值和余弦值，但不限于此。

[0079] 例如，为了确定起始点，文本定位装置620可首先根据指示预测出的文本框的位置的四个拐点的坐标值分别计算文本框的四条边的正弦值和余弦值，然后，根据计算出的文本框的四条边的正弦值和余弦值以及预测出的与文本框的角度有关的值来确定文本框的四个拐点中的哪个点为起始点。具体地，如图7所示，假设预测出的文本框的四个拐点1、2、3和4的坐标值分别是(x1,y1)、(x2,y2)、(x3,y3)和(x4,y4)，则可以计算得到文本框的四条边的方向向量分别是(x2-x1,y2-y1)、(x3-x2,y3-y2)、(x4-x3,y4-y3)和(x1-x4,y1-y4)，并且可进一步计算出四条边的cos值：

$(x2-x1)/\sqrt{(x2-x1)^2+(y2-y1)^2}$、$(x3-x2)/\sqrt{(x3-x2)^2+(y3-y2)^2}$、$(x4-x3)/\sqrt{(x4-x3)^2+(y4-y3)^2}$和$(x1-x4)/\sqrt{(x1-x4)^2+(y1-y4)^2}$，并且可计算出四条边的sin值：$(y2-y1)/\sqrt{(x2-x1)^2+(y2-y1)^2}$、$(y3-y2)/\sqrt{(x3-x2)^2+(y3-y2)^2}$、$(y4-y3)/\sqrt{(x4-x3)^2+(y4-y3)^2}$和$(y1-y4)/\sqrt{(x1-x4)^2+(y1-y4)^2}$。在计算出文本框的四条边的正弦值和余弦值之后，文本定位装置620可根据计算出的文本框的四条边的正弦值和余弦值以及利用文本方向预测分支预测出的与文本框的角度有关的值来确定文本框的四个拐点中的哪个点为起始点。具体地，例如，文本定位装置620可对预测出的文本框的角度的正弦值和余弦值进行处理以使其取值范围与计算出的文本框的四条边的正弦值和余弦值的取值范围相同，随后，可分别计算处理后的所述文本框的角度的正弦值与四条边中的每条边的正弦值之间的差值以及处理后的所述文本框的角度的余弦值与四条边中的每条边的余弦值之间的差值，将计算出的差值求和，并将最小求和值所对应的边的第一个点作为起始点。

[0080] 如上所述，例如，根据本发明示例性实施例的文本定位装置可以是基于深度神经网络的文本定位模型，在这种情况下，利用文本方向预测分支预测出的文本框的正弦值和余弦值的取值范围往往被压缩到0到1之间，因此，文本定位装置620需要对预测出的文本框的角度的正弦值和余弦值进行处理以使其取值范围与计算出的文本框的四条边的正弦值和余弦值的取值范围相同。由于计算出的文本框的四条边的正弦值和余弦值的取值范围在-1到1之间，因此，可对预测出的文本框的角度的余弦值cos进行处理，使得处理后的余弦值COS＝2*cos-1，同样地，可对预测出的文本框的角度的正弦值sin进行处理，使得处理后的正弦值SIN＝2*sin-1。经过上述处理后的正弦值SIN和余弦值COS的取值范围将在-1到1之间。随后，文本定位装置620可分别计算处理后的正弦值SIN与根据文本框的坐标信息得到的四条边中的每条边的正弦值之间的差值以及处理后的余弦值COS与根据文本框的坐标信息得到的四条边中的每条边的余弦值之间的差值，将计算出的差值求和，并将最小求和值所对应的边的第一个点作为起始点。例如，假设根据文本框的坐标信息得到的四条边中的每条边的正弦值分别是sin1、sin2、sin3和sin4，四条边中的每条边的余弦值分别是cos1、cos2、cos3和cos4，则针对四条边计算出的正弦值差值为:SIN-sin1，SIN-sin2，SIN-sin3，SIN-sin4，计算出的余弦值差值为:COS-cos1，COS-cos2，COS-cos3，COS-cos4，并且将每条边对应的正弦值差值与余弦值差值求和所得的求和值分别是:(SIN-sin1)+(COS-cos1)，(SIN-sin2)+(COS-cos2)，(SIN-sin3)+(COS-cos3)和(SIN-sin4)+(COS-cos4)，如果在以上四个求和值中，最大的是(SIN-sin1)+(COS-cos1)，则将(SIN-sin1)+(COS-cos1)所对应的边(即，第一条边)的第一个点确定为起始点，也就是说，图7中的拐点1被确定为起始点。而在确定了起始点的情况下，便可据此确定文本方向。

[0081] 如以上参照图4所描述的，根据本发明示例性实施例，文本定位模型可包括特征提取层、候选区域推荐层、级联的多级文本框分支以及掩膜分支，其中，特征提取层用于提取预测图像样本的特征以生成特征图，候选区域推荐层用于基于生成的特征图在预测图像样本中确定预定数量个候选文本区域，级联的多级文本框分支中的每一级文本框分支包括文本框位置预测分支和文本方向预测分支，并且所述级联的多级文本框分支用于基于特征图中的与每个候选文本区域对应的特征来预测候选水平文本框的位置和与候选水平文本框的角度有关的值，掩膜分支用于基于特征图中与候选水平文本框对应的特征来预测候选水平文本框中的文本的掩膜信息，根据预测出的掩膜信息确定用于在预测图像样本中定位文本的最终文本框。作为示例，预测图像样本的特征可预测图像样本中像素的相关度，但不限于此。此外，作为示例，文本定位模型可以基于Mask-RCNN框架，并且特征提取层对应于Mask-RCNN框架中的深度残差网络，候选区域推荐层对应于Mask-RCNN框架中的区域推荐网络RPN层，级联的多级文本框分支中的每一级文本框分支包括Mask-RCNN框架中的RolAlign层和全连接层，掩膜分支可以包括一系列卷积层。以上参照图4关于文本定位模型的描述均适应于这里，这里不再赘述。

[0082] 由于同一张图像中可能同时存在长文本和短文本，而如果始终将图像放大或缩小到一定尺寸后输入文本定位模型，则可能不能够同时较好地检测到长文本和短文本。这是因为，如果将图像放大到较大尺寸，则短文本的定位性能较好，而如果将图像缩小到较小尺寸，则长文本的定位性能较好。因此，在本发明中，可对图像进行多尺度预测。具体地，预测图像样本获取装置610可首先获取图像，然后对获取的图像进行多尺度缩放来获取与所述

图像对应的不同尺寸的多个预测图像样本。随后，文本定位装置620可针对不同尺寸的多个预测图像样本分别利用预先训练的文本定位模型来确定用于在预测图像样本中定位文本位置的最终文本框，最后，将针对每种尺寸的预测图像样本确定的文本框进行合并来得到最终的结果。这里，图像可来源于任何数据源，本申请对图像的来源、图像的具体获取方式等均无限制。

[0083] 针对每种尺寸的预测图像样本，文本定位装置620可通过执行以下操作来确定用于在预测图像样本中定位文本的最终文本框：利用特征提取层提取预测图像样本的特征以生成特征图；利用候选区域推荐层基于生成的特征图在预测图像样本中确定预定数量个的候选文本区域；利用级联的多级文本框分支基于特征图中的与每个候选文本区域对应的特征预测初始候选水平文本框，并且通过第一非极大值抑制操作从初始候选水平文本框中筛选出文本框重合度小于第一重合度阈值的水平文本框作为候选水平文本框；利用掩膜分支，基于特征图中与候选水平文本框对应的特征来预测候选水平文本框中的文本的掩膜信息，根据预测出的文本的掩膜信息确定初选文本框，并且通过第二非极大值抑制操作从确定的初选文本框中筛选出文本框重合度小于第二重合度阈值的文本框作为所述最终文本框，其中，第一重合度阈值大于第二重合度阈值；将所述最终文本框的位置确定为用于在预测图像样本中定位文本的文本框的位置，并根据所述最终文本框的位置以及预测出的与候选水平文本框的角度有关的值确定文本的方向。

[0084] 接下来，文本定位装置620可将针对不同尺寸的预测图像样本确定的文本框进行合并。具体地，针对第一尺寸的预测图像样本，在利用所述文本定位模型确定了用于在第一尺寸的预测图像样本中定位文本的的文本框之后从该文本框中选择尺寸大于第一阈值的第一文本框，并且针对第二尺寸的预测图像样本，在利用所述文本定位模型确定了用于在第二尺寸的预测图像样本中定位文本的文本框之后从该文本框中选择尺寸小于第二阈值的第二文本框，其中，第一尺寸小于第二尺寸文本定位装置。也就是说，在合并的时候，对于较大尺寸的图像预测样本，保留小尺寸的文本框，而对于较小尺寸的图像预测样本，保留大尺寸的文本框。例如，如果先前获取的预测图像样本的尺寸分别是800像素大小和1600像素大小，则在将800像素大小和1600像素大小的预测图像样本分别输入文本定位模型而分别得到在预测图像样本中定位文本的文本框之后，对于800像素大小的预测图像样本，文本定位装置620可保留相对大的文本框而过滤掉相对小的文本框(具体地可通过以上提及的第一阈值的设置来进行保留)，然而，对于1600像素大小的预测图像样本，文本定位装置620可保留相对小的文本框而过滤掉相对大的文本框(具体地，可通过以上提及的第二阈值的设置来进行保留)。接下来，文本定位装置620可将过滤后的结果进行合并。具体地，文本定位装置620可利用第三非极大值抑制操作对选择的第一文本框和第二文本框进行筛选，以得到用于在所述图像中定位文本的最终文本框。例如，文本定位装置620可将所有选择的第一文本框和第二文本框按照其置信度进行排名并选择置信度最大的一个文本框，然后计算其余文本框与该文本框的重叠度，如果重叠度大于阈值则删除，否则保留，而最终保留的文本框即为在图像中定位文本位置的最终文本框。

[0085] 下面，具体地对文本定位装置620针对每个预测图像样本执行的操作所涉及的一些细节进行描述。需要说明的是，在接下来的描述中，为了避免对公知的功能和结构的描述会用不必要的细节模糊本发明的构思，因此将省略对公知的功能、结构和术语的描述。

[0086] 首先，如上所述，为了确定在预测图像样本中定位文本的文本框，文本定位装置620可利用特征提取层提取预测图像样本的特征以生成特征图，具体地，例如可以利用Mask-RCNN框架中的深度残差网络(例如，resnet101)提取预测图像样本的像素之间的相关度作为特征来生成特征图。然而，本申请对所使用的预测图像样本的特征以及具体的特征提取方式并无任何限制。

[0087] 接下来，文本定位装置620可利用候选区域推荐层基于生成的特征图在预测图像样本中确定预定数量个的候选文本区域，例如，文本定位装置620可利用候选区域推荐层基于生成的特征图预测候选文本区域与预先设置的锚点框之间的差异，根据该差异和锚点框确定初始候选文本区域，并利用第四非极大值抑制操作从初始候选文本区域中筛选出所述预定数量个候选文本区域。这里，所述锚点框的宽高比可以是以上描述的通过在所述文本定位模型的训练阶段对训练图像样本集中所标记的文本框的宽高比进行统计而确定的。利用非极大值抑制操作从初始候选文本区域中筛选出所述预定数量个候选文本区域的具体细节已经在参照图1的描述中提及，因此，这里不再赘述。

[0088] 随后，文本定位装置620可利用级联的多级文本框分支基于特征图中的与每个候选文本区域对应的特征预测初始候选水平文本框，并且通过第一非极大值抑制操作从初始候选水平文本框中筛选出文本框重合度小于第一重合度阈值的水平文本框作为候选水平文本框。作为示例，所述级联的多级文本框分支可以是三级文本框分支，下面，以三级文本框为例对利用级联的多级文本框分支基于特征图中的与每个候选文本区域对应的特征预测初始候选水平文本框进行描述。

[0089] 具体地，文本定位装置620可首先利用第一级文本框分支，从特征图中提取与每个候选文本区域对应的特征并预测每个候选文本区域与真实文本区域的位置偏差、每个候选文本区域包括文本的置信度和不包括文本的置信度、以及与每个候选文本区域的角度有关的值，并且根据第一级文本框分支的预测结果确定第一级水平文本框。例如，文本定位装置620可利用第一级文本框分支中的RolAlign层从特征图中提取与每个候选文本区域对应的特征，并利用第一级文本框分支中的全连接层预测每个候选文本区域与真实文本区域的位置偏差、每个候选文本区域包括文本的置信度和不包括文本的置信度、以及与每个候选文本区域的角度有关的值。然后，文本定位装置620可根据预测的置信度去除部分置信度较低的候选文本区域，并根据保留的候选文本区域及其与真实文本区域的位置偏差确定第一级水平文本框。

[0090] 在确定了第一级水平文本框之后，文本定位装置620可利用第二级文本框分支，从特征图中提取与第一级水平文本框对应的特征并预测第一级水平文本框与真实文本区域的位置偏差、第一级水平文本框包括文本的置信度和不包括文本的置信度、以及与第一级水平文本框的角度有关的值，并根据第二级文本框分支的预测结果确定第二级水平文本框。同样地，例如，文本定位装置620可利用第二级文本框分支中的RolAlign层从特征图中提取与第一级水平文本框对应的特征(即，提取与第一级水平文本框中的像素区域对应的特征)，并利用第二级文本框分支中的全连接层预测第一级水平文本框与真实文本区域的位置偏差、第一级水平文本框包括文本的置信度和不包括文本的置信度、以及与第一级水平文本框的角度有关的值。然后，文本定位装置620可根据预测的置信度去除部分置信度较低的第一级水平文本框，并根据保留的第一级水平文本框及其与真实文本区域的位置偏差

确定第二级水平文本框。

[0091] 在确定了第二级水平文本框之后，文本定位装置620可利用第三级文本框分支，从特征图中提取与第二级水平文本框对应的特征并预测第二级水平文本框与真实文本区域的位置偏差、第二级水平文本框包括文本的置信度和不包括文本的置信度、以及与第二级水平文本框的角度有关的值，并根据第三级文本框分支的预测结果确定初始候选水平文本框。同样地，例如，文本定位装置420可利用第三级文本框分支中的RolAlign层从特征图中提取与第二级水平文本框对应的特征(即，提取与第二级水平文本框中的像素区域对应的特征)，并利用第三级文本框分支中的全连接层预测第二级水平文本框与真实文本区域的位置偏差、第二级水平文本框包括文本的置信度和不包括文本的置信度、以及与第二级水平文本框的角度有关的值。然后，文本定位装置620可根据预测的置信度去除部分置信度较低的第二级水平文本框，并根据保留的第二级水平文本框及其与真实文本区域的位置偏差确定初始候选水平文本框。

[0092] 如上所述，在预测出初始候选水平文本框之后，文本定位装置620可通过第一非极大值抑制操作从初始候选水平文本框中筛选出文本框重合度小于第一重合度阈值的水平文本框作为候选水平文本框。具体地，文本定位装置620可首先根据初始候选水平文本框的置信度选择置信度最大的初始候选水平文本框，然后计算其余初始候选水平文本框与置信度最大的初始候选水平文本框的文本框重合度，如果文本框重合度小于第一重合度阈值则保留，否则删除。所有保留的水平文本框被作为候选水平文本框输入掩膜分支。

[0093] 接下来，文本定位装置620可利用掩膜分支，基于特征图中与候选水平文本框对应的特征来预测候选水平文本框中的文本的掩膜信息。具体地，例如，文本定位装置620可基于特征图中与候选水平文本框中的像素对应的像素相关度特征来预测候选水平文本框中的文本的掩膜信息。随后，文本定位装置620可根据预测出的文本的掩膜信息确定初选文本框。具体而言，例如，文本定位装置620可根据预测出的文本的掩膜信息确定包含文本的最小外接矩形，并将确定的最小外接矩形作为初选文本框。例如，文本定位装置620可根据预测出的文本的掩膜信息使用最小外接矩形函数确定包含文本的最小外部矩形。

[0094] 在确定了初选文本框之后，文本定位装置620可通过第二非极大值抑制操作从确定的初选文本框中筛选出文本框重合度小于第二重合度阈值的文本框作为所述最终文本框。具体地，例如，文本定位装置620可首先根据初始候选水平文本框的置信度选择置信度最大的初始候选水平文本框，然后计算其余初始候选水平文本框与置信度最大的初始候选水平文本框的文本框重合度，如果文本框重合度小于第一重合度阈值则保留，否则删除。

[0095] 需要说明的是，以上提及的第一重合度阈值大于第二重合度阈值。传统的Mask-RCNN框架中只有一级非极大值抑制，并且重合度阈值被固定设置为0.5，也就是说，在筛选时会删除重合度高于0.5的水平文本框。然而，对于旋转角度较大的密集文字，如果重合度阈值设置为0.5，则会导致部分文本框的漏检。而如果提高重合度阈值(例如，将重合度阈值设置为0.8，即，删除重合度高于0.8的文本框)，则会导致最后预侧的水平文本框重叠较多。针对此，本发明还提出了两级非极大值抑制的构思。即，如上所述，在利用级联的多级文本框分支预测出初始候选水平文本框，先通过第一非极大值抑制操作从初始候选水平文本框中筛选出文本框重合度小于第一重合度阈值的水平文本框作为候选水平文本框。随后，在利用掩膜分支预测出候选水平文本框中的文本的掩膜信息并根据预测出的文本的掩膜信

息确定了初选文本框之后，通过第二非极大值抑制操作从确定的初选文本框中筛选出文本框重合度小于第二重合度阈值的文本框作为所述最终文本框。而通过将第一重合度阈值大于第二重合度阈值(例如，第一重合度阈值可设置为0.8，第二重合度阈值可设置为0.2)，可实现先利用第一非极大值抑制操作对通过级联的多级文本框分支确定的文本框进行粗筛，然后，利用第二非极大值抑制操作对通过掩膜分支确定的文本框进行细筛。最终，经过两级非极大值抑制操作和调整两级非极大值抑制操作所使用的重合度阈值，不仅可以定位水平文本而且可以定位旋转文本。

[0096] 此外，图6所示的文本定位系统600还可以包括显示装置(未示出)。显示装置可在图像上显示用于在所述图像中定位文本的最终文本框，从而可方便用户直观地确定文本定位结果。这里，所述最终文本框包括水平文本框和/或旋转文本框。此外，可选地，文本定位系统600还可以包括文本识别装置(未示出)，例如，文本识别装置可对根据文本位置信息和文本方向裁剪出的文本区域进行文本识别。

[0097] 根据示例性实施例的文本定位系统由于使用包括文本框位置预测分支和文本方向预测分支的文本定位模型进行文本定位，因此，可同时定位文本位置和文本方向，提供更好的文本定位效果，便于进行后续文本识别。此外，通过利用包括级联的多级文本框分支的文本定位模型，可提高文本定位性能，而且由于引入了两级非极大值抑制操作可有效防止漏检和文本框重叠，使得不仅可以定位水平文本而且可以定位旋转文本。此外，通过对获取的图像进行多尺度变换之后针对同一图像的不同尺寸的预测图像样本进行预测并将针对不同尺寸的预测图像样本确定的文本框进行合并，可进一步提高文本定位效果，使得即使在图像中同时存在不同尺寸的文本时，也可提供较好的文本定位效果。

[0098] 另外，需要说明的是，尽管以上在描述文本定位系统600时将其划分为用于分别执行相应处理的装置(例如，预测图像样本获取装置610和文本定位装置620)，然而，本领域技术人员清楚的是，上述各装置执行的处理也可以在文本定位系统600不进行任何具体装置划分或者各装置之间并无明确划界的情况下执行。此外，以上参照图6所描述的文本定位系统600并不限于包括以上描述的预测图像样本获取装置610、文本定位装置620、显示装置和文本识别装置，而是还可以根据需要增加一些其他装置(例如，存储装置、数据处理装置等)，或者以上装置也可被组合。而且，作为示例，以上参照图1描述的模型训练系统100和文本定位系统600也可被组合为一个系统，或者它们可以是彼此独立的系统，本申请对此并无限制。

[0099] 图8是示出根据本申请示例性实施例的在图像中定位文本的方法(以下，为描述方便，将其简称为“文本定位方法”)的流程图。

[0100] 这里，作为示例，图8所示的文本定位方法可由图6所示的文本定位系统600来执行，也可完全通过计算机程序或指令以软件方式实现，还可通过特定配置的计算系统或计算装置来执行，例如，可通过包括至少一个计算装置和至少一个存储指令的存储装置的系统来执行，其中，所述指令在被所述至少一个计算装置运行时，促使所述至少一个计算装置执行上述文本定位方法。为了描述方便，假设图8所示的文本定位方法由图6所示的文本定位系统600来执行，并假设文本定位系统600可具有图6所示的配置。

[0101] 参照图8，在步骤S810，预测图像样本获取装置610可获取预测图像样本。例如，在步骤S810，预测图像样本获取装置610可首先获取图像，然后对获取的图像进行多尺度缩放

来获取与所述图像对应的不同尺寸的多个预测图像样本。接下来，在步骤S820，文本定位装置620可基于预测图像样本的特征，利用预先训练的文本定位模型，确定用于在预测图像样本中定位文本的文本框的位置并确定所述文本框中的文本的方向。这里，所述文本定位模型可包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。例如，所述文本定位模型可所述文本定位模型包括特征提取层、候选区域推荐层、级联的多级文本框分支以及掩膜分支，其中，特征提取层用于提取预测图像样本的特征以生成特征图，候选区域推荐层用于基于生成的特征图在预测图像样本中确定预定数量个候选文本区域，级联的多级文本框分支中的每一级文本框分支包括文本框位置预测分支和文本方向预测分支，并且所述级联的多级文本框分支用于基于特征图中的与每个候选文本区域对应的特征来预测候选水平文本框的位置和与候选水平文本框的角度有关的值，掩膜分支用于基于特征图中与候选水平文本框对应的特征来预测候选水平文本框中的文本的掩膜信息，根据预测出的掩膜信息确定用于在预测图像样本中定位文本的最终文本框

[0102] 由于以上已经参照图4对文本定位模型进行了介绍，并参照图6对文本定位系统所执行的上述步骤中所涉及的操作进行了描述，因此，这里为简洁起见不再赘述，相关内容可参见以上关于图4和图6的相关描述。事实上，由于图8所示的文本定位方法可由图6所示的文本定位系统400来执行，因此，关于以上步骤中所涉及的任何相关细节以及文本定位系统除了以上两个步骤可额外执行的操作，均可参见关于图6的相应描述，这里都不再赘述。

[0103] 以上已参照图1至图8描述了根据本申请示例性实施例模型训练系统和模型训练方法以及文本定位系统和文本定位方法。然而，应理解的是：图1和图6所示出的系统及其装置可被分别配置为执行特定功能的软件、硬件、固件或上述项的任意组合。例如，这些系统或装置可对应于专用的集成电路，也可对应于纯粹的软件代码，还可对应于软件与硬件相结合的模块。此外，这些系统或装置所实现的一个或多个功能也可由物理实体设备(例如，处理器、客户端或服务器等)中的组件来统一执行。

[0104] 此外，上述方法可通过记录在计算机可读存储介质上的指令来实现，例如，根据本申请的示例性实施例，可提供一种存储指令的计算机可读存储介质，其中，当所述指令被至少一个计算装置运行时，促使所述至少一个计算装置执行以下步骤：获取训练图像样本集，其中，训练图像样本中对文本进行了文本框标记，其中，所述文本框标记包括文本框位置标记和文本框方向标记两者；基于训练图像样本集训练所述文本定位模型，其中，所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

[0105] 此外，根据本申请的另一示例性实施例，可提供一种存储指令的计算机可读存储介质，其中，当所述指令被至少一个计算装置运行时，促使所述至少一个计算装置执行以下步骤：获取预测图像样本；基于预测图像样本的特征，利用预先训练的文本定位模型，确定用于在预测图像样本中定位文本的文本框的位置并确定所述文本框中的文本的方向，其中，所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

[0106] 上述计算机可读存储介质中存储的指令可在诸如客户端、主机、代理装置、服务器等计算机设备中部署的环境中运行，应注意，所述指令还可在执行上述步骤时执行更为具

体的处理，这些进一步处理的内容已经在参照图1至图8的描述中提及，因此这里为了避免重复将不再进行赘述。

[0107] 应注意，根据本公开示例性实施例的模型训练系统和文本定位系统可完全依赖计算机程序或指令的运行来实现相应的功能，即，各个装置在计算机程序的功能架构中与各步骤相应，使得整个系统通过专门的软件包(例如，lib库)而被调用，以实现相应的功能。

[0108] 另一方面，当图1和图6所示的系统和装置以软件、固件、中间件或微代码实现时，用于执行相应操作的程序代码或者代码段可以存储在诸如存储介质的计算机可读介质中，使得至少一个处理器或至少一个计算装置可通过读取并运行相应的程序代码或者代码段来执行相应的操作。

[0109] 例如，根据本申请示例性实施例，可提供一种包括至少一个计算装置和存储指令的至少一个存储装置的系统，其中，所述指令在被所述至少一个计算装置运行时，促使所述至少一个计算装置执行下述步骤：获取训练图像样本集，其中，训练图像样本中对文本进行了文本框标记，其中，所述文本框标记包括文本框位置标记和文本框方向标记两者；基于训练图像样本集训练所述文本定位模型，其中，所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

[0110] 例如，根据本申请另一示例性实施例，可提供一种包括至少一个计算装置和存储指令的至少一个存储装置的系统，其中，所述指令在被所述至少一个计算装置运行时，促使所述至少一个计算装置执行下述步骤：获取预测图像样本；基于预测图像样本的特征，利用预先训练的文本定位模型，确定用于在预测图像样本中定位文本的文本框的位置并确定所述文本框中的文本的方向，其中，所述文本定位模型包括用于确定文本位置的文本框位置预测分支和用于预测与所述文本框的角度有关的值以确定文本方向的文本方向预测分支。

[0111] 具体说来，上述系统可以部署在服务器或客户端中，也可以部署在分布式网络环境中的节点上。此外，所述系统可以是PC计算机、平板装置、个人数字助理、智能手机、web应用或其他能够执行上述指令集合的装置。此外，所述系统还可包括视频显示器(诸如，液晶显示器)和用户交互接口(诸如，键盘、鼠标、触摸输入装置等)。另外，所述系统的所有组件可经由总线和/或网络而彼此连接。

[0112] 这里，所述系统并非必须是单个系统，还可以是任何能够单独或联合执行上述指令(或指令集)的装置或电路的集合体。所述系统还可以是集成控制系统或系统管理器的一部分，或者可被配置为与本地或远程(例如，经由无线传输)以接口互联的便携式电子装置。

[0113] 在所述系统中，所述至少一个计算装置可包括中央处理器(CPU)、图形处理器(GPU)、可编程逻辑装置、专用处理器系统、微控制器或微处理器。作为示例而非限制，所述至少一个计算装置还可包括模拟处理器、数字处理器、微处理器、多核处理器、处理器阵列、网络处理器等。计算装置可运行存储在存储装置之一中的指令或代码，其中，所述存储装置还可以存储数据。指令和数据还可经由网络接口装置而通过网络被发送和接收，其中，所述网络接口装置可采用任何已知的传输协议。

[0114] 存储装置可与计算装置集成为一体，例如，将RAM或闪存布置在集成电路微处理器等之内。此外，存储装置可包括独立的装置，诸如，外部盘驱动、存储阵列或任何数据库系统可使用的其他存储装置。存储装置和计算装置可在操作上进行耦合，或者可例如通过I/O端

口、网络连接等互相通信，使得计算装置能够读取存储在存储装置中的指令。

[0115] 以上描述了本申请的各示例性实施例，应理解，上述描述仅是示例性的，并非穷尽性的，本申请不限于所披露的各示例性实施例。在不偏离本申请的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。因此，本申请的保护范围应该以权利要求的范围为准。

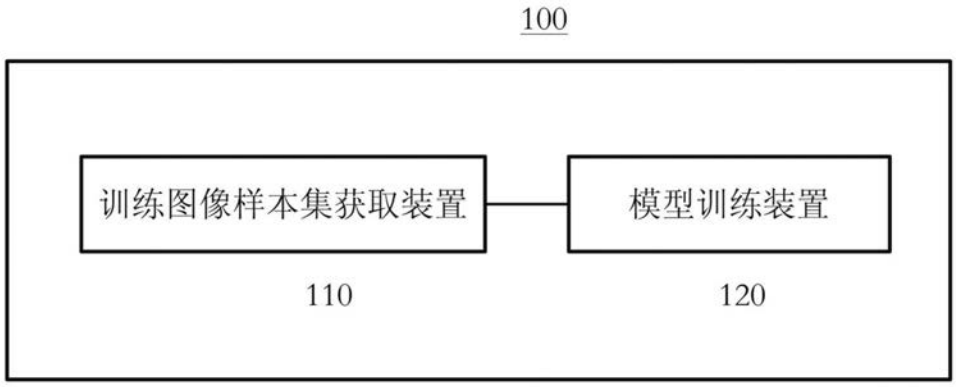

100
训练图像样本集获取装置
模型训练装置
110
120

图1

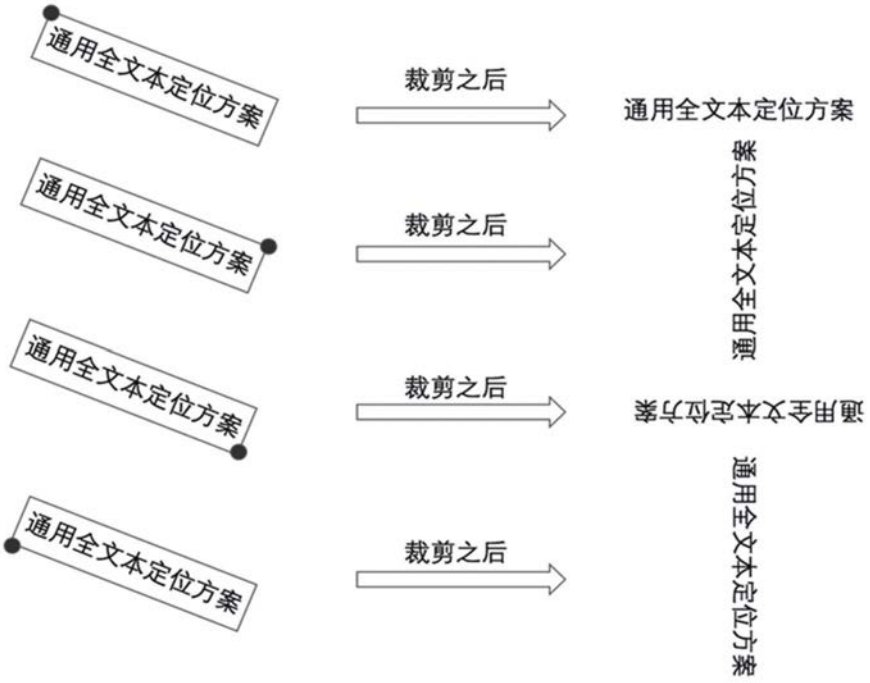

通用全文本定位方案
裁剪之后
通用全文本定位方案
通用全文本定位方案
裁剪之后
通用全文本定位方案
通用全文本定位方案
裁剪之后
通用全文本定位方案
通用全文本定位方案
裁剪之后
通用全文本定位方案

图2

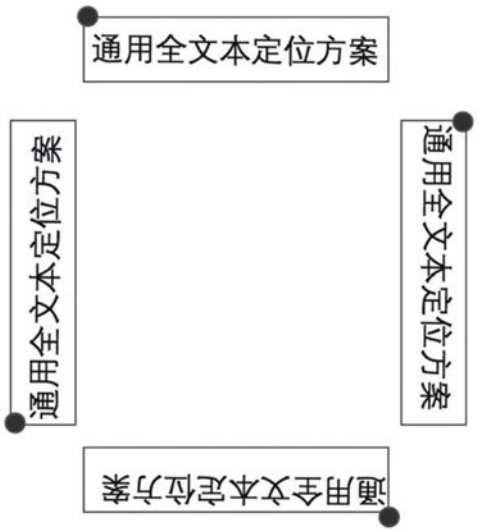

通用全文本定位方案
通用全文本定位方案
通用全文本定位方案
通用全文本定位方案

图3

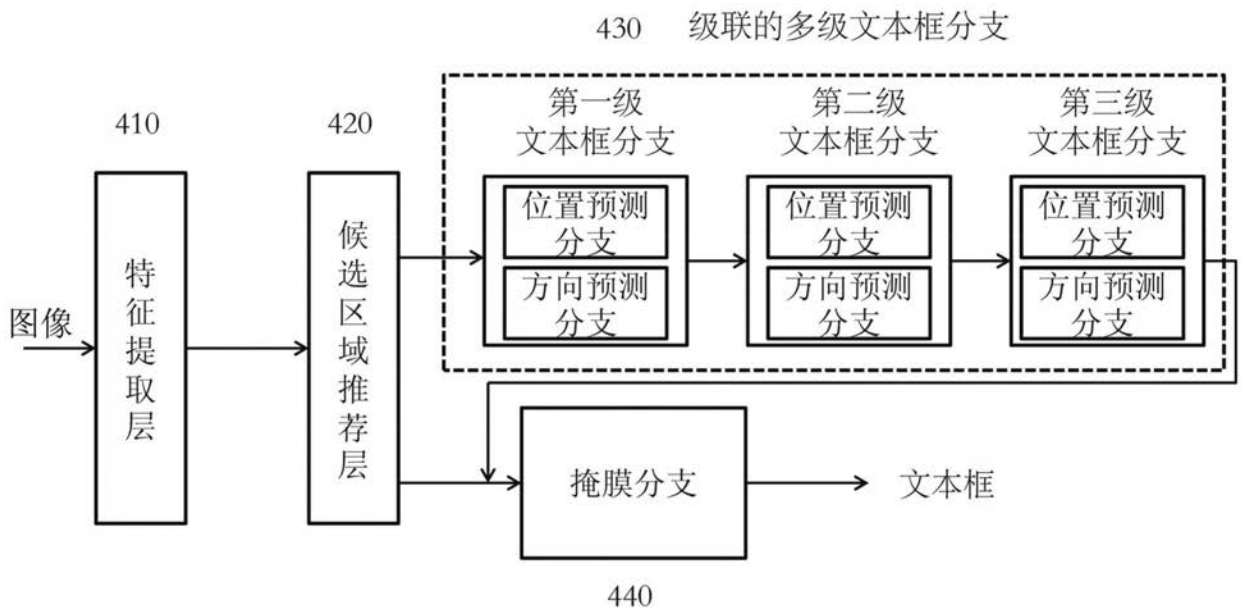

430 级联的多级文本框分支
410
420
第一级
文本框分支
第二级
文本框分支
第三级
文本框分支
图像
特征提取层
候选区域推荐层
位置预测分支
方向预测分支
位置预测分支
方向预测分支
位置预测分支
方向预测分支
掩膜分支
文本框
440

图4

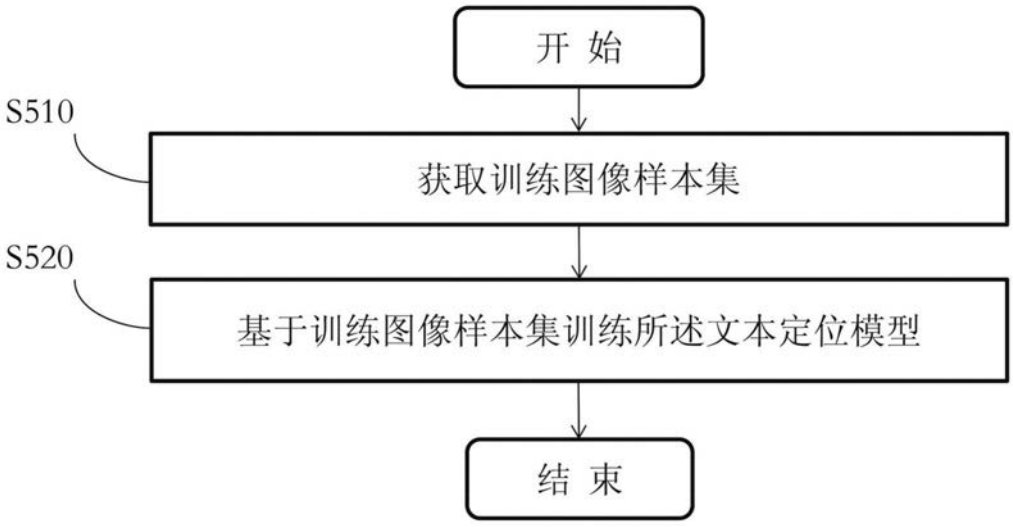

开 始
S510
获取训练图像样本集
S520
基于训练图像样本集训练所述文本定位模型
结 束

图5

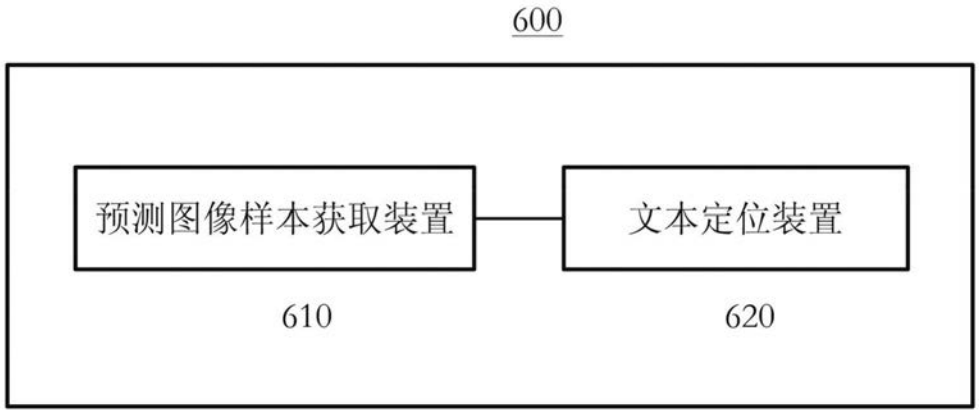

600
预测图像样本获取装置
文本定位装置
610
620

图6

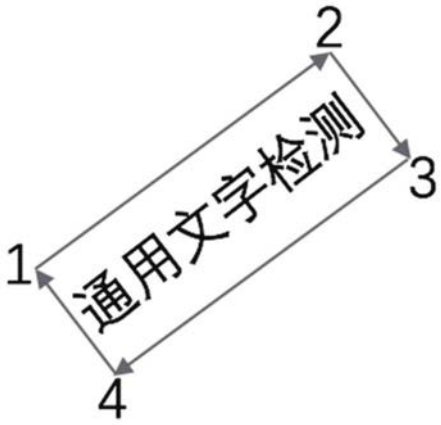

2
通用文字检测
3
1
4

图7

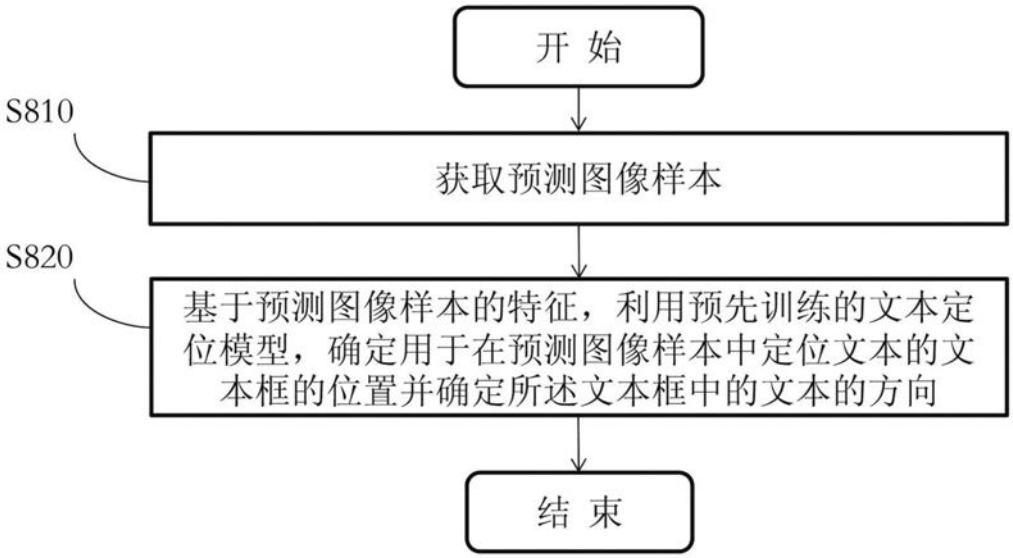

开 始
S810
获取预测图像样本
S820
基于预测图像样本的特征，利用预先训练的文本定位模型，确定用于在预测图像样本中定位文本的文本框的位置并确定所述文本框中的文本的方向
结 束

图8