

用于识别文本的方法和装置

5 本专利申请要求于 2021 年 06 月 07 日提交的、申请号为 202110632180.5、发明名称为“用于识别文本的方法和装置”的中国专利申请的优先权，该申请的全文以引用的方式并入本申请中。

技术领域

10 本公开的实施例涉及计算机技术领域，具体涉及用于识别文本的方法和装置。

背景技术

 现有的文字识别可以划分为两类，一类是 OCR (Optical Character
15 Recognition, 光学字符识别), 另一类是 STR (Scene Text Recognition, 场景文字识别)。其中，OCR 通常指对高质量的扫描文档图像中的文字进行识别，STR 通常指对自然场景图像中的文字进行识别。

 一般地，OCR 的识别场景中图像背景简单、文字排列整齐、字体标准等，而 STR 的识别场景中图像背景较复杂、文字排列随意、字体
20 多样。因此，STR 的识别难度远远大于 OCR 的识别难度。

 STR 在辅助视觉障碍的导航、自动驾驶应用、扩增现实中的文本阅读和翻译等许多领域中都具有重要的实用性，在计算机视觉界引起了越来越多的关注。目前的 STR 的识别方法通常都是先从图像中定位文本区域，然后再识别文本区域中的文字。

25

发明内容

 本公开的实施例提出了用于识别文本的方法和装置。

 第一方面，本公开的实施例提供了一种用于识别文本的方法，该方法包括：获取特征图，其中，特征图通过对呈现有待识别文本的图
30 像进行文本实例分割得到；根据特征图构建关系图，其中，关系图中

5 的节点表示特征图中的像素点，边表示所连接的两个节点的空间语义特征的相似度大于目标阈值，空间语义特征包括节点指示的像素点的位置特征和类别特征；利用预先训练的图卷积网络对关系图进行处理，得到图像对应的第一文本特征；根据第一文本特征，生成图像的文本识别结果。

第二方面，本公开的实施例提供了一种用于识别文本的装置，该装置包括：特征图获取单元，被配置成获取特征图，其中，特征图通过对呈现有待识别文本的图像进行文本实例分割得到；关系图构建单元，被配置成根据特征图构建关系图，其中，关系图中的节点表示特征图
10 中的像素点，边表示所连接的两个节点的空间语义特征的相似度大于目标阈值，空间语义特征包括节点指示的像素点的位置特征和类别特征；图卷积处理单元，被配置成利用预先训练的图卷积网络对关系图进行处理，得到图像对应的第一文本特征；识别单元，被配置成根据第一文本特征，生成图像的文本识别结果。

15 第三方面，本公开的实施例提供了一种电子设备，该电子设备包括：一个或多个处理器；存储装置，用于存储一个或多个程序；当一个或多个程序被一个或多个处理器执行，使得一个或多个处理器实现如第一方面中任一实现方式描述的方法。

第四方面，本公开的实施例提供了一种计算机可读介质，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如第一方面中任一实现方式描述的方法。
20

第五方面，本公开的实施例提供了一种包括计算机程序的计算机程序产品，该计算机程序在被处理器执行时能够实现如第一方面中任一实现方式描述的方法。

25 附图说明

通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本公开的其它特征、目的和优点将会变得更明显：

图 1 是本公开的一个实施例可以应用于其中的示例性系统架构图；
30

图 2 是根据本公开的用于识别文本的方法的一个实施例的流程图；

图 3 是根据本公开的用于识别文本的方法中特征图的生成方式的一个实施例的流程图；

5 图 4 是根据本公开的用于识别文本的方法的又一个实施例的流程图；

图 5a、5b、5c 是根据本公开的用于识别文本的方法的一个示例性的应用场景的示意图；

图 6 是根据本公开的用于识别文本的方法中的图卷积网络、语言
10 模型和分割网络的训练方法的一个实施例的流程图；

图 7 是根据本公开的用于识别文本的装置的一个实施例的结构示意图；

图 8 是适于用来实现本公开的实施例的电子设备的结构示意图。

15 具体实施方式

下面结合附图和实施例对本公开作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与有关发明相关的部分。

20 需要说明的是，在不冲突的情况下，本公开中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本公开。图 1 示出了可以应用本公开的用于识别文本的方法或用于识别文本的装置的实施例的示例性架构 100。

如图 1 所示，系统架构 100 可以包括终端设备 101、102、103，
25 网络 104 和服务器 105。网络 104 用以在终端设备 101、102、103 和服务器 105 之间提供通信链路的介质。网络 104 可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

终端设备 101、102、103 通过网络 104 与服务器 105 交互，以接收或发送消息等。终端设备 101、102、103 上可以安装有各种客户端
30 应用。例如，浏览器类应用、搜索类应用、图像处理类应用、深度学

习框架等等。

终端设备 101、102、103 可以是硬件，也可以是软件。当终端设备 101、102、103 为硬件时，可以是各种电子设备，包括但不限于智能手机、平板电脑、电子书阅读器、膝上型便携计算机和台式计算机
5 等等。当终端设备 101、102、103 为软件时，可以安装在上述所列举的电子设备中。其可以实现成多个软件或软件模块（例如用来提供分布式服务的多个软件或软件模块），也可以实现成单个软件或软件模块。在此不做具体限定。

服务器 105 可以是提供各种服务的服务器，例如为终端设备 101、
10 102、103 上安装的客户端应用提供后端支持的服务器。服务器 105 可以是硬件，也可以是软件。当服务器 105 为硬件时，可以实现成多个服务器组成的分布式服务器集群，也可以实现成单个服务器。当服务器 105 为软件时，可以实现成多个软件或软件模块（例如用来提供分布式服务的多个软件或软件模块），也可以实现成单个软件或软件模
15 块。在此不做具体限定。

需要说明的是，本公开的实施例所提供的用于识别文本的方法一般由服务器 105 执行，相应地，用于识别文本的装置一般设置于服务器 105 中。

还需要指出的是，终端设备 101、102、103 中也可以安装有图像处理类应用，终端设备 101、102、103 也可以基于图像处理类应用对
20 呈现有待识别文本的图像进行处理，此时，用于识别文本的方法也可以由终端设备 101、102、103 执行，相应地，用于识别文本的装置也可以设置于终端设备 101、102、103 中。此时，示例性系统架构 100 可以不存在服务器 105 和网络 104。

25 应该理解，图 1 中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备、网络和服务器的。

本公开的实施例提供的用于识别文本的方法和装置，通过获取对呈现有待识别文本的图像进行文本实例分割得到的特征图，然后以特征图中的像素点为节点，以节点的空间语义特征的相似度建立边，从
30 而构建特征图对应的关系图，然后利用图卷积网络对关系图进行处理，

以提取图像中的待识别文本的第一文本特征，再利用第一文本特征生成图像对应的文本识别结果。这种基于图的文本识别方法可以考虑到图像中的文本的二维空间信息，避免直接将图像中的文本特征压缩成一维的特征而忽略了二维空间信息，有助于提升文本识别效果。

5 继续参考图 2，其示出了根据本公开的用于识别文本的方法的一个实施例的流程 200。该用于识别文本的方法包括以下步骤：

步骤 201，获取特征图。

在本实施例中，特征图可以通过对呈现有待识别文本的图像进行文本实例分割得到。其中，待识别文本可以是各种内容的文本。例如，
10 待识别文本可以包括一种或多种文字（如字母、数字、特殊符号、汉字等等）。呈现有待识别文本的图像可以是各种类型的图像。

根据实际应用场景的不同，呈现有待识别文本的图像的质量可以不同，图像所呈现的待识别文本的位置、书写方式等各种属性可以不同。

15 文本实例分割可以指从图像中检测出待识别文本包括的各个文字并区分各个文字。一般地，经过文本实例分割后得到的特征图(MASK)可以对应于识别文本所包括的文字。特征图的数目可以根据实际的应用场景灵活设置。例如，图像中的待识别文本仅包括一个文字，则特征图的数目可以为一。

20 可选地，特征图的数目为至少两个。或者，可以由技术人员预先指定特征图的数目。一般地，可以根据实际的应用场景估计图像中的待识别文本包括的文字数目，设置的特征图的数目可以大于预估的文字数目，以避免漏识别等情况。

具体地，可以由上述执行主体或其他电子设备采用现有的各种实例分割方法对呈现有待识别文本的图像进行文本实例分割，以得到特征图。
25

在本实施例中，用于识别文本的方法（如图 1 所示的服务器 105 等）本地可以预先存储有对呈现有待识别文本的图像进行文本实例分割得到的特征图，此时，执行主体可以直接从本地获取特征图。执行
30 主体也可以从其他存储设备（如连接的数据库、第三方数据平台、如

图 1 所示的终端设备 101、102、103 等等) 获取特征图。

步骤 202, 根据特征图构建关系图。

在本实施例中, 在得到特征图之后, 可以以特征图中的各个像素点分别作为节点, 根据像素点的空间语义特征之间的相似度构建边,

5 从而得到基于特征图构建的关系图。

其中, 像素点的空间语义特征可以包括像素点的位置特征和类别特征。像素点的位置特征可以用于表示像素点在特征图中的位置。像素点的类别特征可以用于表示像素点所属的文字类别。

文字类别可以根据实际的应用场景预先进行设置。例如, 若待识别文本为数字, 则可以预先划分 11 种文字类别, 分别用于表示 0-9 以及背景。

一般地, 像素点的位置特征可以使用像素点在特征图中的坐标(如横坐标和纵坐标)来表征。像素点的类别特征可以由一个向量来表征, 该向量可以表示像素点分别属于各个预设的文字类别的概率。

15 在构建关系图时, 若两个像素点的空间语义特征的相似度大于目标阈值, 则可以在两个像素点之间建立边。反之, 若两个像素点的空间语义特征的相似度不大于目标阈值, 则不在两个像素点之间建立边。其中, 目标阈值可以由技术人员预先设置, 也可以在关系图建立中灵活确定。例如, 可以根据各个像素点两两之间的空间语义相似度, 确定目标阈值。

空间语义特征的相似度可以灵活采用各种方法确定。例如, 可以分别计算两个像素点的位置特征的相似度和类别特征的相似度, 然后对位置特征的相似度和类别特征的相似度加权求和作为两个像素点的空间语义特征的相似度。

25 可选地, 在特征图的数目为至少两个时, 可以针对每个特征图构建对应的关系子图, 然后合并各个特征图分别对应的关系子图得到最终的关系图。关系图的合并可以按照各个特征图分别对应图像中的位置关系依次连接各个关系子图来实现。具体地, 对于两个关系子图来说, 可以从两个关系子图中分别选取一个节点进行连接来实现这两个关系子图的连接。从关系子图中选取节点的方法可以灵活设置, 如

选取根节点等。

步骤 203, 利用预先训练的图卷积网络对关系图进行处理, 得到图像对应的第一文本特征。

在本实施例中, 图卷积网络 (GCN, Graph Convolutional Networks) 一般可以看作是一种利用图拉普拉斯矩阵的切比雪夫一阶多项式近似谱卷积操作的模型。从谱图卷积的角度看, 图卷积网络可以看做一种特殊形式的图拉普拉斯平滑。图卷积网络的卷积操作可以视为将每个节点的特征信息经过变换后发送给该节点的邻居节点, 然后将邻居节点的特征信息进行融合, 以更新每个节点的特征信息。

具体地, 在利用图卷积网络更新每个节点的特征信息 (即空间语义特征) 之后, 可以更新后的各个节点的特征信息, 采用各种方法生成图像对应的第一文本特征。第一文本特征可以用于表征图像中的待识别文本的特征。

例如, 可以对各个节点的特征信息取平均或取最大值等处理, 然后将处理结果作为第一文本特征。又例如, 可以选取特征图中的目标像素点对应的节点的更新后的特征信息作为第一文本特征。其中, 目标像素点可以灵活设置。

作为示例, 目标像素点可以由技术人员预先指定, 如目标像素点可以是特征图的几何中心点。作为又一示例, 目标像素点可以是对应的节点分别与各个邻居节点之间的相似度均大于预设阈值的像素点。

步骤 204, 根据第一文本特征, 生成图像的文本识别结果。

在本实施例中, 在得到第一文本特征之后, 可以利用现有的各种文本识别方法 (如基于循环神经网络和联接时间分类器的文本识别方法等) 生成第一文本特征对应的文本识别结果作为呈现有待识别文本的图像对应的文本识别结果。

在本实施例的一些可选的实现方式中, 对呈现有待识别文本的图像进行文本实例分割得到的特征图可以表征待识别文本中的各个文字分别所在的图像区域的图像特征和与其他特征图之间的顺序特征。其中, 顺序特征可以表示各个特征图之间的顺序关系, 以便于后续识别过程中可以更准确地结合每个文字的上下文进行识别, 有助于提升识

别结果的准确度。

下面参考图 3，图 3 示出了根据本公开的实施例的用于识别文本的方法中特征图的生成方式的一个实施例的流程图 300。如图 3 所示，可以通过如下步骤生成对呈现有待识别文本的图像进行文本实例分割得到的特征图：

步骤 301，将图像输入至预先训练的卷积神经网络，得到初始特征图。

在本实施例中，卷积神经网络可以用于对图像进行卷积操作以提取图像的各种特征（如纹理特征、颜色特征等等），得到初始特征图。

10 可选地，卷积神经网络可以基于特征金字塔网络 FPN 和残差网络 ResNet 实现，且残差网络的输出层之前的至少一个卷积层的步长可以设置为 1，残差网络输出的特征图可以通过可变形卷积生成特征金字塔网络的输入特征图。

以 ResNet50 作为示例，ResNet50 一般对输入图像通过 5 个阶段的处理得到输出特征图。设输入图像依次经过 5 个阶段输出的特征图分别为 S1-S5，则可以设置阶段 4 和阶段 5 的卷积步长为 1，以使 S4 和 S5 可以保留更多关于文字纹理、文字边界等底层图像信息。

20 然后，可以对特征图 S5 进行可变形卷积处理，并将可变形卷积处理后的特征图输入 FPN，以及将 FPN 最终输出的特征图作为初始特征图。利用可变形卷积可以使得卷积神经网络更好地适应待处理文本的不规则边界，从而提升后续文本识别的准确度。

步骤 302，对初始特征图进行文本实例分割，得到实例特征图。

在本实施例中，可以利用现有的各种实例分割方法对初始特征图进行文本实例分割以得到实例特征图。

25 可选地，可以先利用 PPM（Pyramid Pooling Module）结构的网络分别提取初始特征图不同感受野的特征，然后再串联不同感受野的特征并进行维度转换。例如，可以使用并行的 1*1、3*3、5*5 的卷积层分别对初始特征图进行特征提取，然后将各个卷积层提取的特征进行串联，再使用 1*1 的卷积层进行维度转换以便于后续处理。

30 可选地，对于不同感受野的特征进行串联和维度转换等操作后，

还可以基于空间注意力机制，使用多个堆叠的卷积层进一步进行特征转换，以得到空间位置加强后的特征。例如，对于上述提取的不同感受野的特征串联和维度转换后，可以进一步使用 3×3 和 1×1 的卷积层获取空间注意力特征图。然后，再基于得到的空间位置加强后的特征进行文本实例分割。

步骤 303，对初始特征图进行文本顺序分割，得到顺序特征图。

在本实施例中，为了保持图像中的待识别文本中的各个文字的阅读顺序，可以利用现有的各种文本顺序分割方法对初始特征图进行处理，以得到顺序特征图。

例如，可以将初始特征图输入至基于卷积编码器-解码器的结构构建的网络，以对初始特征图进行简单的卷积下采样和反卷积上采样，从而得到顺序特征图。

步骤 304，融合实例特征图和顺序特征图，得到对呈现有待识别文本的图像进行文本实例分割得到的特征图。

在本实施例中，可以对得到的实例特征图和顺序特征图进行融合，以得到对呈现有待识别文本的图像进行文本实例分割得到的特征图。

在实例特征图中的文字实例为至少两个时，可以将实例特征图分别与各文字实例对应的顺序特征图进行融合，以得到至少两个融合后的特征图。

具体地，可以采用各种特征融合方法融合实例特征图和顺序特征图。例如，可以通过将实例特征图和顺序特征图中的对应像素相乘来得到实现实例特征图和顺序特征图的融合。

在本实施例的一些可选的实现方式中，基于特征图构建的关系图中的节点的空间语义特征还可以包括节点指示的像素点所在的特征图的顺序特征。

此时，对于每个特征图对应的关系子图中的节点，该节点的空间语义特征可以通过如下步骤生成：

步骤一、获取该节点对应的特征图在各个特征图中的顺序值。

在本步骤中，由于特征图是基于对呈现有待识别文本的图像进行文本实例分割得到的，因此各个特征图之间是具有顺序关系的，即对

应于分割出的各个实例之间的顺序。

顺序值可以表示特征图在各个特征图中的排序位置。例如，预设 20 张特征图，则可以将每个特征图的排序编号（1-20 之一）作为该特征图的顺序值。

5 可以理解的是，同一特征图中的各个像素点对应的顺序值相同。

步骤二、根据该节点对应的特征图的顺序值，确定该节点对应的特征图的顺序特征。

在本步骤中，特征图的顺序特征可以使用向量表示。因此，可以通过将各个顺序值映射到同一向量空间来得到各特征图分别对应的顺序特征。具体地映射方法可以采用现有的各种映射方法。

例如，可以使用不同波长的正弦函数和余弦函数对顺序值进行向量表征。作为示例，可以利用下述公式得到每个顺序值的向量表征：

$$\varepsilon_{2i}(z) = \cos\left(\frac{z}{1000^{\frac{2i}{C_\varepsilon}}}\right), i \in (0, \frac{C_\varepsilon}{2} - 1)$$

$$\varepsilon_{2i+1}(z) = \sin\left(\frac{z}{1000^{\frac{2i}{C_\varepsilon}}}\right), i \in (0, \frac{C_\varepsilon}{2} - 1)$$

15

其中， z 可以表示顺序值。 C_ε 表示向量维度，具体可以由技术人员预先设置。 i 表示向量中的元素的序号。 ε_{2i} 表示向量中的序号为偶数的元素， ε_{2i+1} 表示向量中序号为奇数的元素。

步骤三、拼接该节点指示的像素点的位置特征、类别特征和确定的顺序特征，得到该节点的空间语义特征。

在本步骤中，可以顺序拼接该节点指示的像素点的位置特征、类别特征和确定的顺序特征，得到该节点的空间语义特征。其中，像素点的位置特征和类别特征都可以使用向量进行表征。

可选地，对于每个节点，可以先将像素点的位置特征和类别特征进行向量表征，然后拼接该节点的位置特征和类别特征的向量表征和该节点的顺序特征的向量表征，得到该节点的空间语义特征。

作为示例，对于任一节点，该节点的横坐标为 X ，纵坐标为 Y ，类别特征为 P ，然后分别将 X 、 Y 和 P 映射为向量表征，然后串联 X 、

Y、P 分别对应的向量表征作为该节点的位置特征和类别特征对应的向量表征。其中，将 X、Y 和 P 映射为向量表征的方法可以采用现有的各种映射方法。

需要说明的是，为了便于计算，各个向量表征的维度可以保持一致。例如，可以使用 1×1 的卷积灵活调整各个向量表征的维度。

可选地，在拼接该节点的位置特征和类别特征对应的向量表征和该节点的顺序特征的向量表征之前，可以先利用各种标准化方法对该节点的位置特征和类别特征对应的向量表征进行标准化处理，以便于减小向量元素，从而便于后续计算。

作为示例，可以通过对该节点的向量表征减去对应特征图中的目标节点的位置特征和类别特征对应的向量表征来实现标准化。可以理解的是，目标节点的位置特征和类别特征对应的向量表征在经过标准化处理后为 0。

其中，目标节点可以由技术人员预先制定，也可以在计算过程中灵活设置。例如，目标节点可以是特征图的根节点。一般地，可以设置特征图的几何中心点为根节点。

可选地，特征图的根节点可以通过如下步骤确定：对于特征图中的节点，确定该节点分别与其他各个节点的交并比，响应于确定该节点对应的各个交并比不大于预设阈值，确定该节点为根节点。其中，交并比可以表示两个节点的邻居节点的交集所包括的元素数目与邻居节点的并集所包括的元素数目的比值。预设阈值可以由技术人员预先设置。通过这种方法可以有效地筛选出每个文字中心的像素点作为根节点，同时还可以平衡该特征图对应的关系子图中的节点分布。

在本实施例的一些可选的实现方式中，对于两个节点来说，空间语义特征的相似度可以通过如下步骤确定：

步骤一、确定两个节点之间的欧式距离，以及根据所确定的欧式距离确定第一相似度。

在本步骤中，两个节点之间的欧式距离一般与节点的空间语义特征的相似度成反比。第一相似度可以基于此采用各种方法确定。作为示例，可以利用如下公式确定第一相似度：

$$E_s(p, q) = 1 - \frac{D(p, q)}{\max(H_m, W_m)}$$

其中, p 和 q 分别为两个节点。 $D(p, q)$ 表示 p 和 q 之间的欧式距离。 H_m 和 W_m 分别表示特征图的高度和宽度。 $E_s(p, q)$ 表示第一相似度。

步骤二、确定两个节点分别对应的类别特征的余弦相似度。

- 5 在本步骤中, 可以利用现有的余弦相似度计算方法, 利用两个节点分别对应的类别特征的向量表征计算余弦相似度。

步骤三、根据第一相似度和余弦相似度, 确定两个节点的空间语义特征的相似度。

- 10 在本步骤中, 空间语义特征相似度一般可以与第一相似度成正比, 与余弦相似度也成正比。基于此可以采用各种方法确定两个节点的空间语义特征的相似度。例如, 可以直接计算第一相似度和余弦相似度的乘积作为两个节点的空间语义特征的相似度。由此, 可以全面的考虑节点之间在空间位置、类别等各方面的相似度, 从而有助于构建更准确的关系图。

- 15 在本实施例的一些可选的实现方式中, 在构建特征图对应的关系图时, 对于每个节点, 可以分别计算该节点与其他各节点之间的空间语义特征的相似度, 然后按照相似度从大到小的顺序, 以此选取目标数目个节点作为该节点的邻居节点, 即在选取的节点和该节点之间建立边。其中, 目标数目可以根据具体的应用场景灵活设置。例如, 目标数目可以为 8。由此, 可以比较灵活地控制构建的关系图的复杂度和准确度, 以辅助于后续的计算。

- 20 在本实施例的一些可选的实现方式中, 上述图卷积网络可以包括第一图卷积网络和输出网络。其中, 第一图卷积网络可以用于对构建的关系图的特征矩阵进行转换。输出网络可以用于根据第一图卷积网络的输出, 从各关系子图中分别选取节点, 以及聚合从各个关系子图中分别选取的节点对应的转换后的特征以得到第一文本特征。

- 25 对于每个关系子图, 可以使用关系图的特征矩阵和邻接矩阵来表示该关系子图。其中, 特征矩阵中的元素用于表示关系子图中的节点的特征。邻接矩阵用于表示关系子图中各节点之间的连接关系 (如是否有边等)。
- 30

第一图卷积网络可以采用现有的各种卷积网络结构以实现对关系图的特征矩阵的转换。

可选地，第一图卷积网络可以包括第一图卷积子网络和第二图卷积子网络。其中，第一图卷积子网络可以用于对关系图的特征矩阵进行转换。第二图卷积子网络可以用于根据基于第一图卷积子网络的输出构建的关系子图，对第一图卷积子网络输出的特征矩阵进行转换。

这种情况下，第一图卷积子网络对关系图的特征矩阵进行转换之后，可以按照转换后的特征矩阵中各节点的特征，重新计算节点之间的相似度并重新建立边，即更新关系子图。第二图卷积子网络可以对更新后的关系子图进行处理。

第一图卷积子网络和第二图卷积子网络的结构可以由技术人员根据实际的应用需求灵活设置。

作为示例，第一图卷积子网络的处理过程如下述公式所示：

$$Y^l = \sigma \left(\left(X^l \oplus \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \right) X^l \right) W^l \right)$$

$$\tilde{A} = A + I_N$$

$$\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$$

其中， l 为第一图卷积子网络的卷积层的数目。 Y^l 表示第 l 层的输出。 X^l 表示第 l 层的输入。 W^l 为第一图卷积子网络所学习的网络参数。 A 表示关系图的邻接矩阵。 I_N 表示与 A 相同大小的矩阵，且主对角线元素都为1。 $\tilde{D}$ 为对角阵。 i 和 j 分别表示行和列的序号。 σ 表示非线性激活函数。 $\oplus$ 代表矩阵按维度连接。

第二图卷积子网络的处理过程如下述公式所示：

$$Y_c^l = \sigma \left(\left(\tilde{D}^{-\frac{1}{2}} \tilde{A}_c \tilde{D}^{-\frac{1}{2}} \right) W_c^l \right)$$

其中， l 为第二图卷积子网络的卷积层的数目。 Y_c^l 表示第 l 层的输出。 W_c^l 为第二图卷积子网络所学习的网络参数。需要说明的是，其中的 $\tilde{A}_c$ 和 $\tilde{D}$ 都是基于更新后的关系图得到的矩阵，具体的计算方法如上述第一图卷积子网络中的 $\tilde{A}$ 和 $\tilde{D}$ 相同。

具体地，输出网络可以根据第一图卷积网络的输出所表示的对关系图中的各节点的特征信息的更新，重新计算节点与节点之间的距离，然后按照距离从小到大的距离丢弃部分节点（如丢弃一半节点），从而

实现池化操作，减小对应的特征图的大小。

可选地，可以迭代执行上述第一图卷积网络和输出网络的特征更新和筛选节点的过程，直至每个关系子图中仅剩一个节点。然后可以按照关系子图之间的顺序，依次拼接每个关系子图中筛选出的节点的特征信息形成第一文本特征。

通过第一图卷积网络和输出网络的特征更新和筛选节点的过程的迭代，可以从每个关系子图中筛选出空间关系中最能表征对应文字实例的节点，从而利用该节点的特征信息进行后续文本识别，有助于提升文本识别的效率和准确度。

本公开的上述实施例提供的方法提出了一种基于图的文本识别方法。具体通过对文本实例分割得到的特征图构建关系图，以使用图结构来表达文本的空间语义信息，然后对关系图进行图卷积处理，以提取待识别文本在图像中的二维空间特征信息，并结合各个文字实例之间的顺序关系，实现对图像中的待识别文本的识别。

下面参考图 4，其示出了根据本公开的用于识别文本的方法的一个又一个实施例的流程 400。该用于识别文本的方法包括以下步骤：

步骤 401，获取对呈现有待识别文本的图像进行文本实例分割得到的特征图。

步骤 402，根据特征图构建关系图。

步骤 403，利用预先训练的图卷积网络对关系图进行处理，得到图像对应的第一文本特征。

步骤 404，利用预先训练的语言模型对特征图进行处理，得到图像对应的第二文本特征。

在本实施例中，语言模型可以是现有的各种用于文本识别的语言模型。如 N-Gram 模型、HMM (Hidden Markov Model, 隐马尔科夫模型)、BERT (Bidirectional Encoder Representations from Transformers) 等等。一般地，可以选取语言模型在输出层之前生成的文字序列的特征表示作为第二文本特征。

可选地，可以先利用预先训练的语义特征提取网络对特征图进行

处理，得到特征图对应的语义特征。然后将语义特征输入至所述语言模型，得到所述第二文本特征。其中，语义特征提取网络可以采用现有的各种特征提取网络的结构。

作为示例，语义特征提取网络可以包括对特征图进行池化处理的池化层和对池化层的输出进行线性变换的线性层。具体地，先对输入的特征图进行池化操作，然后对池化结果进行线性变换，以降低图像空间的分辨率并生成语义特征。

语言模型可以采用现有的各种基于自然语言处理的模型的结构。例如，翻译模型等以利用每个文字实例前后若干个文字实例来预测该文字的语义，从而实现文本识别。

步骤 405，根据第一文本特征和第二文本特征，生成图像的文本识别结果。

在本实施例中，可以结合第一文本特征和第二文本特征，采用各种方法生成图像的文本识别结果。例如，可以采用现有的各种特征融合方法融合第一文本特征和第二文本特征，得到融合后的文本特征，然后对融合后的文本特征进行识别得到识别结果。

由于语言模型的文本识别过程是使用一维压缩的特征信息，忽略了二维空间特征信息，因此，通过结合基于图的文本识别方法和基于语言模型的文本识别方法，以利用更丰富的特征信息实现更可靠的文本识别。

在本实施例的一些可选的实现方式中，根据第一文本特征、第二文本特征和特征图，生成图像的文本识别结果。

由于特征图本身就可以表征待识别文本的图像特征，因此，在利用第一文本特征和第二文本特征进行文本识别的同时，结合通过文本实例分割得到的特征图，可以进一步增强文本特征的代表能力，从而提升文本识别效果。

具体的识别方法可以根据实际的应用需求灵活设置。例如，可以先利用现有的各种特征融合方法先融合第一文本特征、第二文本特征和特征图，然后利用融合后的特征预测文本识别结果。

可选地，可以将第一文本特征、第二文本特征和特征图输入至预

先训练的特征融合网络，生成图像的文本识别结果。

其中，特征融合网络可以用于拼接第一文本特征、第二文本特征和特征图，然后对得到的拼接结果进行线性变换，以得到图像的文本识别结果。特征融合网络的结构可以由技术人员预先设置。

5 作为示例，特征融合网络的处理过程如下述公式所示：

$$f_t = W_t \odot (W_f \cdot [v_t; l_t; g_t])$$

$$W_t = \text{Sigmoid}(W_z \cdot [v_t; l_t; g_t])$$

10 其中， f_t 表示特征融合网络的融合结果。 v_t 、 l_t 和 g_t 分别表示特征图、第二文本特征和第一文本特征。 W_z 和 W_f 表示特征融合网络要学习的线性变换参数。 $\odot$ 表示元素间的乘法运算。“;”表示按照维度拼接操作。Sigmoid 为激活函数。 t 表示对呈现有待识别文本的图像进行文本实例分割得到的特征图的数目。

需要说明的是，特征图是具有顺序关系的，而第一文本特征和第二文本特征也是基于这种顺序关系生成的文字序列的特征表示，因此，
15 每次融合都是将特征图、第二文本特征和第一文本特征中对应各个文字特征进行融合，从而得到最终的融合结果。

另外，还需要说明的是，由于对应的文字实例之间是具有上下文顺序关系的，因此特征图之间也具有上下文顺序关系，但本公开在对特征图的处理过程中，可以并行地识别各特征图分别对应的文字实例，
20 以提升处理效率。

本实施例中未具体说明的内容可参考图 2 对应实施例中的相关说明，在此不再赘述。

25 继续参考图 5a、图 5b 和图 5c 是根据本实施例的用于识别文本的方法的一个示例性的应用场景的示意图。具体如图 5a 所示，可以先获取呈现有待识别文本的图像，即文本图像，然后将文本图像输入至预先训练的卷积网络以提取初始特征图，然后，分别将初始特征图输入至文本实例分割网络和文本顺序分割网络，以提取实例特征图和顺序特征图，再融合得到的实例特征图和顺序特征图，形成文本图像对应的多个特征图。

30 之后，可以根据得到的多个特征图构建关系图，再利用预先训练

的图卷积网络对构建的关系图进行处理，得到文本图像的第一文本特征。同时，利用语言模型对得到的多个特征图进行处理，得到文本图像的第二文本特征。然后，综合得到的多个特征图、第一文本特征和第二文本特征进行文本识别，得到文本图像对应的文本识别结果。

- 5 其中，利用图卷积网络和语言模型对得到的多个特征图进行处理的过程具体可以参考图 5b。具体地，可以先构建每个特征图对应的关系子图，然后合并各个关系子图得到关系图，再利用图卷积网络对合并得到的关系图进行处理，得到关系图中的节点的特征表示。同时，可以先将特征图输入语义特征提取网络以提取语义特征，然后利用翻译模型根据语义特征形成待识别文本对应的文字序列的特征表示。
- 10

之后，为了融合得到的关系图中的节点的特征表示和文字序列的特征表示，可以利用线性层进行维度变换等处理，以融合关系图中的节点的特征表示和文字序列的特征表示，并根据融合结果生成文字图像的文本识别结果。

- 15 其中，针对每个特征图构建对应的关系子图的具体过程可以参考图 5c。具体地，可以先确定特征图的顺序特征，以及其中每个像素点的位置特征和类别特征。然后，可以融合顺序特征和位置特征并通过映射等方法形成各节点的向量表征，从而得到由各节点的向量表征组成的关系子图的特征矩阵。同时，可以根据类别特征确定各像素点指示的节点之间的相似度，并根据节点之间的相似度为每个节点搜索邻居节点以构建边，从而形成关系子图的邻接矩阵。之后，即可以使用得到的特征矩阵和邻接矩阵来表示特征子图对应的关系子图。
- 20

- 本公开的上述实施例提供的方法通过基于图的文本识别方法和基于语言模型的文本识别方法地动态融合，使得两种识别方法可以互相学习，以利用更多的信息来进行文本识别，从而提升文本识别效果，以更好地适应自然场景文本识别中图片复杂背景、光照不均、图片模糊、文字形态各异等多样的文字识别情况。本公开上述实施例描述的方法可应用于辅助视觉障碍的导航、自动驾驶、扩增现实中的文本阅读和翻译等实际业务场景中的文本识别。对于本公开实施例中的方法还可以应用的其他实际业务场景，本说明书不再一一列举。
- 25
- 30

下面参考图 6，其示出了根据本公开的用于识别文本的方法中的图卷积网络、语言模型和分割网络的训练方法的一个实施例的流程 600。如图 6 所示，图卷积网络、语言模型和分割网络可以通过如下步骤训练得到：

步骤 601，获取有标注训练数据集和无标注训练数据集。

在本实施例中，需要说明的是，图卷积网络、语言模型和分割网络的训练可以由上述用于识别文本的方法的执行主体执行，也可以由其他电子设备执行。训练图卷积网络、语言模型和分割网络的执行主体可以从本地或其他存储设备获取有标注训练数据集和无标注训练数据集。

其中，有标注训练数据集由具有标注的训练数据组成，无标注训练数据集则由无标注的训练数据组成。有标注训练数据集中可以包括不同粒度的标注。例如，对于字符类文本的识别，标注可以包括字符集标注和单词级标注。

步骤 602，构建图卷积网络、语言模型和分割网络对应的教师-学生网络，以及利用有标注训练数据集、无标注训练数据集和预设的损失函数，对教师-学生网络进行训练。

在本实施例中，教师-学生网络是迁移学习中的一种网络结构。教师网络和学生网络的结构可以完全相同，也可以不同。在教师网络和学生网络的结构相同时，可以基于 EMA（指数移动平均数等）技术实现教师网络和学生网络的结构参数共享。

在本实施例中，学生网络的输入可以包括有标注训练数据和无标注训练数据，教师网络的输入可以仅包括无标注训练数据。损失函数可以包括第一损失函数、第二损失函数和第三损失函数。

其中，第一损失函数和第二损失函数可以分别表示学生网络针对有标注训练数据的输出结果与不同粒度的标注之间的差异。例如，第一损失函数可以用于衡量学生网络输出的字符级识别结果与真实的字符级标注之间的差异，第二损失函数可以用于衡量学生网络输出的单词级识别结果与真实的单词级标注之间的差异。

第三损失函数可以表示学生网络和教师网络分别针对无标注训练数据的输出结果之间的差异。具体地，可以综合第一损失函数、第二损失函数和第三损失函数（如优化三个损失函数的和）来调节图卷积网络、语言模型和分割网络的参数，以完成图卷积网络、语言模型和分割网络的训练。

另外，需要说明的是，本公开中的各种网络或模型的训练都可以基于现有的各种机器学习的训练方法完成，对此不再赘述。

本公开的上述实施例提供的针对图卷积网络、语言模型和分割网络的训练方法利用有标注训练数据和无标注训练数据共同训练图卷积网络、语言模型和分割网络，解决了实际中真实训练数据少和真实数据难标注等问题，也可以提升由图卷积网络、语言模型和分割网络组成的整体网络的泛化性和鲁棒性，从而有助于提升文本识别结果准确性。

进一步参考图 7，作为对上述各图所示方法的实现，本公开提供了用于识别文本的装置的一个实施例，该装置实施例与图 2 所示的方法实施例相对应，该装置具体可以应用于各种电子设备中。

如图 7 所示，本实施例提供的用于识别文本的装置 700 包括特征图获取单元 701、关系图构建单元 702、图卷积处理单元 703 和识别单元 704。其中，特征图获取单元 701 被配置成获取特征图，其中，特征图通过对呈现有待识别文本的图像进行文本实例分割得到；关系图构建单元 702 被配置成根据特征图构建关系图，其中，关系图中的节点表示特征图中的像素点，边表示所连接的两个节点的空间语义特征的相似度大于目标阈值，空间语义特征包括节点指示的像素点的位置特征和类别特征；图卷积处理单元 703 被配置成利用预先训练的图卷积网络对关系图进行处理，得到图像对应的第一文本特征；识别单元 704 被配置成根据第一文本特征，生成图像的文本识别结果。

在本实施例中，用于识别文本的装置 700 中：特征图获取单元 701、关系图构建单元 702、图卷积处理单元 703 和识别单元 704 的具体处理及其所带来的技术效果可分别参考图 2 对应实施例中的步骤 201、

步骤 202、步骤 203 和步骤 204 的相关说明，在此不再赘述。

在本实施例的一些可选的实现方式中，用于识别文本的装置 700 还包括：语言模型处理单元（图中未示出），被配置成利用预先训练的语言模型对特征图进行处理，得到图像对应的第二文本特征；以及上述识别单元 704 进一步被配置成根据第一文本特征和第二文本特征，生成图像的文本识别结果。

在本实施例的一些可选的实现方式中，上述识别单元 704 进一步被配置成：根据第一文本特征、第二文本特征和特征图，生成图像的文本识别结果。

在本实施例的一些可选的实现方式中，特征图为至少两个特征图；以及上述关系图构建单元 702 进一步被配置成：对于至少两个特征图中的特征图，构建该特征图对应的关系子图；合并各个特征图分别对应的关系子图，得到关系图。

在本实施例的一些可选的实现方式中，至少两个特征图中的特征图用于表征待识别文本中的各个文字分别所在的图像区域的图像特征和与其他特征图之间的顺序特征。

在本实施例的一些可选的实现方式中，特征图通过如下步骤生成：将图像输入至预先训练的卷积神经网络，得到初始特征图；对初始特征图进行文本实例分割，得到实例特征图；对初始特征图进行文本顺序分割，得到顺序特征图；融合实例特征图和顺序特征图，得到对呈现有待识别文本的图像进行文本实例分割得到的特征图。

在本实施例的一些可选的实现方式中，卷积神经网络基于特征金字塔网络和残差网络实现，其中，残差网络的输出层之前的至少一个卷积层的步长为 1，残差网络输出的特征图通过可变形卷积生成特征金字塔网络的输入特征图。

在本实施例的一些可选的实现方式中，上述语言模型处理单元进一步被配置成：利用预先训练的语义特征提取网络对特征图进行处理，得到特征图对应的语义特征，其中，语义特征提取网络包括对特征图进行池化处理的池化层和对池化层的输出进行线性变换的线性层；将语义特征输入至语言模型，得到第二文本特征。

在本实施例的一些可选的实现方式中，空间语义特征还包括节点指示的像素点所在的特征图的顺序特征；以及对于关系子图中的节点，该节点的空间语义特征通过如下步骤生成：获取该节点对应的特征图在至少两个特征图中的顺序值；根据顺序值，确定该节点对应的特征图的顺序特征；拼接该节点指示的像素点的位置特征、类别特征和确定的顺序特征，得到该节点的空间语义特征。

在本实施例的一些可选的实现方式中，两个节点的空间语义特征的相似度通过如下步骤确定：确定两个节点之间的欧式距离，以及根据所确定的欧式距离确定第一相似度；确定两个节点分别对应的类别特征的余弦相似度；根据第一相似度和余弦相似度，确定两个节点的空间语义特征的相似度。

在本实施例的一些可选的实现方式中，上述关系图构建单元 702 进一步被配置成：按照各个关系子图之间的顺序关系，依次连接相邻关系子图的根节点。

在本实施例的一些可选的实现方式中，特征图的根节点通过如下步骤确定：对于特征图中的节点，确定该节点分别与其他各个节点的交并比，其中，交并比表示两个节点的邻居节点的交集所包括的元素数目与邻居节点的并集所包括的元素数目的比值；响应于确定该节点对应的各个交并比不大于预设阈值，确定该节点为根节点。

在本实施例的一些可选的实现方式中，图卷积网络包括第一图卷积网络和输出网络，其中，第一图卷积网络用于对关系图的特征矩阵进行转换，输出网络用于根据第一图卷积网络的输出，从各关系子图中分别选取节点，以及聚合从各个关系子图中分别选取的节点对应的转换后的特征以得到第一文本特征。

在本实施例的一些可选的实现方式中，第一图卷积网络包括第一图卷积子网络和第二图卷积子网络，其中，第一图卷积子网络用于对关系图的特征矩阵进行转换，第二图卷积子网络用于根据基于第一图卷积子网络的输出构建的关系图，对第一图卷积子网络输出的特征矩阵进行转换。

在本实施例的一些可选的实现方式中，上述识别单元 704 进一步

被配置成：将第一文本特征、第二文本特征和特征图输入至预先训练的特征融合网络，生成图像的文本识别结果，特征融合网络用于拼接第一文本特征、第二文本特征和特征图，以及对得到的拼接结果进行线性变换，得到图像的文本识别结果。

- 5 在本实施例的一些可选的实现方式中，对初始特征图进行文本实例分割，得到实例特征图，包括：利用预先训练的文本实例分割网络对初始特征图进行文本实例分割，得到实例特征图；以及对初始特征图进行文本顺序分割，得到顺序特征图，包括：利用预先训练的文本顺序分割网络对初始特征图进行文本顺序分割，得到顺序特征图；以及上述图卷积网络、语言模型和分割网络通过如下步骤训练得到，其中，分割网络包括卷积神经网络、文本实例分割网络和文本顺序分割网络：获取有标注训练数据集和无标注训练数据集，其中，有标注训练数据集中的训练数据包括不同粒度的标注；构建图卷积网络、语言模型和分割网络对应的教师-学生网络，以及利用有标注训练数据集、
10 无标注训练数据集和预设的损失函数，对教师-学生网络进行训练，其中，学生网络的输入包括有标注训练数据和无标注训练数据，教师网络的输入包括无标注训练数据，损失函数包括第一损失函数、第二损失函数和第三损失函数，第一损失函数和第二损失函数分别表示学生网络针对有标注训练数据的输出结果与不同粒度的标注之间的差异，
15 第三损失函数表示学生网络和教师网络分别针对无标注训练数据的输出结果之间的差异。
20 第三损失函数表示学生网络和教师网络分别针对无标注训练数据的输出结果之间的差异。

- 本公开的上述实施例提供的装置，通过特征图获取单元获取对呈现有待识别文本的图像进行文本实例分割得到的特征图；关系图构建单元根据特征图构建关系图，其中，关系图中的节点表示特征图中的
25 像素点，边表示所连接的两个节点的空间语义特征的相似度大于目标阈值，空间语义特征包括节点指示的像素点的位置特征和类别特征；图卷积处理单元利用预先训练的图卷积网络对关系图进行处理，得到图像对应的第一文本特征；识别单元根据第一文本特征，生成图像的文本识别结果，实现了基于图的文本识别，这种方式可以考虑到图像
30 中的文本的二维空间信息，有助于提升文本识别效果。

下面参考图 8，其示出了适于用来实现本公开的实施例的电子设
备（例如图 1 中的服务器）800 的结构示意图。本公开的实施例中的
终端设备可以包括但不限于诸如移动电话、笔记本电脑、数字广播接
5 收器、PDA（个人数字助理）、PAD（平板电脑）、PMP（便携式多媒体
播放器）、车载终端（例如车载导航终端）等等的移动终端以及诸如
数字 TV、台式计算机等等的固定终端。图 8 示出的服务器仅仅是一个
示例，不应对本公开的实施的功能和使用范围带来任何限制。

如图 8 所示，电子设备 800 可以包括处理装置（例如中央处理器、
10 图形处理器等）801，其可以根据存储在只读存储器（ROM）802 中的
程序或者从存储装置 808 加载到随机访问存储器（RAM）803 中的程
序而执行各种适当的动作和处理。在 RAM 803 中，还存储有电子设备
800 操作所需的各种程序和数据。处理装置 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
15 804。

通常，以下装置可以连接至 I/O 接口 805：包括例如触摸屏、触摸
板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置
806；包括例如液晶显示器（LCD）、扬声器、振动器等等的输出装置 807；
包括例如磁带、硬盘等的存储装置 808；以及通信装置 809。通信装置
20 809 可以允许电子设备 800 与其他设备进行无线或有线通信以交换数
据。虽然图 8 示出了具有各种装置的电子设备 800，但是应理解的是，
并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或
更少的装置。图 8 中示出的每个方框可以代表一个装置，也可以根据
需要代表多个装置。

25 特别地，根据本公开的实施例，上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，本公开的实施例包括一种计算机程
序产品，其包括承载在计算机可读介质上的计算机程序，该计算机程
序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，
该计算机程序可以通过通信装置 809 从网络上被下载和安装，或者从
30 存储装置 808 被安装，或者从 ROM 802 被安装。在该计算机程序被处

理装置 801 执行时，执行本公开的实施例的方法中限定的上述功能。

需要说明的是，本公开的实施例所述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开的实施例中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开的实施例中，计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读信号介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：电线、光缆、RF（射频）等等，或者上述的任意合适的组合。

上述计算机可读介质可以是上述电子设备中所包含的；也可以是单独存在，而未装配入该电子设备中。上述计算机可读介质承载有一个或者多个程序，当上述一个或者多个程序被该电子设备执行时，使得该电子设备：获取特征图，其中，特征图通过对呈现有待识别文本的图像进行文本实例分割得到；根据特征图构建关系图，其中，关系图中的节点表示特征图中的像素点，边表示所连接的两个节点的空间语义特征的相似度大于目标阈值，空间语义特征包括节点指示的像素点的位置特征和类别特征；利用预先训练的图卷积网络对关系图进行处理，得到图像对应的第一文本特征；根据第一文本特征，生成图像

的文本识别结果。

可以以一种或多种程序设计语言或其组合来编写用于执行本公开的实施例的操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如 Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网（LAN）或广域网（WAN）——连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

附图中的流程图和框图，图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分，该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本公开的实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中，例如，可以描述为：一种处理器包括特征图获取单元、关系图构建单元、图卷积处理单元和识别单元。其中，这些单元的名称在某种情况下并不构成对该单元本身的限定，例如，特征图获取单元还可以被描述为“获取特征图，其中，特征图通过对呈现有待识别文本的图像进行文本实例分割得到”。

以上描述仅为本公开的较佳实施例以及对所运用技术原理的说

- 明。本领域技术人员应当理解，本公开的实施例中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本公开的实施例中公开的（但不限于）具有类似功能的技术特征进行互相替换而形成的技术方案。
- 5

权 利 要 求 书

1、一种用于识别文本的方法，包括：

5 获取特征图，其中，所述特征图通过对呈现有待识别文本的图像进行文本实例分割得到；

根据所述特征图构建关系图，其中，所述关系图中的节点表示所述特征图中的像素点，边表示所连接的两个节点的空间语义特征的相似度大于目标阈值，空间语义特征包括节点指示的像素点的位置特征和类别特征；

10 利用预先训练的图卷积网络对所述关系图进行处理，得到所述图像对应的第一文本特征；以及

根据所述第一文本特征，生成所述图像的文本识别结果。

2、根据权利要求 1 所述的方法，其中，所述方法还包括：

15 利用预先训练的语言模型对所述特征图进行处理，得到所述图像对应的第二文本特征；以及

所述根据所述第一文本特征，生成所述图像的文本识别结果，包括：

20 根据所述第一文本特征和第二文本特征，生成所述图像的文本识别结果。

3、根据权利要求 2 所述的方法，其中，所述根据所述第一文本特征和第二文本特征，生成所述图像的文本识别结果，包括：

25 根据所述第一文本特征、第二文本特征和特征图，生成所述图像的文本识别结果。

4、根据权利要求 3 所述的方法，其中，所述特征图为至少两个特征图；以及

所述根据所述特征图构建关系图，包括：

30 对于所述至少两个特征图中的特征图，构建该特征图对应的关系

子图；以及

合并各个特征图分别对应的关系子图，得到所述关系图。

5 5、根据权利要求4所述的方法，其中，所述至少两个特征图中的特征图用于表征所述待识别文本中的各个文字分别所在的图像区域的图像特征和与其他特征图之间的顺序特征。

6、根据权利要求5所述的方法，其中，所述特征图通过如下步骤生成：

10 10 将所述图像输入至预先训练的卷积神经网络，得到初始特征图；
对所述初始特征图进行文本实例分割，得到实例特征图；
对所述初始特征图进行文本顺序分割，得到顺序特征图；以及
融合所述实例特征图和顺序特征图，得到所述对呈现有待识别文本的图像进行文本实例分割得到的特征图。

15

7、根据权利要求6所述的方法，其中，所述卷积神经网络基于特征金字塔网络和残差网络实现，其中，所述残差网络的输出层之前的至少一个卷积层的步长为1，所述残差网络输出的特征图通过可变形卷积生成所述特征金字塔网络的输入特征图。

20

8、根据权利要求2所述的方法，其中，所述利用预先训练的语言模型对所述特征图进行处理，得到所述图像对应的第二文本特征，包括：

25 25 利用预先训练的语义特征提取网络对所述特征图进行处理，得到所述特征图对应的语义特征，其中，所述语义特征提取网络包括对所述特征图进行池化处理的池化层和对所述池化层的输出进行线性变换的线性层；以及

将所述语义特征输入至所述语言模型，得到所述第二文本特征。

30 9、根据权利要求4所述的方法，其中，所述空间语义特征还包括

节点指示的像素点所在的特征图的顺序特征；以及

对于关系子图中的节点，该节点的空间语义特征通过如下步骤生成：

- 获取该节点对应的特征图在所述至少两个特征图中的顺序值；
- 5 根据所述顺序值，确定该节点对应的特征图的顺序特征；以及
- 拼接该节点指示的像素点的位置特征、类别特征和确定的顺序特征，得到该节点的空间语义特征。

- 10 10、根据权利要求 4 所述的方法，其中，两个节点的空间语义特征的相似度通过如下步骤确定：

确定两个节点之间的欧式距离，以及根据所确定的欧式距离确定第一相似度；

确定两个节点分别对应的类别特征的余弦相似度；以及

- 15 根据所述第一相似度和余弦相似度，确定两个节点的空间语义特征的相似度。

11、根据权利要求 4 所述的方法，其中，所述合并各个特征图分别对应的关系子图，包括：

- 20 按照各个关系子图之间的顺序关系，依次连接相邻关系子图的根节点。

12、根据权利要求 11 所述的方法，其中，特征图的根节点通过如下步骤确定：

- 25 对于所述特征图中的节点，确定该节点分别与其他各个节点的交并比，其中，交并比表示两个节点的邻居节点的交集所包括的元素数目与邻居节点的并集所包括的元素数目的比值；以及

响应于确定该节点对应的各个交并比不大于预设阈值，确定该节点为根节点。

- 30 13、根据权利要求 4 所述的方法，其中，所述图卷积网络包括第

一图卷积网络和输出网络，其中，第一图卷积网络用于对所述关系图的特征矩阵进行转换，输出网络用于根据第一图卷积网络的输出，从各关系子图中分别选取节点，以及聚合从各个关系子图中分别选取的节点对应的转换后的特征以得到所述第一文本特征。

5

14、根据权利要求 13 所述的方法，其中，所述第一图卷积网络包括第一图卷积子网络和第二图卷积子网络，其中，第一图卷积子网络用于对所述关系图的特征矩阵进行转换，第二图卷积子网络用于根据基于所述第一图卷积子网络的输出构建的关系图，对所述第一图卷积子网络输出的特征矩阵进行转换。

15、根据权利要求 3 所述的方法，其中，所述根据所述第一文本特征、第二文本特征和特征图，生成所述图像的文本识别结果，包括：
将所述第一文本特征、第二文本特征和特征图输入至预先训练的特征融合网络，生成所述图像的文本识别结果，特征融合网络用于拼接所述第一文本特征、第二文本特征和特征图，以及对得到的拼接结果进行线性变换，得到所述图像的文本识别结果。

16、根据权利要求 6 所述的方法，其中，所述对所述初始特征图进行文本实例分割，得到实例特征图，包括：利用预先训练的文本实例分割网络对所述初始特征图进行文本实例分割，得到实例特征图；以及

所述对所述初始特征图进行文本顺序分割，得到顺序特征图，包括：利用预先训练的文本顺序分割网络对所述初始特征图进行文本顺序分割，得到顺序特征图；以及

所述图卷积网络、语言模型和分割网络通过如下步骤训练得到，其中，所述分割网络包括所述卷积神经网络、文本实例分割网络和文本顺序分割网络：

获取有标注训练数据集和无标注训练数据集，其中，有标注训练数据集中的训练数据包括不同粒度的标注；以及

构建所述图卷积网络、语言模型和分割网络对应的教师-学生网络，以及利用所述有标注训练数据集、无标注训练数据集和预设的损失函数，对所述教师-学生网络进行训练，其中，学生网络的输入包括有标注训练数据和无标注训练数据，教师网络的输入包括无标注训练数据，所述损失函数包括第一损失函数、第二损失函数和第三损失函数，第一损失函数和第二损失函数分别表示学生网络针对有标注训练数据的输出结果与不同粒度的标注之间的差异，第三损失函数表示学生网络和教师网络分别针对无标注训练数据的输出结果之间的差异。

10 17、一种用于识别文本的装置，其中，所述装置包括：

特征图获取单元，被配置成获取特征图，其中，所述特征图通过对呈现有待识别文本的图像进行文本实例分割得到；

关系图构建单元，被配置成根据所述特征图构建关系图，其中，所述关系图中的节点表示所述特征图中的像素点，边表示所连接的两个节点的空间语义特征的相似度大于目标阈值，空间语义特征包括节点指示的像素点的位置特征和类别特征；

图卷积处理单元，被配置成利用预先训练的图卷积网络对所述关系图进行处理，得到所述图像对应的第一文本特征；

20 识别单元，被配置成根据所述第一文本特征，生成所述图像的文本识别结果。

18、一种电子设备，包括：

一个或多个处理器；

存储装置，其上存储有一个或多个程序；

25 当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求 1-16 中任一所述的方法。

19、一种计算机可读介质，其上存储有计算机程序，其中，该程序被处理器执行时实现如权利要求 1-16 中任一所述的方法。

20. 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 1-16 中任一项所述的方法。