

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2017年9月8日 (08.09.2017)



(10) 国际公布号
WO 2017/148265 A1

- (51) 国际专利分类号:
G06K 9/34 (2006.01)
- (21) 国际申请号: PCT/CN2017/073718
- (22) 国际申请日: 2017年2月16日 (16.02.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610112463.6 2016年2月29日 (29.02.2016) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 英属开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。
- (72) 发明人: 周文猛 (ZHOU, Wenmeng); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。程孟力 (CHENG, Mengli); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。毛旭东 (MAO, Xudong); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴

巴集团法务部, Zhejiang 311121 (CN)。施兴 (SHI, Xing); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。褚巍 (CHU, Wei); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

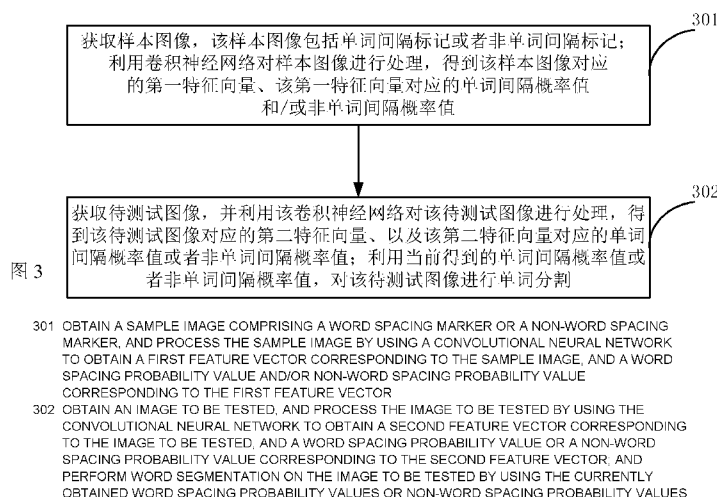
(74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

[见续页]

(54) Title: WORD SEGMENTATION METHOD AND APPARATUS

(54) 发明名称: 一种单词的分割方法和装置



(57) Abstract: A word segmentation method and apparatus. The method comprises: obtaining a sample image comprising a word spacing marker or a non-word spacing marker; processing the sample image by using a convolutional neural network to obtain a first feature vector corresponding to the sample image, and a word spacing probability value and/or non-word spacing probability value corresponding to the first feature vector (301); obtaining an image to be tested and processing the image to be tested by using the convolutional neural network to obtain a second feature vector corresponding to the image to be tested, and a word spacing probability value or a non-word spacing probability value corresponding to the second feature vector; and performing word segmentation on the image to be tested by using the currently obtained word spacing probability values or non-word spacing probability values (302). The technical solution implements precise word segmentation so that word segmentation efficiency is improved, and increases the speed of word segmentation, thereby improving user experience.

(57) 摘要:

[见续页]



WO 2017/148265 A1



(84) **指定国** (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ,

CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种单词的分割方法和装置, 该方法包括: 获取样本图像, 所述样本图像包括单词间隔标记或者非单词间隔标记; 利用卷积神经网络对所述样本图像进行处理, 得到所述样本图像对应的第一特征向量、所述第一特征向量对应的单词间隔概率值和/或非单词间隔概率值 (301); 获取待测试图像, 并利用所述卷积神经网络对所述待测试图像进行处理, 得到所述待测试图像对应的第二特征向量、所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值; 利用当前得到的单词间隔概率值或者非单词间隔概率值, 对所述待测试图像进行单词分割 (302)。通过该技术方案, 精准的进行单词分割, 使得单词分割的准确率得到提高, 提高单词分割的速度, 提高用户使用体验。

一种单词的分割方法和装置

本申请要求 2016 年 02 月 29 日递交的申请号为 201610112463.6、发明名称为“一种单词的分割方法和装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5

技术领域

本申请涉及互联网技术领域，尤其涉及一种单词的分割方法和装置。

背景技术

10

OCR（Optical Character Recognition，光学字符识别）是指使用电子设备检查纸上打印的字符，并通过检测暗、亮的模式确定其形状，然后使用字符识别方法将形状翻译成计算机文字的过程；即，针对印刷体字符，采用光学的方式将纸质文档中的文字转换为黑白点阵的图像文件，并通过识别软件将该图像中的文字转换成文本格式，以供文字处理软件进一步进行编辑加工的技术。因此，基于 OCR 技术，可以将文本图像转换成数字形式。

15

尽管 OCR 技术已经较为成熟，但由于图像背景复杂，图像中的文字字体、大小及颜色等变化繁多，再加上拍摄视角的变化，以及不同光照的影响等诸多因素，使得快速、准确地检测出图像中的文字区域的难度较大。

20

发明内容

本申请提供一种单词的分割方法和装置，以对图像进行单词分割，并快速、准确地检测出图像中的文字区域，提高用户使用体验。技术方案如下：

本申请提供一种单词的分割方法，所述方法包括以下步骤：

25

获取样本图像，所述样本图像包括单词间隔标记或者非单词间隔标记；利用卷积神经网络对所述样本图像进行处理，得到所述样本图像对应的第一特征向量、所述第一特征向量对应的单词间隔概率值和/或非单词间隔概率值；

30

获取待测试图像，并利用所述卷积神经网络对所述待测试图像进行处理，得到所述待测试图像对应的第二特征向量、所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值；利用当前得到的单词间隔概率值或者非单词间隔概率值，对所述待测试图像进行单词分割。

所述获取样本图像，包括：从字体库中抽取至少两个字母图像，并对所述至少两个字母图像进行拼接，得到拼接图像，并为所述拼接图像设定单词间隔标记或者非单词间隔标记；将所述拼接图像标记为所述样本图像；或者，对所述拼接图像进行以下操作之一或者任意组合：平移操作、旋转操作、拉伸操作、添加高斯噪声，并将进行所述操作之后的图像标记为所述样本图像。

所述方法进一步包括：在利用卷积神经网络对所述样本图像进行处理之前，将所述样本图像转换成灰度的样本图像，并将所述灰度的样本图像转换成大小为 $M*N$ 的样本图像，并对所述大小为 $M*N$ 的样本图像进行归一化处理；在利用所述卷积神经网络对所述待测试图像进行处理之前，将所述待测试图像转换成灰度的待测试图像，并将所述灰度的待测试图像转换成大小为 $M*L$ 的待测试图像，并对所述大小为 $M*L$ 的待测试图像进行归一化处理；

其中，所述 M 为正整数、所述 N 为正整数、所述 L 为正整数，且所述 M 和所述 L 的比值，与所述待测试图像的长宽比相同。

利用所述卷积神经网络对所述待测试图像进行处理，得到所述待测试图像对应的第二特征向量，包括：

将所述待测试图像向左平移 K 次或者向右平移 K 次，每次平移时的平移位置不同，以得到 $K+1$ 个不同的待测试图像，所述 K 为正整数；

利用卷积神经网络对 $K+1$ 个不同的待测试图像进行处理，得到 $K+1$ 个不同的待测试图像分别对应的特征向量，并将所述 $K+1$ 个不同的待测试图像分别对应的特征向量组合在一起，得到所述待测试图像对应的第二特征向量。

所述得到所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值的过程，具体包括：

利用所述卷积神经网络获得所述第二特征向量匹配的多个第一特征向量，并将所述多个第一特征向量对应的单词间隔概率值或者非单词间隔概率值，作为所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值。

利用当前得到的单词间隔概率值或者非单词间隔概率值，对所述待测试图像进行单词分割，包括：

将所述第二特征向量对应的待测试图像位置作为横坐标，将所述第二特征向量对应的单词间隔概率值作为纵坐标，得到概率值结果图；从所述概率值结果图中找到单词间隔概率值的极大值位置，并利用所述极大值位置对所述待测试图像进行单词分割；或者，

将所述第二特征向量对应的待测试图像位置作为横坐标，将所述第二特征向量对应的非单词间隔概率值作为纵坐标，得到概率值结果图；从所述概率值结果图中找到非单词间隔概率值的极小值位置，并利用所述极小值位置对所述待测试图像进行单词分割。

本申请提供一种单词的分割装置，所述装置具体包括：

- 5 训练模块，用于获取样本图像，所述样本图像包括单词间隔标记或者非单词间隔标记；利用卷积神经网络对所述样本图像进行处理，得到所述样本图像对应的第一特征向量、所述第一特征向量对应的单词间隔概率值和/或非单词间隔概率值；

预测模块，用于获取待测试图像，并利用所述卷积神经网络对所述待测试图像进行处理，得到所述待测试图像对应的第二特征向量、所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值；利用当前得到的单词间隔概率值或者非单词间隔概率值，
10 对所述待测试图像进行单词分割。

所述训练模块，具体用于在获取所述样本图像的过程中，从字体库中抽取至少两个字母图像，并对所述至少两个字母图像进行拼接，得到拼接图像，并为所述拼接图像设定单词间隔标记或者非单词间隔标记；

- 15 将所述拼接图像标记为所述样本图像；或者，对所述拼接图像进行以下操作之一或者任意组合：平移操作、旋转操作、拉伸操作、添加高斯噪声，并将进行所述操作之后的图像标记为所述样本图像。

所述训练模块，还用于在利用卷积神经网络对样本图像进行处理之前，将所述样本图像转换成灰度的样本图像，并将所述灰度的样本图像转换成大小为 $M \times N$ 的样本图像，
20 并对所述大小为 $M \times N$ 的样本图像进行归一化处理；

所述预测模块，还用于在利用卷积神经网络对待测试图像进行处理之前，将待测试图像转换成灰度的待测试图像，并将灰度的待测试图像转换成大小为 $M \times L$ 的待测试图像，并对所述大小为 $M \times L$ 的待测试图像进行归一化处理；

- 其中，所述 M 为正整数、所述 N 为正整数、所述 L 为正整数，且所述 M 和所述 L 的比值，与所述待测试图像的长宽比相同。
25

所述预测模块，具体用于在利用所述卷积神经网络对所述待测试图像进行处理，得到所述待测试图像对应的第二特征向量的过程中，将所述待测试图像向左平移 K 次或者向右平移 K 次，每次平移时的平移位置不同，以得到 $K+1$ 个不同的待测试图像，所述 K 为正整数；

- 30 利用卷积神经网络对 $K+1$ 个不同的待测试图像进行处理，得到 $K+1$ 个不同的待测试

图像分别对应的特征向量，并将所述 $K+1$ 个不同的待测试图像分别对应的特征向量组合在一起，得到所述待测试图像对应的第二特征向量。

所述预测模块，具体用于在得到所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值的过程中，利用所述卷积神经网络获得所述第二特征向量匹配的多个第一特征向量，并将所述多个第一特征向量对应的单词间隔概率值或者非单词间隔概率值，作为所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值。

所述预测模块，具体用于在利用当前得到的单词间隔概率值或者非单词间隔概率值，对所述待测试图像进行单词分割的过程中，将所述第二特征向量对应的待测试图像位置作为横坐标，将所述第二特征向量对应的单词间隔概率值作为纵坐标，得到概率值结果图；从所述概率值结果图中找到单词间隔概率值的极大值位置，并利用所述极大值位置对所述待测试图像进行单词分割；或者，将所述第二特征向量对应的待测试图像位置作为横坐标，将所述第二特征向量对应的非单词间隔概率值作为纵坐标，得到概率值结果图；从所述概率值结果图中找到非单词间隔概率值的极小值位置，并利用所述极小值位置对所述待测试图像进行单词分割。

基于上述技术方案，本申请实施例中，通过卷积神经网络对图像进行单词分割，可以有效地将包含多个单词的图像分割成若干个小图像，保证每个小图像仅包含一个单词，并且不会将一个单词分割到多个小图像中，从而精准的进行单词分割，使得单词分割的准确率得到提高，并且提高单词分割的速度，可以快速、准确地检测出图像中的文字区域，提高用户使用体验。

附图说明

为了更加清楚地说明本申请实施例或者现有技术中的技术方案，下面将对本申请实施例或者现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请中记载的一些实施例，对于本领域普通技术人员来讲，还可以根据这些附图获得其他的附图。

图 1 和图 2 是本申请一种实施方式中的卷积神经网络的结构示意图；

图 3 是本申请一种实施方式中的单词的分割方法的流程图；

图 4 是本申请一种实施方式中的样本图像的示意图；

图 5 是本申请一种实施方式中的概率值结果图的示意图；

图 6 是本申请一种实施方式中的电子设备的硬件结构图；

图 7 是本申请一种实施方式中的单词的分割装置的结构图。

具体实施方式

在本申请使用的术语仅仅是出于描述特定实施例的目的，而非限制本申请。本申请
5 和权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，
除非上下文清楚地表示其它含义。还应当理解，本文中使用的术语“和/或”是指包含一个
个或多个相关联的列出项目的任何或所有可能组合。

应当理解，尽管在本申请可能采用术语第一、第二、第三等来描述各种信息，但这些
10 些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不
脱离本申请范围的情况下，第一信息也可以被称为第二信息，类似地，第二信息也可以
被称为第一信息。取决于语境，此外，所使用的词语“如果”可以被解释成为“在……
时”或“当……时”或“响应于确定”。

为了从图像中检测出文字区域，如从图像的单行文字中分割出英文单词，在申请人
已知的一种技术中，可以采用如下步骤：步骤 1、将图像转换为灰度图像。步骤 2、对灰
15 度图像进行中值滤波。步骤 3、对灰度图像进行中值滤波后的图像进行二值化处理，并使用连通分量
进行标记，将各联通区作为文字初步候选区域。步骤 4、根据图像中文字的先验知识设
置剔除判决条件，以文字候选区域的文字相似度、文字的笔画宽度、高度等特征，剔除
不是文字的候选区域。步骤 5、对步骤 4 中未被提出的候选区域进行形态学闭处理，并
使用卷积神经网络判断该文字区域，剔除不是文字的区域。步骤 6、根据文字区域的间
20 隔进行统计，若相邻区域间隔大于人工设定的距离阈值，则将两个区域分开。

在上述方式中，二值化过程对于复杂背景的图像来说，可靠性非常差，很容易失败，
导致后续结果错误。而且，该方法引入了很多先验的人工知识和规则，例如，利用文字
的先验知识设置判决条件，利用文字笔画的宽度高度来剔除候选文字区域等，这些特征
都随着场景的变化而变化，通过人工规则的方式很难适应所有场景。而且，该方法对单
25 词的分割，是基于字符间距离大小设定人工阈值来判断，对于字符间距稍大单词，就容易
把其都分割开。

针对上述发现，本申请实施例提出一种单词的分割方法，通过卷积神经网络对图像
进行单词分割，可有效地将包含多个单词的图像分割成若干个小图像，保证每个小图像
仅包含一个单词，且不会将一个单词分割到多个小图像中，从而精准的进行单词分割，
30 单词分割的准确率得到提高，并提高单词分割的速度，可以快速、准确地检测出图像中

的文字区域，提高用户使用体验。与上述已知技术相比，不需要对图像进行二值化处理，提高算法的可靠性，避免二值化过程导致的结果错误问题。此外，不需要引入很多先验的人工知识和规则，通过给卷积神经网络提供大量的样本图像，可以自动辨别单词与单词之间的间隔，这种方式完全由数据驱动，不需要人工干预。而且，通过卷积神经网络进行加速，在保证识别准确率的前提下，识别速度很快。

其中，本申请实施例中所指的卷积神经网络均是全卷积神经网络。

本申请实施例中，为了完成单词的分割，需要设计有效的卷积神经网络，并使用该卷积神经网络进行单词分割。如图 1 所示，为卷积神经网络的结构示意图，卷积神经网络的输入是二维图像，卷积神经网络的输出是单词间隔概率值和/或非单词间隔概率值，例如，卷积神经网络的输出可以是两个神经元，一个神经元输出单词间隔概率值，其表示当前图像是单词间隔的概率值，另一个神经元输出非单词间隔概率值，其表示当前图像不是单词间隔的概率值。此外，卷积神经网络的结构单元可以包括卷积层，池化层，全连接层等。

在卷积层中，通过使用卷积核对图像进行卷积运算，可以使图像特征增强，卷积层在空间范围内使用卷积核进行卷积运算，该卷积核可以是一个 $m \times n$ 大小的矩阵，卷积层的输入与卷积核进行卷积，即可得到卷积层的输出。卷积运算实际上是一个滤波的过程，在卷积运算中，是将图像上点 (x, y) 的灰度值 $f(x, y)$ 与卷积核 $w(x, y)$ 进行卷积。例如，提供一个 4×4 的卷积核，该 4×4 的卷积核中包含 16 个数值，这 16 个数值的大小可以根据实际需要任意配置。按照 4×4 的大小在图像上依次滑动，得到多个 4×4 的滑动窗口，将该 4×4 的卷积核与每个滑动窗口进行卷积，得到多个卷积特征，这些卷积特征，就是卷积层的输出，且被提供给池化层。

在池化层中，实际上是一个降采样过程，通过对多个卷积特征（卷积层的输出）进行取最大、取最小、取平均值等操作，从而减少计算量，并保持特征不变性。在池化层中，利用图像局部相关性的原理，对图像进行子抽样，从而可以减少数据处理量，并保留有用信息。在一个例子中，可以利用如下公式对卷积特征进行池化处理：

$$y_{j,k}^i = \max_{0 \leq m, n < s} \{x_{j \cdot s + m, k \cdot s + n}^i\},$$

并得到池化处理后的特征。其中， s 表示对应的池化处理时的窗口大小（ $s \times s$ ）， m 和 n 均为设定的数值， j 和 k 为卷积层输出的卷积特征， i 表示针对第 i 个图像。

在池化层之后，通常可以使用激活函数（如非线性函数）对池化层输出的特征进行

映射，从而引入非线性因素，使得卷积神经网络通过非线性的组合而增强表达能力。其中，激活函数可以包括但不限于 sigmoid（S 型）函数、ReLU（Rectified Linear Units，整流线性单元）函数等。以 ReLU 函数为例进行说明，则该 ReLU 函数可以将池化层输出的所有特征中，小于 0 的特征置 0，而大于 0 的特征保持不变。

5 在全连接层中，全连接层用于将输入给本全连接层的所有特征进行全连接处理，从而得到一个特征向量，且该特征向量中会包括多个特征。进一步的，全连接层还可以采用 1*1 的卷积层，以构成全卷积的网络。

在图 1 中，卷积神经网络是以两个卷积层、两个池化层和两个全连接层为例，当然，在实际应用中，卷积层的数量、池化层的数量和全连接层的数量还可以为其它数值。例如，在一种优选的实施方式中，卷积层的数量可以为三个，池化层的数量可以为两个，全连接层的数量可以为两个，具体的网络结构还可以如图 2 所示。在图 2 中，卷积神经网络由三个卷积层、两个池化层以及两个全连接层构成，且每个池化层之后还可以包括一个 ReLU。

15 基于上述卷积神经网络，本申请实施例中提出的单词的分割方法，可以包括两个阶段，第一阶段为训练阶段，第二阶段为预测阶段。在训练阶段中，通过使用大量的样本图像，对卷积神经网络进行训练，从而得到一个有识别单词间隔能力的模型。在预测阶段中，通过训练得到的卷积神经网络，对当前输出的待测试图像进行预测，并利用预测结果对待测试图像进行单词分割。

20 参见图 3 所示的单词的分割方法的流程图，训练阶段的处理流程可以如步骤 301 所示，预测阶段的处理流程可以步骤 302 所示。在图 3 中：

步骤 301，获取样本图像，该样本图像包括单词间隔标记或者非单词间隔标记；利用卷积神经网络对样本图像进行处理，得到该样本图像对应的第一特征向量、该第一特征向量对应的单词间隔概率值和/或非单词间隔概率值。

25 步骤 302，获取待测试图像，并利用该卷积神经网络对该待测试图像进行处理，得到该待测试图像对应的第二特征向量、以及该第二特征向量对应的单词间隔概率值或者非单词间隔概率值；利用当前得到的单词间隔概率值或者非单词间隔概率值，对该待测试图像进行单词分割。

30 本申请实施例中，第一特征向量和第二特征向量只是为了区分样本图像对应的特征向量和待测试图像对应的特征向量，将样本图像对应的特征向量称为第一特征向量，并将待测试图像对应的特征向量称为第二特征向量。

针对步骤 301, 在获取样本图像的过程中, 可以从字体库中抽取至少两个字母图像, 并对至少两个字母图像进行拼接, 得到拼接图像, 并为该拼接图像设定单词间隔标记或者非单词间隔标记。之后, 可以将该拼接图像标记为样本图像; 或者, 对该拼接图像进行以下操作之一或者任意组合: 平移操作、旋转操作、拉伸操作、添加高斯噪声, 并将

5 进行操作之后的图像标记为样本图像。

其中, 在设计好卷积神经网络之后, 为了对卷积神经网络进行训练, 需要获取大量的样本图像, 每个样本图像的处理过程均相同, 即步骤 301, 因此, 以一个样本图像的处理为例进行说明。在样本图像中, 需要设定单词间隔标记或者非单词间隔标记, 用以表示样本图像的每个区域是否是单词间区域。

例如, 针对由单词 a 和单词 b 组成的样本图像, 该样本图像可以为 ab, 且 a 和 b 之间没有空格, 这样, 可以为该样本图像设定非单词间隔标记, 如标记 0。或者, 针对由单词 a 和单词 b 组成的样本图像, 该样本图像可以为 a b, 且 a 和 b 之间有空格, 这样, 可以为该样本图像设定单词间隔标记。

由于对卷积神经网络进行训练时, 需要获取大量的样本图像, 如需要上百万的样本

15 图像, 针对如此多的样本图像, 无法通过人工方式, 在每个样本图像中设定单词间隔标记或者非单词间隔标记。因此, 本申请实施例中, 可以从字体库中抽取至少两个字母图像, 并对该至少两个字母图像进行拼接, 得到拼接图像, 并直接为拼接图像设定单词间隔标记或者非单词间隔标记, 从而不需要通过人工方式在每个样本图像中设定单词间隔标记或者非单词间隔标记。例如, 可以从字体库中抽取字母图像 a 和字母图像 b, 并将

20 这两个字母图像拼接成单词 ab, 假设 ab 之间没有空格, 则为该拼接成的图像设定非单词间隔标记。

进一步的, 可以直接将拼接图像标记为样本图像。或者, 考虑到字体库中的图像质量很高, 没有任何形变和噪声干扰, 与实际采集到的真实图像有一定差距, 因此, 还可以对拼接图像进行随机的平移操作、旋转操作、拉伸操作, 并添加一定程度的高斯噪声,

25 从而使得上述操作之后的图像尽可能地接近真实图像, 并可以将进行上述操作之后的图像标记为样本图像, 如图 4 所示。

针对步骤 301, 在利用卷积神经网络对样本图像进行处理之前, 还可以将样本图像转换成灰度的样本图像, 并将灰度的样本图像转换成大小为 $M*N$ 的样本图像, 对大小为 $M*N$ 的样本图像进行归一化处理, 并将归一化处理后的样本图像输入到卷积神经网络,

30 利用卷积神经网络对样本图像进行处理。

其中， M 为正整数、且 N 为正整数。在一个例子中， M 和 N 可以均为 32，基于此，步骤 301 中获取的样本图像可以为一个大小为 $32*32$ 的样本图像，或者，当步骤 301 中获取的样本图像不是大小为 $32*32$ 的样本图像时，则需要将步骤 301 中获取的样本图像转换成大小为 $32*32$ 的样本图像。

5 针对大小为 $32*32$ 的样本图像，还可以进行归一化处理，其公式可以为：

$$\mu = \frac{1}{n} \sum_{i=1, j=1}^{i=32, j=32} p(i, j), \quad \sigma = \sqrt{\frac{1}{n} \sum_{i=1, j=1}^{i=32, j=32} (p(i, j) - \mu)^2}, \quad p'(i, j) = \frac{p(i, j) - \mu}{\sigma} * 128。$$

公式中， $p(i, j)$ 表示样本图像（灰度的样本图像）的一个像素值， i 表示样本图像中的横坐标， j 表示样本图像中的纵坐标， $p(i, j)$ 就是在像素位置 (i, j) 的像素值。 n 表示样本图像预处理后的像素个数。 $p'(i, j)$ 为归一化后的像素值。

10 其中，进行归一化处理的原因是：由于样本图像的输入数据的单位不同，有些输入数据的范围可能特别大，其导致的结果是收敛慢、训练时间长。而且，数据范围大的输入数据在模式分类中的作用可能会偏大，而数据范围小的输入数据在模式分类中的作用可能会偏小，因此，可以通过对输入数据进行归一化处理，将输入数据映射到 $[-128, +128]$ 区间、或其它区间。

15 针对步骤 301，可以利用卷积神经网络对样本图像进行处理，最终得到该样本图像对应的第一特征向量、该第一特征向量对应的单词间隔概率值和/或非单词间隔概率值。例如，基于图 2 所示的卷积神经网络，将样本图像输入到卷积层，并对样本图像进行卷积处理，并将卷积处理结果输出给池化层，对卷积处理结果进行池化处理，并将池化处理结果输出给 ReLU 函数，利用 ReLU 函数对池化处理结果进行映射，并将映射结果输出给卷积层，以此类推，一直到进行全连接处理，得到第一特征向量以及该第一特征向量对应的单词间隔概率值和/或非单词间隔概率值。针对卷积神经网络内的处理过程，本申请实施例中不再赘述，只要通过卷积神经网络，得到第一特征向量以及该第一特征向量对应的单词间隔概率值和/或非单词间隔概率值即可。

25 由于卷积神经网络会对大量的样本图像进行处理，如对上百万的样本图像进行处理，因此，利用卷积神经网络对样本图像进行处理的过程，可以通过调用 CAFFE（Convolutional Architecture for Fast Feature Embedding，卷积神经网络框架），在 GPU（Graphics Processing Unit，图形处理器）集群上对样本图像进行处理，即同时对多个样本图像进行处理，从而加快处理速度。

基于上述过程，即可以训练得到卷积神经网络，通过该卷积神经网络可以对当前输出的待测试图像进行预测，并对待测试图像进行单词分割。

针对步骤 302，在获取待测试图像的过程中，待测试图像是指需要进行单词分割的图像，对于如何获取待测试图像的方式，在此不再赘述。

5 针对步骤 302，在利用卷积神经网络对待测试图像进行处理之前，还可以将该待测试图像转换成灰度的待测试图像，并将灰度的待测试图像转换成大小为 $M \times L$ 的待测试图像，并对大小为 $M \times L$ 的待测试图像进行归一化处理，之后，将归一化处理后的待测试图像输入到卷积神经网络。

其中， M 为正整数、 L 为正整数， M 和 L 的比值与待测试图像的长宽比相同。在一个例子中， M 可以为 32。假设步骤 302 中获取的待测试图像的长为 64，宽为 60，则将待测试图像转换成大小为 32×30 的待测试图像，即 $L=30$ 。

针对大小为 $M \times L$ 的待测试图像，还可以进行归一化处理，其公式可以为：

$$\mu = \frac{1}{n} \sum_{i=1, j=1}^{i=M, j=L} p(i, j), \sigma = \sqrt{\frac{1}{n} \sum_{i=1, j=1}^{i=M, j=L} (p(i, j) - \mu)^2}, p'(i, j) = \frac{p(i, j) - \mu}{\sigma} * 128. p(i, j) \text{ 表示待}$$

15 测试图像（灰度的待测试图像）的一个像素值， i 表示待测试图像中的横坐标， j 表示待测试图像中的纵坐标， $p(i, j)$ 就是在像素位置 (i, j) 的像素值。 n 表示待测试图像预处理后的像素个数。 $p'(i, j)$ 为归一化后的像素值。

其中，进行归一化处理的原因是：由于待测试图像的输入数据的单位不同，有些输入数据的范围可能特别大，其导致的结果是收敛慢、训练时间长。而且，数据范围大的输入数据在模式分类中的作用可能会偏大，而数据范围小的输入数据在模式分类中的作用可能会偏小，因此，可以通过对输入数据进行归一化处理，将输入数据映射到 $[-128, +128]$ 区间、或其它区间。

在实际应用中，针对获取的待测试图像，可以将待测试图像的存储类型设置为浮点型，使得待测试图像中的像素值可以包括小数点后的位置。例如，当像素值为 1.2 时，使用浮点型存储待测试图像时，像素值是真实的像素值 1.2，而使用其它类型存储待测试图像时，像素值可能是不真实的像素值 1。

针对步骤 302，可以利用卷积神经网络对待测试图像进行处理，最终得到该待测试图像对应的第二特征向量、该第二特征向量对应的单词间隔概率值或者非单词间隔概率值。例如，基于图 2 所示的卷积神经网络，将待测试图像输入到卷积层，并对待测试图像进行卷积处理，并将卷积处理结果输出给池化层，对卷积处理结果进行池化处理，并

将池化处理结果输出给 ReLU 函数，利用 ReLU 函数对池化处理结果进行映射，并将映射结果输出给卷积层，以此类推，一直到进行全连接处理，得到第二特征向量、该第二特征向量对应的单词间隔概率值或者非单词间隔概率值。针对卷积神经网络内的处理过程，本申请实施例中不再赘述，只要通过卷积神经网络，得到第二特征向量、该第二特征向量对应的单词间隔概率值或者非单词间隔概率值即可。

本申请实施例中，针对得到第二特征向量对应的单词间隔概率值或者非单词间隔概率值的过程，在利用卷积神经网络得到第二特征向量后，还可以利用卷积神经网络获得该第二特征向量匹配的多个第一特征向量，并将这多个第一特征向量对应的单词间隔概率值或者非单词间隔概率值，作为该第二特征向量对应的单词间隔概率值或者非单词间隔概率值。具体的，将第一特征向量对应的单词间隔概率值作为该第二特征向量对应的单词间隔概率值，将第一特征向量对应的非单词间隔概率值作为该第二特征向量对应的非单词间隔概率值。

其中，针对获得该第二特征向量匹配的多个第一特征向量的过程，由卷积神经网络自身的算法决定，本申请实施例中不再详加赘述。

其中，如果在步骤 301 中得到并记录第一特征向量对应的单词间隔概率值，则在步骤 302 中，得到第二特征向量对应的单词间隔概率值。如果在步骤 301 中得到并记录第一特征向量对应的非单词间隔概率值，则在步骤 302 中，得到第二特征向量对应的非单词间隔概率值。如果在步骤 301 中得到并记录第一特征向量对应的单词间隔概率值和非单词间隔概率值，则在步骤 302 中，得到第二特征向量对应的单词间隔概率值或者非单词间隔概率值。

其中，第二特征向量对应的单词间隔概率值表示第二特征向量对应的待测试图像位置（如第二特征向量对应的多个第一特征向量在待测试图像中对应的位置）是单词间隔的概率值，第二特征向量对应的非单词间隔概率值表示第二特征向量对应的待测试图像位置（如第二特征向量对应的多个第一特征向量在待测试图像中对应的位置）不是单词间隔的概率值。例如，第二特征向量对应于待测试图像上的位置 1（水平方向 $x=16$ 的垂直位置，该位置表示一条竖线区域，此处仅以一个位置为例），位置 1 是待测试图像上的单词 best 中的 e 所在位置，假设第二特征向量对应的单词间隔概率值为 5%，表示位置 1 是单词间隔的概率值仅为 5%，假设第二特征向量对应的非单词间隔概率值为 98%，表示位置 1 不是单词间隔的概率值为 98%。

针对步骤 302，在利用卷积神经网络对待测试图像进行处理，得到待测试图像对应

的第二特征向量的过程中,可以将待测试图像向左平移 K 次或者向右平移 K 次,每次平移时的平移位置不同,得到 $K+1$ 个不同的待测试图像, K 为正整数。利用卷积神经网络对 $K+1$ 个不同的待测试图像进行处理,得到 $K+1$ 个不同的待测试图像分别对应的特征向量,并将 $K+1$ 个不同的待测试图像分别对应的特征向量组合在一起,得到待测试图像对应的第二特征向量。

在实际应用中,由于池化处理过程实际上是一个降采样(下采样)过程,通过对多个特征进行取最大、取最小、取平均值等操作,将待测试图像的大量特征简化为少量特征,使得输出特征比待测试图像的特征数量少很多。例如,待测试图像有 1000 个特征,但是经过卷积神经网络对该待测试图像进行处理后,得到的第二特征向量中只包括 50 个特征。这样,最终得到的第二特征向量无法完全体现出待测试图像的所有特征。基于此,本申请实施例中,可以将待测试图像向左平移 K 次或者向右平移 K 次,得到 $K+1$ 个不同的待测试图像,在利用卷积神经网络对 $K+1$ 个不同的待测试图像进行处理后,可以得到 $K+1$ 个不同的待测试图像分别对应的特征向量,并将 $K+1$ 个不同的待测试图像分别对应的特征向量组合在一起,得到待测试图像对应的第二特征向量,这样增加了特征数量,使得第二特征向量可以体现出待测试图像的所有特征。例如,当 K 为 19 时,则第二特征向量中可以包括 $50*20$ 个特征。

其中,以将待测试图像向左平移 K 次, K 为 5 为例进行说明,针对一个 $32*30$ 的待测试图像 1,向左平移一个像素,变成一个 $31*30$ 的图像,在该 $31*30$ 的图像最右侧之后添加一列数据,得到一个 $32*30$ 的待测试图像 2,在最右侧之后添加的一列数据可以与 $31*30$ 的图像中的第 31 列数据相同。针对该 $32*30$ 的待测试图像 1,向左平移两个像素,变成一个 $30*30$ 的图像,在该 $30*30$ 的图像最右侧之后添加两列数据,得到一个 $32*30$ 的待测试图像 3,在最右侧之后添加的两列数据可以与 $30*30$ 的图像中的第 30 列数据相同。针对该 $32*30$ 的待测试图像 1,向左平移三个像素,以此类推,一直到向左平移五个像素,这样可以得到待测试图像 1、待测试图像 2、待测试图像 3、待测试图像 4、待测试图像 5、待测试图像 6,利用卷积神经网络对待测试图像 1、待测试图像 2、待测试图像 3、待测试图像 4、待测试图像 5、待测试图像 6 分别进行处理,得到 6 个不同的待测试图像分别对应的特征向量,并将 6 个不同的待测试图像分别对应的特征向量组合在一起,得到第二特征向量。

由于利用卷积神经网络对待测试图像 1、待测试图像 2、待测试图像 3、待测试图像 4、待测试图像 5、待测试图像 6 分别进行处理的过程,共享了大量的卷积操作,从而使

得计算速度得到大大提高，加速了预测过程。

针对步骤 302，在利用当前得到的单词间隔概率值或者非单词间隔概率值，对待测试图像进行单词分割的过程中，方式一、将第二特征向量对应的待测试图像位置作为横坐标，将第二特征向量对应的单词间隔概率值作为纵坐标，得到概率值结果图；从概率值结果图中找到单词间隔概率值的极大值位置，并利用极大值位置对待测试图像进行单词分割。或者，方式二、将第二特征向量对应的待测试图像位置作为横坐标，将第二特征向量对应的非单词间隔概率值作为纵坐标，得到概率值结果图；从概率值结果图中找到非单词间隔概率值的极小值位置，并利用极小值位置对待测试图像进行单词分割。

其中，将第二特征向量对应的待测试图像位置作为横坐标，将第二特征向量对应的单词间隔概率值/非单词间隔概率值作为纵坐标的过程，实际上是将第二特征向量匹配的多个第一特征向量在待测试图像对应的位置作为横坐标，并将这多个第一特征向量对应的单词间隔概率值/非单词间隔概率值作为纵坐标。

上述方式一和方式二的处理类似，后续以方式一的处理为例进行说明。

针对方式一，在得到第二特征向量对应的单词间隔概率值之后，可以将第二特征向量对应的待测试图像位置作为横坐标，将第二特征向量对应的单词间隔概率值作为纵坐标，具体的，将第二特征向量匹配的多个第一特征向量在待测试图像中的位置作为横坐标，并将每个第一特征向量对应的单词间隔概率值作为纵坐标，得到如图 5 所示的概率值结果图。在图 5 中，概率值表示当前位置是否是单词间隔的概率，概率值越大，则表明当前位置越有可能是单词间隔。基于此，可以从概率值结果图中找到单词间隔概率值的极大值位置，并利用该极大值位置对待测试图像进行单词分割。

在实际应用中，可以使用非极大值抑制算法找到这些极大值位置，即图 5 中竖线的位置，该非极大值抑制算法在计算机视觉领域中已经被广泛应用，在此不再赘述。在找到这些极大值位置后，可以在极大值位置处将待测试图像分割开，即可得到单个单词的小图像，这样，待测试图像被分割成多个小图像，每个小图像仅包含一个单词，且不会将一个单词分割到多个小图像。

当上述方法应用在文本识别任务时，如果需要对单词句子进行翻译，则采用上述方法将待测试图像分割成多个小图像后，可以对每个小图像进行单词识别，将若干个单词识别结果加上空格拼接在一起，完成单词句子的翻译，

本申请实施例的上述过程中，样本图像和待测试图像均是指单行图像，且该单行中

包含多个单词的图像。此外，单词可以包括但不限于：英文单词、日文单词、德文单词、俄文单位、法文单词、少数民族单词等。

基于上述技术方案，本申请实施例中，通过卷积神经网络对图像进行单词分割，可以有效地将包含多个单词的图像分割成若干个小图像，保证每个小图像仅包含一个单词，并且不会将一个单词分割到多个小图像中，从而精准的进行单词分割，使得单词分割的准确率得到提高，并且提高单词分割的速度，可以快速、准确地检测出图像中的文字区域，提高用户使用体验。

基于与上述方法同样的申请构思，本申请实施例提供一种单词的分割装置，该单词的分割装置应用在电子设备上。该单词的分割装置可通过软件实现，也可通过硬件或者软硬件结合的方式实现。以软件实现为例，作为一个逻辑意义上的装置，是通过其所在的电子设备的处理器，读取非易失性存储器中对应的计算机程序指令形成的。从硬件层面而言，如图 6 所示，为本申请提出的单词的分割装置所在的电子设备的硬件结构图，除了图 6 所示的处理器、非易失性存储器外，电子设备还可以包括其他硬件，如负责处理报文的转发芯片、网络接口、内存等；从硬件结构上来讲，该电子设备还可能是分布式设备，可能包括多个接口卡，以便在硬件层面进行报文处理的扩展。

如图 7 所示，为本申请提出的单词的分割装置的结构图，所述装置包括：

训练模块 11，用于获取样本图像，所述样本图像包括单词间隔标记或者非单词间隔标记；利用卷积神经网络对所述样本图像进行处理，得到所述样本图像对应的第一特征向量、所述第一特征向量对应的单词间隔概率值和/或非单词间隔概率值；

预测模块 12，用于获取待测试图像，并利用所述卷积神经网络对所述待测试图像进行处理，得到所述待测试图像对应的第二特征向量、所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值；利用当前得到的单词间隔概率值或者非单词间隔概率值，对所述待测试图像进行单词分割。

所述训练模块 11，具体用于在获取所述样本图像的过程中，从字体库中抽取至少两个字母图像，并对所述至少两个字母图像进行拼接，得到拼接图像，并为所述拼接图像设定单词间隔标记或者非单词间隔标记；

将所述拼接图像标记为所述样本图像；或者，对所述拼接图像进行以下操作之一或者任意组合：平移操作、旋转操作、拉伸操作、添加高斯噪声，并将进行所述操作之后的图像标记为所述样本图像。

所述训练模块 11，还用于在利用卷积神经网络对样本图像进行处理之前，将所述样

本图像转换成灰度的样本图像，并将所述灰度的样本图像转换成大小为 $M*N$ 的样本图像，并对所述大小为 $M*N$ 的样本图像进行归一化处理；所述预测模块 12，还用于在利用卷积神经网络对待测试图像进行处理之前，将待测试图像转换成灰度的待测试图像，并将灰度的待测试图像转换成大小为 $M*L$ 的待测试图像，并对所述大小为 $M*L$ 的待测试图像进行归一化处理；其中，所述 M 为正整数、所述 N 为正整数、所述 L 为正整数，且所述 M 和所述 L 的比值，与所述待测试图像的长宽比相同。

所述预测模块 12，具体用于在利用所述卷积神经网络对所述待测试图像进行处理，得到所述待测试图像对应的第二特征向量的过程中，将所述待测试图像向左平移 K 次或者向右平移 K 次，每次平移时的平移位置不同，以得到 $K+1$ 个不同的待测试图像，所述 K 为正整数；

利用卷积神经网络对 $K+1$ 个不同的待测试图像进行处理，得到 $K+1$ 个不同的待测试图像分别对应的特征向量，并将所述 $K+1$ 个不同的待测试图像分别对应的特征向量组合在一起，得到所述待测试图像对应的第二特征向量。

所述预测模块 12，具体用于在得到所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值的过程中，利用所述卷积神经网络获得所述第二特征向量匹配的多个第一特征向量，并将所述多个第一特征向量对应的单词间隔概率值或者非单词间隔概率值，作为所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值。

所述预测模块 12，具体用于在利用当前得到的单词间隔概率值或者非单词间隔概率值，对所述待测试图像进行单词分割的过程中，将所述第二特征向量对应的待测试图像位置作为横坐标，将所述第二特征向量对应的单词间隔概率值作为纵坐标，得到概率值结果图；从所述概率值结果图中找到单词间隔概率值的极大值位置，并利用所述极大值位置对所述待测试图像进行单词分割；或者，将所述第二特征向量对应的待测试图像位置作为横坐标，将所述第二特征向量对应的非单词间隔概率值作为纵坐标，得到概率值结果图；从所述概率值结果图中找到非单词间隔概率值的极小值位置，并利用所述极小值位置对所述待测试图像进行单词分割。

基于上述技术方案，本申请实施例中，通过卷积神经网络对图像进行单词分割，可以有效地将包含多个单词的图像分割成若干个小图像，保证每个小图像仅包含一个单词，并且不会将一个单词分割到多个小图像中，从而精准的进行单词分割，使得单词分割的准确率得到提高，并且提高单词分割的速度，可以快速、准确地检测出图像中的文字区域，提高用户使用体验。

其中，本申请装置的各个模块可以集成于一体，也可以分离部署。上述模块可以合并为一个模块，也可以进一步拆分成多个子模块。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本申请各个实施例所述的方法。本领域技术人员可以理解附图只是一个优选实施例的示意图，附图中的模块或流程并不一定是实施本申请所必须的。

10

本领域技术人员可以理解实施例中的装置中的模块可以按照实施例描述进行分布于实施例的装置中，也可以进行相应变化位于不同于本实施例的一个或多个装置中。上述实施例的模块可以合并为一个模块，也可进一步拆分成多个子模块。上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。

15

以上公开的仅为本申请的几个具体实施例，但是，本申请并非局限于此，任何本领域的技术人员能思之的变化都应落入本申请的保护范围。

权利要求书

1. 一种单词的分割方法，其特征在于，所述方法包括以下步骤：

获取样本图像，所述样本图像包括单词间隔标记或者非单词间隔标记；利用卷积神经网络对所述样本图像进行处理，得到所述样本图像对应的第一特征向量、所述第一特征向量对应的单词间隔概率值和/或非单词间隔概率值；

获取待测试图像，并利用所述卷积神经网络对所述待测试图像进行处理，得到所述待测试图像对应的第二特征向量、所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值；利用当前得到的单词间隔概率值或者非单词间隔概率值，对所述待测试图像进行单词分割。

2. 根据权利要求 1 所述的方法，其特征在于，所述获取样本图像，包括：

从字体库中抽取至少两个字母图像，并对所述至少两个字母图像进行拼接，得到拼接图像，并为所述拼接图像设定单词间隔标记或者非单词间隔标记；

将所述拼接图像标记为所述样本图像；或者，对所述拼接图像进行以下操作之一或者任意组合：平移操作、旋转操作、拉伸操作、添加高斯噪声，并将进行所述操作之后的图像标记为所述样本图像。

3. 根据权利要求 1 所述的方法，其特征在于，所述方法进一步包括：

在利用卷积神经网络对所述样本图像进行处理之前，将所述样本图像转换成灰度的样本图像，并将所述灰度的样本图像转换成大小为 $M*N$ 的样本图像，并对所述大小为 $M*N$ 的样本图像进行归一化处理；

在利用所述卷积神经网络对所述待测试图像进行处理之前，将所述待测试图像转换成灰度的待测试图像，并将所述灰度的待测试图像转换成大小为 $M*L$ 的待测试图像，并对所述大小为 $M*L$ 的待测试图像进行归一化处理；

其中，所述 M 为正整数、所述 N 为正整数、所述 L 为正整数，且所述 M 和所述 L 的比值，与所述待测试图像的长宽比相同。

4. 根据权利要求 1 所述的方法，其特征在于，利用所述卷积神经网络对所述待测试图像进行处理，得到所述待测试图像对应的第二特征向量，包括：

将所述待测试图像向左平移 K 次或者向右平移 K 次，每次平移时的平移位置不同，以得到 $K+1$ 个不同的待测试图像，所述 K 为正整数；

利用卷积神经网络对 $K+1$ 个不同的待测试图像进行处理，得到 $K+1$ 个不同的待测试图像分别对应的特征向量，并将所述 $K+1$ 个不同的待测试图像分别对应的特征向量组

合在一起，得到所述待测试图像对应的第二特征向量。

5. 根据权利要求 1 或 4 所述的方法，其特征在于，所述得到所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值的过程，具体包括：

利用所述卷积神经网络获得所述第二特征向量匹配的多个第一特征向量，并将所述
5 多个第一特征向量对应的单词间隔概率值或者非单词间隔概率值，作为所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值。

6. 根据权利要求 1 或 4 所述的方法，其特征在于，利用当前得到的单词间隔概率值或者非单词间隔概率值，对所述待测试图像进行单词分割，包括：

将所述第二特征向量对应的待测试图像位置作为横坐标，将所述第二特征向量对应的
10 的单词间隔概率值作为纵坐标，得到概率值结果图；从所述概率值结果图中找到单词间隔概率值的极大值位置，并利用所述极大值位置对所述待测试图像进行单词分割；或者，将所述第二特征向量对应的待测试图像位置作为横坐标，将所述第二特征向量对应的非单词间隔概率值作为纵坐标，得到概率值结果图；从所述概率值结果图中找到非单词间隔概率值的极小值位置，并利用所述极小值位置对所述待测试图像进行单词分割。

15 7. 一种单词的分割装置，其特征在于，所述装置具体包括：

训练模块，用于获取样本图像，所述样本图像包括单词间隔标记或者非单词间隔标记；利用卷积神经网络对所述样本图像进行处理，得到所述样本图像对应的第一特征向量、所述第一特征向量对应的单词间隔概率值和/或非单词间隔概率值；

预测模块，用于获取待测试图像，并利用所述卷积神经网络对所述待测试图像进行
20 处理，得到所述待测试图像对应的第二特征向量、所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值；利用当前得到的单词间隔概率值或者非单词间隔概率值，对所述待测试图像进行单词分割。

8. 根据权利要求 7 所述的装置，其特征在于，

所述训练模块，具体用于在获取所述样本图像的过程中，从字体库中抽取至少两个
25 字母图像，并对所述至少两个字母图像进行拼接，得到拼接图像，并为所述拼接图像设定单词间隔标记或者非单词间隔标记；

将所述拼接图像标记为所述样本图像；或者，对所述拼接图像进行以下操作之一或者任意组合：平移操作、旋转操作、拉伸操作、添加高斯噪声，并将进行所述操作之后的图像标记为所述样本图像。

30 9. 根据权利要求 7 所述的装置，其特征在于，

所述训练模块，还用于在利用卷积神经网络对样本图像进行处理之前，将所述样本图像转换成灰度的样本图像，并将所述灰度的样本图像转换成大小为 $M*N$ 的样本图像，并对所述大小为 $M*N$ 的样本图像进行归一化处理；

5 所述预测模块，还用于在利用卷积神经网络对待测试图像进行处理之前，将待测试图像转换成灰度的待测试图像，并将灰度的待测试图像转换成大小为 $M*L$ 的待测试图像，并对所述大小为 $M*L$ 的待测试图像进行归一化处理；

其中，所述 M 为正整数、所述 N 为正整数、所述 L 为正整数，且所述 M 和所述 L 的比值，与所述待测试图像的长宽比相同。

10. 根据权利要求 7 所述的装置，其特征在于，

10 所述预测模块，具体用于在利用所述卷积神经网络对所述待测试图像进行处理，得到所述待测试图像对应的第二特征向量的过程中，将所述待测试图像向左平移 K 次或者向右平移 K 次，每次平移时的平移位置不同，以得到 $K+1$ 个不同的待测试图像，所述 K 为正整数；

15 利用卷积神经网络对 $K+1$ 个不同的待测试图像进行处理，得到 $K+1$ 个不同的待测试图像分别对应的特征向量，并将所述 $K+1$ 个不同的待测试图像分别对应的特征向量组合在一起，得到所述待测试图像对应的第二特征向量。

11. 根据权利要求 7 或 10 所述的装置，其特征在于，

20 所述预测模块，具体用于在得到所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值的过程中，利用所述卷积神经网络获得所述第二特征向量匹配的多个第一特征向量，并将所述多个第一特征向量对应的单词间隔概率值或者非单词间隔概率值，作为所述第二特征向量对应的单词间隔概率值或者非单词间隔概率值。

12. 根据权利要求 7 或 10 所述的装置，其特征在于，

25 所述预测模块，具体用于在利用当前得到的单词间隔概率值或者非单词间隔概率值，对所述待测试图像进行单词分割的过程中，将所述第二特征向量对应的待测试图像位置作为横坐标，将所述第二特征向量对应的单词间隔概率值作为纵坐标，得到概率值结果图；从所述概率值结果图中找到单词间隔概率值的极大值位置，并利用所述极大值位置对所述待测试图像进行单词分割；或者，将所述第二特征向量对应的待测试图像位置作为横坐标，将所述第二特征向量对应的非单词间隔概率值作为纵坐标，得到概率值结果图；从所述概率值结果图中找到非单词间隔概率值的极小值位置，并利用所述极小值位置对所述待测试图像进行单词分割。

30

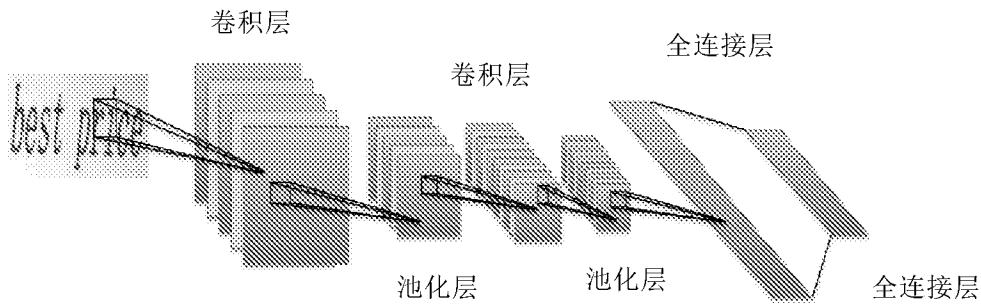


图 1

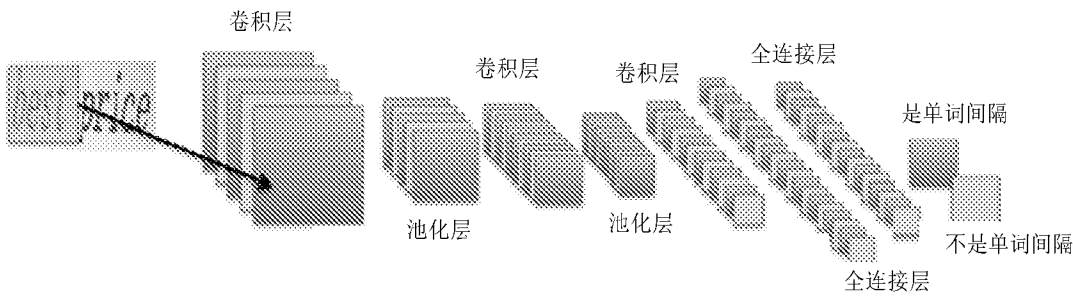


图 2

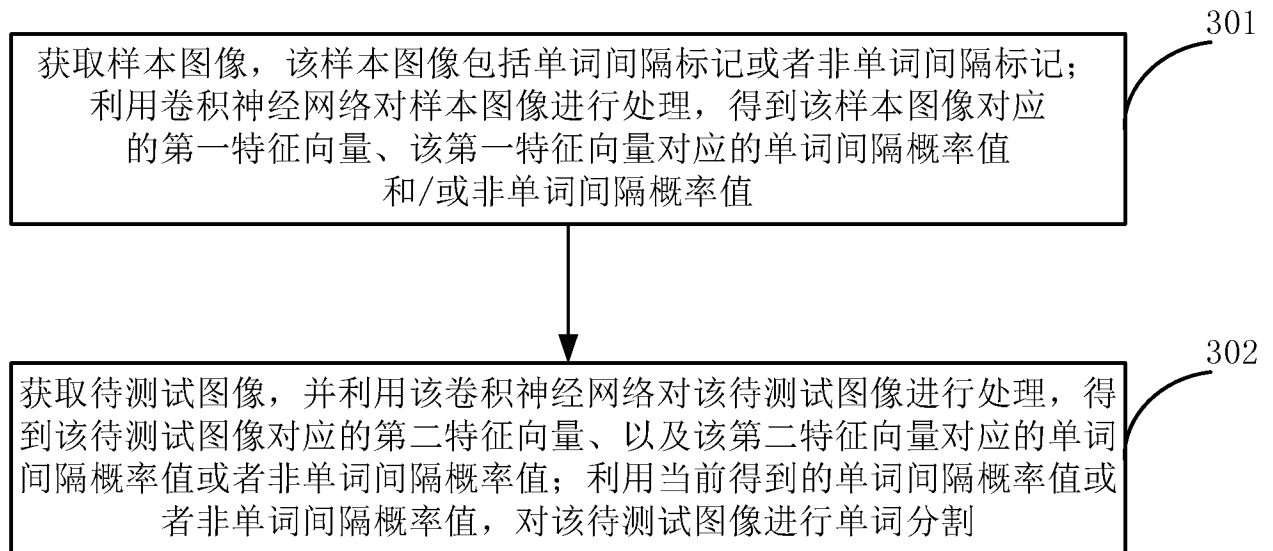


图 3

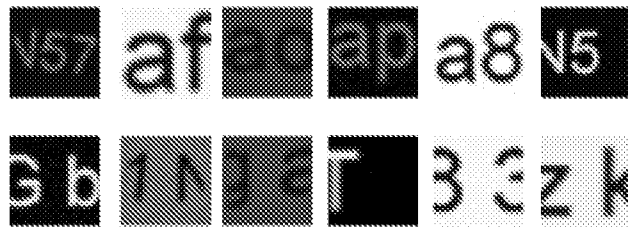


图 4

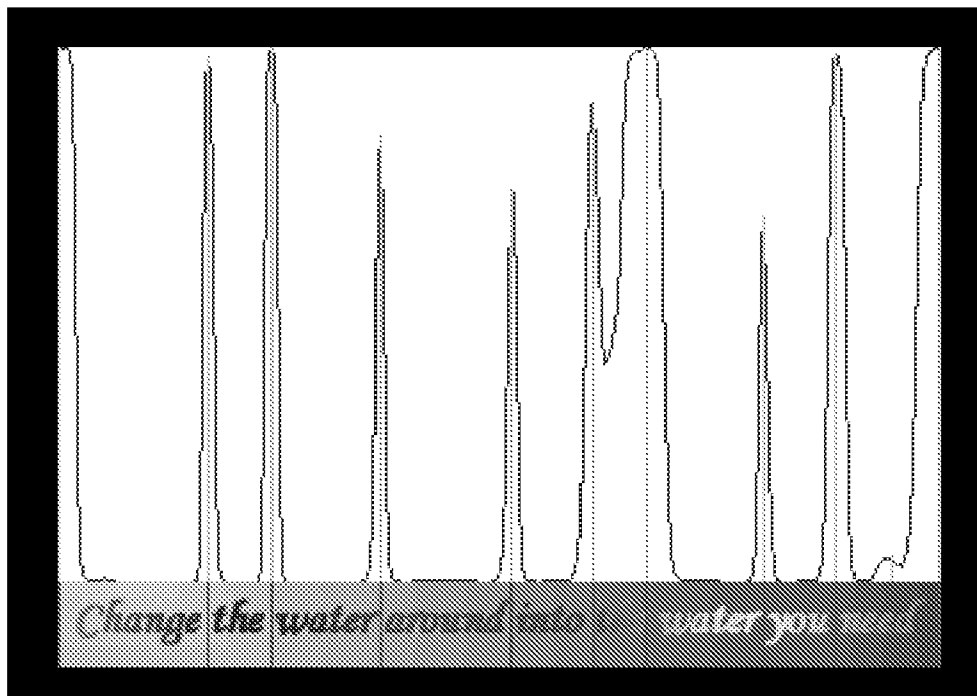


图 5

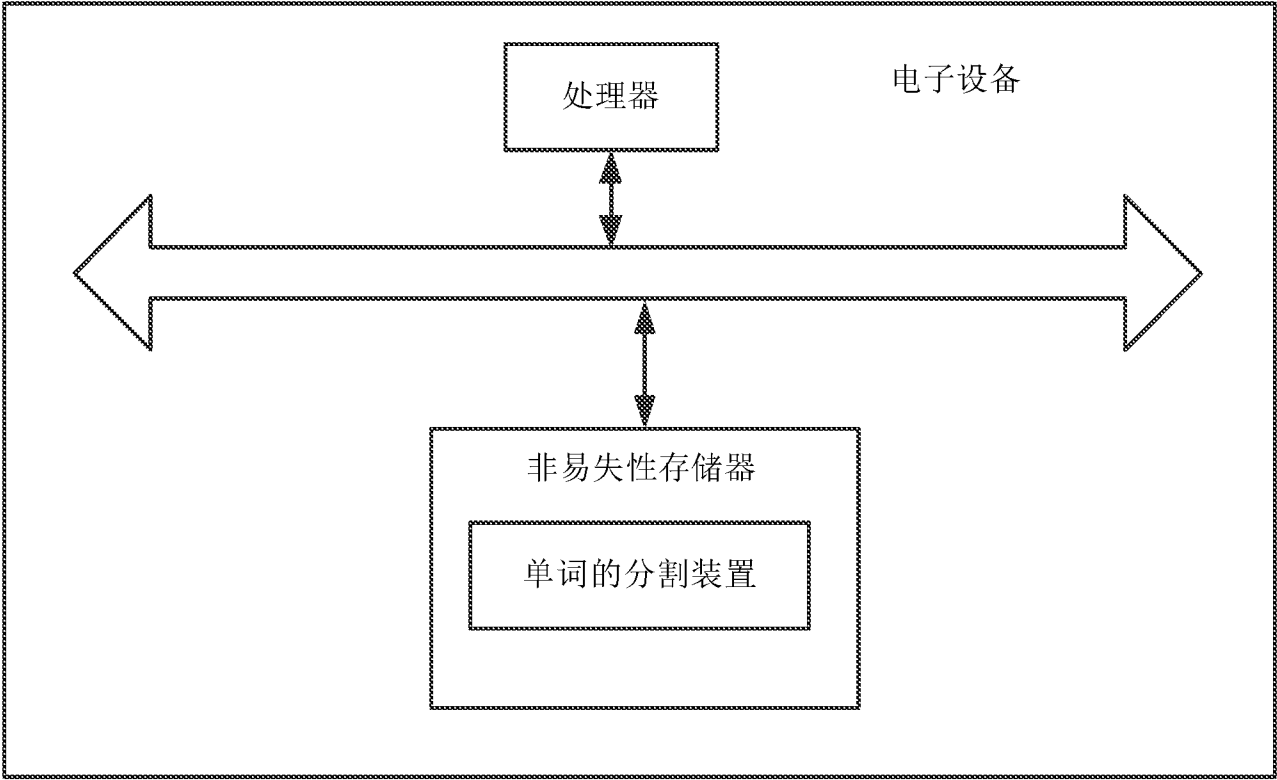


图 6

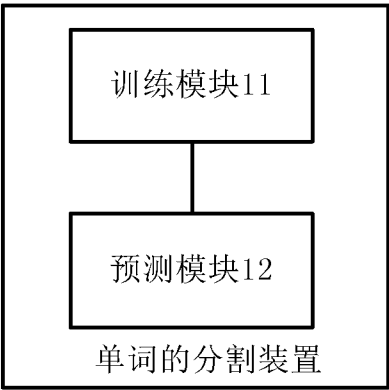


图 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/073718

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/34 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, WPI, EPODOC: CNN, margin, OCR, character recognition, image, character?, recogni+, word?, text, cut+, separat+,
divid+, division, space, convolutional, neural, network, probability

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102822846 A (MICROSOFT CORPORATION) 12 December 2012 (12.12.2012) claim 1, and description, paragraphs [0003] and [0027]	1-12
A	CN 103530600 A (NEUSOFT CORPORATION) 22 January 2014 (22.01.2014) the whole document	1-12
A	CN 102254157 A (BEIJING WENTONG IMAGE RECOGNITION TECHNOLOGY RESEARCH CENTER CO., LTD.) 23 November 2011 (23.11.2011) the whole document	1-12
A	CN 104050471 A (HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY) 17 September 2014 (17.09.2014) the whole document	1-12
A	US 2005027511 A1 (OHGURO, YOSHIHISA) 03 February 2005 (03.02.2005) the whole document	1-12

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 28 April 2017	Date of mailing of the international search report 23 May 2017
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer WANG, Wei Telephone No. (86-10) 62413695

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/073718

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 102822846 A	12 December 2012	US 2011243445 A1	06 October 2011
		HK 1179387 A1	02 December 2016
		EP 2553625 A2	06 February 2013
		WO 2011126755 A2	13 October 2011
CN 103530600 A	22 January 2014	None	
CN 102254157 A	23 November 2011	None	
CN 104050471 A	17 September 2014	None	
US 2005027511 A1	03 February 2005	JP 2005063419 A	10 March 2005

A. 主题的分类 G06K 9/34 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																				
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06K 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNKI, CNPAT, WPI, EPDOC: 单词, 字符, 分割, 切分, 切割, 卷积, 神经, 网络, CNN, 间隔, 空格, 间距, 分隔, 空白, OCR, 文字识别, 字符识别, 图像, 概率, character?, recogni+, word?, text, cut+, separat+, divid+, division, space, convolutional, neural, network, probability																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 102822846 A (微软公司) 2012年 12月 12日 (2012 - 12 - 12) 权利要求1, 说明书第[0003], [0027]段</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 103530600 A (东软集团股份有限公司) 2014年 1月 22日 (2014 - 01 - 22) 全文</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 102254157 A (北京文通图像识别技术研究中心有限公司) 2011年 11月 23日 (2011 - 11 - 23) 全文</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 104050471 A (华中科技大学) 2014年 9月 17日 (2014 - 09 - 17) 全文</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>US 2005027511 A1 (OHGURO, YOSHIHISA) 2005年 2月 3日 (2005 - 02 - 03) 全文</td> <td>1-12</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 102822846 A (微软公司) 2012年 12月 12日 (2012 - 12 - 12) 权利要求1, 说明书第[0003], [0027]段	1-12	A	CN 103530600 A (东软集团股份有限公司) 2014年 1月 22日 (2014 - 01 - 22) 全文	1-12	A	CN 102254157 A (北京文通图像识别技术研究中心有限公司) 2011年 11月 23日 (2011 - 11 - 23) 全文	1-12	A	CN 104050471 A (华中科技大学) 2014年 9月 17日 (2014 - 09 - 17) 全文	1-12	A	US 2005027511 A1 (OHGURO, YOSHIHISA) 2005年 2月 3日 (2005 - 02 - 03) 全文	1-12
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
X	CN 102822846 A (微软公司) 2012年 12月 12日 (2012 - 12 - 12) 权利要求1, 说明书第[0003], [0027]段	1-12																		
A	CN 103530600 A (东软集团股份有限公司) 2014年 1月 22日 (2014 - 01 - 22) 全文	1-12																		
A	CN 102254157 A (北京文通图像识别技术研究中心有限公司) 2011年 11月 23日 (2011 - 11 - 23) 全文	1-12																		
A	CN 104050471 A (华中科技大学) 2014年 9月 17日 (2014 - 09 - 17) 全文	1-12																		
A	US 2005027511 A1 (OHGURO, YOSHIHISA) 2005年 2月 3日 (2005 - 02 - 03) 全文	1-12																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
<table border="0"> <tr> <td style="vertical-align: top;"> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td style="vertical-align: top;"> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																			
国际检索实际完成的日期 2017年 4月 28日		国际检索报告邮寄日期 2017年 5月 23日																		
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10) 62019451		受权官员 王伟 电话号码 (86-10) 62413695																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/073718

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	102822846	A	2012年 12月 12日	US	2011243445	A1	2011年 10月 6日
				HK	1179387	A1	2016年 12月 2日
				EP	2553625	A2	2013年 2月 6日
				WO	2011126755	A2	2011年 10月 13日
CN	103530600	A	2014年 1月 22日	无			
CN	102254157	A	2011年 11月 23日	无			
CN	104050471	A	2014年 9月 17日	无			
US	2005027511	A1	2005年 2月 3日	JP	2005063419	A	2005年 3月 10日