

(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
2025 年 1 月 23 日 (23.01.2025)

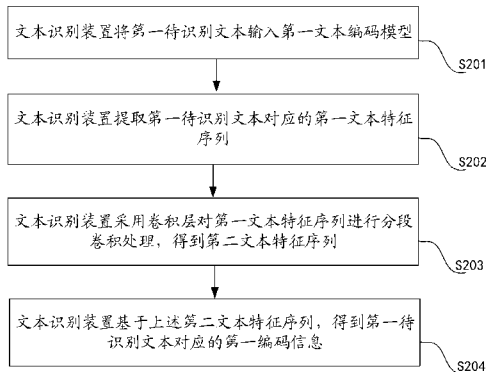


(10) 国际公布号
WO 2025/016237 A1

- (51) 国际专利分类号:
G06F 40/126 (2020.01)
- (21) 国际申请号: PCT/CN2024/104175
- (22) 国际申请日: 2024 年 7 月 8 日 (08.07.2024)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202310873503.9 2023年7月14日 (14.07.2023) CN
- (71) 申请人: 维沃移动通信有限公司(VIVO MOBILE COMMUNICATION CO., LTD.) [CN/CN]; 中国广东省东莞市长安镇维沃路 1 号, Guangdong 523863 (CN)。
- (72) 发明人: 冉佳诺(RAN, Jianuo); 中国广东省东莞市长安镇维沃路1号, Guangdong 523863 (CN)。
- (74) 代理人: 北京远志博慧知识产权代理有限公司(BOHUI INTELLECTUAL PROPERTY); 中国北京市海淀区交大东路31号东区10号楼等17幢31幢108, Beijing 100044 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA,

(54) **Title:** TEXT RECOGNITION METHOD AND APPARATUS, ELECTRONIC DEVICE, AND READABLE STORAGE MEDIUM

(54) 发明名称: 文本识别方法、装置、电子设备及可读存储介质



S201 A text recognition apparatus inputs a first text to be recognized into a first text encoding model
S202 The text recognition apparatus extracts a first text feature sequence corresponding to said first text
S203 The text recognition apparatus performs segment-based convolution processing on the first text feature sequence by means of the convolution layer to obtain a second text feature sequence
S204 The text recognition apparatus obtains, on the basis of the second text feature sequence, first code information corresponding to said first text

图 1

(57) **Abstract:** The present application relates to the field of natural language processing, and discloses a text recognition method and apparatus, an electronic device, and a readable storage medium. The method comprises: inputting a first text to be recognized into a first text encoding model, wherein said first text comprises a first polyphone, and the first text encoding model comprises a convolution layer; extracting a first text feature sequence corresponding to said first text; performing segment-based convolution processing on the first text feature sequence by means of the convolution layer to obtain a second text feature sequence, wherein the second text feature sequence at least comprises global semantic information of said first text and local semantic information of a first text segment in said first text, and the first text segment comprises the first polyphone; and on the basis of the second text feature sequence, obtaining first code information corresponding to said first text, and obtaining a recognition result of said first text on the basis of the first code information.

(57) **摘要:** 本申请公开了一种文本识别方法、装置、电子设备及可读存储介质, 属于自然语言处理领域。该方法包括: 将第一待识别文本输入第一文本编码模型, 第一待识别文本包含第一多音字, 第一文本编码模型包括卷积层; 提取第一待识别文本对应的第一文本特征序列; 采用卷积层对第一文本特征序列进行分段卷积处理, 得到第二文本特征序列, 第二文本特征序列至少包含第一待识别文本的全局语义信息和第一待识别文本中的第一文本片段的局部语义信息, 第一文本片段包含第一多音字; 基于第二文本特征序列, 得到第一待识别文本对应的第一编码信息, 并基于上述第一编码信息获得第一待识别文本的识别结果。

PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))。

文本识别方法、装置、电子设备及可读存储介质

相关申请的交叉引用

5 本申请主张在 2023 年 07 月 14 日在中国提交的中国专利申请号 No.202310873503.9 的优先权，其全部内容通过引用包含于此。

技术领域

本申请属于自然语言处理领域，具体涉及一种文本识别方法、装置、电子设备及可读存储介质。

背景技术

10 语音合成是一种将文本转换为语音的技术，在用户与电子设备交互的场景中，语音合成技术为用户提供了一种新的信息交互方式。目前，在中文语音合成播报场景中，由于汉字里包含大量多音字，语音合成在处理多音字时可能会出现读错的情况。

15 在相关技术中，为了让电子设备正确地读出一段文本，通常需要使用多音字模型来预测文本读音，多音字模型需要根据文本编码后的结果来进行读音预测，文本编码指将自然语言文本转换成向量的形式便于计算机理解和处理。目前在进行文本编码时通常采用的字编码方式，即逐字进行编码，只考虑当前字符的内容。

20 然而，在采用字编码方式处理一些语言结构比较复杂的文本时，可能会出现多音字读音预测错误，从而导致对文本中的多音字的读音预测的准确率较低。

发明内容

25 本申请实施例的目的是提供一种文本识别方法、装置、电子设备及可读存储介质，能够更好地利用局部语义信息和全局语义信息分析长难句、复杂文本读音，进而提高对文本中的多音字读音预测的准确率。

30 第一方面，本申请实施例提供了一种文本识别方法，该方法包括：将第一待识别文本输入第一文本编码模型，第一待识别文本包含第一多音字，第一文本编码模型包括卷积层；提取第一待识别文本对应的第一文本特征序列；采用卷积层对第一文本特征序列进行分段卷积处理，得到第二文本特征序列，第二文本特征序列至少包含第一待识别文本的全局语义信息和第一待识别文本中的第一文本片段的局部语义信息，第一文本片段包含第一多音字；基于第二文本特征序列，得到第一待识别文本对应的第一编码信息，并基于上述第一编码信息获得第一待识别文本的识别结果。

35 第二方面，本申请实施例提供了一种文本识别装置，该装置包括：处理模块和提取模块，其中：处理模块，用于将第一待识别文本输入第一文

本编码模型，第一待识别文本包含第一多音字，第一文本编码模型包括卷积层；提取模块，用于提取第一待识别文本对应的第一文本特征序列；处理模块，还用于采用卷积层对第一文本特征序列进行分段卷积处理，得到第二文本特征序列，第二文本特征序列至少包含第一待识别文本的全局语义信息和第一待识别文本中的第一文本片段的局部语义信息，第一文本片段包含第一多音字；处理模块，还用于基于第二文本特征序列，得到第一待识别文本对应的第一编码信息，并基于上述第一编码信息获得第一待识别文本的识别结果。

第三方面，本申请实施例提供了一种电子设备，该电子设备包括处理器和存储器，所述存储器存储可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如第一方面所述的方法的步骤。

第四方面，本申请实施例提供了一种可读存储介质，所述可读存储介质上存储程序或指令，所述程序或指令被处理器执行时实现如第一方面所述的方法的步骤。

第五方面，本申请实施例提供了一种芯片，所述芯片包括处理器和通信接口，所述通信接口和所述处理器耦合，所述处理器用于运行程序或指令，实现如第一方面所述的方法。

第六方面，本申请实施例提供一种计算机程序产品，该程序产品被存储在存储介质中，该程序产品被至少一个处理器执行以实现如第一方面所述的方法。

在本申请实施例中，文本识别装置将第一待识别文本输入第一文本编码模型，该第一待识别文本包含第一多音字，该第一文本编码模型包括卷积层，然后，提取第一待识别文本对应的第一文本特征序列，并采用所述卷积层对上述第一文本特征序列进行分段卷积处理，得到第二文本特征序列，上述第二文本特征序列至少包含所述第一待识别文本的全局语义信息和第一待识别文本中的第一文本片段的局部语义信息，上述第一文本片段包含第一多音字，最后，基于上述第二文本特征序列，得到第一待识别文本对应的第一编码信息，并基于上述第一编码信息获得第一待识别文本的识别结果。通过该方法，文本识别装置通过第一文本编码模型中的卷积层对文本的全局文本特征序列进行卷积操作获得包含全局语义信息和局部语义信息的文本特征序列，并基于该文本特征序列得到更为准确的识别结果，从而加强对文本序列局部信息建模能力，能够弥补现有模型对于局部特征信息提取的不足，从而能够更为准确地分析长难句、复杂文本中多音字的读音，进而提高对文本中的多音字读音预测的准确率。

附图说明

图 1 为本申请实施例提供的文本识别方法的流程示意图；

图 2 为本申请实施例提供的改进后的 DeBERTa 模型中的 CNN 层的结构示意图；

图 3 为本申请实施例提供的改进的 DeBERTa 模型的结构示意图；

图 4 为本申请实施例提供的文本识别方法的流程图；
图 5 为本申请实施例提供的文本识别装置的结构示意图；
图 6 为本申请实施例提供的电子设备的结构示意图；
图 7 为本申请实施例提供的电子设备的硬件结构示意图。

5 具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员获得的所有其他实施例，都属于本申请保护的范围。

10 本申请的说明书和权利要求书中的术语“第一”、“第二”等是用于区别类似的对象，而不用来描述特定的顺序或先后次序。应该理解这样使用的术语在适当情况下可以互换，以便本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施，且“第一”、“第二”等所区分的对象通常为一类，并不限定对象的个数，例如第一对象可以是一个，也可以是多个。此外，说明书以及权利要求中“和/或”表示所连接对象的至少其中之一，字符“/”，一般表示前后关联对象是一种“或”的关系。

15 语音合成是一种将文本转换为语音的技术，在用户与手机交互的场景中，语音合成技术为用户提供了一种新的信息交互方式。相比于视觉上的信息交互，在一些特定场景，例如，驾驶场景或者运动场景，语音合成以声音的形式能够更方便快捷地传递信息。其次，语音合成通过模拟语言的语调、语气和情感帮助用户更好地理解信息，为用户提供更加真实的交互体验。

20 在中文语音合成播报场景中，由于汉字里包含大量多音字，语音合成在处理多音字时可能会出现读错的情况，这种误读向用户传递了错误信息进而对信息产生了误解，降低了语音信息传递的可靠性和准确性。因此对错误读音进行纠正非常重要，尤其是针对多音字等重难点字词读音的识别与校正。

30 在相关技术中，为了让电子设备正确地读出一段文本，通常需要使用词典规则和多音字模型来预测文本读音，其中多音字模型需要根据文本编码后的结果来进行读音预测，文本编码指将自然语言文本转换成向量的形式便于计算机理解和处理。目前有两种编码方式：字编码和句编码。字编码指的是逐字进行编码只考虑当前字符的内容，采取字编码方式计算量比较小，效率高，但在一些复杂、有歧义的场景下由于上下文的联系可能会出现多音字读音预测错误。句编码考虑了前后语境、语义，因此能够更准确地预测多音字的读音，但句编码方式计算速度较为缓慢，需要大量的语料库进行训练，以及，现有的句编码方式通过基于全局的自注意力机制捕捉序列中的上下文关系，输入序列的每个位置能够关注到序列中其他所有位置的信息，但也因此在处理局部信息时容易受到其他位置的影响，难以提取细粒度的局部特征，如此，基于句编码对长难句等复杂文本中的多音

字进行读音预测的准确性低。

针对上述问题，本申请提出一种文本识别方法，通过在 DeBERTa (Decoding-enhanced BERT with disentangled attention) 模型新增卷积模块，即，卷积层，并通过该卷积层基于待识别文本的全局文本特征序列，提取待识别文本的局部特征信息，从而加强对文本序列局部信息建模能力，卷积网络通过局部连接和卷积核权值共享的方式，对序列分段进行卷积操作获得局部特征。在 Transformer 层后新增卷积模块，能够弥补 Transformer 层对于局部特征提取的不足，从而能够更好地利用局部信息和全局信息分析长难句、复杂文本读音，进而能够提高对长难句等复杂文本中的多音字进行读音预测的准确性。

需要说明的是，DeBERTa，即，带有分散注意力的解码增强型 BERT。DeBERTa 模型为基于 BERT (Bidirectional Encoder Representation from Transformers) 模型改进后得到的模型，BERT，即，双向 Transformer 的编码。

需要说明的是，BERT 模型和 DeBERTa 模型是为不同的自然语言处理 (Natural Language Processing, NLP) 任务提供支持的通用的新型语言模型。DeBERTa 模型的编码器结构为多层 Transformer 叠加，Transformer 是一种在文本、时间序列建模任务中广泛使用的模型，它对整个序列进行建模，通过基于全局的自注意力机制捕捉序列中的上下文关系，输入序列的每个位置能够关注到序列中其他所有位置的信息。

下面结合附图，通过具体的实施例及其应用场景对本申请实施例提供的文本识别方法进行详细地说明。

图 1 为本申请实施例提供的文本识别方法的流程示意图，如图 1 所示，该文本识别方法可以包括以下步骤 S201 至步骤 S204：

步骤 S201：文本识别装置将第一待识别文本输入第一文本编码模型。

其中，上述第一待识别文本包含第一多音字，上述第一文本编码模型包括卷积层。

可选地，在本申请实施例中，上述第一文本编码模型可以为包括 CNN (卷积神经网络，Convolutional Neural Networks) 层的 DeBERTa 模型。

需要说明的是，对 DeBERTa 模型的解释说明可以参见上文描述，此处不再赘述。

可选地，在本申请实施例中，上述第一待识别文本可以为用户输入的文本，或者页面中已有的文本等，本申请实施例对此不作限定。

可选地，在本申请实施例中，上述第一待识别文本可以包括：至少一个句子或者至少一个段落。

可选地，在本申请实施例中，上述第一待识别文本可以为需要进行语音合成的文本。

可选地，在本申请实施例中，上述第一多音字可以包括一个或者两个以上多音字。

在本申请实施例中，上述卷积层包括卷积核（convolutional kernel）。

可选地，在本申请实施例中，上述卷积核对应的卷积窗口的窗长可以为3，所述卷积核的移动步长可以为1。

需要说明的是，上述卷积核对应的卷积窗口指的是卷积核执行卷积操作时，基于步长移动时的窗口长度。例如，卷积核尺寸为 3×1 ，即卷积核每移动一次对应的窗口长度为3。

进一步可选地，上述卷积核用于对第一待识别文本进行局部特征提取。

进一步可选地，上述卷积核可以包括至少一个尺寸固定的卷积核。

需要说明的是，上述卷积窗口，又名卷积层窗口，卷积窗等，上述卷积窗口用于提取输入数据中的特征信息。

步骤 S202：文本识别装置提取第一待识别文本对应的第一文本特征序列。

可选地，在本申请实施例中，文本识别装置可以将第一待识别文本输入至 DeBERTa 模型的 Transformer 层，对该第一待识别文本进行文本特征提取，得到该待识别文本对应的第一文本特征序列。

需要说明的是，上述 Transformer 层是一个基于多头自注意力机制的双向长短期记忆网络，包括编码和解码两个部分。

可选地，在本申请实施例中，上述第一文本特征序列可以为：通过 DeBERTa 模型的 Transformer 层基于第一待识别文本的上下文语义特征得到的注意力结果，该注意力结果可以用 H_0 表示。

进一步可选地，在通过 DeBERTa 模型的 Transformer 层对输入的文本进行处理时，将文本中的字符的内容信息和位置信息分别用两个向量表示，用以计算不同位置的字符间的注意力权重，然后，通过注意力权重，计算字符序列中每个位置的加权和，并在序列不同位置上分配不同的权重，得到上述注意力结果。

进一步地，文本中的每个字符可以转换成一个固定维度的向量形式，用以表示文本中每个字符对应的内容信息，并使用绝对位置编码记录每个字符的位置信息。

需要说明的是，在 BERT 中，输入文本的每个字符都转换成一个固定维度的向量形式，用以表示文本中每个字符对应的内容信息，使用绝对位置编码记录每个字符的位置信息。而 DeBERTa 模型考虑了字符与字符间的相对位置的重要性，例如“深度”和“学习”两个词在句子中相邻出现时，比它们分别出现在句子其他不同位置时具有更强的关联性，因此将字符的内容信息和位置信息分别用两个向量表示，用以计算注意力权重。

以下通过具体的实施例对通过 DeBERTa 模型的 Transformer 层，对第一待识别文本进行文本特征提取的过程进行解释说明。

示例性地，以处理文本中第 i 个字符为例， $\{H_{ij}\}$ 表示第 i 个字符的嵌入向量，即将字符转换为一个固定维度的向量。 $\{P_{ij}\}$ 表示第 i 个字符和第 j 个字符的相对位置，由此两个字符之间的注意力权重可用以下公式进行表

示：

$$\begin{aligned} A_{i,j} &= \{H_i, P_{ij}\} \times \{H_j, P_{ji}\} \\ &= H_i H_j^T + H_i P_{ji}^T + P_{ij} H_j^T + P_{ij} P_{ji}^T \end{aligned} \quad (1)$$

由上式可知，字符与字符间的注意力权重 $A_{i,j}$ 可以分解为“内容到内容”，“内容到位置”，“位置到内容”，“位置到位置”四个部分，由于使用相对位置编码，“位置到位置”本身不具备意义，因此在后续计算中略去。相比于相关技术中通过 BERT 计算注意力权重时仅考虑“内容到内容”和“内容到位置”，本申请实施例基于 DeBERTa 模型计算注意力权重，考虑了“位置到内容”的信息，通过位置信息帮助模型更好的理解序列中的语义内容。

需要说明的是，上述“内容到内容”指的是一个字符和另一个字符的语义之间的联系，例如，待识别的文本中的字符 A 的语义和字符 B 的语义之间的联系；上述“内容到位置”指的是一个字符的内容和另一个字符的位置之间的联系，能够反映字符之间的联系紧密程度；上述“位置到内容”指的是一个字符的位置与另一个字符的内容之间的联系；上述“位置到位置”指的是一个字符的位置与另一个字符之间的位置之间的关系，例如，待识别的文本中的字符 A 的位置和字符 B 的在文本中的位置的远近。

结合上述公式 (1)，标准的注意力机制计算公式如下：

$$\begin{aligned} Q &= HW_q \\ K &= HW_k \\ V &= HW_v \\ A &= QK^T/\sqrt{d} \\ H_0 &= \text{softmax}(A)V \end{aligned} \quad (2)$$

其中，H 表示输入向量，Q、K 和 V 为通过对 H 作映射得到的内容向量， H_0 为自注意力机制输出结果， W_q, W_k, W_v 为可学习的映射矩阵，A 为注意力权重矩阵，N 为输入向量长度，d 为输入向量的维度。

需要说明的是，H 为一个维度为 $N \times d$ 的实数矩阵，可以表示为 $H \in \mathbb{R}^{N \times d}$ ； W_v 是一个维度为 $d \times d$ 的实数矩阵，可以表示为 $W_v \in \mathbb{R}^{d \times d}$ ；A 是一个维度为 $N \times N$ 的实数矩阵，可以表示为 $A \in \mathbb{R}^{N \times N}$ 。

对于上述公式 (2)，使用公式 (1) 中的方式将其进一步分解公式 (3)。其中 Q_c, K_c, V_c 表示由矩阵 $W_{q,c}, W_{k,c}, W_{v,c}$ 映射后的内容向量， Q_r, K_r 表示 $W_{q,r}, W_{k,r}$ 由映射后的相对位置向量，公式如下：

$$\begin{aligned} Q_c &= HW_{q,c}, K_c = HW_{k,c}, V_c = HW_{v,c}, Q_r = PW_{q,r}, K_r = PW_{k,r}, \hat{A}_{i,j} \\ &= Q_i^c K_j^{cT} + Q_i^c K_{\delta(i|j)}^{rT} + K_j^c Q_{\delta(j|i)}^{rT} \\ H_0 &= \text{softmax}(\hat{A}/\sqrt{3d})V_c \end{aligned} \quad (3)$$

其中， $\hat{A}_{i,j}$ 为注意力矩阵 $\hat{A}$ 中的元素，表示第 i 个字符到第 j 个字符间的注意力权重， $Q_i^c K_j^{cT}, K_j^c Q_{\delta(j|i)}^{rT}, Q_i^c K_{\delta(i|j)}^{rT}$ 对应公式 (1) 中前三项的计算公式， Q_i^c 为 Q_c 的第 i 行向量， K_j^c 为 K_c 的第 j 行向量， $\delta(i|j)$ 表示文本中的字符 i 和字

符 j 间的相对距离, $\sqrt{3d}$ 为尺度因子, d 为输入向量的维度。

需要说明的是,通过注意力权重,可以计算序列中每个位置的加权值,从而帮助模型更好地捕捉序列中的相关性,得到准确的文本特征序列。

示例性地,在通过上述公式(2)计算得到注意力权重后,将序列中每个元素与该元素对应的注意力权重相乘,得到序列中每个元素的加权值,也就是说,通过在序列的不同位置上分配不同的权重,获得注意力结果 H_0 ,从而得到更加准确的文本特征序列。

可选地,在本申请实施例中,第一待识别文本可以为将多个第一文本进行拼接后得到的文本,换言之,上述多个第一文本可以为第一待识别文本中的文本片段。

可选地,在本申请实施例中第一文本特征序列可以包括多个文本特征序列,一个文本特征序列对应第一待识别文本中的一个文本片段。

进一步可选地,在通过 DeBERTa 模型的 Transformer 层对上述拼接后得到的文本进行处理时,文本识别装置在第一待识别文本输入至 DeBERTa 模型后,通过该 DeBERTa 模型将该第一待识别文本转换为向量,得到第一待识别文本中的每个文本片段对应的字符嵌入向量以及类型嵌入向量,并将字符嵌入向量与类型嵌入向量的和作为 Transformer 层的输入,得到注意力结果 H_0 。

进一步地,上述字符嵌入向量为字符内容的向量表达形式;类型嵌入向量为人工给每个拼接的句打上标记,然后将标记转换为向量,而得到的向量,类型嵌入向量可以使模型能够区分不同的句子。

需要说明的是,输入向量为多个句子拼接输入至模型,通过类型嵌入向量给拼接句子里的不同句子打标记,使模型能够区分不同句子。

进一步地,上述 H_0 可以包括 M 个文本特征序列,每个序列中包括至 N 个元素,即,每个序列的长度可以为 N 。其中,一个元素可以对应一个字符嵌入向量,每个字符嵌入向量的维度可以相同。

进一步地,每个文本特征序列包含上述第一待识别文本中的一个文本片段对应的文本特征序列。

举例说明,以第一待识别文本为对文本 1、文本 2 和文本 3 进行拼接后得到的文本,在通过 Transformer 层对该文本处理后,得到的注意力结果 H_0 包括文本 1 对应的序列,文本 2 对应的序列和文本 3 对应的序列,其中每个序列可以包括 8 个元素,即序列的长度为 8。

需要说明的是,上述 M 、 N 的取值仅为一种示例,并不构成对本申请的限定, M 和 N 的取值的具体可以根据实际需求确定。

步骤 S203: 文本识别装置采用卷积层对第一文本特征序列进行分段卷积处理,得到第二文本特征序列。

其中,上述第二文本特征序列至少包含第一待识别文本的全局语义信息和第一待识别文本中的第一文本片段的局部语义信息,上述第一文本片段包含上述第一多音字。

可选地，在本申请实施例中，文本识别装置可以将上述第一文本特征序列输入至 DeBERTa 模型的 CNN 层，并通过 CNN 层对该第一文本特征序列进行分段卷积处理，得到第二文本特征序列。

示例性地，以第一文本特征序列为 Transformer 层的输出向量 H_0 为例，文本识别装置在通过 CNN 层对该第一文本特征序列进行分段卷积处理时，可以通过一个固定尺寸的卷积核对上述输出向量 H_0 的序列进行分段卷积处理。具体地，将固定大小的卷积核与输出向量 H_0 分段进行卷积计算获得局部的卷积结果，当计算结束后，前移卷积核与下一段序列进行卷积，重复该卷积操作遍历整个序列，最后输出整体的卷积结果。

以下结合附图对本申请实施例中的分段卷积处理过程进行举例说明。

图 2 为本申请实施例提供的改进后的 DeBERTa 模型中的 CNN 层的结构示意图，如图 2 所示，上述 DeBERTa 模型至少包括 CNN 输入层、卷积层和输出层，其中，上述 CNN 输入层用于输入上述第一文本特征序列，即，Transformer 层的输出向量 H_0 ，该输出向量 H_0 的一个序列包括 X_1 、 X_2 、 X_3 、 X_4 、 X_5 、 X_6 、 X_7 以及 X_8 以上八个元素，每个元素表示一个字符嵌入向量，每个字符嵌入向量的维度为 3。上述卷积层用于通过卷积核对上述第一文本特征序列进行分段卷积，上述输出层用于输出卷积结果。

需要说明的是，图 2 中的输入层矩阵中的每一列表示对应字符嵌入向量的维度。

结合上述图 2，在将上述输出向量 H_0 输入至卷积层后，通过固定大小的卷积核与输出向量 H_0 分段进行卷积计算获得局部的卷积结果，当计算结束后，前移卷积核与下一段序列进行卷积，重复该卷积操作遍历整个序列，最后输出整体的卷积结果。比如，卷积核的大小远小于输出向量 H_0 的序列长度，因此使用卷积核与向量做卷积时，可以分段进行卷积。

举例说明，卷积核大小为 3×1 ， H_0 大小为 42×1 ，那么第一次卷积运算时，卷积核与 H_0 的第 1、2、3 个位置做乘法并求和，得到一个卷积结果。之后前向移动卷积核，通过卷积核与序列的第 2、3、4 个位置的元素做乘法，分别得到对应的卷积结果。然后基于每个卷积结果的和，得到最终的卷积结果。如此，通过控制卷积层中的卷积核大小，在做卷积运算时能够仅关注指定范围内的内容，从而获得准确的局部语义信息。

可选地，在本申请实施例中，文本识别装置可以根据得到的局部语义信息，确定与该局部语义信息对应的文本片段，从而确定多音字所在的文本片段的语义信息。

示例性地，以第一待识别文本为“我们一起同行的人都是同行”为例，则通过 Transformer 层获取到该文本的全局语义信息可以为“我们在行走”，进一步通过 CNN 层获取到该文本的全局语义信息可以为“我们在行走”，获取到的局部语义信息可以包括“我们同行”和“是同行”。

可以理解的是，后续通过该文本的全局语义信息进行多音字预测时，可能会将该文本中的“行”的读音均预测为“xing2”，导致准确性较差。

而通过局部语义信息进行多音字预测时,可以根据第一个位置出现的“行”所在的文本片段的语义信息“我们同行”预测该处的“行”的读音为“xing2”,根据第二个位置出现的“行”所在的文本片段的语义信息预测该处的“行”的读音为“hang2”。

5 步骤 S204: 文本识别装置基于上述第二文本特征序列,得到第一待识别文本对应的第一编码信息。

可选地,在本申请实施例中,文本识别装置可以将上述第二文本特征序列输入至 DeBERTa 模型的编码模块,得到上述第一编码信息。

以下结合附图对本申请实施例提供的编码方法进行解释说明。

10 图 3 为本申请实施例提供的改进的 DeBERTa 模型的结构示意图,如图 3 所示,该 DeBERTa 模型至少包括: Transformer 层、CNN 层和编码模块,即 Mask Decoder。如图 3 所示,文本识别装置将多个不同的文本拼接,作为模型输入信息输入至 DeBERTa 模型,并得到拼接文本对应的字符嵌入向量和类型嵌入向量,然后将字符嵌入向量与类型嵌入向量的和作为
15 Transformer 层输入,通过 Transformer 层结合相对位置嵌入向量计算得到拼接文本对应的表征全局语义的文本特征向量后,将该文本特征向量输入至 CNN 层进行分段卷积处理,得到表征局部语义和全局语义的文本特征向量,即第二文本特征序列,最后将该表征局部语义和全局语义的文本特征向量输入至编码模块,得到最终的句编码结果。如此,通过将 Transformer
20 与 CNN 相结合,使模型不但能从 Transformer 获得全局语义信息,也能通过 CNN 更好地对局部序列信息进行建模,考虑了序列语义信息的可能存在的局部强相关信息,使模型在预测多音字时能够综合全局和局部信息进行判断。将 CNN 输出层输出的结果经过 Decoder 获得最终的句编码结果,在已有标注的验证集测试中,分别对比了基于字、句编码的读音预测结果
25 和标注的读音结果,统计结果表明,基于改进的 DeBERTa 生成句编码的多音字预测结果准确率远高于基于字编码的多音字读音结果。

本申请实施例提供的文本识别方法,文本识别装置将第一待识别文本输入第一文本编码模型,该第一待识别文本包含第一多音字,该第一文本编码模型包括卷积层,然后,提取第一待识别文本对应的第一文本特征序列,并采用所述卷积层对上述第一文本特征序列进行分段卷积处理,得到
30 第二文本特征序列,上述第二文本特征序列至少包含所述第一待识别文本的全局语义信息和第一待识别文本中的第一文本片段的局部语义信息,上述第一文本片段包含第一多音字,最后,基于上述第二文本特征序列,得到第一待识别文本对应的第一编码信息。通过该方法,文本识别装置通过
35 第一文本编码模型中的卷积层对文本的全局文本特征序列进行卷积操作获得包含全局语义信息和局部语义信息的文本特征序列,并基于该文本特征序列得到更为准确的识别结果,从而加强对文本序列局部信息建模能力,能够弥补现有模型对于局部特征信息提取的不足,从而能够更为准确地分析长难句、复杂文本中多音字的读音,进而提高对文本中的多音字读音预

测的准确率。

可选地，在本申请实施例中，上述步骤 S203 可以包括以下步骤 S203a 和步骤 S203b：

5 步骤 S203a：文本识别装置通过上述卷积层中的卷积核对应的卷积窗口，对上述第一文本特征序列进行分段卷积处理，得到 N 个序列片段。

其中，任一序列片段对应的文本特征序列包含上述任一序列片段对应文本片段的局部语义信息。

可选地，在本申请实施例中，上述第二文本特征序列由上述 N 个序列片段组成。

10 可选地，在本申请实施例中，上述第一文本特征序列可以包括多个文本特征序列，上述每个文本特征序列包含至少一个按序排列的文本特征元素。

需要说明的是，对于卷积层的相关解释说明可以参见上文，此处不再赘述，

15 可选地，在本申请实施例中，文本识别装置可以根据卷积窗口的窗长，每次选择一个第一文本特征序列中的至少两个特征元素进行卷积处理，并且根据卷积步长移动卷积窗口，以选择特征元素进行下一次的卷积处理。

可选地，在本申请实施例中，一个序列片段中包含的文本特征元素的数量是基于卷积窗口的窗长确定的，

20 示例性地，结合上述实施例，以文本特征序列包括 X1、X2、X3、X4、X5、X6、X7 以及 X8 以上八个元素、卷积核的尺寸为 3*1，卷积步长为 1 为例，在进行分段卷积处理时，卷积核与文本特征序列的第 1 个位置的元素 X1、第 2 个位置的元素 X2 和第 3 个位置的元素 X3 做乘法并求和，得到一个卷积结果 R1。之后前向移动卷积核，通过卷积核与序列的第 2、3、
25 4 个位置的元素（即，X2、X3、X4）做乘法并求和，得到对应的卷积结果 R1，以此类推，通过卷积核与序列的第 6、7、8 个位置的元素（即，X6、X7、X8）做乘法并求和，得到对应的卷积结果 R6，然后基于每个卷积结果的和，得到最终的卷积结果。

30 如此，通过控制卷积层中的卷积核大小，在做卷积运算时能够仅关注指定范围内的内容，从而获得准确的局部语义信息。

可选地，在本申请实施例中，上述步骤 S201 之前，本申请实施例提供的文本识别方法还包括以下步骤 C1 至步骤 C3：

步骤 C1：文本识别装置接收用户的第一输入。

35 步骤 C2：文本识别装置响应于上述第一输入，将第一输入对应的待识别文本存储至第一文件中。

步骤 C3：文本识别装置在第一文件中存储的待识别文本的数量超过第一阈值的情况下，将第一文件中存储的待识别文本合成为上述第一待识别文本。

可选地，在本申请实施例中，上述第一输入可以为用户的触控输入、

语音输入或者手势输入等任意具备可行性的输入，本申请实施例对此不作限定。

进一步地，上述触控输入可以为用户的点击输入、长按输入或者滑动输入等。

5 可选地，在本申请实施例中，上述第一阈值可以为 50 条、100 条或者 150 条等，本申请实施例对此不作限定。

可选地，在本申请实施例中，文本识别装置可以将上述第一文件中存储的待识别文本进行文本拼接，得到上述第一待识别文本。

10 示例性地，以第一输入为用户对待识别文本的点击输入为例，用户点击电子设备的界面上显示的需要进行语音合成的文本后，文本识别装置获取该文本，并将该文本自动写入固定目录文件 D，当该目录文件 D 中文本数量达到预先设定的数量时，将会自动启动读音纠正系统，并将所积累的文本进行拼接得到上述第一待识别文本，并将该第一待识别文本输入至第一文本编码模型进行处理。

15 本申请实施例提供的文本识别方法，用户在手机端请求朗读文本时，编码装置可以将每次请求朗读的文本存储至文件中，并在文本数量超过一定阈值后，将文本输入至第一文本编码模型进行处理，从而能够第一文本编码模型对用户惯常需要朗读的文本进行准确地语义识别和编码，得到更为准确的编码信息，从而在后续能够基于编码更为准确地预测文本中多音字的读音。

20 可选地，在本申请实施例中，上述步骤 S204 之后，本申请实施例提供的文本识别方法还包括以下步骤 S205：

步骤 S205：文本识别装置将上述第一编码信息输入至第一多音字模型，对该第一多音字进行读音预测，得到该第一多音字的第一读音预测信息。

25 可选地，在本申请实施例中，上述第一多音字模型可以由双向长短时记忆神经网络和线性神经网络训练获得。

可选地，在本申请实施例中，文本识别装置将上述第一编码信息输入至第一多音字模型后，基于句编码的多音字模型推理，得到对应的多音字读音预测结果。

30 举例说明，以第一待识别文本包括“差点着了他的道”为例，在基于上述第一文本编码模型得到该文本的编码信息后，将该编码信息输入至第一多音字模型，通过第一多音字模型预测该文本中的多音字“着”发音为“zhao1”的概率为 85%，大于发音为“zhe5”的概率，则输出该文本的读音预测信息“zhao1”。需要说明的是，读音预测信息“zhao1”中的“zhao”表示拼音，“1”表示声调，即一声。

35 进一步可选地，在本申请实施例中，上述步骤 S205 之后，本申请实施例提供的文本识别方法还包括以下步骤 S206：

步骤 S206：文本识别装置将上述第一待识别文本作为预设文本，与该第一读音预测信息关联存储至文本读音库。

其中，上述文本读音库中包含多个预设文本以及每个预设文本对应的读音预测信息，每个预设文本中包含至少一个多音字。

为了便于理解，上述文本读音库可以称为纠错干预库。

5 可选地，在本申请实施例中，上述文本读音库中存储的数据的数据格式为：<文本>|<带目标多音字拼音的文本>。例如，文本读音库中存储的数据包括：“他的穿着很普通|他的穿着（zhuo2）很普通”，“差点着了他的道|差点着 zhao1 了他的道”，“家里着火了|家里着（zhao2）火了”。

10 如此，通过对累积的多个待识别文本基于句编码结果进行读音预测，从而得到更为准确的读音预测结果，并且，将得到的更为准确的读音预测信息预存至文本读音库中，使得后续进行在线语音合成时，可以通过文本读音库对基于字编码结果进行读音预测后得到的读音预测结果进行纠正，从而在保证处理速度的同时保证了读音预测的高准确率。

15 进一步可选地，在本申请实施例中，文本识别装置得到该第一多音字的第一读音预测信息，并将第一待识别文本作为预设文本，与该第一读音预测信息关联存储至文本读音库之后，可以获取文本读音库中错误的读音预测信息，并将错误的读音预测信息添加到上述第一多音字模型的训练数据集，以通过错误的读音预测信息对该第一多音字模型进行训练，从而对模型迭代更新优化。

20 可选地，文本识别装置可以对文本读音库中保存的预设文本对应的读音预测信息进行检查，在检查到预设文本对应的读音预测信息为错误信息时，将该错误的读音预测信息添加到上述第一多音字模型的训练数据集，以通过错误的读音预测信息对该第一多音字模型进行训练。

25 示例性地，文本识别装置将错误的读音预测信息分别加入训练集中训练，通过积累的数据集迭代训练多音字模型，修复原本在线朗读系统中读音出错的例子，增强模型的泛化能力，从而实现周期性地提高多音字模型读音预测准确率。

可选地，在本申请实施例中，本申请实施例提供的一种读音预测方法，该读音预测方法可以包括以下步骤 A1 至步骤 A4：

30 步骤 A1：文本识别装置将第二待识别文本输入第二文本编码模型，提取第二待识别文本对应的第三文本特征序列。

其中，上述第二待识别文本包含第二多音字，上述第二文本编码模型与上述第一文本编码模型不同。

步骤 A2：文本识别装置基于上述第三文本特征序列，得到上述第二待识别文本对应的第二编码信息。

35 步骤 A3：文本识别装置将上述第二编码信息输入至第二多音字模型，对上述第二多音字进行读音预测，得到上述第二多音字的第三读音预测信息。

步骤 A4：文本识别装置在文本读音库中存在目标预设文本的情况下，将上述第三读音预测信息更新为目标预设文本对应的第四读音预测信息。

其中，上述目标预设文本为上述文本读音库中与第二待识别文本匹配的文本。

可选地，在本申请实施例中，上述第二文本编码模型可以为字编码模型。

5 可选地，在本申请实施例中，该多音字模型可以由双向长短时记忆神经网络和线性神经网络训练获得。

示例性地，在用户需要对文本进行朗读时，用户点击该文本后，编码装置获取文本请求并将该文本请求实时发送到语音合成引擎端，通过语音合成引擎端采用字编码方式对文本编码处理，并将编码后的编码结果输入到第二多音字模型中。

10 可选地，在本申请实施例中，在进行字编码处理时，语音合成引擎端使用 Focal Loss 损失函数以减缓数据不平衡带来的影响，从而提高多音字读音准确率，其公式如下：

$$\text{Focal Loss} = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (4)$$

15 其中，通过权重因子 α 抑制正负样本的数量失衡的情况，权重因子 α 的取值可以为 $\{0,1\}$ ，即， $\alpha \in [0,1]$ ； $(1 - p_t)^\gamma$ 为调制因子，辅助难、易样本的区分。通过 Focal Loss 损失函数，改善了样本类别不均衡问题。

需要说明的是，由于多音字读音的使用频率及场景不同，因此在训练数据中每种读音样本数据难以保持均衡，例如训练集中把的读音“ba3”数据量是“ba4”的十倍，这导致模型预测结果会倾向于预测“ba3”，将部分“ba4”读音预测错误。

可选地，在本申请实施例中，上述第二多音字模型的模型结构及损失函数与上述第一多音字模型相同。

25 可选地，在本申请实施例中，在得到第二待识别文本基于字编码得到的第三读音预测信息的情况下，将该第二待识别文本与文本读音库中的预设文本进行匹配，在文本读音库中存在与上述第二待识别文本匹配的目标预设文本的情况下，将上述第三读音预测信息更新文本读音库中目标预设文本对应的第四读音预测信息。

30 进一步可选地，在文本读音库中存在与上述第二待识别文本匹配的目标预设文本的情况下，可以判断该第二待识别文本的第三读音预测信息与目标预设文本对应的第四读音预测信息是否相同，若相同，则输出上述第三读音预测信息，若不同，则将第三读音预测信息更新为第四读音预测信息。

35 举例说明，以第二待识别文本为“他的穿着很普通”为例，假设第二待识别文本基于字编码得到读音预测结果为“他的穿着(zhe5)很普通”，将该第二待识别文本与文本读音库中的预设文本进行匹配，得到与其匹配的文本对应的读音预测结果为“他的穿着(zhuo2)很普通”，则将该第二待识别文本的读音预测结果纠正为“他的穿着(zhuo2)很普通”。

需要说明的是，读音“zhe5”中的“zhe”为拼音，“5”为声调，即

轻声。同理，读音“zhuo2”中的“zhuo”为拼音，“2”为声调，表示二声。

需要说明的是，上述步骤 A1 至步骤 A4 可以在步骤 S201 之前执行，或者在上述步骤 S201 之后执行。

5 本申请实施例提供的文本识别方法，结合了句编码和字编码的优势，在保证低延迟语音朗读服务的同时，能够及时快速纠正错误读音；根据纠错系统的纠错数据，周期性地迭代多音字模型，提高多音字读音准确率，极大减小了人工查错纠错所需的人力和时间成本。

10 进一步可选地，在本申请实施例中，上述步骤 A3 之后，本申请实施例提供的文本识别方法还包括以下步骤 B1：

步骤 B1：文本识别装置在检测到上述第三读音预测信息为错误读音预测信息的情况下，将上述第三读音预测信息添加至训练数据集。

其中，上述训练数据集用于训练上述第二多音字模型。

15 可选地，在本申请实施例中，在第二待识别文本的第三读音预测信息与从文本读音库中匹配到的第四读音预测信息不同的情况下，将该第三读音预测信息作为错误的读音预测信息添加至上述第二多音字模型的训练集，以对第二多音字模型进行模型迭代更新优化。

20 示例性地，文本识别装置将基于字编码得到的错误读音预测信息加入根据字编码的编码结果预测读音的多音字模型的训练集中，并基于积累的错误读音预测信息训练该多音字模型，以修复原本在线朗读系统中读音出错的例子，增强模型的泛化能力，从而实现周期性地提高多音字模型读音预测准确率。

以下结合流程图通过具体的实施例对本申请实施例提供的文本识别方法进行说明。

25 图 4 为本申请实施例提供的文本识别方法的流程图，如图 4 所示，该文本识别方法可以包括以下步骤 101 至步骤 109：

步骤 101：通过第二文本编码模型对用户请求识别的文本进行字编码。

30 示例性地，用户点击文本进行语音朗读，此时该文本请求会实时发送到语音合成引擎端，对文本进行字编码处理，并将编码后的结果输入到步骤 102 的多音字模型中。

步骤 102：通过第二多音字模型获得读音预测结果。

需要说明的是，对于该步骤的解释说明可以参见上文描述，此处不再赘述。

步骤 103：干预库内匹配查询目标文本。

35 示例性地，若干预库中存在该文本，匹配结果为 True，则使用干库对应的结果，并跳转至步骤 104；若干预库中不存在该文本，则匹配结果为 False，跳转至步骤 105。

步骤 104：读音结果干预修复。

示例性地，若步骤 104 匹配结果为 True，则使用干预库中所对应的文

本读音结果，跳转至步骤 105，完成错误读音的纠正。

步骤 105：输出最终的读音。

步骤 106：文本积累。

5 示例性地，读音纠正系统非实时运行，用户在手机端请求朗读的文本会自动写入固定目录文件 D，当请求文本数量达到预先设定的数量时，将会自动启动读音纠正系统，并将所积累的文本送入步骤 107。

步骤 107：基于大模型 DeBERTa 进行文本句编码。

10 示例性地，读取步骤 106 中积累的文本，使用预训练大模型 DeBERTa 批量输出文本句编码结果，预训练大模型 DeBERTa，即上述第一文本编码模型。

步骤 108：基于句编码的第二音字模型获得读音预测结果。

步骤 109a：对第一多音字模型进行迭代更新。

步骤 109b：对第二多音字模型进行迭代更新。

15 需要说明的是，对该实施例中的解释说明可以参见上文，此处不再赘述。

本申请实施例提供的文本识别方法，一方面，通过多音字读音自动化纠错系统，能够自动纠错线上语音合成多音字读音，通过改进的 DeBERTa 预训练模型生成的句编码能够更好地表达和捕捉语境和语义信息，保证了读音的高准确率。另一方面，通过预训练大模型句编码训练的多音字模型指导线上语音合成多音字模型迭代更新，提高线上多音字模型自身的读音准确率。再一方面，通过自动纠错系统进行纠错，避免了人工纠错所需要的高昂时间和人力成本，同时提高了纠错效率。

20 本申请实施例提供的文本识别方法，执行主体可以为文本识别装置。本申请实施例中以文本识别装置执行文本识别方法为例，说明本申请实施例提供的文本识别装置。

图 5 为本申请实施例提供的文本识别装置的结构示意图，如图 5 所示，该文本识别装置 500 包括执行模块 501 和处理模块 502，其中：

上述执行模块 501，用于将第一待识别文本输入第一文本编码模型，上述第一待识别文本包含第一多音字，上述第一文本编码模型包括卷积层；

30 上述处理模块 502，用于提取第一待识别文本对应的第一文本特征序列；

上述处理模块 502，还用于采用上述卷积层对上述第一文本特征序列进行分段卷积处理，得到第二文本特征序列，上述第二文本特征序列至少包含上述第一待识别文本的全局语义信息和上述第一待识别文本中的第一文本片段的局部语义信息，上述第一文本片段包含上述第一多音字；

35 上述处理模块 502，还用于基于上述第二文本特征序列，得到上述第一待识别文本对应的第一编码信息，并基于上述第一编码信息获得上述第一待识别文本的识别结果。

可选地，在本申请实施例中，上述处理模块，具体用于通过上述卷积

层中的卷积核对应的卷积窗口，对上述第一文本特征序列进行分段卷积处理，得到N个序列片段对应的文本特征序列，一个上述序列片段中包含的文本特征元素的数量是基于上述卷积窗口的窗长确定的，任一序列片段对应的文本特征序列包含上述任一序列片段对应文本片段的局部语义信息；
5 其中，上述第二文本特征序列由上述N个序列片段组成。

可选地，在本申请实施例中，上述处理模块，具体用于将上述第一编码信息输入至第一多音字模型，对上述第一多音字进行读音预测，得到上述第一多音字的第一读音预测信息。

可选地，在本申请实施例中，上述装置还包括：存储模块；

10 上述存储模块，用于在上述处理模块将上述第一编码信息输入至第一多音字模型，对上述第一多音字进行读音预测，得到上述第一多音字的第一读音预测信息之后，将上述第一待识别文本作为预设文本，与上述第一读音预测信息关联存储至文本读音库；

15 其中，上述文本读音库中包含多个预设文本以及每个预设文本对应的读音预测信息，每个上述预设文本中包含至少一个多音字。

可选地，在本申请实施例中，上述处理模块，还用于将第二待识别文本输入第二文本编码模型，提取上述第二待识别文本对应的第三文本特征序列，第二待识别文本包含第二多音字，上述第二文本编码模型与上述第一文本编码模型不同；上述处理模块，还用于基于上述第三文本特征序列，
20 得到上述第二待识别文本对应的第二编码信息；上述处理模块，还用于将上述第二编码信息输入至第二多音字模型，对上述第二多音字进行读音预测，得到上述第二多音字的第三读音预测信息；上述装置还包括：更新模块；上述更新模块，用于在上述文本读音库中存在目标预设文本的情况下，将上述第三读音预测信息更新为上述目标预设文本对应的第四读音预测信息；其中，上述目标预设文本为上述文本读音库中与上述第二待识别文本
25 匹配的文本。

可选地，在本申请实施例中，上述处理模块，还用于在检测到上述第三读音预测信息为错误读音预测信息的情况下，将上述第三读音预测信息添加至训练数据集；其中，上述训练数据集用于训练上述第二多音字模型。

30 可选地，在本申请实施例中，上述装置还包括：接收模块和存储模块，其中：上述接收模块，用于接收用户的第一输入；上述存储模块，用于响应于上述接收模块接收的上述第一输入，将上述第一输入对应的待识别文本存储至第一文件中；上述处理模块，还用于在上述第一文件中存储的待识别文本的数量超过第一阈值的情况下，将上述第一文件中存储的待识别
35 文本合成为上述第一待输入文本。

本申请实施例提供的文本识别装置，文本识别装置将第一待识别文本输入第一文本编码模型，该第一待识别文本包含第一多音字，该第一文本编码模型包括卷积层，然后，提取第一待识别文本对应的第一文本特征序列，并采用上述卷积层对上述第一文本特征序列进行分段卷积处理，得到

第二文本特征序列，上述第二文本特征序列至少包含上述第一待识别文本的全局语义信息和第一待识别文本中的第一文本片段的局部语义信息，上述第一文本片段包含第一多音字，最后，基于上述第二文本特征序列，得到第一待识别文本对应的第一编码信息，并基于上述第一编码信息获得第一待识别文本的识别结果。通过该方法，文本识别装置通过第一文本编码模型中的卷积层对文本的全局文本特征序列进行卷积操作获得包含全局语义信息和局部语义信息的文本特征序列，并基于该文本特征序列得到更为准确的识别结果，从而加强对文本序列局部信息建模能力，能够弥补现有模型对于局部特征信息提取的不足，从而能够更为准确地分析长难句、复杂文本中多音字的读音，进而提高对文本中的多音字读音预测的准确率。

本申请实施例中的文本识别装置可以是电子设备，也可以是电子设备中的部件，例如集成电路或芯片。该电子设备可以是终端，也可以为除终端之外的其他设备。示例性的，电子设备可以为手机、平板电脑、笔记本电脑、掌上电脑、车载电子设备、移动上网装置（Mobile Internet Device, MID）、增强现实(augmented reality, AR)/虚拟现实(virtual reality, VR)设备、机器人、可穿戴设备、超级移动个人计算机（ultra-mobile personal computer, UMPC）、上网本或者个人数字助理（personal digital assistant, PDA）等，还可以为服务器、网络附属存储器（Network Attached Storage, NAS）、个人计算机（personal computer, PC）、电视机（television, TV）、柜员机或者自助机等，本申请实施例不作具体限定。

本申请实施例中的文本识别装置可以为具有操作系统的装置。该操作系统可以为安卓（Android）操作系统，可以为ios操作系统，还可以为其他可能的操作系统，本申请实施例不作具体限定。

本申请实施例提供的文本识别装置能够实现图1至图4的方法实施例实现的各个过程，为避免重复，这里不再赘述。

可选地，如图6所示，本申请实施例还提供一种电子设备600，包括处理器601和存储器602，存储器602上存储有可在上述处理器601上运行的程序或指令，该程序或指令被处理器601执行时实现上述文本识别方法实施例的各个步骤，且能达到相同的技术效果，为避免重复，这里不再赘述。

需要说明的是，本申请实施例中的电子设备包括上述的移动电子设备和非移动电子设备。

图7为实现本申请实施例的一种电子设备的硬件结构示意图。

该电子设备1000包括但不限于：射频单元1001、网络模块1002、音频输出单元1003、输入单元1004、传感器1005、显示单元1006、用户输入单元1007、接口单元1008、存储器1009、以及处理器1010等部件。

本领域技术人员可以理解，电子设备1000还可以包括给各个部件供电的电源（比如电池），电源可以通过电源管理系统与处理器1010逻辑相连，从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。图7

中示出的电子设备结构并不构成对电子设备的限定，电子设备可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置，在此不再赘述。

5 其中，上述处理器 1010，用于将第一待识别文本输入第一文本编码模型，上述第一待识别文本包含第一多音字，上述第一文本编码模型包括卷积层；

上述处理器 1010，用于提取第一待识别文本对应的第一文本特征序列；

10 上述处理器 1010，还用于采用上述卷积层对上述第一文本特征序列进行分段卷积处理，得到第二文本特征序列，上述第二文本特征序列至少包含上述第一待识别文本的全局语义信息和上述第一待识别文本中的第一文本片段的局部语义信息，上述第一文本片段包含上述第一多音字；

上述处理器 1010，还用于基于上述第二文本特征序列，得到上述第一待识别文本对应的第一编码信息，并基于上述第一编码信息获得上述第一待识别文本的识别结果。

15 可选地，在本申请实施例中，上述处理器 1010，具体用于通过上述卷积层中的卷积核对应的卷积窗口，对上述文本特征序列进行分段卷积处理，得到 N 个序列片段对应的文本特征序列，任一序列片段对应的文本特征序列包含上述任一序列片段对应文本片段的局部语义信息；其中，上述第二文本特征序列由上述 N 个序列片段组成。

20 可选地，在本申请实施例中，上述处理器 1010，具体用于将上述第一编码信息输入至第一多音字模型，对上述第一多音字进行读音预测，得到上述第一多音字的第一读音预测信息。

25 可选地，在本申请实施例中，上述存储器 1009，用于在将上述第一编码信息输入至第一多音字模型，对上述第一多音字进行读音预测，得到上述第一多音字的第一读音预测信息之后，将上述第一待识别文本作为预设文本，与上述第一读音预测信息关联存储至文本读音库；

其中，上述文本读音库中包含多个预设文本以及每个预设文本对应的读音预测信息，每个上述预设文本中包含至少一个多音字。

30 可选地，在本申请实施例中，上述处理器 1010，还用于将第二待识别文本输入第二文本编码模型，提取上述第二待识别文本对应的第三文本特征序列，第二待识别文本包含第二多音字，上述第二文本编码模型与上述第一文本编码模型不同；上述处理器 1010，还用于基于上述第三文本特征序列，得到上述第二待识别文本对应的第二编码信息；上述处理器 1010，还用于将上述第二编码信息输入至第二多音字模型，对上述第二多音字进行读音预测，得到上述第二多音字的第三读音预测信息；上述装置还包括：更新模块；上述更新模块，用于在上述文本读音库中存在目标预设文本的情况下，将上述第三读音预测信息更新为上述目标预设文本对应的第四读音预测信息；其中，上述目标预设文本为上述文本读音库中与上述第二待识别文本匹配的文本。

35

可选地，在本申请实施例中，上述处理器 1010，还用于在检测到上述第三读音预测信息为错误读音预测信息的情况下，将上述第三读音预测信息添加至训练数据集；其中，上述训练数据集用于训练上述第二多音字模型。

5 可选地，在本申请实施例中，上述用户输入单元 1007，用于接收用户的第一输入；上述存储器 1009，用于响应于上述用户输入单元 1007 接收的上述第一输入，将上述第一输入对应的待识别文本存储至第一文件中；上述处理器 1010，还用于在上述第一文件中存储的待识别文本的数量超过第一阈值的情况下，将上述第一文件中存储的待识别文本合成为上述第一待输入文本。

10 本申请实施例提供的电子设备，电子设备将第一待识别文本输入第一文本编码模型，该第一待识别文本包含第一多音字，该第一文本编码模型包括卷积层，然后，提取第一待识别文本对应的第一文本特征序列，并采用上述卷积层对上述第一文本特征序列进行分段卷积处理，得到第二文本特征序列，上述第二文本特征序列至少包含上述第一待识别文本的全局语义信息和第一待识别文本中的第一文本片段的局部语义信息，上述第一文本片段包含第一多音字，最后，基于上述第二文本特征序列，得到第一待识别文本对应的第一编码信息，并基于上述第一编码信息获得第一待识别文本的识别结果。通过该方法，文本识别装置通过第一文本编码模型中的

15 卷积层对文本的全局文本特征序列进行卷积操作获得包含全局语义信息和局部语义信息的文本特征序列，并基于该文本特征序列得到更为准确的识别结果，从而加强对文本序列局部信息建模能力，能够弥补现有模型对于局部特征信息提取的不足，从而能够更为准确地分析长难句、复杂文本中多音字的读音，进而提高对文本中的多音字读音预测的准确率。

20 应理解的是，本申请实施例中，输入单元 1004 可以包括图形处理器（Graphics Processing Unit, GPU）1041 和麦克风 1042，图形处理器 1041 对在视频捕获模式或图像捕获模式中由图像捕获装置（如摄像头）获得的静态图片或视频的图像数据进行处理。显示单元 1006 可包括显示面板 1061，可以采用液晶显示器、有机发光二极管等形式来配置显示面板 1061。用户

25 输入单元 1007 包括触控面板 1071 以及其他输入设备 1072 中的至少一种。触控面板 1071，也称为触摸屏。触控面板 1071 可包括触摸检测装置和触摸控制器两个部分。其他输入设备 1072 可以包括但不限于物理键盘、功能键（比如音量控制按键、开关按键等）、轨迹球、鼠标、操作杆，在此不再赘述。

30 存储器 1009 可用于存储软件程序以及各种数据。存储器 1009 可主要包括存储程序或指令的第一存储区和存储数据的第二存储区，其中，第一存储区可存储操作系统、至少一个功能所需的应用程序或指令（比如声音播放功能、图像播放功能等）等。此外，存储器 1009 可以包括易失性存储器或非易失性存储器，或者，存储器 1009 可以包括易失性和非易失性存储

器两者。其中，非易失性存储器可以是只读存储器(Read-Only Memory, ROM)、可编程只读存储器(Programmable ROM, PROM)、可擦除可编程只读存储器(Erasable PROM, EPROM)、电可擦除可编程只读存储器(Electrically EPROM, EEPROM)或闪存。易失性存储器可以是随机存取存储器(Random Access Memory, RAM)、静态随机存取存储器(Static RAM, SRAM)、动态随机存取存储器(Dynamic RAM, DRAM)、同步动态随机存取存储器(Synchronous DRAM, SDRAM)、双倍数据速率同步动态随机存取存储器(Double Data Rate SDRAM, DDRSDRAM)、增强型同步动态随机存取存储器(Enhanced SDRAM, ESDRAM)、同步连接动态随机存取存储器(Synch link DRAM, SLDRAM)和直接内存总线随机存取存储器(Direct Rambus RAM, DRRAM)。本申请实施例中的存储器 1009 包括但不限于这些和任意其它适合类型的存储器。

处理器 1010 可包括一个或多个处理单元；可选的，处理器 1010 集成应用处理器和调制解调处理器，其中，应用处理器主要处理涉及操作系统、用户界面和应用程序等的操作，调制解调处理器主要处理无线通信信号，如基带处理器。可以理解的是，上述调制解调处理器也可以不集成到处理器 1010 中。

本申请实施例还提供一种可读存储介质，所述可读存储介质上存储有程序或指令，该程序或指令被处理器执行时实现上述文本识别方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

其中，所述处理器为上述实施例中所述的电子设备中的处理器。所述可读存储介质，包括计算机可读存储介质，如计算机只读存储器 ROM、随机存取存储器 RAM、磁碟或者光盘等。

本申请实施例另提供了一种芯片，所述芯片包括处理器和通信接口，所述通信接口和所述处理器耦合，所述处理器用于运行程序或指令，实现上述文本识别方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

应理解，本申请实施例提到的芯片还可以称为系统级芯片、系统芯片、芯片系统或片上系统芯片等。

本申请实施例提供一种计算机程序产品，该程序产品被存储在存储介质中，该程序产品被至少一个处理器执行以实现如上述文本识别方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

需要说明的是，在本文中，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。此外，需要指出的是，本申请实施方式中的方法和装置的范围不限按示出或讨论的

顺序来执行功能，还可包括根据所涉及的功能按基本同时的方式或按相反的顺序来执行功能，例如，可以按不同于所描述的次序来执行所描述的方法，并且还可以添加、省去、或组合各种步骤。另外，参照某些示例所描述的特征可在其他示例中被组合。

- 5 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上述实施例方法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计算机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（如
- 10 ROM/RAM、磁碟、光盘）中，包括若干指令用以使得一台终端（可以是手机，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。

- 上面结合附图对本申请的实施例进行了描述，但是本申请并不局限于上述的具体实施方式，上述的具体实施方式仅仅是示意性的，而不是限制性的，本领域的普通技术人员在本申请的启示下，在不脱离本申请宗旨和
- 15 权利要求所保护的范围情况下，还可做出很多形式，均属于本申请的保护之内。

权 利 要 求 书

1.一种文本识别方法，所述方法包括：

将第一待识别文本输入第一文本编码模型，所述第一待识别文本包含第一多音字，所述第一文本编码模型包括卷积层；

5 提取第一待识别文本对应的第一文本特征序列，所述第一文本特征序列包含所述第一待识别文本的语义信息；

采用所述卷积层对所述第一文本特征序列进行分段卷积处理，得到第二文本特征序列，所述第二文本特征序列至少包含所述第一待识别文本的全局语义信息和所述第一待识别文本中的第一文本片段的局部语义信息，
10 所述第一文本片段包含所述第一多音字；

基于所述第二文本特征序列，得到所述第一待识别文本对应的第一编码信息，并基于所述第一编码信息获得所述第一待识别文本的识别结果。

2.根据权利要求1所述的方法，其中，所述采用所述卷积层对所述第一文本特征序列进行卷积处理，得到第二文本特征序列，包括：

15 通过所述卷积层中的卷积核对应的卷积窗口，对所述第一文本特征序列进行分段卷积处理，得到N个序列片段，任一序列片段对应的文本特征序列包含所述任一序列片段对应文本片段的局部语义信息；

其中，所述第二文本特征序列由所述N个序列片段组成。

3.根据权利要求1或2所述的方法，其中，所述基于所述第一编码信息获得所述第一待识别文本的识别结果，包括：

20 将所述第一编码信息输入至第一多音字模型，对所述第一多音字进行读音预测，得到所述第一多音字的第一读音预测信息。

4.根据权利要求3所述的方法，其中，所述将所述第一编码信息输入至第一多音字模型，对所述第一多音字进行读音预测，得到所述第一多音字的第一读音预测信息之后，所述方法还包括：

25 将所述第一待识别文本作为预设文本，与所述第一读音预测信息关联存储至文本读音库；

其中，所述文本读音库中包含多个预设文本以及每个预设文本对应的读音预测信息，每个所述预设文本中包含至少一个多音字。

30 5.根据权利要求4所述的方法，其中，所述方法还包括：

将第二待识别文本输入第二文本编码模型，提取所述第二待识别文本对应的第三文本特征序列，第二待识别文本包含第二多音字，所述第二文本编码模型与所述第一文本编码模型不同；

35 基于所述第三文本特征序列，得到所述第二待识别文本对应的第二编码信息；

将所述第二编码信息输入至第二多音字模型，对所述第二多音字进行读音预测，得到所述第二多音字的第三读音预测信息；

在所述文本读音库中存在目标预设文本的情况下，将所述第三读音预测信息更新为所述目标预设文本对应的第四读音预测信息；其中，所述目

标预设文本为所述文本读音库中与所述第二待识别文本匹配的文本。

6.根据权利要求 5 所述的方法，其中，所述基于所述第二编码信息，得到所述第二多音字的第三读音预测信息之后，所述方法还包括：

5 在检测到所述第三读音预测信息为错误读音预测信息的情况下，将所述第三读音预测信息添加至训练数据集；

其中，所述训练数据集用于训练所述第二多音字模型。

7.根据权利要求 1 所述的方法，其中，所述将第一待识别文本输入文本编码模型之前，所述方法还包括：

接收用户的第一输入；

10 响应于所述第一输入，将所述第一输入对应的待识别文本存储至第一文件中；

在所述第一文件中存储的待识别文本的数量超过第一阈值的情况下，将所述第一文件中存储的待识别文本合成为所述第一待输入文本。

8.一种文本识别装置，所述装置包括：执行模块和处理模块，其中：

15 所述执行模块，用于将第一待识别文本输入第一文本编码模型，所述第一待识别文本包含第一多音字，所述第一文本编码模型包括卷积层；

所述处理模块，用于提取第一待识别文本对应的第一文本特征序列；

20 所述处理模块，还用于采用所述卷积层对所述第一文本特征序列进行分段卷积处理，得到第二文本特征序列，所述第二文本特征序列至少包含所述第一待识别文本的全局语义信息和所述第一待识别文本中的第一文本片段的局部语义信息，所述第一文本片段包含所述第一多音字；

所述处理模块，还用于基于所述第二文本特征序列，得到所述第一待识别文本对应的第一编码信息，并基于所述第一编码信息获得所述第一待识别文本的识别结果。

25 9.根据权利要求 8 所述的装置，其中，所述处理模块，具体用于通过所述卷积层中的卷积核对应的卷积窗口，对所述第一文本特征序列进行分段卷积处理，得到 N 个序列片段，任一序列片段对应的文本特征序列包含所述任一序列片段对应文本片段的局部语义信息；

其中，所述第二文本特征序列由所述 N 个序列片段组成。

30 10.根据权利要求 8 或 9 所述的装置，其中，所述处理模块，具体用于将所述第一编码信息输入至第一多音字模型，对所述第一多音字进行读音预测，得到所述第一多音字的第一读音预测信息。

11.根据权利要求 10 所述的装置，其中，所述装置还包括：存储模块；

35 所述存储模块，用于在所述处理模块将所述第一编码信息输入至第一多音字模型，对所述第一多音字进行读音预测，得到所述第一多音字的第一读音预测信息之后，将所述第一待识别文本作为预设文本，与所述第一读音预测信息关联存储至文本读音库；

其中，所述文本读音库中包含多个预设文本以及每个预设文本对应的读音预测信息，每个所述预设文本中包含至少一个多音字。

12.根据权利要求 11 所述的装置,其中,所述处理模块,还用于将第二待识别文本输入第二文本编码模型,提取所述第二待识别文本对应的第三文本特征序列,第二待识别文本包含第二多音字,所述第二文本编码模型与所述第一文本编码模型不同;

5 所述处理模块,还用于基于所述第三文本特征序列,得到所述第二待识别文本对应的第二编码信息;

所述处理模块,还用于将所述第二编码信息输入至第二多音字模型,对所述第二多音字进行读音预测,得到所述第二多音字的第三读音预测信息;

10 所述装置还包括:更新模块;

所述更新模块,用于在所述文本读音库中存在目标预设文本的情况下,将所述第三读音预测信息更新为所述目标预设文本对应的第四读音预测信息;

15 其中,所述目标预设文本为所述文本读音库中与所述第二待识别文本匹配的文本。

13.根据权利要求 12 所述的装置,其中,所述处理模块,还用于在检测到所述第三读音预测信息为错误读音预测信息的情况下,将所述第三读音预测信息添加至训练数据集;

其中,所述训练数据集用于训练所述第二多音字模型。

20 14.根据权利要求 8 所述的装置,其中,所述装置还包括:接收模块和存储模块,其中:

所述接收模块,用于接收用户的第一输入;

所述存储模块,用于响应于所述接收模块接收的所述第一输入,将所述第一输入对应的待识别文本存储至第一文件中;

25 所述处理模块,还用于在所述第一文件中存储的待识别文本的数量超过第一阈值的情况下,将所述第一文件中存储的待识别文本合成为所述第一待输入文本。

30 15.一种电子设备,包括处理器和存储器,所述存储器存储可在所述处理器上运行的程序或指令,所述程序或指令被所述处理器执行时实现如权利要求 1-7 任一项所述的文本识别方法的步骤。

16.一种可读存储介质,所述可读存储介质上存储程序或指令,所述程序或指令被处理器执行时实现如权利要求 1-7 任一项所述的文本识别方法的步骤。

35 17.一种计算机程序产品,所述计算机程序产品被存储在存储介质中,所述计算机程序产品被至少一个处理器执行以实现如权利要求 1 至 7 任一项所述的文本识别方法。

18.一种芯片,所述芯片包括处理器和通信接口,所述通信接口和所述处理器耦合,所述处理器用于运行程序或指令,实现如权利要求 1 至 7 任一项所述的文本识别方法。

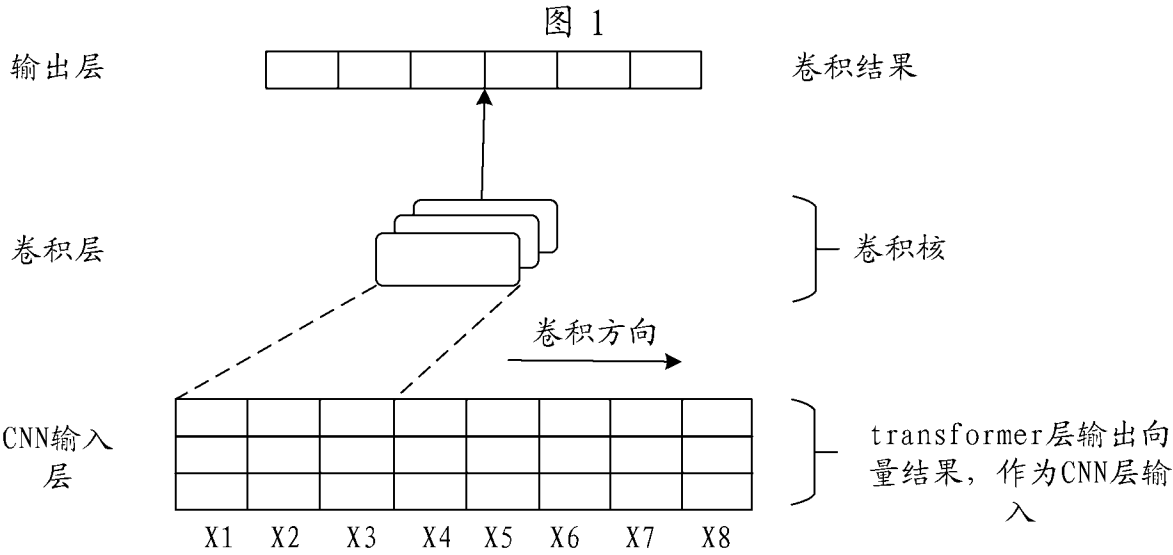
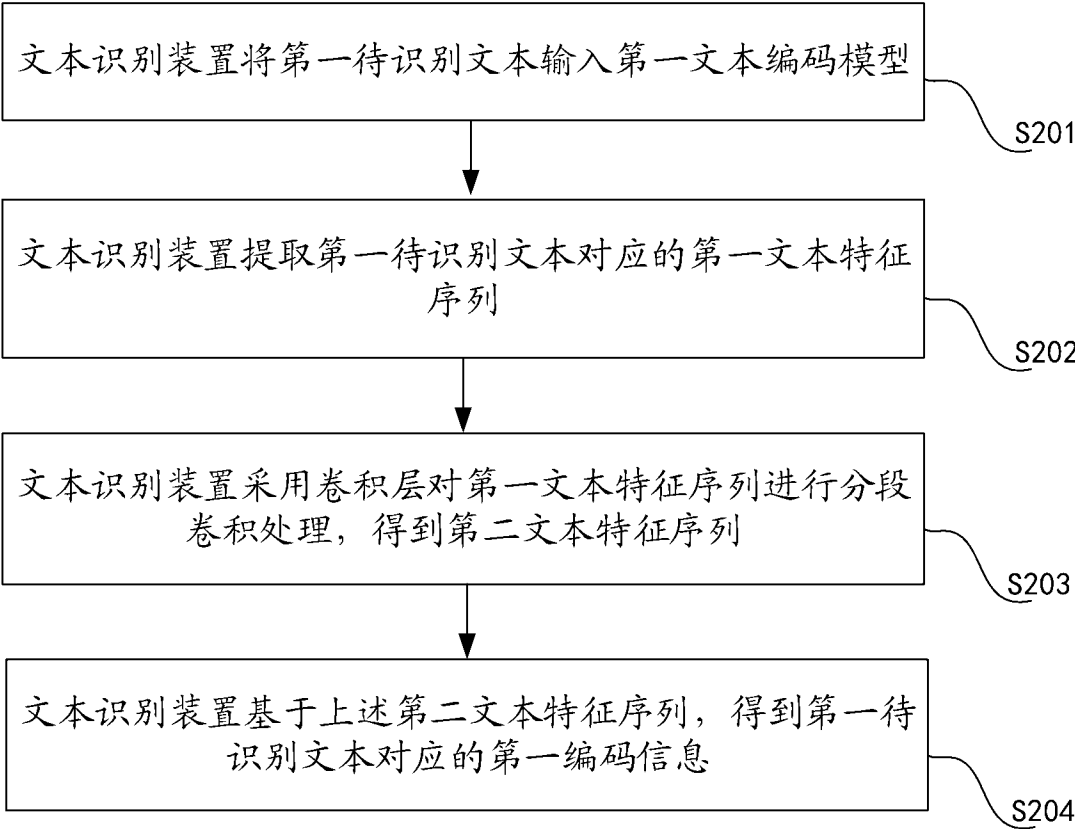


图 2

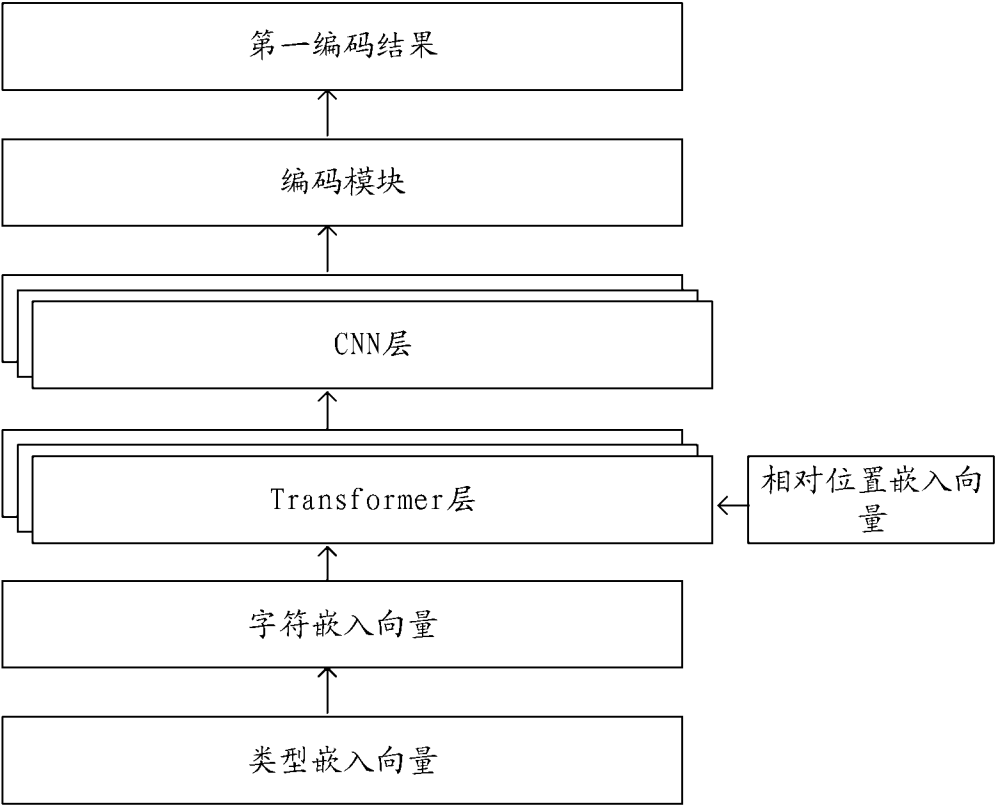


图 3

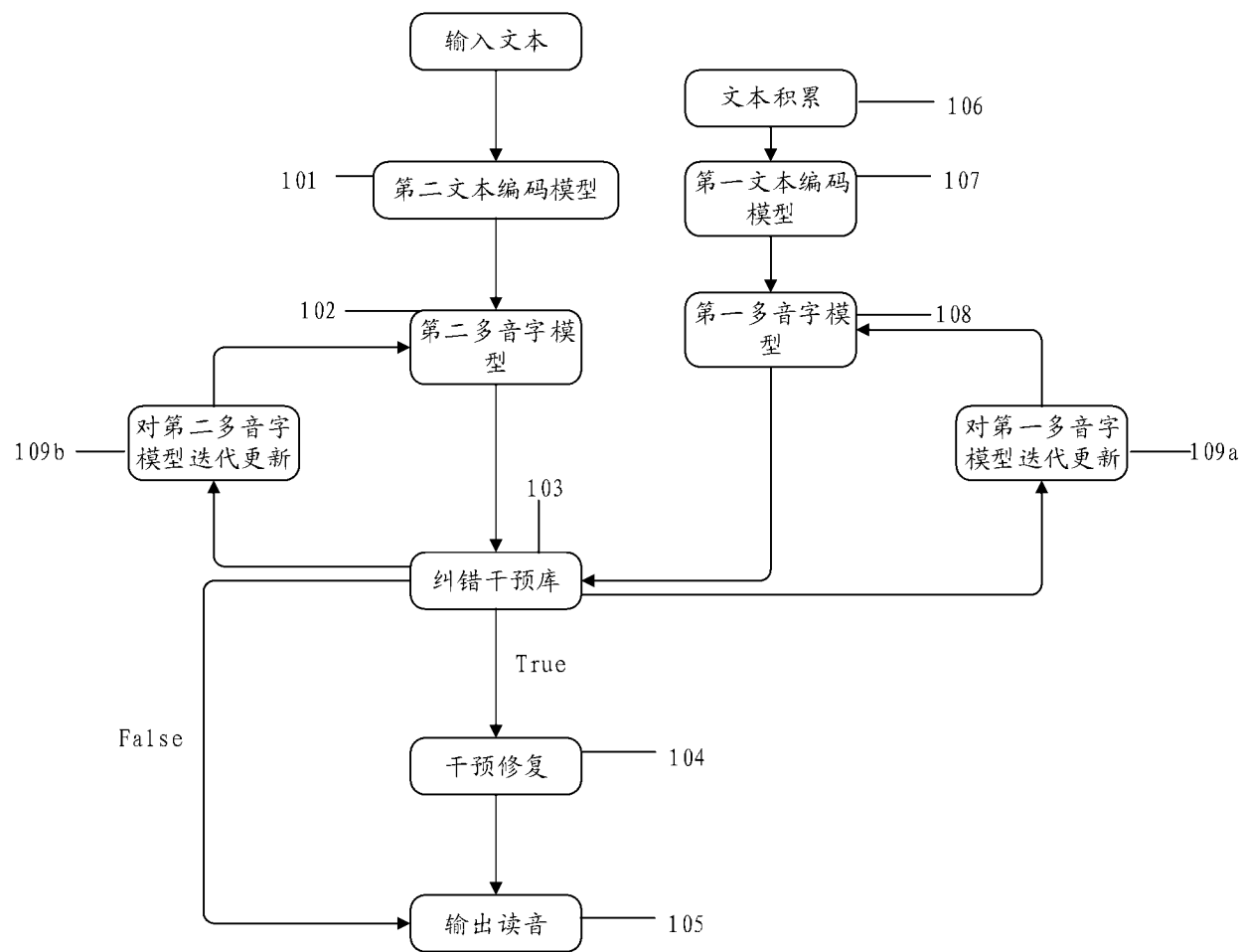


图 4

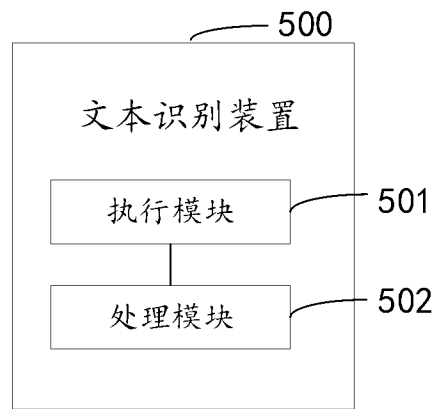


图 5

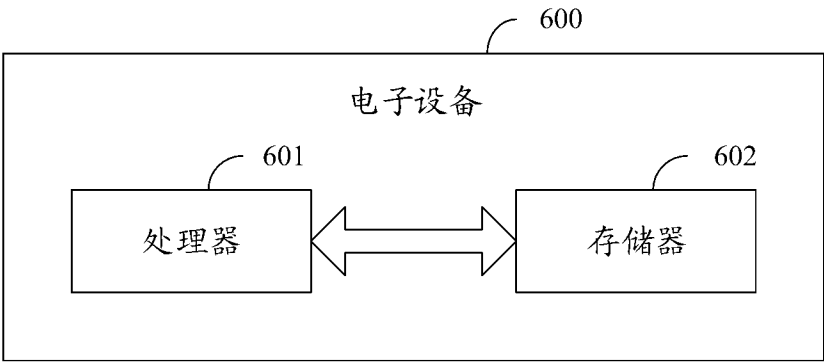


图 6

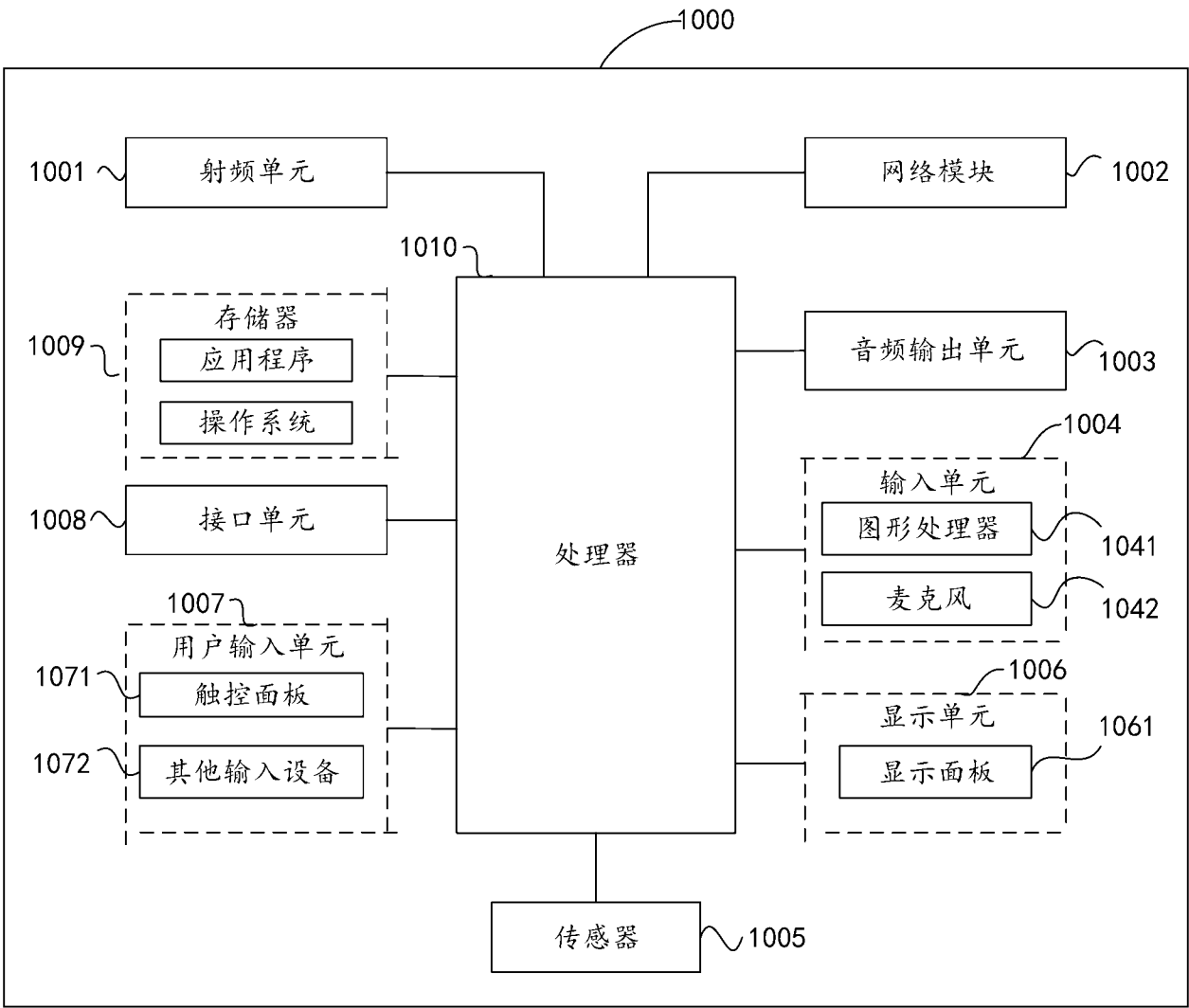


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/104175

A. CLASSIFICATION OF SUBJECT MATTER

G06F40/126(2020.01)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)

IPC: G06F

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)

CNABS; CNTXT; VEN; USTXT; WOTXT; EPTXT; IEEE; CNKI: 文本, 文档, 文件, 多音字, 识别, 卷积, 编码, 语义, 全局, 局部, text, document, file, polyphonic, recognition, convolution, encode, semantic, global, local

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 110277085 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 24 September 2019 (2019-09-24) description, paragraphs 43-47, 62-71, 87-102, and 232-245	1-18
X	CN 113823259 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD. et al.) 21 December 2021 (2021-12-21) description, paragraphs 37-120, and figures 1-3	1-18
PX	CN 116861851 A (VIVO COMMUNICATION TECHNOLOGY CO., LTD.) 10 October 2023 (2023-10-10) claims 1-16, and description, paragraphs 5-12	1-18
A	CN 112989821 A (BEIJING CENTURY TAL EDUCATION TECHNOLOGY CO., LTD.) 18 June 2021 (2021-06-18) entire document	1-18
A	US 2005192807 A1 (EMAM, Ossama et al.) 01 September 2005 (2005-09-01) entire document	1-18

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“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)

“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

“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

“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

“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

“&” document member of the same patent family

Date of the actual completion of the international search

16 October 2024

Date of mailing of the international search report

19 October 2024

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2024/104175

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110277085	A	24 September 2019	CN	110277085	B	24 August 2021
CN	113823259	A	21 December 2021	CN	113823259	B	02 July 2024
CN	116861851	A	10 October 2023	None			
CN	112989821	A	18 June 2021	CN	112989821	B	13 August 2021
US	2005192807	A1	01 September 2005	US	8069045	B2	29 November 2011

A. 主题的分类

G06F40/126(2020.01)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)

IPC: G06F

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))

CNABS;CNTXT;VEN;USTXT;WOTXT;EPTXT;IEEE;CNKI:文本, 文档, 文件, 多音字, 识别, 卷积, 编码, 语义, 全局, 局部, text, document, file, polyphonic, recognition, convolution, encode, semantic, global, local

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 110277085 A (腾讯科技(深圳)有限公司) 2019年9月24日 (2019 - 09 - 24) 说明书第43-47、62-71、87-102、232-245段	1-18
X	CN 113823259 A (腾讯科技(深圳)有限公司等) 2021年12月21日 (2021 - 12 - 21) 说明书第37-120段, 图1-3	1-18
PX	CN 116861851 A (维沃移动通信有限公司) 2023年10月10日 (2023 - 10 - 10) 权利要求1-16, 说明书5-12段	1-18
A	CN 112989821 A (北京世纪好未来教育科技有限公司) 2021年6月18日 (2021 - 06 - 18) 全文	1-18
A	US 2005192807 A1 (EMAM,Ossama 等) 2005年9月1日 (2005 - 09 - 01) 全文	1-18

☐ 其余文件在C栏的续页中列出。

☒ 见同族专利附件。

* 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

“D” 申请人在国际申请中引证的文件

“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

2024年10月16日

国际检索报告邮寄日期

2024年10月19日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

李宝

电话号码 (+86) 020-28958958

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2024/104175

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	110277085	A	2019年9月24日	CN 110277085 B	2021年8月24日
CN	113823259	A	2021年12月21日	CN 113823259 B	2024年7月2日
CN	116861851	A	2023年10月10日	无	
CN	112989821	A	2021年6月18日	CN 112989821 B	2021年8月13日
US	2005192807	A1	2005年9月1日	US 8069045 B2	2011年11月29日