

(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
2023 年 1 月 5 日 (05.01.2023)



(10) 国际公布号
WO 2023/273612 A1

(51) 国际专利分类号:
G10L 15/06 (2013.01) **G10L 15/26** (2006.01)

(21) 国际申请号: PCT/CN2022/091481

(22) 国际申请日: 2022 年 5 月 7 日 (07.05.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110735675.0 2021年6月30日 (30.06.2021) CN

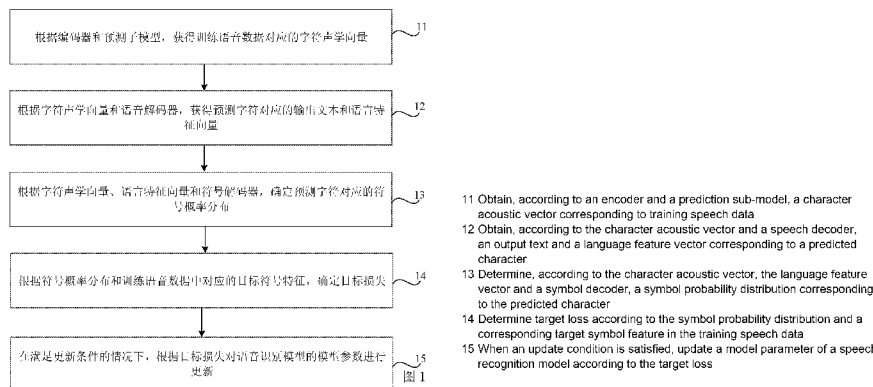
(71) 申请人: 北京有竹居网络技术有限公司
(**BEIJING YOUZHUJU NETWORK TECHNOLOGY CO., LTD.**) [CN/CN]; 中国北京市平谷区林荫北街 13 号信息大厦 802 室, Beijing 101299 (CN).

(72) 发明人: 董林昊 (**DONG, Linhao**); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。 梁镇麟 (**LIANG, Zhenlin**); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。 蔡猛 (**CAI, Meng**); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。 马泽君 (**MA, Zejun**); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。

(74) 代理人: 北京知帆远景知识产权代理有限公司 (**ZHIFAN & PARTNERS**); 中国北京市海淀区阜成路 73 号裕惠大厦 B 座 805, Beijing 100142 (CN)。

(54) **Title:** TRAINING METHOD AND APPARATUS FOR SPEECH RECOGNITION MODEL, SPEECH RECOGNITION METHOD AND APPARATUS, MEDIUM, AND DEVICE

(54) 发明名称: 语音识别模型的训练方法、语音识别方法、装置、介质及设备



(57) **Abstract:** A training method and apparatus for a speech recognition model, a speech recognition method and apparatus, a medium, and a device. The training method comprises: obtaining, according to an encoder and a prediction sub-model, a character acoustic vector corresponding to training speech data (11); obtaining, according to the character acoustic vector and a speech decoder, an output text and a language feature vector corresponding to a predicted character (12); determining, according to the character acoustic vector, the language feature vector and a symbol decoder, a symbol probability distribution corresponding to the predicted character (13); determining target loss according to the symbol probability distribution and a corresponding target symbol feature in the training speech data (14); and when an update condition is satisfied, updating a model parameter of a speech recognition model according to the target loss (15).

(57) **摘要:** 一种语音识别模型的训练方法、语音识别方法、装置、介质及设备, 该训练方法包括: 根据编码器和预测子模型, 获得训练语音数据对应的字符声学向量 (11); 根据字符声学向量和语音解码器, 获得预测字符对应的输出文本和语言特征向量 (12); 根据字符声学向量、语言特征向量和符号解码器, 确定预测字符对应的符号概率分布 (13); 根据符号概率分布和训练语音数据中对应的目标符号特征, 确定目标损失 (14); 在满足更新条件的情况下, 根据目标损失对语音识别模型的模型参数进行更新 (15)。

(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, IQ, IR, IS, IT, JM, JO, 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, 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, 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))。

模型的训练方法、语音识别方法、装置、介质及设备 相关申请的交叉引用

本申请要求于 2021 年 06 月 30 日提交的，申请名称为“模型的训练方法、语音识别方法、装置、介质及设备”的、中国专利申请号为“202110735675.0”的优先权，该中国专利申请的全部内容通过引用结合在本申请中。

技术领域

本公开涉及计算机技术领域，具体地，涉及一种模型的训练方法、语音识别方法、装置、介质及设备。

背景技术

随着深度学习的兴起，各种完全依赖于神经网络进行端到端建模的方法逐渐兴起，逐步发展为自动语音识别（ASR）技术中的主流。通过自动语音识别，可直接将原始的语音数据转换为对应的文本结果。为了进一步提高文本结果的可读性，通常需要该文本结果中增加标点符号。相关技术中，可以采用离线打标点的后处理方式，即在完成语音识别后，对获得的识别结果进行标点处理。该方式需要等识别结果识别完成后，将识别结果输入一个后处理模型进行处理，使得最终的识别结果的延时较大。

技术解决方案

提供该公开内容部分以便以简要的形式介绍构思，这些构思将在后面的具体实施方式部分被详细描述。该公开内容部分并不旨在标识要求保护的技术方案的关键特征或必要特征，也不旨在用于限制所要求的保护的技术方案的范围。

第一方面，本公开提供一种语音识别模型的训练方法，所述语音识别模型包括编码器、预测子模型、语音解码器以及符号解码器，所述方法包括：

根据所述编码器和所述预测子模型，获得训练语音数据对应的字符声学向量，其中，每一所述字符声学向量对应于所述训练语音数据的一个预测字符；

根据所述字符声学向量和所述语音解码器，获得所述预测字符对应的输出文本和语言特征向量；

根据所述字符声学向量、所述语言特征向量和所述符号解码器，确定所述预测字符对应的符号概率分布；

根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失；

在满足更新条件的情况下，根据所述目标损失对所述语音识别模型的模型参数进行更新。

可选地，所述目标符号特征为独热编码特征，所述根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失，包括：

根据所述符号概率分布和所述预测字符对应的目标符号特征，确定交叉熵损失；

将确定出的所述交叉熵损失作为所述目标损失。

可选地，所述根据所述目标损失对所述语音识别模型的模型参数进行更新，包括：

根据所述目标损失对所述语音识别模型中、所述符号解码器的模型参数进行更新。

可选地，所述目标符号特征为独热编码特征，所述根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失，包括：

根据所述符号概率分布和所述预测字符对应的目标符号特征，确定第一交叉熵损失；

根据所述预测字符对应的目标文本与所述输出文本，确定所述语音解码器对应的第二交叉熵损失；

根据所述第一交叉熵损失和所述第二交叉熵损失确定所述目标损失。

可选地，通过以下方式确定所述训练语音数据对应的文本中每一字符对应的符号特征：

获取所述训练语音数据对应的包含标点符号的文本数据；

针对所述文本数据中的每一字符，若该字符后相邻存在标点符号，则将该标点符号作为该字符对应的标注符号；若该字符后相邻未存在标点符号，则将预设符号作为该字符对应的标注符号；

针对所述文本数据中的每一字符，根据该字符对应的标注符号与符号空间，生成该字符对应的符号特征。

第二方面，本公开提供一种语音识别方法，所述方法包括：

接收待识别的语音数据；

根据所述语音数据和语音识别模型，确定所述语音数据对应的待添加的符号和识别文本，其中，所述语音识别模型是基于第一方面所述的语音识别模型的训练方法进行训练获得的；

在所述待添加的符号为目标符号的情况下，根据所述待添加的符号和所述识别文本生成所述待识别的语音数据对应的文本结果。

第三方面，本公开提供一种语音识别模型的训练装置，所述语音识别模型包括编码器、预测子模型、语音解码器以及符号解码器，所述装置包括：

第一处理模块，用于根据所述编码器和所述预测子模型，获得训练语音数据对应的字符声学向量，其中，每一所述字符声学向量对应于所述训练语

音数据的一个预测字符；

第二处理模块，用于根据所述字符声学向量和所述语音解码器，获得所述预测字符对应的输出文本和语言特征向量；

第一确定模块，用于根据所述字符声学向量、所述语言特征向量和所述符号解码器，确定所述预测字符对应的符号概率分布；

第二确定模块，用于根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失；

更新模块，用于在满足更新条件的情况下，根据所述目标损失对所述语音识别模型的模型参数进行更新。

第四方面，本公开提供一种语音识别装置，所述装置包括：

接收模块，用于接收待识别的语音数据；

第三确定模块，用于根据所述语音数据和语音识别模型，确定所述语音数据对应的待添加的符号和识别文本，其中，所述语音识别模型是基于第一方面所述的语音识别模型的训练方法进行训练获得的；

生成模块，用于在所述待添加的符号为目标符号的情况下，根据所述待添加的符号和所述识别文本生成所述待识别的语音数据对应的文本结果。

第五方面，本公开提供一种计算机可读介质，其上存储有计算机程序，该程序被处理装置执行时实现第一方面或第二方面所述方法的步骤。

第六方面，本公开提供一种电子设备，包括：

存储装置，其上存储有计算机程序；

处理装置，用于执行所述存储装置中的所述计算机程序，以实现第一方面或第二方面所述方法的步骤。

在上述技术方案中，根据编码器和预测子模型，获得训练语音数据对应的字符声学向量，从而可以根据所述字符声学向量和所述语音解码器，获得所述预测字符对应的输出文本和语言特征向量，并根据所述字符声学向量、所述语言特征向量和所述符号解码器，确定所述预测字符对应的符号概率分布，之后根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失，以在满足更新条件的情况下，根据所述目标损失对所述语音识别模型的模型参数进行更新。由此，通过上述技术方案，针对语音数据中预测字符的声学向量进行处理，并且通过结合符号解码器以对每一预测字符进行语音识别及标点预测，一方面可以提高语音识别模型的处理精度，另一方面可以有效降低语音识别过程中确定标点符号的时延，使得训练所得的语音识别模型可以应用于流式数据处理过程，为流式识别产品的用户体验提供支撑。并且，在符号解码器进行解码的过程中可以复用预测子模型中的字符声学向量以及语音解码器中的语言特征向量，从而可以综合利用声学信息和语言信息进行标点符号的预测，进一步训练所得的语音识别模型的准确性，提升用户使用体验。

本公开的其他特征和优点将在随后的具体实施方式部分予以详细说明。

附图说明

结合附图并参考以下具体实施方式，本公开各实施例的上述和其他特征、优点及方面将变得更加明显。贯穿附图中，相同或相似的附图标记表示相同或相似的元素。应当理解附图是示意性的，原件和元素不一定按照比例绘制。在附图中：

图1是根据本公开的一种实施方式提供的语音识别模型的训练方法的流程图；

图2是根据本公开的一种实施方式提供的语音识别方法的流程图；

图3为根据本公开的一种实施方式提供的语音识别模型的训练方法的流程图；

图4是根据本公开的一种实施方式提供的语音识别装置的框图；

图5示出了适于用来实现本公开实施例的电子设备的结构示意图。

具体实施方式

下面将参照附图更详细地描述本公开的实施例。虽然附图中显示了本公开的某些实施例，然而应当理解的是，本公开可以通过各种形式来实现，而且不应该被解释为限于这里阐述的实施例，相反提供这些实施例是为了更加透彻和完整地理解本公开。应当理解的是，本公开的附图及实施例仅用于示例性作用，并非用于限制本公开的保护范围。

应当理解，本公开的方法实施方式中记载的各个步骤可以按照不同的顺序执行，和/或并行执行。此外，方法实施方式可以包括附加的步骤和/或省略执行示出的步骤。本公开的范围在此方面不受限制。

本文使用的术语“包括”及其变形是开放性包括，即“包括但不限于”。术语“基于”是“至少部分地基于”。术语“一个实施例”表示“至少一个实施例”；术语“另一实施例”表示“至少一个另外的实施例”；术语“一些实施例”表示“至少一些实施例”。其他术语的相关定义将在下文描述中给出。

需要注意，本公开中提及的“第一”、“第二”等概念仅用于对不同的装置、模块或单元进行区分，并非用于限定这些装置、模块或单元所执行的功能的顺序或者相互依存关系。

需要注意，本公开中提及的“一个”、“多个”的修饰是示意性而非限制性的，本领域技术人员应当理解，除非在上下文另有明确指出，否则应该理解为“一个或多个”。

本公开实施方式中的多个装置之间所交互的消息或者信息的名称仅用于说明性的目的，而并不是用于对这些消息或信息的范围进行限制。

图1所示，为根据本公开的一种实施方式提供的语音识别模型的训练方

法的流程图，所述语音识别模型包括编码器、预测子模型、语音解码器以及符号解码器，如图 1 所示，所述方法包括：

在步骤 11 中，根据编码器和预测子模型，获得训练语音数据对应的字符声学向量，其中，每一所述字符声学向量对应于所述训练语音数据的一个预测字符。

在该实施例中，可以根据编码器对所述训练语音数据进行编码，获得所述训练语音数据的音频帧的声学向量。

通常情况下，可以将每秒的语音数据切分为多个音频帧，从而基于音频帧进行数据处理，示例地，可以将每秒的语音数据切分为 100 个音频帧进行处理。相应地，通过该编码器对该训练语音数据的音频帧进行编码，获得的声学向量序列 H 可以表示为：

$H: \{H_1, H_2, \dots, H_U\}$ ，其中， U 用于表示该输入训练语音数据中的音频帧的数量，即该声学向量序列的长度。

之后，可以根据声学向量和所述预测子模型，获得训练语音数据对应的字符声学向量。

示例地，可以将所述声学向量输入预测子模型，则该预测子模型可以对声学向量进行信息量预测，获得该音频帧对应的信息量。之后可以根据多个音频帧的信息量对音频帧的声学向量进行合并，获得字符声学向量序列。

示例地，该预测子模型可以为 CIF（Continuous Integrate-and-Fire，连续整合发放）模型，在本公开实施例中默认每一预测字符对应的信息量是相同的，因此可以将音频帧对应的信息量从左到右的方式进行累加，信息量累加至预设阈值时，认为此时该累加的信息量对应的音频帧形成一个预测字符，一个预测字符对应于一个或多个音频帧。其中，该预设阈值可以根据实际应用场景和经验进行设置，示例地该预设阈值可以设置为 1，本公开对此不进行限定。

在一种可能的实施例中，可以通过如下方式根据多个音频帧的信息量对音频帧的声学向量进行合并：

按照信息量的顺序，依次获取一音频帧 i 的信息量 W_i ；

若 W_i 小于预设阈值 β ，则获取下一音频帧作为当前的音频帧，即 $i=i+1$ ，并对遍历到的音频帧的信息量进行累加，若累加和大于该预设阈值，可以认为此时出现了字符边界，即该当前遍历到的音频帧中部分属于当前的预测字符，另一部分属于下一预测字符。

示例地，若 W_1+W_2 大于 β ，则可以认为此时出现了字符边界，即第 1 个音频帧和第 2 个音频帧的部分可以对应于一个预测字符，该预测字符的边界处于第 2 个音频帧中。此时可以将该第 2 个音频帧的信息量切分为两部分，即一部分信息量属于当前的预测字符，剩余一部分的信息量属于下一预测字符。

相应地, 第 2 个音频帧的信息量 W_2 中属于当前的预测字符的信息量 W_{21} 可以表示为: $W_{21}=\beta-W_1$; 属于下一预测字符的信息量 W_{22} 可以表示为: $W_{22}=W_1-W_{21}$ 。

之后继续遍历音频帧的信息量, 并从该第 2 个音频帧的剩余一部分的信息量继续进行信息量的累加, 即将第 2 个音频帧中的信息量 W_{22} 和第 3 个音频帧中的信息量 W_3 进行累加, 直至累加至预设阈值 β , 获得下一预测字符对应的音频帧。针对后续的音频帧的信息量以此类推, 通过上述方式进行合并, 获得该多个音频帧对应的各个预测字符。

基于此, 在确定出该语音数据中预测字符和音频帧的对应关系后, 针对每一预测字符, 可以将该预测字符对应的每一音频帧的声学向量的加权和确定为该预测字符对应的声学向量。其中, 该预测字符对应的每一音频帧的声学向量的权重为该音频帧在该预测字符中对应的信息量。若该音频帧全部属于该预测字符, 则该音频帧的声学向量的权重为该音频帧的信息量, 若该音频帧部分属于该预测字符, 则该音频帧的声学向量的权重为该音频帧中该部分的信息量。

如上文所述示例, 针对第一个预测字符, 其包含第 1 个音频帧和第 2 个音频帧的部分, 则该预测字符对应的声学向量 C_1 可以表示为:

$$C_1=W_1*H_1+W_{21}*H_2;$$

又如示例, 针对第二个预测字符, 其包含第 2 个音频帧的部分和第 3 个音频帧, 则该预测字符对应的声学向量 C_2 可以表示为:

$$C_2=W_{22}*H_2+W_3*H_3。$$

转回图 1, 在步骤 12 中, 根据字符声学向量和语音解码器, 获得预测字符对应的输出文本和语言特征向量。

其中, 在语音解码器中可以对输入的每一预测字符的声学向量进行特征提取, 从而基于提取出的特征进行 softmax 处理, 获得该预测字符对应于各个真实字符的概率, 即该预测字符对应的文本概率分布。之后, 则可以基于该文本概率分布, 通过贪心搜索 (Greedy Search) 的算法或者集束搜索 (Beam Search) 的算法确定出每一预测字符对应的识别字符, 以获得该输出文本。

在该实施例, 该语言特征向量可以是该语音解码器中间某一特征层输出的特征向量, 如可以是进行 softmax 处理时对应的特征。

在步骤 13 中, 根据字符声学向量、语言特征向量和符号解码器, 确定预测字符对应的符号概率分布。

在该步骤中, 在对符号进行解码时, 其对应的输入包含预测字符的字符声学向量以及语言特征向量, 因此可以结合语音特征和语言特征对标点符号进行预测, 保证确定出的符号概率分布的准确性。

在步骤 14 中, 根据符号概率分布和训练语音数据中对应的目标符号特征, 确定目标损失。

其中，可以对训练语音数据进行预先标注，从而可以获得该训练语音数据对应的目标文本，以及该训练语音数据对应的符号特征，其中，训练语音数据对应的符号的符号特征可以用于表征训练语音数据对应的带标点符号的文本中各个字符后相邻的标点符号。

在步骤 15 中，在满足更新条件的情况下，根据目标损失对语音识别模型的模型参数进行更新。

作为示例，该更新条件可以为目标损失大于预设的损失阈值，此时表示语音识别模型的识别准确性不足。作为另一示例，该更新条件可以是迭代次数小于预设的次数阈值，此时认为语音识别模型迭代次数较少，其识别准确性不足。相应地，在满足更新条件的情况下，可以根据该目标损失对该语音识别模型的模型参数进行更新。其中，基于确定出的损失对模型参数进行更新的方式可以采用本领域中常用的更新方式，如梯度下降法，在此不再赘述。

在不满足该更新条件的情况下，则可以认为该语音识别模型的识别精确性达到训练要求，此时可以停止训练过程，获得训练完成的语音识别模型。

在上述技术方案中，根据编码器和预测子模型，获得训练语音数据对应的字符声学向量，从而可以根据所述字符声学向量和所述语音解码器，获得所述预测字符对应的输出文本和语言特征向量，并根据所述字符声学向量、所述语言特征向量和所述符号解码器，确定所述预测字符对应的符号概率分布，之后根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失，以在满足更新条件的情况下，根据所述目标损失对所述语音识别模型的模型参数进行更新。由此，通过上述技术方案，针对语音数据中预测字符的声学向量进行处理，并且通过结合符号解码器以对每一预测字符进行语音识别及标点预测，一方面可以提高语音识别模型的处理精度，另一方面可以有效降低语音识别过程中确定标点符号的时延，使得训练所得的语音识别模型可以应用于流式数据处理过程，为流式识别产品的用户体验提供支撑。并且，在符号解码器进行解码的过程中可以复用预测子模型中的字符声学向量以及语音解码器中的语言特征向量，从而可以综合利用声学信息和语言信息进行标点符号的预测，进一步训练所得的语音识别模型的准确性，提升用户使用体验。

在一种可能的实施例中，可以通过以下方式确定所述训练语音数据对应的文本中每一字符对应的符号特征：

获取所述训练语音数据对应的包含标点符号的文本数据。

针对所述文本数据中的每一字符，若该字符后相邻存在标点符号，则将该标点符号作为该字符对应的标注符号；若该字符后相邻未存在标点符号，则将预设符号作为该字符对应的标注符号。

示例地，所述文本数据可以为：你好，初次见面，请多多关照。则针对所述文本数据中的字符，在字符“好”、“面”、“照”后相邻存在标点符号，

则针对字符“好”，可以将其后相邻的标点符号“，”作为其对应的标注符号，针对字符“面”，可以将其后相邻的标点符号“，”作为其对应的标注符号，针对字符“照”，可以将其后相邻的标点符号“。”作为其对应的标注符号。针对文本数据中的其他字符，则可以将预设符号，如“*”，作为其对应的标注符号，则针对该文本数据，可以获得标注符号序列“**，****，*****。”。

针对所述文本数据中的每一字符，根据该字符对应的标注符号与符号空间，生成该字符对应的符号特征。

其中，符号空间中可以包含常用的标点符号以及该预设符号，则在确定出字符对应的标注符号后，则可以将该标注符号以 one-hot 独热编码的形式确定出标注符号映射到该符号空间中的符号特征。

由此，通过上述技术方案，可以通过对带有标点符号的文本数据中的标点进行单独编码，从而可以获得支持标点预测与语音识别进行联合训练的训练语音数据，与语音识别模型的训练过程相匹配，为进行语音识别模型的联合训练提供数据支持。

在一种可能的实施例中，所述目标符号特征为独热编码特征，在步骤 14 中，根据符号概率分布和训练语音数据中预测字符对应的目标符号特征，确定目标损失的示例性实现方式如下，该步骤可以包括：

根据所述符号概率分布和所述预测字符对应的目标符号特征，确定交叉熵损失。

其中，通过符号解码器进行解码可以获得当前预测字符后相邻的符号为各个标点符号的概率，所述预测字符对应的目标符号特征为独热编码特征，即该预测字符在训练语音数据中的对应的标点符号的维度的取值为 1，其他维度的取值为 0，相当于标注符号对应的特殊概率分布。

因此，在该实施例中，可以基于预测字符对应的符号概率分布和标注符号对应的特殊概率分布，计算交叉熵损失。其中，交叉熵损失用于度量两个概率分布间的差异性，交叉熵损失越小，表征预测的概率分布和真实概率分布之间则更相似，即训练所得的语音识别模型的预测效果越好。

之后，将确定出的所述交叉熵损失作为所述目标损失。

由此，通过上述技术方案，可以直接根据符号解码器输出的符号概率分布确定对应的目标损失，从而可以直接对符号解码器对应的交叉熵损失进行优化，有效提高符号解码器的准确性，从而提高训练所得的语音识别模型对标点符号预测的准确性，在一定程度上提高语音识别模型输出文本结果的可读性，提升用户使用体验。

在一种可能的实施例中，在将基于符号解码器的输出确定出的交叉熵损失作为所述目标损失的情况下，在步骤 15 中，根据目标损失对语音识别模型的模型参数进行更新的示例性实现方式如下，该步骤可以包括：

根据所述目标损失对所述语音识别模型中、所述符号解码器的模型参数

进行更新。

如上文所示，在该实施例中，根据符号解码器输出的符号概率分布确定对应的目标损失，即在进行模型优化时，可以仅对语音识别模型中的符号解码器的模型参数进行优化，而语音识别模型中的其他部分的参数保持不变。

由此，通过上述方案，可以在现有训练完成的模型中增加符号解码器，并通过训练语音数据对该模型进行训练以获得该语音识别模型，以在现有训练完成的模型中增加符号识别能力，同时保证文本识别能力不变，从而提高该语音识别模型的训练方法的适用范围和训练效率。

在一种可能的实施例中，所述目标符号特征为独热编码特征，在步骤 14 中，根据符号概率分布和训练语音数据中预测字符对应的目标符号特征，确定目标损失的示例性实现方式如下，该步骤可以包括：

根据所述符号概率分布和所述预测字符对应的目标符号特征，确定第一交叉熵损失。其中，第一交叉熵损失的确定方式已在上文进行详述，在此不再赘述。

根据所述预测字符对应的目标文本与所述输出文本，确定所述语音解码器对应的第二交叉熵损失。

其中，语音识别模型中包含用于将语音转换成文本的语音解码器，从而可以通过优化该第二交叉熵损失，以提高该语音识别模型中语音转换成文本的识别准确性。

根据所述第一交叉熵损失和所述第二交叉熵损失确定所述目标损失。

其中，可以将第一交叉熵损失和所述第二交叉熵损失进行加权求和的方式确定该目标损失，其中第一交叉熵损失和第二交叉熵损失分别对应的权重可以根据实际应用场景进行设置，本公开对此不进行限定。

由此，通过上述技术方案，在确定语音识别模型的目标损失时，可以基于语音解码器和符号解码器分别对应的交叉熵损失确定目标损失，通过构建多任务学习的方式对该语音解码器和符号解码器进行联合训练，从而可以在一定程度上提高训练所得的语音识别模型识别出的文本和符号的准确性，提升用户使用体验。

本公开还提供一种语音识别方法，如图 2 所示，所述方法可以包括：

在步骤 21 中，接收待识别的语音数据。

在步骤 22 中，根据语音数据和语音识别模型，确定语音数据对应的待添加的符号和识别文本，其中，所述语音识别模型是基于上文任一所述的语音识别模型的训练方法进行训练获得的。

如上文所述，基于该语音识别模型可以对输入的语音数据进行流式处理，即针对该语音数据进行字符级别的符号概率分布和识别文本的确定。其中，语音数据流式输入该语音识别模型，则经过编码器进行编码，在预测子模型确定出一个预测字符的字符声学向量时，则可以分别通过语音解码器和符号

解码器对该字符声学向量进行解码，进而获得待添加的符合和识别文本。在该实施例中，针对语音数据中确定出的每一字符，可以将其对应的符号概率分布中对应概率最大的符号确定为该待添加的符号。

在步骤 23 中，在待添加的符号为目标符号的情况下，根据待添加的符号和识别文本生成待识别的语音数据对应的文本结果，该文本结果为包含标点符号信息的文本。

在语音识别模型的符号解码器中，其解码出的符号可能是预设符号“*”，其表示在该预测字符后无相邻的标点符号，因此，在该实施例中，目标符号可以为符号空间中除该预设符号之外的任一符号。在确定出待添加的符号为目标符号的情况下，表示该预测字符对应的文本后应存在相邻的标点符号，则可以将该待添加的符号添加至识别文本中该预测字符对应的文本之后，从而获得待识别的语音数据对应的文本结果。

由此，通过上述技术方案，语音识别模型可以应用于流式数据处理过程，基于该语音识别模型可以在字符级别对输入的语音数据进行处理，从而可以实时确定出每一字符后是否需要添加标点符号，提升标点符号预测的准确性和实时性，同时也可以提高获得的文本结果的准确性，可以应用于在线语音识别场景中，有效降低文本结果显示的时延，提升用户使用体验。

图 3 所示，为根据本公开的一种实施方式提供的语音识别模型的训练装置的框图。所述语音识别模型包括编码器、预测子模型、语音解码器以及符号解码器，所述装置 30 包括：

第一处理模块 31，用于根据所述编码器和所述预测子模型，获得训练语音数据对应的字符声学向量，其中，每一所述字符声学向量对应于所述训练语音数据的一个预测字符；

第二处理模块 32，用于根据所述字符声学向量和所述语音解码器，获得所述预测字符对应的输出文本和语言特征向量；

第一确定模块 33，用于根据所述字符声学向量、所述语言特征向量和所述符号解码器，确定所述预测字符对应的符号概率分布；

第二确定模块 34，用于根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失；

更新模块 35，用于在满足更新条件的情况下，根据所述目标损失对所述语音识别模型的模型参数进行更新。

可选地，所述目标符号特征为独热编码特征，所述第二确定模块包括：

第一确定子模块，用于根据所述符号概率分布和所述预测字符对应的目标符号特征，确定交叉熵损失，并将确定出的所述交叉熵损失作为所述目标损失。

可选地，所述更新模块包括：

更新子模块，用于根据所述目标损失对所述语音识别模型中、所述符号

解码器的模型参数进行更新。

可选地，所述目标符号特征为独热编码特征，所述第二确定模块包括：第二确定子模块，用于根据所述符号概率分布和所述预测字符对应的目标符号特征，确定第一交叉熵损失；

第二确定子模块，用于根据所述预测字符对应的目标文本与所述输出文本，确定所述语音解码器对应的第二交叉熵损失；

第三确定子模块，用于根据所述第一交叉熵损失和所述第二交叉熵损失确定所述目标损失。

可选地，通过以下方式确定所述训练语音数据对应的文本中每一字符对应的符号特征：

获取所述训练语音数据对应的包含标点符号的文本数据；

针对所述文本数据中的每一字符，若该字符后相邻存在标点符号，则将该标点符号作为该字符对应的标注符号；若该字符后相邻未存在标点符号，则将预设符号作为该字符对应的标注符号；

针对所述文本数据中的每一字符，根据该字符对应的标注符号与符号空间，生成该字符对应的符号特征。

图 4 所示，为根据本公开的一种实施方式提供的语音识别装置的框图。所述装置 40 包括：

接收模块 41，用于接收待识别的语音数据；

第三确定模块 42，用于根据所述语音数据和语音识别模型，确定所述语音数据对应的待添加的符号和识别文本，其中，所述语音识别模型是基于上文任一所述的语音识别模型的训练方法进行训练获得的；

生成模块 43，用于在所述待添加的符号为目标符号的情况下，根据所述待添加的符号和所述识别文本生成所述待识别的语音数据对应的文本结果。

下面参考图 5，其示出了适于用来实现本公开实施例的电子设备 600 的结构示意图。本公开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 5 示出的电子设备仅仅是一个示例，不应对本公开实施例的功能和使用范围带来任何限制。

如图 5 所示，电子设备 600 可以包括处理装置（例如中央处理器、图形处理器等）601，其可以根据存储在只读存储器（ROM）602 中的程序或者从存储装置 608 加载到随机访问存储器（RAM）603 中的程序而执行各种适当的动作和处理。在 RAM 603 中，还存储有电子设备 600 操作所需的各种程序和数据。处理装置 601、ROM 602 以及 RAM 603 通过总线 604 彼此相连。输入/输出（I/O）接口 605 也连接至总线 604。

通常，以下装置可以连接至 I/O 接口 605：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 606；包括例如液晶显示器(LCD)、扬声器、振动器等的输出装置 607；包括例如磁带、硬盘等的存储装置 608；以及通信装置 609。通信装置 609 可以允许电子设备 600 与其他设备进行无线或有线通信以交换数据。虽然图 5 示出了具有各种装置的电子设备 600，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。

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，其包括承载在非暂态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 609 从网络上被下载和安装，或者从存储装置 608 被安装，或者从 ROM 602 被安装。在该计算机程序被处理装置 601 执行时，执行本公开实施例的方法中限定的上述功能。

需要说明的是，本公开上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。

在一些实施方式中，客户端、服务器可以利用诸如 HTTP(HyperText Transfer Protocol, 超文本传输协议)之类的任何当前已知或未来研发的网络协议进行通信，并且可以与任意形式或介质的数字数据通信(例如，通信网络)互连。通信网络的示例包括局域网(“LAN”), 广域网(“WAN”), 网际网(例如，互联网)以及端对端网络(例如，ad hoc 端对端网络)，以及任何当前已知或未来研发的网络。

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，使得该电子设备执行语音识别模型的训练方法，所述语音识别模型包括编码器、预测子模型、语音解码器以及符号解码器，所述方法包括：根据所述编码器和所述预测子模型，获得训练语音数据对应的字符声学向量，其中，每一所述字符声学向量对应于所述训练语音数据的一个预测字符；根据所述字符声学向量和所述语音解码器，获得所述预测字符对应的输出文本和语言特征向量；根据所述字符声学向量、所述语言特征向量和所述符号解码器，确定所述预测字符对应的符号概率分布；根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失；在满足更新条件的情况下，根据所述目标损失对所述语音识别模型的模型参数进行更新。

或者，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，使得该电子设备执行语音识别方法，所述方法包括：接收待识别的语音数据；根据所述语音数据和语音识别模型，确定所述语音数据对应的待添加的符号和识别文本，其中，所述语音识别模型是基于上文任一所述的语音识别模型的训练方法进行训练获得的；在所述待添加的符号为目标符号的情况下，根据所述待添加的符号和所述识别文本生成所述待识别的语音数据对应的文本结果。

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操作的计算机程序代码，上述程序设计语言包括但不限于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。

描述于本公开实施例中涉及到的模块可以通过软件的方式实现，也可以通过硬件的方式来实现。其中，模块的名称在某种情况下并不构成对该模块本身的限定，例如，第一处理模块还可以被描述为“根据所述编码器和所述预测子模型，获得训练语音数据对应的字符声学向量的模块”。

本文中以上描述的功能可以至少部分地由一个或多个硬件逻辑部件来执行。例如，非限制性地，可以使用的示范类型的硬件逻辑部件包括：现场可编程门阵列（FPGA）、专用集成电路（ASIC）、专用标准产品（ASSP）、片上系统（SOC）、复杂可编程逻辑设备（CPLD）等等。

在本公开的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦除可编程只读存储器（EPROM 或快闪存储器）、光纤、便捷式紧凑盘只读存储器（CD-ROM）、光学储存设备、磁储存设备、或上述内容的任何合适组合。

根据本公开的一个或多个实施例，示例 1 提供了一种语音识别模型的训练方法，所述语音识别模型包括编码器、预测子模型、语音解码器以及符号解码器，所述方法包括：

根据所述编码器和所述预测子模型，获得训练语音数据对应的字符声学向量，其中，每一所述字符声学向量对应于所述训练语音数据的一个预测字符；

根据所述字符声学向量和所述语音解码器，获得所述预测字符对应的输出文本和语言特征向量；

根据所述字符声学向量、所述语言特征向量和所述符号解码器，确定所述预测字符对应的符号概率分布；

根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失；

在满足更新条件的情况下，根据所述目标损失对所述语音识别模型的模型参数进行更新。

根据本公开的一个或多个实施例，示例 2 提供了示例 1 的方法，所述目标符号特征为独热编码特征，所述根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失，包括：

根据所述符号概率分布和所述预测字符对应的目标符号特征，确定交叉

熵损失；

将确定出的所述交叉熵损失作为所述目标损失。

根据本公开的一个或多个实施例，示例 3 提供了示例 2 的方法，所述根据所述目标损失对所述语音识别模型的模型参数进行更新，包括：

根据所述目标损失对所述语音识别模型中、所述符号解码器的模型参数进行更新。

根据本公开的一个或多个实施例，示例 4 提供了示例 1 的方法，所述目标符号特征为独热编码特征，所述根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失，包括：

根据所述符号概率分布和所述预测字符对应的目标符号特征，确定第一交叉熵损失；

根据所述预测字符对应的目标文本与所述输出文本，确定所述语音解码器对应的第二交叉熵损失；

根据所述第一交叉熵损失和所述第二交叉熵损失确定所述目标损失。

根据本公开的一个或多个实施例，示例 5 提供了示例 1-4 中任一示例的方法，通过以下方式确定所述训练语音数据对应的文本中每一字符对应的符号特征：

获取所述训练语音数据对应的包含标点符号的文本数据；

针对所述文本数据中的每一字符，若该字符后相邻存在标点符号，则将该标点符号作为该字符对应的标注符号；若该字符后相邻未存在标点符号，则将预设符号作为该字符对应的标注符号；

针对所述文本数据中的每一字符，根据该字符对应的标注符号与符号空间，生成该字符对应的符号特征。

根据本公开的一个或多个实施例，示例 6 提供了一种语音识别方法，所述方法包括：

接收待识别的语音数据；

根据所述语音数据和语音识别模型，确定所述语音数据对应的待添加的符号和识别文本，其中，所述语音识别模型是基于示例 1-5 中任一示例所述的语音识别模型的训练方法进行训练获得的；

在所述待添加的符号为目标符号的情况下，根据所述待添加的符号和所述识别文本生成所述待识别的语音数据对应的文本结果。

根据本公开的一个或多个实施例，示例 7 提供了一种语音识别模型的训练装置，所述语音识别模型包括编码器、预测子模型、语音解码器以及符号解码器，所述装置包括：

第一处理模块，用于根据所述编码器和所述预测子模型，获得训练语音数据对应的字符声学向量，其中，每一所述字符声学向量对应于所述训练语音数据的一个预测字符；

第二处理模块，用于根据所述字符声学向量和所述语音解码器，获得所述预测字符对应的输出文本和语言特征向量；

第一确定模块，用于根据所述字符声学向量、所述语言特征向量和所述符号解码器，确定所述预测字符对应的符号概率分布；

第二确定模块，用于根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失；

更新模块，用于在满足更新条件的情况下，根据所述目标损失对所述语音识别模型的模型参数进行更新。

根据本公开的一个或多个实施例，示例 8 提供了一种语音识别装置，所述装置包括：

接收模块，用于接收待识别的语音数据；

第三确定模块，用于根据所述语音数据和语音识别模型，确定所述语音数据对应的待添加的符号和识别文本，其中，所述语音识别模型是基于示例 1-5 中任一示例所述的语音识别模型的训练方法进行训练获得的；

生成模块，用于在所述待添加的符号为目标符号的情况下，根据所述待添加的符号和所述识别文本生成所述待识别的语音数据对应的文本结果。

根据本公开的一个或多个实施例，示例 9 提供了一种计算机可读介质，其上存储有计算机程序，该程序被处理装置执行时实现示例 1-6 中任一示例所述方法的步骤。

根据本公开的一个或多个实施例，示例 10 提供了一种电子设备，包括：

存储装置，其上存储有计算机程序；

处理装置，用于执行所述存储装置中的所述计算机程序，以实现示例 1-6 中任一示例所述方法的步骤。

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，本公开中所涉及的公开范围，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公开构思的情况下，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。

此外，虽然采用特定次序描绘了各操作，但是这不应理解为要求这些操作以所示出的特定次序或以顺序次序执行来执行。在一定环境下，多任务和并行处理可能是有利的。同样地，虽然在上面论述中包含了若干具体实现细节，但是这些不应被解释为对本公开的范围的限制。在单独的实施例的上下文中描述的某些特征还可以组合地实现在单个实施例中。相反地，在单个实施例的上下文中描述的各种特征也可以单独地或以任何合适的子组合的方式实现在多个实施例中。

尽管已经采用特定于结构特征和/或方法逻辑动作的语言描述了本主题，

但是应当理解所附权利要求书中所限定的主题未必局限于上面描述的特定特征或动作。相反，上面所描述的特定特征和动作仅仅是实现权利要求书的示例形式。关于上述实施例中的装置，其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述，此处将不做详细阐述说明。

权 利 要 求 书

1、一种语音识别模型的训练方法，所述语音识别模型包括编码器、预测子模型、语音解码器以及符号解码器，所述方法包括：

根据所述编码器和所述预测子模型，获得训练语音数据对应的字符声学向量，其中，每一所述字符声学向量对应于所述训练语音数据的一个预测字符；

根据所述字符声学向量和所述语音解码器，获得所述预测字符对应的输出文本和语言特征向量；

根据所述字符声学向量、所述语言特征向量和所述符号解码器，确定所述预测字符对应的符号概率分布；

根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失；

在满足更新条件的情况下，根据所述目标损失对所述语音识别模型的模型参数进行更新。

2、根据权利要求 1 所述的方法，其中，所述目标符号特征为独热编码特征，所述根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失，包括：

根据所述符号概率分布和所述预测字符对应的目标符号特征，确定交叉熵损失；

将确定出的所述交叉熵损失作为所述目标损失。

3、根据权利要求 2 所述的方法，其中，所述根据所述目标损失对所述语音识别模型的模型参数进行更新，包括：

根据所述目标损失对所述语音识别模型中、所述符号解码器的模型参数进行更新。

4、根据权利要求 1 所述的方法，其中，所述目标符号特征为独热编码特征，所述根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失，包括：

根据所述符号概率分布和所述预测字符对应的目标符号特征，确定第一交叉熵损失；

根据所述预测字符对应的目标文本与所述输出文本，确定所述语音解码器对应的第二交叉熵损失；

根据所述第一交叉熵损失和所述第二交叉熵损失确定所述目标损失。

5、根据权利要求 1-4 中任一项所述的方法，其中，通过以下方式确定所述训练语音数据对应的文本中每一字符对应的符号特征：

获取所述训练语音数据对应的包含标点符号的文本数据；

针对所述文本数据中的每一字符，若该字符后相邻存在标点符号，则将该标点符号作为该字符对应的标注符号；若该字符后相邻未存在标点符号，则将预设符号作为该字符对应的标注符号；

5 针对所述文本数据中的每一字符，根据该字符对应的标注符号与符号空间，生成该字符对应的符号特征。

6、一种语音识别方法，所述方法包括：

接收待识别的语音数据；

10 根据所述语音数据和语音识别模型，确定所述语音数据对应的待添加的符号和识别文本，其中，所述语音识别模型是基于权利要求 1-5 中任一项所述的语音识别模型的训练方法进行训练获得的；

在所述待添加的符号为目标符号的情况下，根据所述待添加的符号和所述识别文本生成所述待识别的语音数据对应的文本结果。

15

7、一种语音识别模型的训练装置，所述语音识别模型包括编码器、预测子模型、语音解码器以及符号解码器，所述装置包括：

第一处理模块，用于根据所述编码器和所述预测子模型，获得训练语音数据对应的字符声学向量，其中，每一所述字符声学向量对应于所述训练语音数据的一个预测字符；

20

第二处理模块，用于根据所述字符声学向量和所述语音解码器，获得所述预测字符对应的输出文本和语言特征向量；

第一确定模块，用于根据所述字符声学向量、所述语言特征向量和所述符号解码器，确定所述预测字符对应的符号概率分布；

25

第二确定模块，用于根据所述符号概率分布和所述训练语音数据中所述预测字符对应的目标符号特征，确定目标损失；

更新模块，用于在满足更新条件的情况下，根据所述目标损失对所述语音识别模型的模型参数进行更新。

30

8、一种语音识别装置，所述装置包括：

接收模块，用于接收待识别的语音数据；

第三确定模块，用于根据所述语音数据和语音识别模型，确定所述语音数据对应的待添加的符号和识别文本，其中，所述语音识别模型是基于权利要求 1-5 中任一项所述的语音识别模型的训练方法进行训练获得的；

35

生成模块，用于在所述待添加的符号为目标符号的情况下，根据所述待添加的符号和所述识别文本生成所述待识别的语音数据对应的文本结果。

9、一种计算机可读介质，其上存储有计算机程序，该程序被处理装置

执行时实现权利要求 1-6 中任一项所述方法的步骤。

10、一种电子设备，包括：

存储装置，其上存储有计算机程序；

5 处理装置，用于执行所述存储装置中的所述计算机程序，以实现权利要求 1-6 中任一项所述方法的步骤。

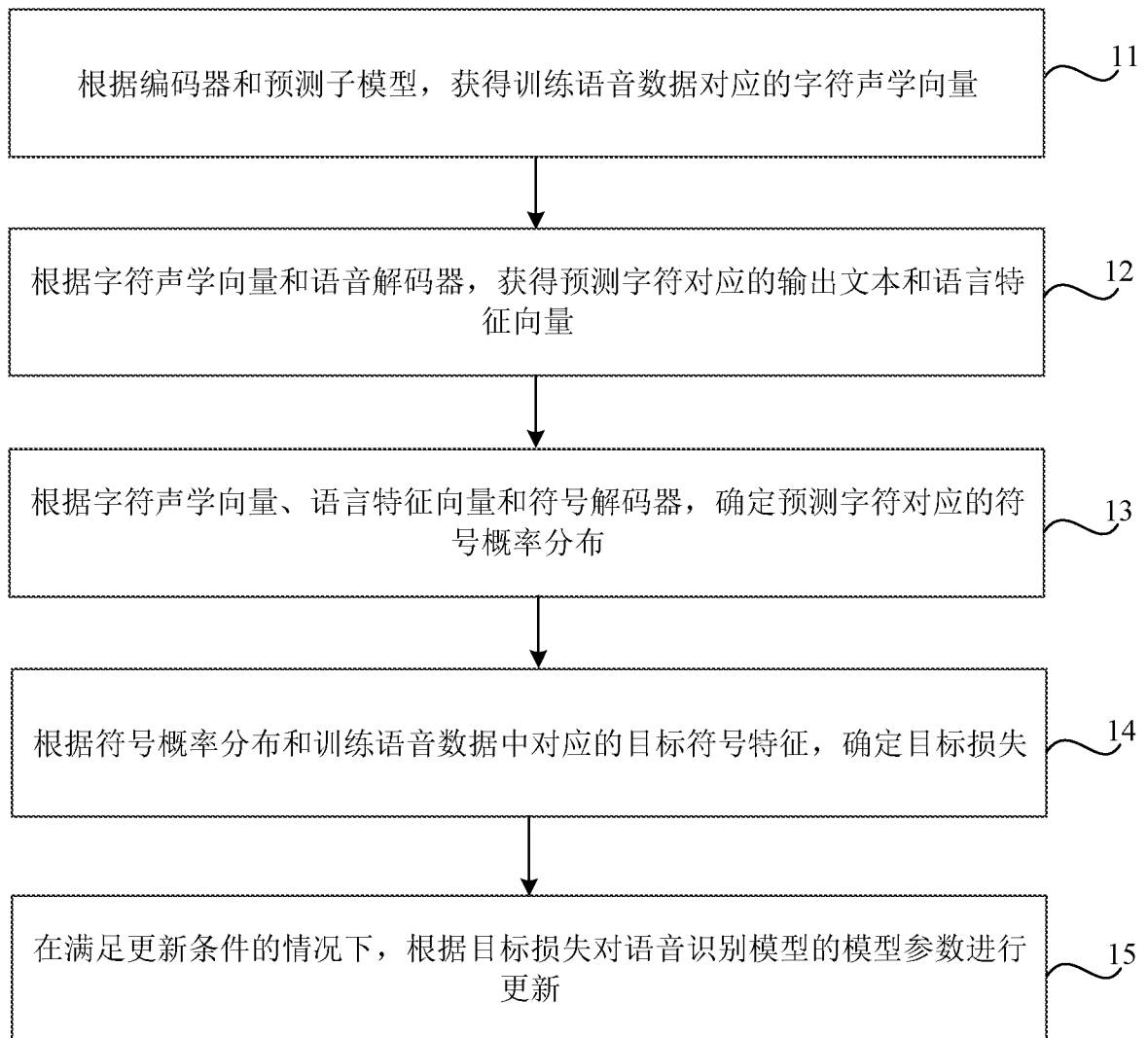


图 1

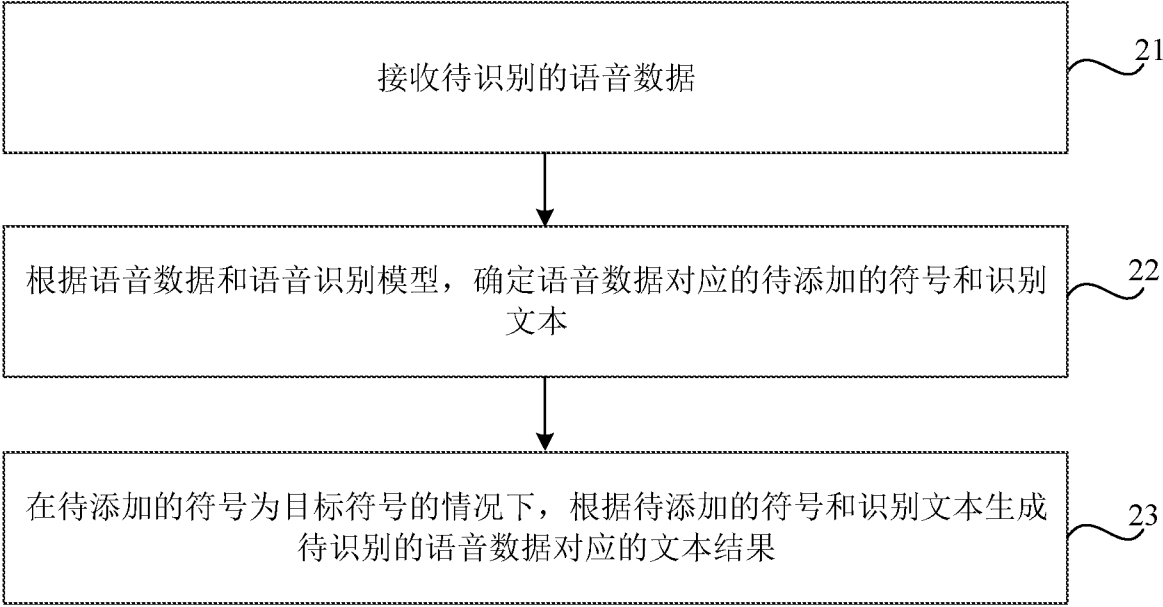


图 2

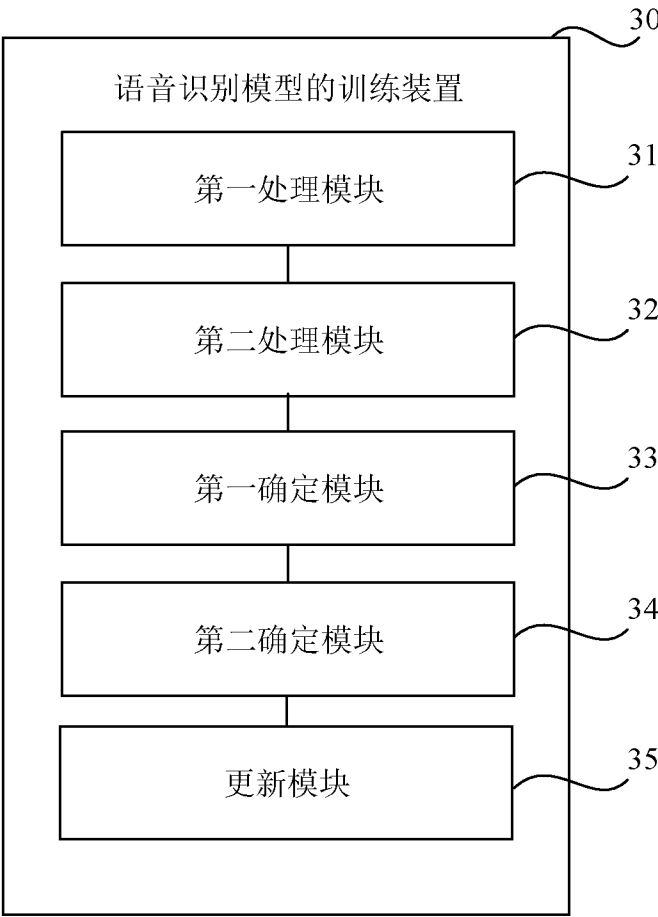


图 3

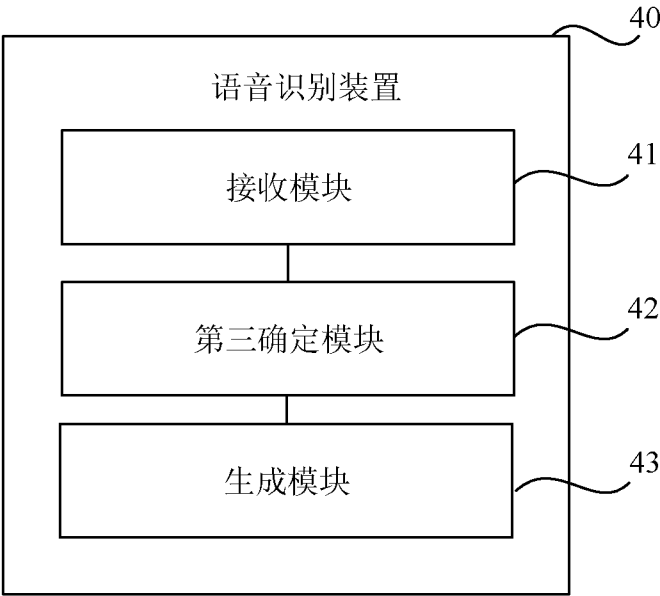


图 4

600

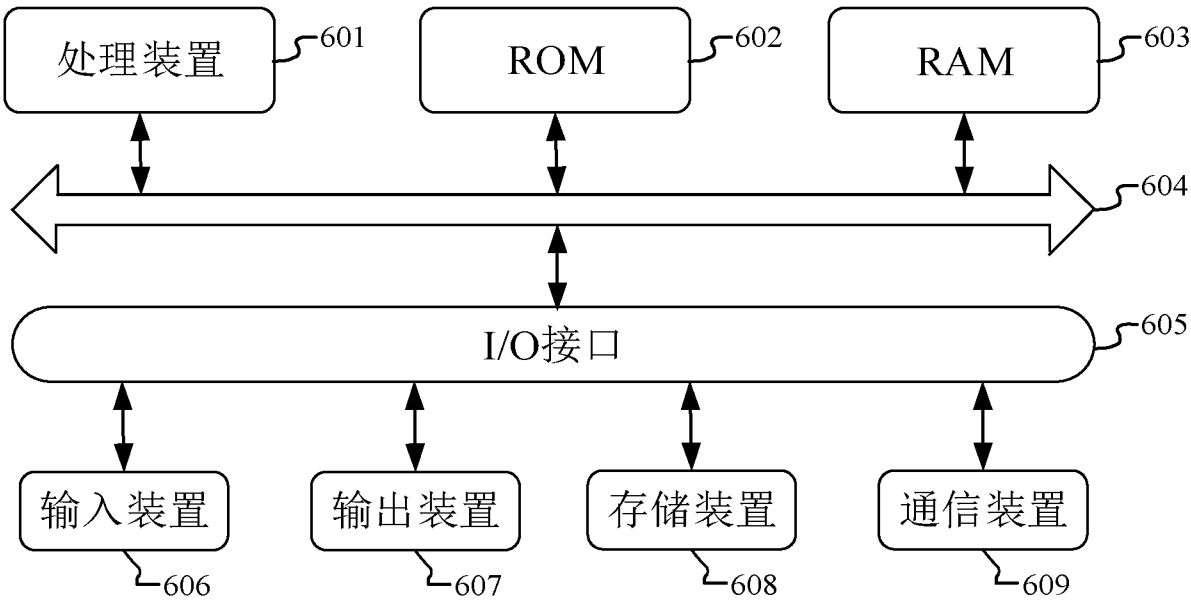


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/091481

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/06(2013.01)i; G10L 15/26(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)

G10L

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)

CNTXT, CNABS, CNKI, ENTXT: 语音识别, 标点, 符号, 字符, 向量, 编码器, 解码器, 损失, 更新, 交叉熵, 概率分布, speech recognition, target loss, symbol, probability, distribution, characteristics, voice recognition model, parameter

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113362811 A (BEIJING YOUZHUJU NETWORK TECHNOLOGY CO., LTD.) 07 September 2021 (2021-09-07) description, paragraphs [0055]-[0188], and figures 1-5	1-10
A	CN 112634876 A (BEIJING YOUZHUJU NETWORK TECHNOLOGY CO., LTD.) 09 April 2021 (2021-04-09) description, paragraphs [0047]-[0056], and figure 3	1-10
A	CN 112927679 A (CHONGDONG INNOVATION PLATFORM (SHENZHEN) CO., LTD.) 08 June 2021 (2021-06-08) entire document	1-10
A	CN 111199727 A (XIAMEN KUAISHANGTONG TECHNOLOGY CO., LTD.) 26 May 2020 (2020-05-26) entire document	1-10
A	CN 111916067 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 10 November 2020 (2020-11-10) entire document	1-10
A	CN 113012701 A (LENOVO (BEIJING) LTD.) 22 June 2021 (2021-06-22) entire document	1-10



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

“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

21 June 2022

Date of mailing of the international search report

29 June 2022

Name and mailing address of the ISA/CN

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

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/091481**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 112712804 A (HARBIN INSTITUTE OF TECHNOLOGY, WEIHAI) 27 April 2021 (2021-04-27) entire document	1-10
A	US 2018174589 A1 (SAMSUNG ELECTRONICS CO., LTD.) 21 June 2018 (2018-06-21) entire document	1-10
A	WO 2020242580 A1 (GOOGLE L.L.C.) 03 December 2020 (2020-12-03) entire document	1-10

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/091481

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113362811	A	07 September 2021	None			
CN	112634876	A	09 April 2021	None			
CN	112927679	A	08 June 2021	None			
CN	111199727	A	26 May 2020	None			
CN	111916067	A	10 November 2020	None			
CN	113012701	A	22 June 2021	None			
CN	112712804	A	27 April 2021	None			
US	2018174589	A1	21 June 2018	KR	20180071029	A	27 June 2018
WO	2020242580	A1	03 December 2020	US	2020380215	A1	03 December 2020
				KR	20220007160	A	18 January 2022
				CN	113924619	A	11 January 2022
				EP	3966816	A1	16 March 2022

国际检索报告

国际申请号

PCT/CN2022/091481

A. 主题的分类 G10L 15/06(2013.01)i; G10L 15/26(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) G10L 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)) CNTXT, CNABS, CNKI, ENTXT: 语音识别, 标点, 符号, 字符, 向量, 编码器, 解码器, 损失, 更新, 交叉熵, 概率分布, speech recognition, target loss, symbol, probability, distribution, characteristics, voice recognition model, parameter																										
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>PX</td> <td>CN 113362811 A (北京有竹居网络技术有限公司) 2021年9月7日 (2021 - 09 - 07) 说明书第[0055]-[0188]段, 图1-5</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 112634876 A (北京有竹居网络技术有限公司) 2021年4月9日 (2021 - 04 - 09) 说明书第[0047]-[0056]段, 图3</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 112927679 A (虫洞创新平台深圳有限公司) 2021年6月8日 (2021 - 06 - 08) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 111199727 A (厦门快商通科技股份有限公司) 2020年5月26日 (2020 - 05 - 26) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 111916067 A (腾讯科技深圳有限公司) 2020年11月10日 (2020 - 11 - 10) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 113012701 A (联想北京有限公司) 2021年6月22日 (2021 - 06 - 22) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 112712804 A (哈尔滨工业大学威海) 2021年4月27日 (2021 - 04 - 27) 全文</td> <td>1-10</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	PX	CN 113362811 A (北京有竹居网络技术有限公司) 2021年9月7日 (2021 - 09 - 07) 说明书第[0055]-[0188]段, 图1-5	1-10	A	CN 112634876 A (北京有竹居网络技术有限公司) 2021年4月9日 (2021 - 04 - 09) 说明书第[0047]-[0056]段, 图3	1-10	A	CN 112927679 A (虫洞创新平台深圳有限公司) 2021年6月8日 (2021 - 06 - 08) 全文	1-10	A	CN 111199727 A (厦门快商通科技股份有限公司) 2020年5月26日 (2020 - 05 - 26) 全文	1-10	A	CN 111916067 A (腾讯科技深圳有限公司) 2020年11月10日 (2020 - 11 - 10) 全文	1-10	A	CN 113012701 A (联想北京有限公司) 2021年6月22日 (2021 - 06 - 22) 全文	1-10	A	CN 112712804 A (哈尔滨工业大学威海) 2021年4月27日 (2021 - 04 - 27) 全文	1-10
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
PX	CN 113362811 A (北京有竹居网络技术有限公司) 2021年9月7日 (2021 - 09 - 07) 说明书第[0055]-[0188]段, 图1-5	1-10																								
A	CN 112634876 A (北京有竹居网络技术有限公司) 2021年4月9日 (2021 - 04 - 09) 说明书第[0047]-[0056]段, 图3	1-10																								
A	CN 112927679 A (虫洞创新平台深圳有限公司) 2021年6月8日 (2021 - 06 - 08) 全文	1-10																								
A	CN 111199727 A (厦门快商通科技股份有限公司) 2020年5月26日 (2020 - 05 - 26) 全文	1-10																								
A	CN 111916067 A (腾讯科技深圳有限公司) 2020年11月10日 (2020 - 11 - 10) 全文	1-10																								
A	CN 113012701 A (联想北京有限公司) 2021年6月22日 (2021 - 06 - 22) 全文	1-10																								
A	CN 112712804 A (哈尔滨工业大学威海) 2021年4月27日 (2021 - 04 - 27) 全文	1-10																								
☒ 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> * 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> “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 2022年6月21日		国际检索报告邮寄日期 2022年6月29日																								
ISA/CN的名称和邮寄地址 中国国家知识产权局(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 张静 电话号码 (86-10)62085727																								

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
A	US 2018174589 A1 (SAMSUNG ELECTRONICS CO., LTD.) 2018年6月21日 (2018 - 06 - 21) 全文	1-10
A	W0 2020242580 A1 (GOOGLE LLC) 2020年12月3日 (2020 - 12 - 03) 全文	1-10

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/091481

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	113362811	A	2021年9月7日	无	
CN	112634876	A	2021年4月9日	无	
CN	112927679	A	2021年6月8日	无	
CN	111199727	A	2020年5月26日	无	
CN	111916067	A	2020年11月10日	无	
CN	113012701	A	2021年6月22日	无	
CN	112712804	A	2021年4月27日	无	
US	2018174589	A1	2018年6月21日	KR	20180071029 A 2018年6月27日
WO	2020242580	A1	2020年12月3日	US	2020380215 A1 2020年12月3日
				KR	20220007160 A 2022年1月18日
				CN	113924619 A 2022年1月11日
				EP	3966816 A1 2022年3月16日