

(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
2020 年 1 月 9 日 (09.01.2020)



(10) 国际公布号
WO 2020/006898 A1

- (51) 国际专利分类号:
G10L 25/51 (2013.01) *G10L 25/30* (2013.01)
G10L 25/18 (2013.01) *G10L 25/45* (2013.01)
G10L 25/24 (2013.01)
- (21) 国际申请号: PCT/CN2018/108239
- (22) 国际申请日: 2018 年 9 月 28 日 (28.09.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201810729847.1 2018 年 7 月 5 日 (05.07.2018) CN
- (71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福

安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

- (72) 发明人: 梅亚琦(MEI, Yaqi); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 刘鼻智(LIU, Aozhi); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 王义文(WANG, Yiwēn); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 王健宗(WANG, Jianzong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 肖京(XIAO, Jing); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(54) Title: METHOD AND DEVICE FOR RECOGNIZING AUDIO DATA OF INSTRUMENT, ELECTRONIC APPARATUS, AND STORAGE MEDIUM

(54) 发明名称: 音频数据的乐器识别方法及装置、电子设备、存储介质

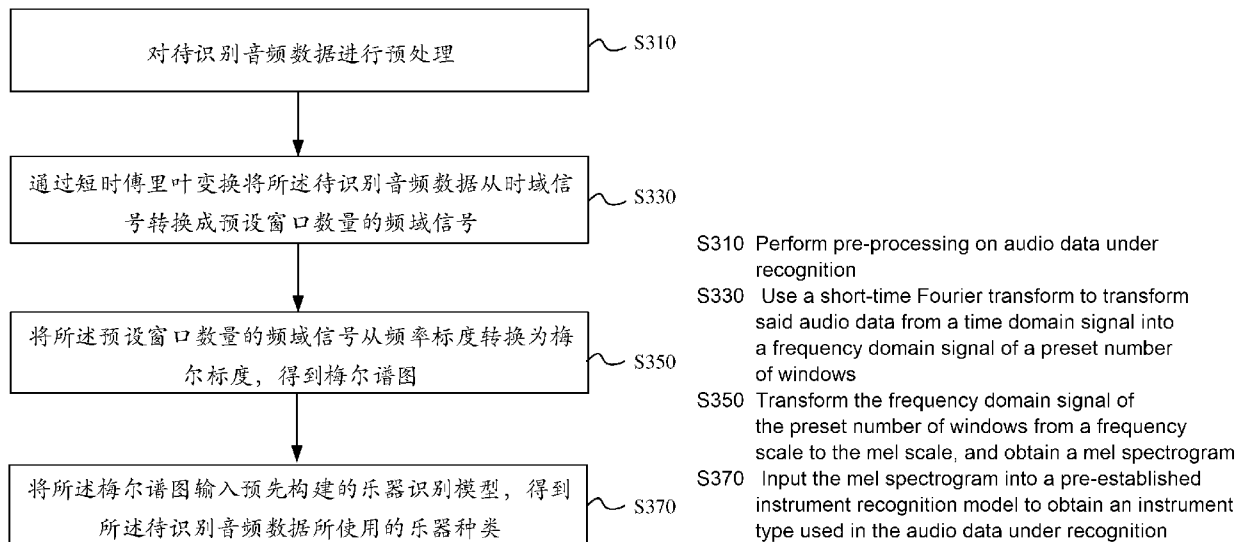


图 3

(57) Abstract: A method and device for recognizing audio data of an instrument, an electronic apparatus, and a computer readable storage medium. The method comprises: performing pre-processing on audio data under recognition (S310); using a short-time Fourier transform to transform said audio data from a time domain signal into a frequency domain signal of a preset number of windows (S330); transforming the frequency domain signal of the preset number of windows from a frequency scale to the mel scale, and obtaining a mel spectrogram (S350); inputting the mel spectrogram into a pre-established instrument recognition model to obtain an instrument type used in the audio data under recognition (S370). In order to resolve the issue of the prior art in which the types of instruments used in polyphonus music cannot be recognized, a pre-constructed instrument recognition model is used to process a mel spectrogram of audio data under recognition, and accordingly to identify the types of instruments used in polyphonus music.

(74) 代理人: 深圳市隆天联鼎知识产权代理有限公司 (SHENZHEN LUNGTIN LIANDING INTELLECTUAL PROPERTY AGENT.LTD.); 中国广东省深圳市福田区南园路上田大厦 4A 刘抗美, Guangdong 518000 (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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))。

(57) 摘要: 一种音频数据的乐器识别方法及装置、电子设备、计算机可读存储介质, 包括: 对待识别音频数据进行预处理 (S310); 通过短时傅里叶变换将待识别音频数据从时域信号转换成预设窗口数量的频域信号 (S330); 将预设窗口数量的频域信号从频率标度转换为梅尔标度, 得到梅尔谱图 (S350); 将梅尔谱图输入预先构建的乐器识别模型, 得到待识别音频数据所使用的乐器种类 (S370)。解决了现有技术无法识别出复调音乐中所使用乐器种类的缺陷, 通过预先构建的乐器识别模型对待识别音频数据的梅尔谱图进行处理, 进而可以识别出复调音乐中所使用的乐器种类。

音频数据的乐器识别方法及装置、电子设备、存储介质

技术领域

- [0001] 本申请要求2018年7月5日递交、发明名称为“音频数据的乐器识别方法及装置、电子设备、存储介质”的中国专利申请201810729847.1的优先权，在此通过引用将其全部内容合并于此。
- [0002] 本申请涉及计算机技术领域，特别涉及一种音频数据的乐器识别方法及装置、电子设备、计算机可读存储介质。

背景技术

- [0003] 乐器识别是音乐内容分析的一个重要问题，它可以为其它声源识别技术提供有效参考，但由于应用领域狭窄，研究尚未成熟。目前已有的技术方案，主要研究对象是西方乐器和中国民族乐器。其中西方乐器包括：小提琴、中提琴、大提琴、低音大提琴、吉他、竖琴、钢琴、小号、短号、法国号、萨克斯、低音号、长笛、双簧管、单簧管、排箫和低音管等。中国民族乐器主要研究了二胡、高胡、中胡、琵琶、古筝、阮、扬琴、鼓、竹笛。
- [0004] 现有的乐器音频识别方法主要配置为识别单调音乐（仅使用一种乐器），通过提取某单调音乐的声学特征（抽取的声学特征大致有时频特征、能量特征、波谱特征、和声特征、感知特征等），将其与训练阶段获得的不同乐器的声学特征进行匹配，进而实现单调音乐的乐器识别。而发明人意识到复调音乐与单调音乐不同，复调音乐是由多种乐器混合演奏产生的音频数据，从复调音乐中所提取的声学特征是多种乐器声学特征的混合，由此难以通过特征匹配的方式，识别出复调音乐中所使用的乐器种类。

发明概述

技术问题

- [0005] 为了解决相关技术中存在的难以识别出复调音乐中所使用的乐器种类的问题，本申请提供了一种音频数据的乐器识别方法。

问题的解决方案

技术解决方案

- [0006] 本申请提供了一种音频数据的乐器识别方法，包括：
- [0007] 对待识别音频数据进行预处理；
- [0008] 通过短时傅里叶变换将所述待识别音频数据从时域信号转换成预设窗口数量的频域信号；
- [0009] 将所述预设窗口数量的频域信号从频率标度转换为梅尔标度，得到梅尔谱图；
- [0010] 将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类。
- [0011] 本申请还提供了一种音频数据的乐器识别装置，所述装置包括：
- [0012] 预处理模块，配置为对待识别音频数据进行预处理；
- [0013] 信号转换模块，配置为通过短时傅里叶变换将所述待识别音频数据从时域信号转换成预设窗口数量的频域信号；
- [0014] 梅尔转换模块，配置为将所述预设窗口数量的频域信号从频率标度转换为梅尔标度，得到梅尔谱图；
- [0015] 乐器识别模块，配置为将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类。
- [0016] 进一步，本申请还提供了一种电子设备，所述电子设备：
- [0017] 处理器；
- [0018] 配置为存储处理器可执行指令的存储器；
- [0019] 其中，所述处理器被配置为执行上述任意一种音频数据的乐器识别方法。
- [0020] 进一步的，本申请还提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序可由处理器执行完成上述任意一种音频数据的乐器识别方法。

发明的有益效果

有益效果

- [0021] 本申请的实施例提供的技术方案可以包括以下有益效果：
- [0022] 本申请提供的技术方案，通过将待识别音频数据转换为梅尔谱图，并利用预先

构建的乐器识别模型对梅尔谱图进行处理，进而可以得到待识别音频数据所使用的乐器种类。由于复调音乐是由多种乐器混合演奏产生的音频数据，从复调音乐中所提取的声学特征是多种乐器声学特征的混合，由此难以通过与乐器声学特征匹配的方式，识别出复调音乐中所使用的乐器种类。本申请提供的方案，解决了现有技术无法识别出复调音乐中所使用乐器种类的缺陷，通过预先构建的乐器识别模型对待识别音频数据的梅尔谱图进行处理，进而可以识别出复调音乐中所使用的乐器种类。

[0023] 应当理解的是，以上的一般描述和后文的细节描述仅是示例性的，并不能限制本申请。

对附图的简要说明

附图说明

[0024] 此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本申请的实施例，并于说明书一起配置为解释本申请的原理。

[0025] 图1是根据本申请所涉及的实施环境的示意图；

[0026] 图2是根据一示例性实施例示出的一种服务器的框图；

[0027] 图3是根据一示例性实施例示出的一种音频数据的乐器识别方法的流程图；

[0028] 图4是图3对应实施例中步骤370的细节流程图；

[0029] 图5为音频数据的乐器识别过程示意图；

[0030] 图6是在图3对应实施例的基础上示出的一种音频数据的乐器识别方法的流程图；

[0031] 图7是图6对应实施例中步骤362的细节流程图；

[0032] 图8是图3对应实施例中步骤310的细节流程图；

[0033] 图9是根据另一示例性实施例示出的一种音频数据的乐器识别装置的框图。

发明实施例

本发明的实施方式

[0034] 里将详细地对示例性实施例执行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本申请相一致的所有实施方式。

相反，它们仅是与如所附权利要求书中所详述的、本申请的一些方面相一致的装置和方法的例子。

[0035] 图1是根据一示例性实施例示出的本申请所涉及的实施环境示意图。本申请所涉及的实施环境包括服务器110。服务器110可以采用本申请提供的方法对音频数据进行乐器识别。

[0036] 根据需要，该实施环境还将包括提供数据，即音频数据的数据来源。具体而言，在本实施环境中，数据来源可以为移动终端130。服务器110可以获取移动终端130上传的待识别音频数据，进而采用本申请提供的方法对该待识别音频数据进行乐器识别处理。

[0037] 应当说明的是，本申请音频数据的乐器识别方法，不限于在服务器110中部署相应的处理逻辑，其也可以是部署于其它机器中的处理逻辑。例如，在具备计算能力的终端设备中部署对音频数据进行乐器识别的处理逻辑等。

[0038] 参见图2，图2是本申请实施例提供的一种服务器结构示意图。该服务器200可因配置或性能不同而产生比较大的差异，可以包括一个或一个以上中央处理器（central processing units，CPU）222（例如，一个或一个以上处理器）和存储器232，一个或一个以上存储应用程序242或数据244的存储介质230（例如一个或一个以上海量存储设备）。其中，存储器232和存储介质230可以是短暂存储或持久存储。存储在存储介质230的程序可以包括一个或一个以上模块（图示未示出），每个模块可以包括对服务器200中的一系列指令操作。更进一步地，中央处理器222可以设置为与存储介质230通信，在服务器200上执行存储介质230中的一系列指令操作。服务器200还可以包括一个或一个以上电源226，一个或一个以上有线或无线网络接口250，一个或一个以上输入输出接口258，和/或，一个或一个以上操作系统241，例如Windows Server™，Mac OS X™，Unix™，Linux™，FreeBSD™等等。下述图3、图4、图6-图8所示实施例中所述的由服务器所执行的步骤可以基于该图2所示的服务器结构。

[0039] 本领域普通技术人员可以理解实现下述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种

计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

[0040] 图3是根据一示例性实施例示出的一种音频数据的乐器识别方法的流程图。该音频数据的乐器识别方法的适用范围和执行主体，例如，该方法用于图1所示实施环境的服务器110。如图3所示，该方法可以由服务器110执行，可以包括以下步骤。

[0041] 在步骤310中，对待识别音频数据进行预处理。

[0042] 其中，待识别音频数据是指所使用乐器种类未知的音频片段，需要对该音频片段多采用的乐器种类进行识别。预处理可以包括但不限于对待识别音频数据从立体声变成单声道、进行降采样以及归一化。

[0043] 在步骤330中，通过短时傅里叶变换将所述预处理后的待识别音频数据从时域信号转换成预设窗口数量的频域信号。

[0044] 其中，短时傅里叶变换(STFT)是和傅里叶变换相关的一种数学变换，用以确定时变信号其局部区域正弦波的频率与相位。简单来说，短时傅里叶变换是给待识别音频数据（时域信号）在时域上加窗，把信号分成一小段一小段，然后对每一段信号分别做傅里叶变换。短时傅里叶变换使用一个固定的窗函数，将按照时间变化的待识别音频数据划分成n个窗口，按照音频数据的时间长短可以调整窗口数量，时间越长窗口越多，当然窗越短越好，尽可能使得窗内信号频率近似不变。在一种实施例中，对于3秒的音频数据，窗口数量可以是43个。具体的，通过进行窗口平移，将待识别音频数据划分成43个窗口的时域信号，然后采用傅里叶变换对每个窗口的时域信号进行傅里叶变换，得到43段时域信号对应的43段频域信号。需要说明的是，傅里叶变换将时域信号转换为频域信号属于现有技术在此不再赘述。

[0045] 在步骤350中，将所述预设窗口数量的频域信号从频率标度转换为梅尔标度，得到梅尔谱图。

[0046] 具体的，梅尔标度（the mel scale）由Stevens, Volkmann和Newman在1937年命名。频率的单位是赫兹（Hz），人耳能听到的频率范围是20-20000Hz，但人耳对Hz这种标度单位并不是线性感知关系。例如如果我们适应了1000Hz的音调，

如果把音调频率提高到2000Hz，我们的耳朵只能觉察到频率提高了一点点，根本察觉不到频率提高了一倍。如果将普通的频率标度（以HZ为标度单位）转化为梅尔标度，映射关系如下式所示：

[]

$$f_{mel}(f) = 2595 \cdot \log \left(1 + \frac{f}{700Hz} \right)$$

[0047] 其中，f代表频率标度，单位为HZ（赫兹），f_{mel}(f)代表梅尔标度。经过上述转化则人耳对频率的感知度就成了线性关系。也就是说，在梅尔标度下，如果两段语音的梅尔频率相差两倍，则人耳可以感知到的音调大概也相差两倍。将每个窗口的频域信号从频率标度转换为梅尔标度，也就是将普通频率值映射为用户对音高的感知（因为频率的变化用户没法准确感知，将频率值转化为mel标度，mel值的变化用户可以准确感知）。

[0048] 根据需要还可以通过自然对数压缩，使每个窗口内只使用128个梅尔频率点，从而既能够充分保留音乐的谐波特性，同时大大降低输入数据的维度。进而将每个窗口的128个梅尔频率点进行拼接，假设窗口数量是43个，可以得到1×43×128数据大小的梅尔谱图。

[0049] 在步骤370中，将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类。

[0050] 其中，乐器识别模型是通过大量已知所使用乐器种类的样本音频片段训练得到的。对于模型的训练将梅尔谱图作为乐器识别模型的输入，通过乐器识别模型对梅尔谱图进行卷积处理，根据处理结果可以得到待识别音频数据所使用的乐器种类。

[0051] 本申请上述示例性实施例提供的技术方案，通过将待识别音频数据转换为梅尔谱图，并利用预先构建的乐器识别模型对梅尔谱图进行处理，进而可以得到待识别音频数据所使用的乐器种类。由于复调音乐是由多种乐器混合演奏产生的音频数据，从复调音乐中所提取的声学特征是多种乐器声学特征的混合，由此难以通过与乐器特征匹配的方式，识别出复调音乐中所使用的乐器种类。本申请提供的方案，解决了现有技术无法识别出复调音乐中所使用乐器种类的缺陷。

，通过预先构建的乐器识别模型对待识别音频数据的梅尔谱图进行处理，可以识别出复调音乐中所使用的乐器种类。

[0052] 在一种示例性实施例中，如图4所示，上述步骤370具体包括：

[0053] 在步骤371中，通过预先构建的乐器识别模型对所述梅尔谱图进行卷积计算，输出卷积计算结果。

[0054] 其中，乐器识别模型可以是经过训练的深度卷积神经网络模型，利用深度卷积神经网络模型训练好的卷积层、池化层和全连接层的权重参数，对梅尔谱图进行卷积计算。图5为音频数据的乐器识别过程示意图。如图5所示，待识别音频数据被转化为梅尔谱图，然后通过训练好的多层卷积层、池化层和全连接层对梅尔谱图进行卷积计算，输出的结果即为卷积计算结果。

[0055] 在步骤372中，对所述卷积计算结果进行归一化处理，得到所述待识别音频数据包含每种乐器的概率。

[0056] 其中，卷积计算结果是一个一维向量，包含多个元素，每个元素代表一种乐器。假设在乐器识别模型训练阶段存在11种乐器，则卷积计算结果包括11项。通过对卷积计算结果进行归一化处理，归一化后的每个元素代表使用每种乐器的概率。归一化处理可以是最大值归一化方法，即将所有元素的最大值转化为1，将其他元素的值线性化地转换到0到

1的范围。归一化公式如下： $X^* = (X - X_{\min}) / (X_{\max} - X_{\min})$ 。其中， $X_{\max}$ 为卷积计算结果中所有元素的最大值， $X_{\min}$ 为卷积计算结果中所有元素的最小值， X^* 为元素 X 的归一化结果。

[0057] 举例来说，对卷积计算结果的所有元素进行归一化后得到11个输出数值，11个输出数值的概率依次为0.4、0.3、0.7、0.1、0.1、0.1、0.8、0、0.1、0.1、0.1。假设在建模阶段11个数值依次代表小提琴、中提琴、大提琴、低音大提琴、吉他、竖琴、钢琴、小号、短号、法国号、萨克斯，则根据归一化的11个输出数值，得到待识别音频数据使用小提琴概率是0.4，中提琴概率是0.3，大提琴概率是0.7、钢琴概率是0.8，其他概率较小可忽略。

[0058] 在步骤373中，将所述待识别音频数据包含每种乐器的概率与预设参数进行比较，确定所述待识别音频数据所使用的乐器种类。

- [0059] 举例来说, 预设参数可以是0.5, 将待识别音频数据包含每种乐器的概率与预设参数进行比较, 如果包含某种乐器的概率大于等于预设参数, 可以认为待识别音频数据使用了该乐器。假设待识别音频数据使用小提琴概率是0.4, 中提琴概率是0.3, 大提琴概率是0.7、钢琴概率是0.8, 其他概率较小可忽略, 通过将上述概率值与0.5进行比较, 使用小提琴和中提琴的概率均小于0.5, 使用大提琴和钢琴的概率大于0.5, 所以可以确定待识别音频数据是由大提琴和钢琴演奏的复调音乐。
- [0060] 在一种示例性实施例中, 如图6所示, 在上述步骤370之前本申请提供的乐器识别方法还包括以下步骤:
- [0061] 在步骤361中, 获取样本音频片段以及所述样本音频片段已知的乐器标签信息。
- [0062] 其中, 样本音频片段可以是从小提琴清晰录音中提取的6000个时长为3秒的音频数据, 并且已知样本音频片段所使用的乐器种类。根据每个样本音频片段所使用的乐器种类对样本音频片段进行标记, 可以得到样本标签信息。举例来说, 假设一共存在11种乐器, 某个样本音频片段使用了小提琴、低音大提琴、钢琴、萨克斯, 使用该乐器就标记为1, 没有使用该乐器就标记为0, 则该样本标签信息为100100100 01。样本音频片段及其对应的乐器标签信息可以提前存储在服务器110的数据库中。
- [0063] 在步骤362中, 通过所述样本音频片段以及对应的乐器标签信息进行神经网络学习, 得到所述乐器识别模型。
- [0064] 其中, 进行神经网络学习是指根据样本音频片段及其对应的乐器标签信息, 通过搭建神经网络架构学习乐器特征提取, 使学习到的乐器特征与其乐器标签信息之间的差异最小, 此时根据搭建的神经网络架构的网络参数即可得到乐器识别模型。该神经网络架构通常包括输入层、卷积层和池化层以及全连接层。
- [0065] 在一种示例性实施例中, 如图7所示, 上述步骤362具体包括:
- [0066] 在步骤3621中, 提取所述样本音频片段的样本梅尔谱图;
- [0067] 具体的, 可以参照对待识别音频数据的处理过程, 对样本音频片段进行预处理, 然后通过短时傅里叶变换对预处理后的样本音频片段从时域信号转换成预设

窗口的频域信号，并对所有窗口的频域信号从普通的频率标度转换成梅尔标度，得到样本音频片段的梅尔谱图（为进行区分，此处称为样本梅尔谱图）。

[0068] 在步骤3622中，通过搭建深度卷积神经网络对所述样本梅尔谱图进行卷积计算，提取所述样本梅尔谱图对应的乐器特征；

[0069] 其中，深度卷积神经网络可以包含4阶段，每一阶段包含双卷积层和池化层，最后经过全连接层进行输出。在深度卷积神经网络中，卷积层可以使用尺寸为 3×3 感受野、步长为1的32个滤波器进行卷积计算，在池化层用尺寸为 3×3 ，步长为1的最大值池化方法进行空间抽象。最大值池化之后需要按照一定概率暂时从网络中丢弃一些神经网络单元，用于防止神经网络过度拟合。其中，随机丢弃的概率可以是0.25。全连接层之后的随机丢弃概率可以增加到0.5，因为全连接层更容易发生过度拟合。表1是深度卷积神经网络卷积计算过程中每层的输入数据大小、以及参数值（除了补零过程）。

[0070] 表1 卷积过程中每层的输入数据大小、参数值

[]

输入数据大小	描述
1×43×128	梅尔谱图
32×45×130	3×3 卷积, 32 个滤波器
32×47×132	3×3 卷积, 32 个滤波器
32×15×44	3×3 最大池化
32×15×44	随机丢弃概率 (0.25)
64×17×46	3×3 卷积, 64 个滤波器
64×19×48	3×3 卷积, 64 个滤波器
64×6×16	3×3 最大池化
64×6×16	随机丢弃概率 (0.25)
128×8×18	3×3 卷积, 128 个滤波器
128×10×20	3×3 卷积, 128 个滤波器
128×3×6	3×3 最大池化
128×3×6	随机丢弃概率 (0.25)
256×5×8	3×3 卷积, 256 个滤波器
256×7×10	3×3 卷积, 256 个滤波器
256×1×1	全局最大池化
1024	扁平化全连接
1024	随机丢弃概率 (0.5)
11	S 型连续输出

[0071] 如上表1所示, 尺寸为1×43×128的梅尔谱图, 经过32个3×3的滤波器进行卷积计算, 得到32×45×130数据大小的中间图像 (即32张尺寸为45×130的图像), 其中43变成45, 由128变成130是因为经过了补零过程 (表中未示出)。参照上述表中所列过程, 对梅尔谱图经过卷积、最大池化和全连接的计算, 最后输出结果可以是11个数据, 得到样本梅尔谱图的乐器特征。

[0072] 在步骤3623中, 根据所述样本音频片段对应的乐器标签信息, 调整所述深度卷积神经网络的权重参数, 使所述乐器特征与所述乐器标签信息之间的差异最小, 调整后的所述深度卷积神经网络作为所述乐器识别模型。

[0073] 具体的, 乐器特征可以是上述深度卷积神经网络最后输出的11个数据, 计算样本音频片段的乐器标签信息 (例如100100100 01) 与乐器特征 (11个数据) 之间

的差异，通过调整深度卷积神经网络的权重参数（如上述表中各滤波器的参数），使乐器标签信息与最终输出的乐器特征之间的差异最小。其中乐器标签信息与乐器特征之间的差异可以通过计算乐器标签信息与乐器特征之间的相似度、距离等，相似度较高时认为差异较小，距离较小时认为差异较小。经过参数调整的深度卷积神经网络作为后续进行乐器识别的乐器识别模型。

[0074] 在一种示例性实施例中，可以通过调整所述深度卷积神经网络的权重参数，使提取的所述乐器特征与所述乐器标签信息之间的欧式距离最小。

[0075] 其中，乐器特征和乐器标签信息可以看成是N维空间（如11维）中的两个点x1、x2，通过以下公式可以计算出这两个点之间的距离d，不断调整深度卷积神经网络的权重参数，使其输出的乐器特征与乐器标签信息之间的距离d最小。

[]

$$d = \sqrt{\sum_{i=1}^N (x_{1i} - x_{2i})^2}$$

[0076] 在一种示例性实施例中，如图8所示，上述步骤310具体包括：

[0077] 在步骤311中，将所述待识别音频数据从立体声转化为单声道；

[0078] 其中，立体声是指包含左右声道，将待识别音频数据从立体声转换成单声道，可以是将待识别音频数据的左声道数据和右声道数据取平均值，从而得到单声道的音频数据。通过将立体声转换成单声道，采样频率可以从原来的44100HZ降为22050HZ。

[0079] 在步骤312中，对单声道的所述待识别音频数据进行降采样；

[0080] 其中，降采样是指降低特定信号（待识别音频数据）的采样率的过程，具体的，可以基于Nyquist（奈奎斯特）采样定理，使22050HZ的单声道音频数据的采样频率降为11025HZ。该采样频率包含了乐器产生的绝大部分和声音响，同时尽可能去除了此频率之上的噪音。

[0081] 在步骤313中，对降采样后的所述待识别音频数据进行归一化。

[0082] 在经过降采样，将待识别音频数据的采样频率变成11025HZ后，对降采样的音频数据进行归一化。其中，归一化方法可以采用最大值归一化方法，即将最大值变换成1，最小值变换成0，将音频数据的幅值线性地转换到0到1的范围。

[0083] 下述为本申请装置实施例，可以配置为执行本申请上述服务器110执行的音频

数据的乐器识别方法实施例。对于本申请装置实施例中未披露的细节，请参照本申请音频数据的乐器识别方法实施例。

[0084] 图9是根据一示例性实施例示出的一种音频数据的乐器识别装置的框图，该音频数据的乐器识别装置可以用于图1所示实施环境的服务器110中，执行图3、图4、图6-图8任一所示的音频数据的乐器识别方法的全部或者部分步骤。如图9所示，该装置包括但不限于：预处理模块910、信号转换模块930、梅尔转换模块950以及乐器识别模块970。

[0085] 预处理模块910，配置为对待识别音频数据进行预处理；

[0086] 信号转换模块930，配置为通过短时傅里叶变换将所述待识别音频数据从时域信号转换成预设窗口数量的频域信号；

[0087] 梅尔转换模块950，配置为将所述预设窗口数量的频域信号从频率标度转换为梅尔标度，得到梅尔谱图；

[0088] 乐器识别模块970，配置为将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类。

[0089] 上述装置中各个模块的功能和作用的实现过程具体详见上述音频数据的乐器识别方法中对应步骤的实现过程，在此不再赘述。

[0090] 预处理模块910、信号转换模块930、梅尔转换模块950以及乐器识别模块970可以是功能模块，配置为执行上述音频数据的乐器识别方法中的对应步骤。可以理解，这些模块可以通过硬件、软件、或二者结合来实现。当以硬件方式实现时，这些模块可以实施为一个或多个硬件模块，例如一个或多个专用集成电路。当以软件方式实现时，这些模块可以实施为在一个或多个处理器上执行的一个或多个计算机程序，例如图2的中央处理器222所执行的存储在存储器232中的程序。

[0091] 在一种示例性实施例中，所述乐器识别模块包括：

[0092] 结果输出单元，配置为通过预先构建的乐器识别模型对所述梅尔谱图进行卷积计算，输出卷积计算结果；

[0093] 结果归一化单元，配置为对所述卷积计算结果进行归一化处理，得到所述待识别音频数据包含每种乐器的概率；

- [0094] 参数比较单元，配置为将所述待识别音频数据包含每种乐器的概率与预设参数进行比较，确定所述待识别音频数据所使用的乐器种类。
- [0095] 在一种示例性实施例中，所述装置还包括：
- [0096] 数据获取模块，配置为获取样本音频片段以及所述样本音频片段已知的乐器标签信息；
- [0097] 模型构建模块，配置为通过所述样本音频片段以及对应的乐器标签信息进行神经网络学习，得到所述乐器识别模型。
- [0098] 在一种示例性实施例中，所述模型构建模块包括：
- [0099] 梅尔谱图提取单元，配置为提取所述样本音频片段的样本梅尔谱图；
- [0100] 乐器特征提取单元，配置为通过搭建深度卷积神经网络对所述样本梅尔谱图进行卷积计算，提取所述样本梅尔谱图对应的乐器特征；
- [0101] 参数调整单元，配置为根据所述样本音频片段对应的乐器标签信息，调整所述深度卷积神经网络的权重参数，使所述乐器特征与所述乐器标签信息之间的差异最小，调整后的所述深度卷积神经网络作为所述乐器识别模型。
- [0102] 在一种示例性实施例中，所述参数调整单元包括：
- [0103] 距离调整子单元，配置为通过调整所述深度卷积神经网络的权重参数，使提取的所述乐器特征与所述乐器标签信息之间的欧式距离最小。
- [0104] 在一种示例性实施例中，所述预处理模块包括：
- [0105] 声道转化单元，配置为将所述待识别音频数据从立体声转化为单声道；
- [0106] 降采样单元，配置为对单声道的所述待识别音频数据进行降采样；
- [0107] 音频归一化单元，配置为对降采样后的所述待识别音频数据进行归一化。
- [0108] 在一种示例性实施例中，所述声道转化单元包括：
- [0109] 数据平均子单元，配置为将所述待识别音频数据的左声道数据和右声道数据取平均值，得到单声道的音频数据。
- [0110] 可选的，本申请还提供一种电子设备，该电子设备可以配置为图1所示实施环境的服务器110中，执行图3、图4、图6-图8任一所示的音频数据的乐器识别方法的全部或者部分步骤。所述电子设备包括：
- [0111] 处理器；

- [0112] 配置为存储处理器可执行指令的存储器；
- [0113] 其中，所述处理器被配置为执行上述示例性实施例所述的音频数据的乐器识别方法。
- [0114] 该实施例中电子设备的处理器执行操作的具体方式已经在有关音频数据的乐器识别方法的实施例中执行了详细描述，此处将不做详细阐述说明。
- [0115] 在示例性实施例中，还提供了一种存储介质，该存储介质为计算机可读存储介质，例如可以为包括指令的临时性和非临时性计算机可读存储介质。该存储介质存储有计算机程序，所述计算机程序可由服务器200的中央处理器222执行以完成上述音频数据的乐器识别方法。
- [0116] 应当理解的是，本申请并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围执行各种修改和改变。本申请的范围仅由所附的权利要求来限制。

权利要求书

- [权利要求 1] 一种音频数据的乐器识别方法，其中，包括：
- 对待识别音频数据进行预处理；
- 通过短时傅里叶变换将所述待识别音频数据从时域信号转换成预设窗口数量的频域信号；
- 将所述预设窗口数量的频域信号从频率标度转换为梅尔标度，得到梅尔谱图；
- 将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类。
- [权利要求 2] 根据权利要求1所述的方法，其中，所述将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类，包括：
- 通过预先构建的乐器识别模型对所述梅尔谱图进行卷积计算，输出卷积计算结果；
- 对所述卷积计算结果进行归一化处理，得到所述待识别音频数据包含每种乐器的概率；
- 将所述待识别音频数据包含每种乐器的概率与预设参数进行比较，确定所述待识别音频数据所使用的乐器种类。
- [权利要求 3] 根据权利要求1-2任意一项所述的方法，其中，所述将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类之前，所述方法包括：
- 获取样本音频片段以及所述样本音频片段已知的乐器标签信息；
- 通过所述样本音频片段以及对应的乐器标签信息进行神经网络学习，得到所述乐器识别模型。
- [权利要求 4] 根据权利要求3所述的方法，其中，所述通过所述样本音频片段以及对应的乐器标签信息进行神经网络学习，得到所述乐器识别模型，包括：
- 提取所述样本音频片段的样本梅尔谱图；

通过搭建深度卷积神经网络对所述样本梅尔谱图进行卷积计算，提取所述样本梅尔谱图对应的乐器特征；

根据所述样本音频片段对应的乐器标签信息，调整所述深度卷积神经网络的权重参数，使所述乐器特征与所述乐器标签信息之间的差异最小，调整后的所述深度卷积神经网络作为所述乐器识别模型。

[权利要求 5] 根据权利要求4所述的方法，其中，根据所述样本音频片段对应的乐器标签信息，调整所述深度卷积神经网络的权重参数，使所述乐器特征与所述乐器标签信息之间的差异最小，包括：
通过调整所述深度卷积神经网络的权重参数，使提取的所述乐器特征与所述乐器标签信息之间的欧式距离最小。

[权利要求 6] 根据权利要求1-5任意一项所述的方法，其中，所述对待识别音频数据进行预处理，包括：
将所述待识别音频数据从立体声转化为单声道；
对单声道的所述待识别音频数据进行降采样；
对降采样后的所述待识别音频数据进行归一化。

[权利要求 7] 根据权利要求6所述的方法，其中，将所述待识别音频数据从立体声转化为单声道，包括：
将所述待识别音频数据的左声道数据和右声道数据取平均值，得到单声道的音频数据。

[权利要求 8] 一种音频数据的乐器识别装置，其中，包括：
预处理模块，配置为对待识别音频数据进行预处理；
信号转换模块，配置为通过短时傅里叶变换将所述待识别音频数据从时域信号转换成预设窗口数量的频域信号；
梅尔转换模块，配置为将所述预设窗口数量的频域信号从频率标度转换为梅尔标度，得到梅尔谱图；
乐器识别模块，配置为将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类。

[权利要求 9] 根据权利要求8所述的装置，其中，所述乐器识别模块包括：

结果输出单元，配置为通过预先构建的乐器识别模型对所述梅尔谱图进行卷积计算，输出卷积计算结果；

结果归一化单元，配置为对所述卷积计算结果进行归一化处理，得到所述待识别音频数据包含每种乐器的概率；

参数比较单元，配置为将所述待识别音频数据包含每种乐器的概率与预设参数进行比较，确定所述待识别音频数据所使用的乐器种类。

[权利要求 10]

根据权利要求8-9任意一项所述的装置，其中，所述装置还包括：

数据获取模块，配置为获取样本音频片段以及所述样本音频片段已知的乐器标签信息；

模型构建模块，配置为通过所述样本音频片段以及对应的乐器标签信息进行神经网络学习，得到所述乐器识别模型。

[权利要求 11]

根据权利要求10所述的装置，其中，所述模型构建模块包括：

梅尔谱图提取单元，配置为提取所述样本音频片段的样本梅尔谱图；

乐器特征提取单元，配置为通过搭建深度卷积神经网络对所述样本梅尔谱图进行卷积计算，提取所述样本梅尔谱图对应的乐器特征；

参数调整单元，配置为根据所述样本音频片段对应的乐器标签信息，调整所述深度卷积神经网络的权重参数，使所述乐器特征与所述乐器标签信息之间的差异最小，调整后的所述深度卷积神经网络作为所述乐器识别模型。

[权利要求 12]

根据权利要求11所述的装置，其中，所述参数调整单元包括：

距离调整子单元，配置为通过调整所述深度卷积神经网络的权重参数，使提取的所述乐器特征与所述乐器标签信息之间的欧式距离最小。

[权利要求 13]

根据权利要求8-12任意一项所述的装置，其中，所述预处理模块包括：

声道转化单元，配置为将所述待识别音频数据从立体声转化为单声道；

降采样单元，配置为对单声道的所述待识别音频数据进行降采样；

音频归一化单元，配置为对降采样后的所述待识别音频数据进行归一

化。

- [权利要求 14] 根据权利要求13所述的方法，其中，所述声道转化单元包括：
数据平均子单元，配置为将所述待识别音频数据的左声道数据和右声道数据取平均值，得到单声道的音频数据。
- [权利要求 15] 一种电子设备，其中，所述电子设备：
处理器；
配置为存储处理器可执行指令的存储器；
其中，所述处理器被配置为执行以下步骤：
对待识别音频数据进行预处理；
通过短时傅里叶变换将所述待识别音频数据从时域信号转换成预设窗口数量的频域信号；
将所述预设窗口数量的频域信号从频率标度转换为梅尔标度，得到梅尔谱图；
将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类。
- [权利要求 16] 根据权利要求15所述的电子设备，其中，所述将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类，包括：
通过预先构建的乐器识别模型对所述梅尔谱图进行卷积计算，输出卷积计算结果；
对所述卷积计算结果进行归一化处理，得到所述待识别音频数据包含每种乐器的概率；
将所述待识别音频数据包含每种乐器的概率与预设参数进行比较，确定所述待识别音频数据所使用的乐器种类。
- [权利要求 17] 根据权利要求15-16任意一项所述的电子设备，其中，所述将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类之前，所述方法包括：
获取样本音频片段以及所述样本音频片段已知的乐器标签信息；

通过所述样本音频片段以及对应的乐器标签信息进行神经网络学习，得到所述乐器识别模型。

[权利要求 18] 根据权利要求17所述的电子设备，其中，所述通过所述样本音频片段以及对应的乐器标签信息进行神经网络学习，得到所述乐器识别模型，包括：

提取所述样本音频片段的样本梅尔谱图；

通过搭建深度卷积神经网络对所述样本梅尔谱图进行卷积计算，提取所述样本梅尔谱图对应的乐器特征；

根据所述样本音频片段对应的乐器标签信息，调整所述深度卷积神经网络的权重参数，使所述乐器特征与所述乐器标签信息之间的差异最小，调整后的所述深度卷积神经网络作为所述乐器识别模型。

[权利要求 19] 根据权利要求18所述的电子设备，其中，根据所述样本音频片段对应的乐器标签信息，调整所述深度卷积神经网络的权重参数，使所述乐器特征与所述乐器标签信息之间的差异最小，包括：

通过调整所述深度卷积神经网络的权重参数，使提取的所述乐器特征与所述乐器标签信息之间的欧式距离最小。

[权利要求 20] 根据权利要求15-19任意一项所述的电子设备，其中，所述对待识别音频数据进行预处理，包括：

将所述待识别音频数据从立体声转化为单声道；

对单声道的所述待识别音频数据进行降采样；

对降采样后的所述待识别音频数据进行归一化。

[权利要求 21] 根据权利要求20所述的电子设备，其中，将所述待识别音频数据从立体声转化为单声道，包括：

将所述待识别音频数据的左声道数据和右声道数据取平均值，得到单声道的音频数据。

[权利要求 22] 一种计算机可读存储介质，其中，所述计算机可读存储介质存储有计算机程序，所述计算机程序可由处理器执行以下步骤：

对待识别音频数据进行预处理；

通过短时傅里叶变换将所述待识别音频数据从时域信号转换成预设窗口数量的频域信号；

将所述预设窗口数量的频域信号从频率标度转换为梅尔标度，得到梅尔谱图；

将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类。

[权利要求 23] 根据权利要求22所述的计算机可读存储介质，其中，所述将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类，包括：

通过预先构建的乐器识别模型对所述梅尔谱图进行卷积计算，输出卷积计算结果；

对所述卷积计算结果进行归一化处理，得到所述待识别音频数据包含每种乐器的概率；

将所述待识别音频数据包含每种乐器的概率与预设参数进行比较，确定所述待识别音频数据所使用的乐器种类。

[权利要求 24] 根据权利要求22-23任意一项所述的计算机可读存储介质，其中，所述将所述梅尔谱图输入预先构建的乐器识别模型，得到所述待识别音频数据所使用的乐器种类之前，所述方法包括：

获取样本音频片段以及所述样本音频片段已知的乐器标签信息；

通过所述样本音频片段以及对应的乐器标签信息进行神经网络学习，得到所述乐器识别模型。

[权利要求 25] 根据权利要求24所述的计算机可读存储介质，其中，所述通过所述样本音频片段以及对应的乐器标签信息进行神经网络学习，得到所述乐器识别模型，包括：

提取所述样本音频片段的样本梅尔谱图；

通过搭建深度卷积神经网络对所述样本梅尔谱图进行卷积计算，提取所述样本梅尔谱图对应的乐器特征；

根据所述样本音频片段对应的乐器标签信息，调整所述深度卷积神经

网络的权重参数，使所述乐器特征与所述乐器标签信息之间的差异最小，调整后的所述深度卷积神经网络作为所述乐器识别模型。

[权利要求 26] 根据权利要求25所述的计算机可读存储介质，其中，根据所述样本音频片段对应的乐器标签信息，调整所述深度卷积神经网络的权重参数，使所述乐器特征与所述乐器标签信息之间的差异最小，包括：
通过调整所述深度卷积神经网络的权重参数，使提取的所述乐器特征与所述乐器标签信息之间的欧式距离最小。

[权利要求 27] 根据权利要求22-26任意一项所述的计算机可读存储介质，其中，所述对待识别音频数据进行预处理，包括：
将所述待识别音频数据从立体声转化为单声道；
对单声道的所述待识别音频数据进行降采样；
对降采样后的所述待识别音频数据进行归一化。

[权利要求 28] 根据权利要求27所述的计算机可读存储介质，其中，将所述待识别音频数据从立体声转化为单声道，包括：
将所述待识别音频数据的左声道数据和右声道数据取平均值，得到单声道的音频数据。

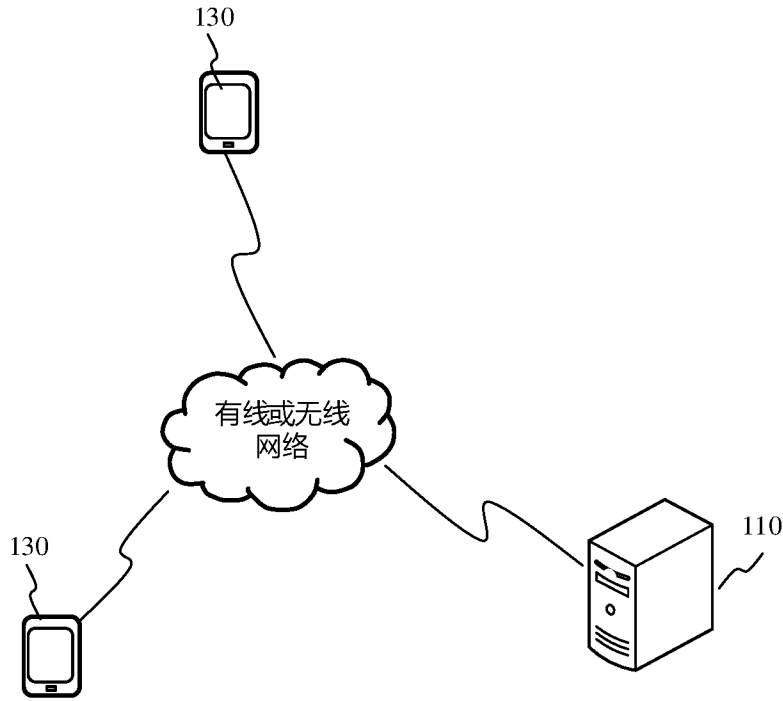


图 1

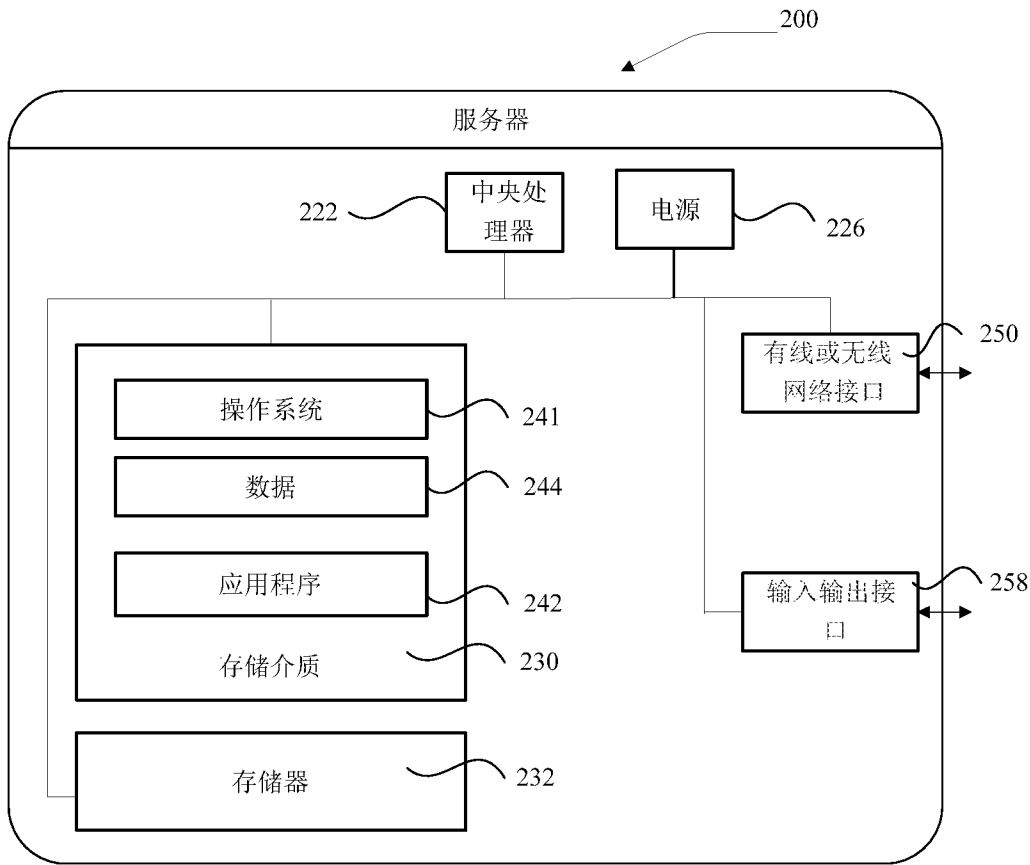


图 2

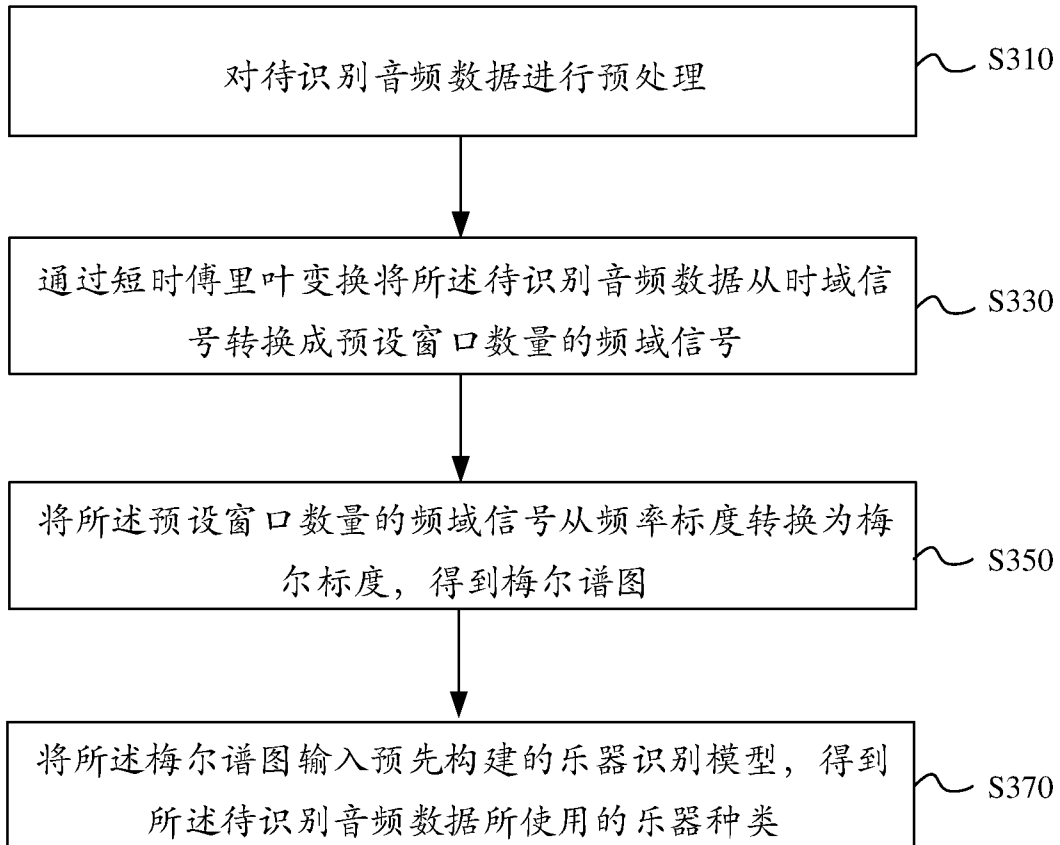


图 3

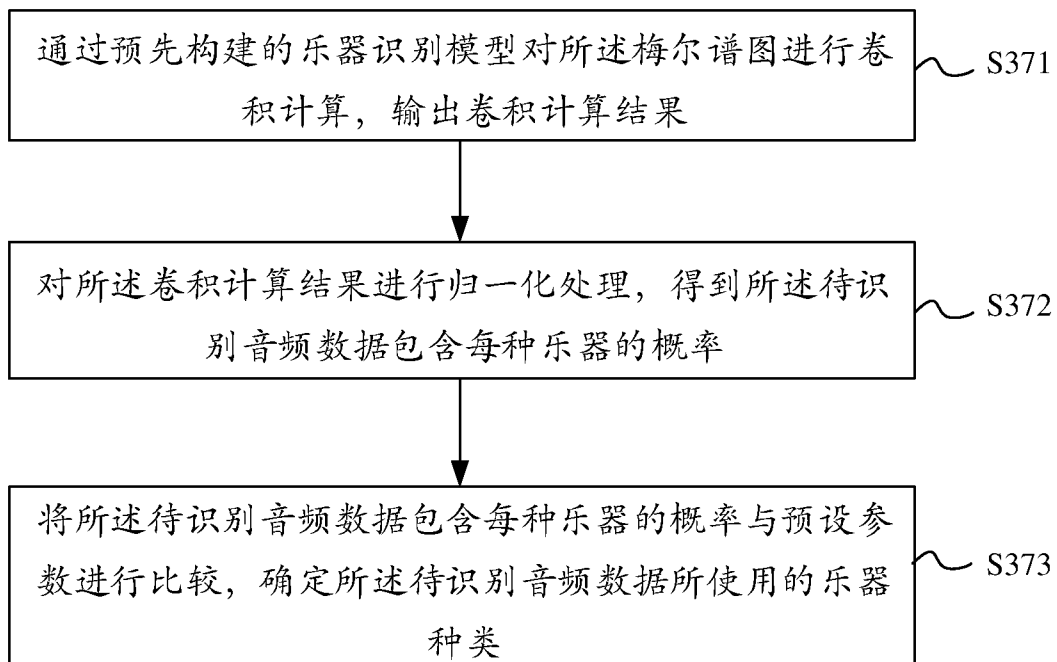


图 4

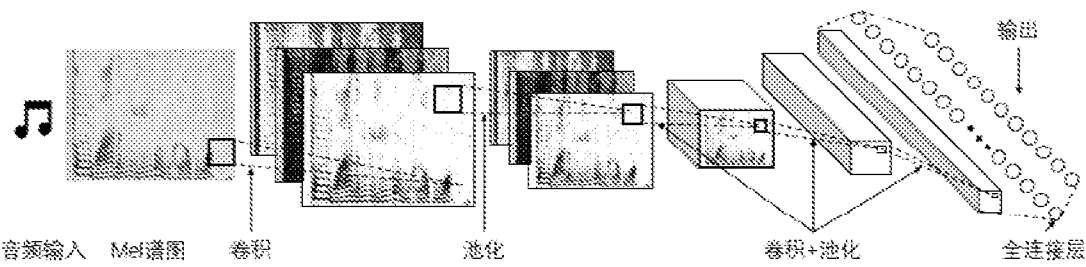


图 5

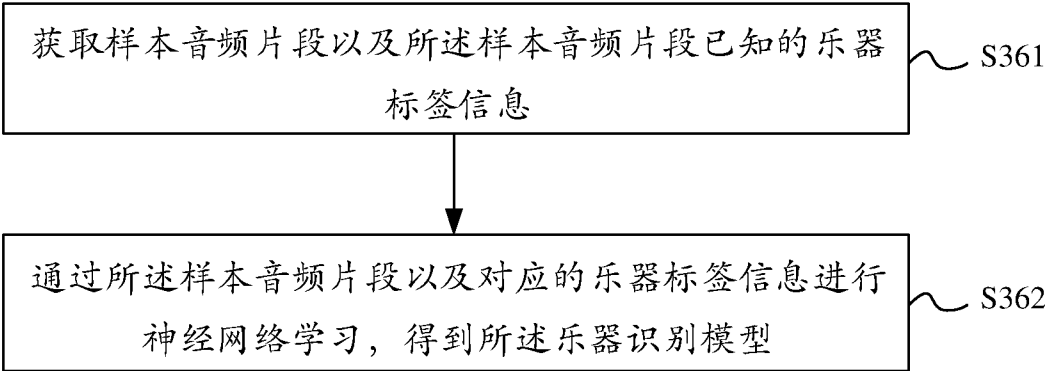


图 6

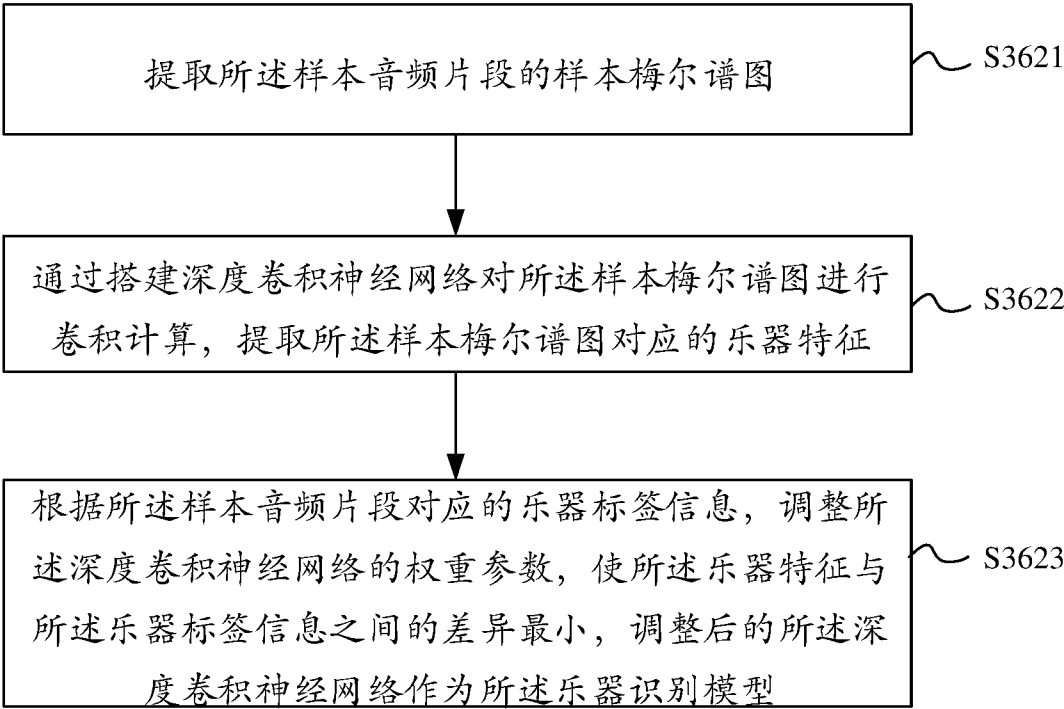


图 7

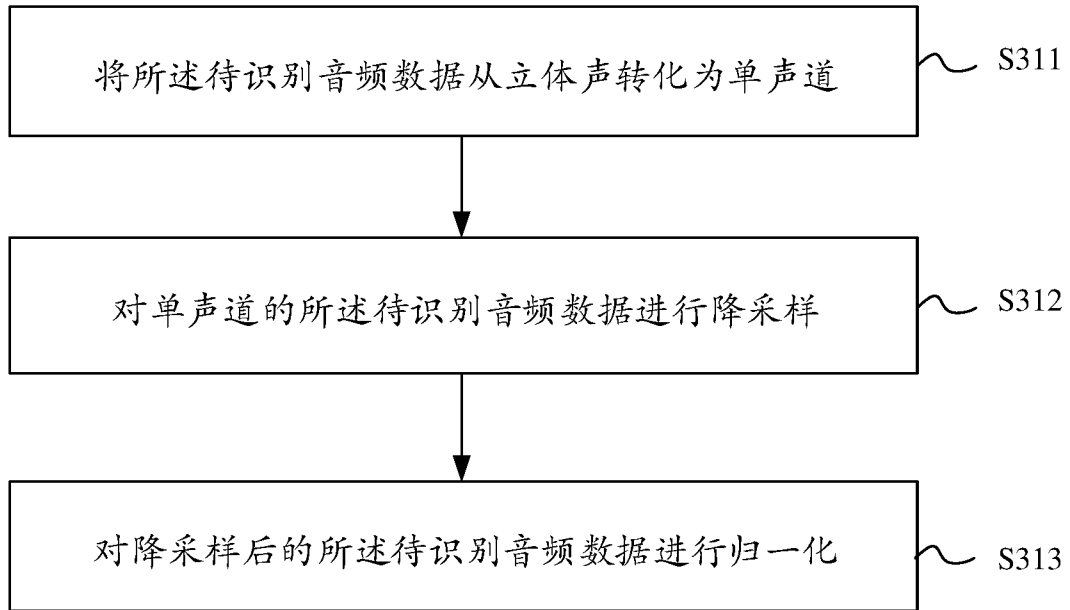


图 8

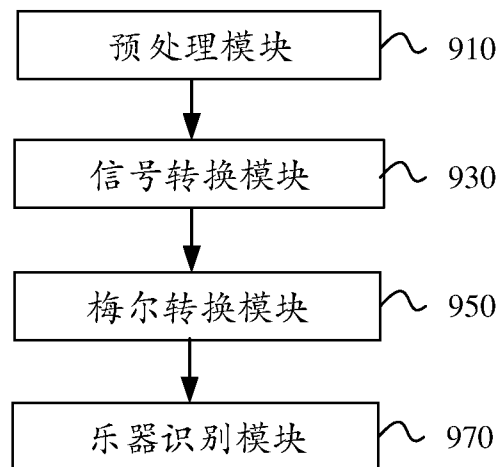


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/108239

A. CLASSIFICATION OF SUBJECT MATTER

G10L 25/51(2013.01)i; G10L 25/18(2013.01)i; G10L 25/24(2013.01)i; G10L 25/30(2013.01)i; G10L 25/45(2013.01)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 25/-; G10L 15/-

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)

CNPAT, WPI, EPODOC, CNKI: 乐器, 识别, 分类, 语音, 声音, 语音, 音频, 梅尔, 美尔, mel, mfcc, 深度, 卷积, 神经, 网络, 权重, 权值, music+, instrument?, recogni+, class+, speech+, voice?, sound?, audio?, dcnn, deep+, convolut+, convolv+, neural +, network+, weight?.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103325382 A (DALIAN NATIONALITIES UNIVERSITY) 25 September 2013 (2013-09-25) abstract, description, paragraphs [0028]-[0041], and figure 1	1, 6-8, 13-15, 20-22, 27-28
Y	CN 103325382 A (DALIAN NATIONALITIES UNIVERSITY) 25 September 2013 (2013-09-25) abstract, description, paragraphs [0028]-[0041], and figure 1	2-5, 9-12, 16-19, 23-26
Y	CN 106328121 A (NANJING UNIVERSITY OF SCIENCE AND TECHNOLOGY) 11 January 2017 (2017-01-11) abstract, and description, paragraphs [0005]-[0025]	2-5, 9-12, 16-19, 23-26
X	JP 05188932 A (SONY CORP.) 30 July 1993 (1993-07-30) abstract, description, paragraphs [0008]-[0038], and figure 1	1, 8, 15, 22
A	CN 103761965 A (TAIYUAN UNIVERSITY OF SCIENCE AND TECHNOLOGY) 30 April 2014 (2014-04-30) entire document	1-28



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

08 March 2019

Date of mailing of the international search report

27 March 2019

Name and mailing address of the ISA/CN

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

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/108239

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106682649 A (CHENGDU RONGHAO ELECTRONIC INFORMATION TECHNOLOGY CO., LTD.) 17 May 2017 (2017-05-17) entire document	1-28
A	CN 106295717 A (NANJING UNIVERSITY OF SCIENCE AND TECHNOLOGY) 04 January 2017 (2017-01-04) entire document	1-28
A	CN 106682694 A (FUDAN UNIVERSITY) 17 May 2017 (2017-05-17) entire document	1-28
A	CN 106546892 A (PDSTARS ELECTRIC (SHANGHAI) CO., LTD.) 29 March 2017 (2017-03-29) entire document	1-28

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/108239

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	103325382	A	25 September 2013	None			
CN	106328121	A	11 January 2017	CN	106328121	B	27 June 2017
JP	05188932	A	30 July 1993	JP	H05188932	A	30 July 1993
CN	103761965	A	30 April 2014	CN	103761965	B	25 May 2016
CN	106682649	A	17 May 2017	None			
CN	106295717	A	04 January 2017	None			
CN	106682694	A	17 May 2017	None			
CN	106546892	A	29 March 2017	None			

国际检索报告

国际申请号

PCT/CN2018/108239

A. 主题的分类

G10L 25/51(2013.01)i; G10L 25/18(2013.01)i; G10L 25/24(2013.01)i; G10L 25/30(2013.01)i; G10L 25/45(2013.01)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 25/-; G10L 15/-

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))

CNPAT, WPI, EPOD00, CNKI: 乐器, 识别, 分类, 语音, 声音, 话音, 音频, 梅尔, 美尔, mel, mfcc, 深度, 卷积, 神经, 网络, 权重, 权值, music+, instrument?, recogni+, class+, speech+, voice?, sound?, audio?, dcnn, deep+, convolut+, convolv+, neural+, network+, weight?.

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 103325382 A (大连民族学院) 2013年 9月 25日 (2013 - 09 - 25) 摘要, 说明书第[0028]-[0041]段, 图1	1, 6-8, 13-15, 20-22, 27-28
Y	CN 103325382 A (大连民族学院) 2013年 9月 25日 (2013 - 09 - 25) 摘要, 说明书第[0028]-[0041]段, 图1	2-5, 9-12, 16-19, 23-26
Y	CN 106328121 A (南京理工大学) 2017年 1月 11日 (2017 - 01 - 11) 摘要, 说明书第[0005]-[0025]段	2-5, 9-12, 16-19, 23-26
X	JP 05188932 A (SONY CORP.) 1993年 7月 30日 (1993 - 07 - 30) 摘要, 说明书第[0008]-[0038]段, 图1	1, 8, 15, 22
A	CN 103761965 A (太原科技大学) 2014年 4月 30日 (2014 - 04 - 30) 全文	1-28
A	CN 106682649 A (成都容豪电子科技有限公司) 2017年 5月 17日 (2017 - 05 - 17) 全文	1-28
A	CN 106295717 A (南京理工大学) 2017年 1月 4日 (2017 - 01 - 04) 全文	1-28

☒ 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

“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

2019年 3月 8日

国际检索报告邮寄日期

2019年 3月 27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

受权官员

康丹丹

传真号 (86-10)62019451

电话号码 86-(10)-53962640

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	CN 106682694 A (复旦大学) 2017年 5月 17日 (2017 - 05 - 17) 全文	1-28
A	CN 106546892 A (华乘电气科技上海股份有限公司) 2017年 3月 29日 (2017 - 03 - 29) 全文	1-28

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/108239

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	103325382	A	2013年 9月 25日	无	
CN	106328121	A	2017年 1月 11日	CN 106328121	B 2017年 6月 27日
JP	05188932	A	1993年 7月 30日	JP H05188932	A 1993年 7月 30日
CN	103761965	A	2014年 4月 30日	CN 103761965	B 2016年 5月 25日
CN	106682649	A	2017年 5月 17日	无	
CN	106295717	A	2017年 1月 4日	无	
CN	106682694	A	2017年 5月 17日	无	
CN	106546892	A	2017年 3月 29日	无	