

(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
2022 年 4 月 14 日 (14.04.2022)



(10) 国际公布号
WO 2022/073285 A1

- (51) 国际专利分类号:
G10L 25/51 (2013.01) *G10L 25/30* (2013.01)
G10L 25/66 (2013.01) *G10L 25/24* (2013.01)
- (21) 国际申请号: PCT/CN2020/130136
- (22) 国际申请日: 2020 年 11 月 19 日 (19.11.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202011070732.X 2020年10月9日 (09.10.2020) CN
- (71) 申请人: 深圳大学(SHENZHEN UNIVERSITY) [CN/
CN]; 中国广东省深圳市南山区南海大道
3688号, Guangdong 518060 (CN)。
- (72) 发明人: 李坚强(LI, Jianqiang); 中国广东省深圳市南山区南海大道3688号, Guangdong 518060 (CN)。 陈杰(CHEN, Jie); 中国广东省深圳市南山区南海大道3688号, Guangdong 518060 (CN)。 明仲(MING, Zhong); 中国广东省深圳市南山区南海大道3688号, Guangdong 518060 (CN)。 梁中明(LIANG, Zhongming); 中国广东省深圳市南山区南海大道3688号, Guangdong 518060 (CN)。 王成(WANG, Cheng); 中国广东省深圳市南山区南海大道3688号, Guangdong 518060 (CN)。
- (74) 代理人: 深圳市君胜知识产权代理事务所(普通合伙)(JOHNSON INTELLECTUAL PROPERTY AGENCY(SHENZHEN)); 中国广东省深圳市南山区粤海街道高新区社区高新南七

(54) **Title:** LUNG SOUND CLASSIFICATION METHOD AND SYSTEM BASED ON KNOWLEDGE DISTILLATION, TERMINAL, AND STORAGE MEDIUM

(54) 发明名称: 基于知识蒸馏的肺音分类方法、系统、终端及存储介质

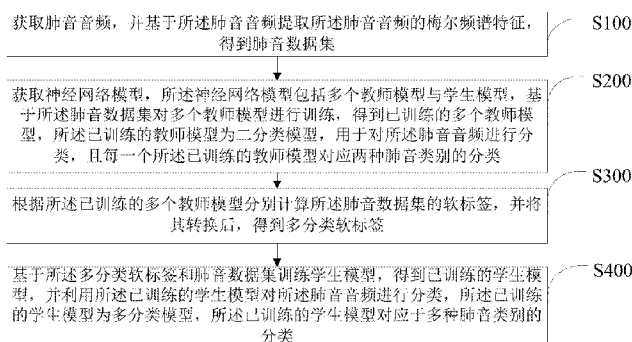


图 1

- S100 Obtain a lung sound audio, extract Mel frequency spectrum features of the lung sound audio on the basis of the lung sound audio, and obtain a lung sound data set
- S200 Obtain a neural network model, the neural network model comprising a plurality of teacher models and a student model; train the plurality of teacher models on the basis of the lung sound data set to obtain a plurality of trained teacher models, the plurality of trained teacher models being binary classification models and being used for classifying the lung sound audio, and each trained teacher model corresponding to two lung sound categories
- S300 Respectively calculate soft labels of the lung sound data set according to the plurality of trained teacher models, and perform conversion to obtain multi-classification soft labels
- S400 Train a student model on the basis of the multi-classification soft labels and the lung sound data set to obtain a trained student model, and classify the lung sound audio by utilizing the trained student model, the trained student model being a multi-classification model, and the trained student model corresponding to a plurality of lung sound categories

(57) **Abstract:** A lung sound classification method and system based on knowledge distillation, a terminal, and a storage medium. The method comprises: obtaining a lung sound audio, extracting Mel frequency spectrum features of the lung sound audio on the basis of the lung sound audio, and obtaining a lung sound data set (S100); training a plurality of teacher models on the basis of the lung sound data set to obtain a plurality of trained teacher models (S200); respectively calculating soft labels of the lung sound data set according to the plurality of trained teacher models, and performing conversion to obtain multi-classification soft labels (S300); and training a student

道 20 号深圳国家工程实验室大楼 A503,
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, IT, 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))。

model on the basis of the multi-classification soft labels and the lung sound data set to obtain a trained student model, and classifying the lung sound audio by utilizing the trained student model (S400). According to the method, on the basis of knowledge distillation, student model learning is guided by utilizing knowledge of a plurality of teacher models, so that multi-party lung sound audios are fully utilized, and the lung sound audios are accurately classified.

(57) 摘要: 一种基于知识蒸馏的肺音分类方法、系统、终端及存储介质, 方法包括: 获取肺音音频, 并基于肺音音频提取肺音音频的梅尔频谱特征, 得到肺音数据集 (S100); 基于肺音数据集对多个教师模型进行训练, 得到已训练的多个教师模型 (S200); 根据已训练的多个教师模型分别计算肺音数据集的软标签, 并经转换后, 得到多分类软标签 (S300); 基于多分类软标签和肺音数据集训练学生模型, 得到已训练的学生模型, 并利用已训练的学生模型对肺音音频进行分类 (S400)。该方法可以基于知识蒸馏, 通过利用多个教师模型的知识, 指导学生模型学习, 从而充分利用多方肺音音频, 并准确对肺音音频进行分类。

基于知识蒸馏的肺音分类方法、系统、终端及存储介质

技术领域

本发明涉及音频分类技术领域，尤其涉及一种基于知识蒸馏的肺音分类方法、系统、终端及存储介质。

背景技术

传统对于肺音的分析是利用传统的听诊器采用人工听诊的方式来实现，这可能由于经验不足和人耳对肺音的低频段敏感性低等因素，导致判断结果不够准确。与主观听诊相比，对肺音的计算机分析能够进行可重复的肺音量化近年来，已有很多学者专注于研究肺音分析和检测中的机器学习技术。机器学习技术应用于肺音分析上，使得肺音识别的准确率得到了提升。

由于肺部声音信号不稳定，现有技术中使用神经网络来对肺音进行分类时，基本都只能实现单一类型的分类，用作肺音的细类区分，仍存在一定的局限性。并且，由于听取肺音是可能出现各种各样不确定音和噪音，使得对肺音的分析不够准确，影响分类的准确性。

因此，现有技术还有待改进和提高。

发明内容

本发明要解决的技术问题在于，针对现有技术的上述缺陷，提供一种基于知识蒸馏的肺音分类方法、系统、终端及存储介质，旨在解决现有技术中用神经网络来对肺音进行分类时，基本都只能实现单一类型的分类，用作肺音的细类区分，仍存在一定的局限性，并且对肺音的分析不够准确，影响分类的准确性。

为了解决上述技术问题，本发明所采用的技术方案如下：

第一方面，本发明提供一种基于知识蒸馏的肺音分类方法，其中，所述方法包括：

获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集；

获取神经网络模型，所述神经网络模型包括多个教师模型与学生模型，基于所述肺音数据集对多个教师模型进行训练，得到已训练的多个教师模型，所述已训练的多个教师模型为二分类模型，用于对所述肺音音频进行分类，且每一个已训练的教师模型对应两种肺音类别的分类；

根据所述已训练的多个教师模型分别计算所述肺音数据集的软标签，并将其转换后，得到多分类软标签；

基于所述多分类软标签和肺音数据集训练所述学生模型，得到已训练的学生模型，并利用所述已训练的学生模型对所述肺音音频进行分类，所述已训练的学生模型为多分类模型，所述已训练的学生模型对应于多种肺音类别的分类。

在一种实现方式中，所述获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集，包括：

采集所述肺音音频，并对所述肺音音频进行音频数据增强处理，得到增强后的肺音音频；
对所述增强后的肺音音频进行切片，得到多个音频片段；

分别从每一个所述音频片段中提取所述梅尔频谱特征，得到所述肺音数据集。

在一种实现方式中，所述音频数据增强处理包括：时间变换处理、音高变换处理和 μ 律压缩处理。

在一种实现方式中，所述获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集之后，包括：

对每一个所述音频片段进行零均值归一化处理。

在一种实现方式中，所述基于所述肺音数据集对多个教师模型进行训练，得到已训练的多个教师模型，包括：

获取所述肺音数据集中的预测标签；

基于所述预测标签与真实标签，对多个教师模型进行训练，得到所述已训练多个的教师模型，所述真实标签为肺音类别。

在一种实现方式中，所述基于所述多分类软标签和肺音数据集训练所述学生模型，得到已训练的学生模型，并利用所述已训练的学生模型对所述肺音音频进行分类，包括：

获取所述多分类软标签；

基于所述多分类软标签与肺音数据集，对所述学生模型进行训练，得到所述已训练的学生模型，所述已训练的学生模型的软标签接近所述已训练的多个教师模型的软标签；

利用所述已训练的学生模型对所述肺音音频进行分类，得到所述肺音音频的类别。

在一种实现方式中，所述教师模型与所述学生模型均为一个 6 层的卷积神经网络模型。

第二方面，本发明还提供一种基于知识蒸馏的肺音分析系统，其中，所述系统包括：

音频特征获取单元，用于获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集；

教师模型训练单元，用于获取神经网络模型，所述神经网络模型包括多个教师模型与学生模型，基于所述肺音数据集对多个教师模型进行训练，得到已训练的多个教师模型，所述已训练的多个教师模型为二分类模型，用于对所述肺音音频进行分类，且每一个已训练的教师模型对应两种肺音类别的分类；

软标签拟合单元，用于根据所述已训练的教师模型分别计算所述肺音数据集的软标签，并将其转换后，得到多分类软标签；

学生模型训练单元，用于基于所述多分类软标签和肺音数据集训练所述学生模型，得到已训练的学生模型，并利用所述已训练的学生模型对所述肺音音频进行分类，所述已训练的学生模型为多分类模型，所述已训练的学生模型对应于多种肺音类别的分类。

第三方面，本发明还提供一种智能终端，其中，包括有存储器，以及一个或者一个以上的程序，其中一个或者一个以上程序存储于存储器中，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于执行上述方案中任意一项所述的方法。

第四方面，本发明还提供一种非临时性计算机可读存储介质，其中，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行上述方案中任意一项所述的方法。

有益效果：与现有技术相比，本发明提供了一种基于知识蒸馏的肺音分类方法，首先获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集；基于所述肺音数据集对多个教师模型进行训练，得到已训练的多个教师模型；根据所述已训练的多个教师模型分别计算所述肺音数据集的软标签，并将其转换后，得到多分类软标签；基于所述多分类软标签和肺音数据集训练所述学生模型，得到已训练的学生模型，并利用所述已训练的学生模型对所述肺音音频进行分类。本发明可以基于知识蒸馏，通过利用多个教师模型的知识，指导学生模型学习，从而充分利用多方肺音音频，并准确对所述肺音音频进行分类。

附图说明

图1为本发明实施例提供的基于知识蒸馏的肺音分类方法的实施例流程图。

图2为本发明实施例提供的基于知识蒸馏的肺音分类系统的原理框图。

图3为本发明实施例提供的智能终端的功能原理图。

具体实施方式

为使本发明的目的、技术方案及效果更加清楚、明确，以下参照附图并举实施例对本发明进一步详细说明。应当理解，此处所描述的具体实施例仅用以解释本发明，并不用于限定

本发明。

传统对于肺音的分析是利用传统的听诊器采用人工听诊的方式来实现，这可能由于经验不足和人耳对肺音的低频段敏感性低等因素，导致判断结果不够准确。与主观听诊相比，对肺音的计算机分析能够进行可重复的肺音量化近年来，已有很多学者专注于研究肺音分析和检测中的机器学习技术。机器学习技术应用于肺音分析上，使得肺音识别的准确率得到了提升。

由于肺部声音信号不稳定，现有技术中也有提出一种使用小波变换分析肺部声音信号以及使用人工神经网络（ANN）进行分类的方法，但是这种方法需要人工标记肺音的周期。异常肺音本身是一种非周期信号，其可能不会在整个记录中出现。此外，目前还开发出一种自动肺音分析仪，该工具在不需要额外的流量信号或标记呼吸周期的情况下，就能识别肺部声音特征，但是这种方法主要是区分正常肺音和异常肺音，用作肺音的细类区分，仍存在一定的局限性。听诊时可能出现的各种各样的不确定音和噪声使肺音检测不够准确。目前，研究人员提出了基于隐马尔科夫模型提出“异常呼吸置信度”检测的两步分类方法，但是该工作也只是关注单一类型的肺音。可见，现有技术中使用神经网络来对肺音进行分类时，基本都只能实现单一类型的分类，用作肺音的细类区分，仍存在一定的局限性。并且，由于听取肺音是可能出现各种各样不确定音和噪音，使得对肺音的分析不够准确，影响分类的准确性。

为解决现有技术中的问题，本实施例提供一种基于知识蒸馏的肺音分类方法，本实施例中的知识蒸馏，是可以将一个网络的知识转移到另一个网络，两个网络可以是同构或者异构。具体实施是，是先训练一个教师网络，然后使用这个教师网络的输出和数据的真实标签去训练学生网络。知识蒸馏可以用来将网络从大网络转化成一个小网络，并保留接近于大网络的性能；也可以将多个网络的学到的知识转移到一个网络中。因此，本实施例可基于知识蒸馏的原理，利用多个教师模型的知识，指导学生模型学习，从而充分利用多方肺音音频，并准确对所述肺音音频进行分类。具体如图1中所示。所述基于知识蒸馏的肺音分类方法具体包括如下步骤：

步骤 S100、获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集。

本实施例中的肺音音频可来自各大医院所采集到的音频数据，这样就可以利用多方数据来进行分析，以便准确地对肺音音频进行分类。在本实施例中，当获取到所述肺音音频后，对所述肺音音频进行音频数据增强处理，得到增强后的肺音音频。具体地，本实施例使用三种数据增强方式来数据增强方法：时间变换、音高变换和 μ 律压缩。时间变换旨在通过改变参

数因子而影响音频的采样速率。而音高变换就是以半音为单位，对音频的音调进行调整。 μ 律压缩是一种动态范围压缩算法，对于输入的信号 s ， $-1 \leq s \leq 1$ ，计算 μ 律压缩可以表述为以下公式：

$$y_s = \text{sgn}(s) \frac{\ln(1 + \text{abs}(s)^\mu)}{\ln(1 + \mu)}。$$

其中， $\text{sgn}(s)$ 表示符号函数， $\text{abs}(s)$ 表示取信号 s 的绝对值， μ 为压缩参数，一般取值为 2 的幂减 1。通过这三种数据增强方法来对数据进行增强，使得采集的肺音音频可以明显的增强，使得整个数据集中的样本数据量增加。

接着，本实施例从所述增强后的肺音音频中获取梅尔频谱特征，提取梅尔频谱特征的作用在于把音频信号中具有辨识性的成分提取出来，然后去除肺音音频的干扰因素，以提高识别的准确性。具体实施时，本实施例中提取梅尔频谱特征的流程主要是包括四个步骤。具体为：首先是先对肺音音频进行预加重、分帧和加窗（加强语音信号性能（信噪比，处理精度等）的一些预处理）。然后，对每一个短时分析窗，通过 FFT（Fast Fourier Transformation，快速傅氏变换）得到对应的频谱（即获得分布在时间轴上不同时间窗内的频谱）。接着将上面的频谱通过 Mel（梅尔）滤波器组得到 Mel 频谱（通过 Mel 频谱，将线形的自然频谱转换为体现人类听觉特性的 Mel 频谱）。当然，本实施例中对于梅尔频谱特征的提取可以基于一个开源的 python 工具包 librosa 来实现。

在一种实现方式中，本实施例可以对所述增强后的肺音音频进行切片，得到多个音频片段。然后分别从每一个所述音频片段中提取所述梅尔频谱特征，得到所述肺音数据集。具体地，由于原始采集的肺音音频是一段较长的时序序列，在 3 至 10 秒之间会有较为明显的变化，因此本实施例中可以按照预先设定好的时间窗口，将增强处理后的肺音音频分割为固定时长的音频片段，然后再对各个音频片段进行梅尔频谱特征，并构成所述数据集，通过所述肺音数据集进行模型的训练。在一种实现方式中，本实施例在得到所述肺音数据集后还对每一个所述音频片段进行零均值归一化处理，能够加快模型的优化过程，本实施例使用的零均值归一化的方法是音频和图像处理任务中比较常用的归一化方法。对于音频片段 s ，零均值归一化就是减去其均值 $\bar{s}$ 之后，再除以 s 的标准差 σ ，如以下公式所示：

$$s_z = \frac{s - \bar{s}}{\sigma}。$$

步骤 S200、获取神经网络模型，所述神经网络模型包括多个教师模型与学生模型，基于所述肺音数据集对多个教师模型进行训练，得到已训练的多个教师模型，所述已训练的多个

教师模型为二分类模型，用于对所述肺音音频进行分类，且每一个已训练的教师模型对应两种肺音类别的分类。

在本实施例中，本实施例首先获取神经网络模型，所述神经网络模型包括多个教师模型与学生模型，然后获取所述肺音数据集中的预测标签。接着基于所述预测标签与真实标签，对所述教师模型进行训练，得到已训练的多个教师模型，所述真实标签为肺音类别。在本实施例中，每一个已训练的教师模型为二分类模型，假设共用 M 种肺音类别，则就需要对应训练 M 个教师模型，以使得每一个已训练的教师模型可以分别对应两种肺音类别的分类。本实施例中的肺音数据集为 X^{T_i} ， X^{T_i} 数据对应的预测标签定义为 Y^{T_i} 。本实施例中，所述教师模型是一个 6 层的 CNN（Convolutional Neural Networks，卷积神经网络）模型，第 1 层是有 24 个滤波器的卷积层，卷积核大小为 (5, 5)，步幅为 (4, 2)。第 2 层同样是卷积层，包含 48 个滤波器，卷积核大小、步幅与第 1 层相同。第 3 层是一个最大池化层，其 pool size 为 (2, 2)。第 4 层卷积层的滤波器数量和卷积核大小与第 2 层相同。第 5 层全连接层有 64 个隐藏单元，最后一层为输出层。

在训练多个教师模型的过程中，每个教师模型分别对应一种肺音类别，拟合来自不同医院采集到的肺音音频数据，训练目标为最小化教师模型的预测标签和真实标签之间的交叉熵损失，对于教师模型 T_i ，其损失函数如下公式所示：

$$\text{Loss}T_i = -\frac{1}{|X^{T_i}|} \sum_{n=1}^{|X^{T_i}|} y_n^{T_i} \log f^{T_i}(x_n^{T_i}),$$

这里 $|X^{T_i}|$ 表示 X^{T_i} 中元素数量， $|X_n^{T_i}|$ 表示 X^{T_i} 中的一个数据点， $y_n^{T_i} \in \{0, 1\}$

是与 $x_n^{T_i}$ 相对应的类别标签， f^{T_i} 表示教师模型 T_i 的预测函数。

步骤 S300、根据所述已训练的多个教师模型分别计算所述肺音数据集的软标签，并将其转换后，得到多分类软标签。

对于每一个训练好的教师模型而言，其已经学习了某一种肺音类别的数据特征。先由教师模型计算学生模型待学习的数据集 X^S 的软标签，然后让学生模型拟合这些软标签，可以将教师模型的知识传授给学生模型，即实现知识蒸馏。在本实施例中，对于数据集 X^S 中的一个数据点 X_n^S ，通过教师模型 T_i 计算其软标签时，如以下公式所示：

$$q_j^{(i)}(x_n^S) = \frac{\exp(z_j^{(i)}(x_n^S)/t)}{\sum_{j=0}^1 \exp(z_j^{(i)}(x_n^S)/t)},$$

这里的 $z_j^{(i)}(\cdot)$ 表示教师模型 T_i 对应音频类别 $j \in \{0,1\}$ 时 logits 层的输出, t

为温控超参数, t 取值越大, 软标签的分布越平滑。

由于每一个教师模型 T_i 都是一个二分类的模型, 其 logits 层的输出只有 2 个类别, 而对于另外的 $M-1$ 个类别, 本实施例采用取 0 值的方式, 以此将多个二分类模型的软标签转换为

一个多分类的软标签 $\pi_j(x_n^S, y_n^S)$, 从而与学生模型的类别维度统一, y_n^S 是与 x_n^S

相对应的类别标签。 $\pi_j(x_n^S, y_n^S)$ 的计算如以下公式所示:

$$\pi_j(x_n^S, y_n^S) = \begin{cases} q_0^{(\xi(y_n^S))}(x_n^S), & j = 0 \\ q_1^{(\xi(y_n^S))}(x_n^S), & j = \xi(y_n^S) \\ 0, & \text{otherwise} \end{cases},$$

其中, $j \in \{0, \dots, M\}$, 对于任意一个 $y_n^S \in \{0, \dots, M\}$, 如果 $y_n^S = 0$, 定义 0

为正常类别, 那么函数 $\xi(y_n^S)$ 的返回值为集合 $\{1, \dots, M\}$ 中任意一个元素, 且满足每个

元素被选中的概率相等; 否则, 函数 $\xi(y_n^S)$ 直接返回 y_n^S 。这样通过训练后的所述学

生模型就可以对肺音音频进行分类，从而得到肺音音频对应的类别。

步骤 S400、基于所述多分类软标签和肺音数据集训练学生模型，得到已训练的学生模型，并利用所述已训练的学生模型对所述肺音音频进行分类，所述已训练的学生模型为多分类模型，所述已训练的学生模型对应于多种肺音类别的分类。

本实施例中获取所述多分类软标签：基于所述多分类软标签与肺音数据集，对所述学生模型进行训练，得到所述已训练的学生模型，所述已训练的学生模型的软标签接近所述已训练的教师模型的软标签；利用所述已训练的学生模型对所述肺音音频进行分类，得到所述肺音音频的类别。

在本实施例中，学生模型对于待学习的数据集 X^S 对应的标签定义为 Y^S 。所述学生模型是在教师模型的基础上进行简化的一个模型。该学生模型共有 6 层，第 1 层是有 6 个滤波器的卷积层，卷积核大小为(5, 5)，步幅为(4, 2)。第 2 层同样是卷积层，包含 18 个滤波器，卷积核大小、步幅与第 1 层相同。第 3 层是一个最大池化层，其 pool size 为(2, 2)。第 4 卷积层有 24 个滤波器，卷积核大小与第 2 层相同。第 5 层全连接层有 32 个隐藏单元，然后连接输出层。

在本实施例中，学生模型的软标签计算与教师模型类似，区别在于学生模型需要考虑 $M+1$ 个类别，如以下公式所示：

$$p_j(x_n^S) = \frac{\exp(g_j(x_n^S)/t)}{\sum_{j=0}^M \exp(g_j(x_n^S)/t)},$$

这里的 $g_j(\cdot)$ 表示学生模型对应类别 $j \in \{0, \dots, M\}$ 时 logits 层的输出，0 表示正常的类别。

除此之外，本实施例的学生模型的学习目标和教师模型也不同，学生模型一方面需要使得预测 $f^S(X_n^S)$ 尽量接近真实标签 y_n^S ，另外一方面需要使得自身的软标签 $p_j(X_n^S)$ 接近教师模型的软标签 $\pi_j(X_n^S, y_n^S)$ 。与真实标签相比，软标签 $\pi_j(X_n^S, y_n^S)$ 具有更为丰富的信息，其包含样本中类与类之间的关联，这些关联由复杂的教师模型学习得到。学生

模型由于自身结构比较简单，对样本的学习能力比教师模型弱，但如果在训练过程中，让学生模型拟合教师模型学习到的关联信息，能够对自身的学习起到辅助作用。

但是，由于教师模型也是存在一定的出错概率，与单纯的向教师模型学习相比，添加真实标签就可以减少学生模型学习错误知识的可能性，学生模型的损失如以下公式所示：

$$\text{Loss}_S = -\frac{1}{|X^S|} \sum_{n=1}^{|X^S|} \left\{ \lambda y_n^S \log f^S(x_n^S) + (1 - \lambda) \sum_{j=0}^M \pi_j(x_n^S, y_n^S) \log(p_j(x_n^S)) \right\}。$$

这里的 $|X^S|$ 表示 X^S 中元素数量， f^S 表示 S 的预测函数。 $\lambda \in [0,1]$ 是一个常数，用于学生模型选择侧重真实标签，还是侧重软标签。

在相同实验条件下，现有技术中的模型准确率为 98%，参数量为 0.28M，预测 8000 条数据的时耗为 3.12 秒，而本实施例中的学生模型准确率为 95%，参数量为 0.06M，预测同样的数据时耗为 2.48 秒。本方案的模型虽然在准确率上比现有技术中的模型低 0.03，但是参数量比其减少了 79%，完成相同任务比其节省了 20% 时间。可见，本发明可以通过多个教师模型的知识蒸馏，指导学生模型学习，从而充分利用多方肺音音频，并准确对所述肺音音频进行分类。

基于上述实施例，本发明还提供一种基于知识蒸馏的肺音分析系统，如图 2 中所示，所述系统包括：音频特征获取单元 10、教师模型训练单元 20、软标签拟合单元 30 以及学生模型训练单元 40。

具体地，所述音频特征获取单元 10，用于获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集。

所述教师模型训练单元 20，用于获取神经网络模型，所述神经网络模型包括多个教师模型与学生模型，基于所述肺音数据集对多个教师模型进行训练，得到已训练的多个教师模型，所述已训练的多个教师模型为二分类模型，用于对所述肺音音频进行分类，且每一个已训练的教师模型对应两种肺音类别的分类。

所述软标签拟合单元 30，用于根据所述已训练的多个教师模型分别计算所述肺音数据集的软标签，并将其转换后，得到多分类软标签。

所述学生模型训练单元 40，用于基于所述多分类软标签和肺音数据集训练学生模型，得到已训练的学生模型，并利用所述已训练的学生模型对所述肺音音频进行分类，所述已训练的学生模型为多分类模型，所述已训练的学生模型对应于多种肺音类别的分类。

基于上述实施例，本发明还提供了一种智能终端，其原理框图可以如图3所示。该智能终端包括通过系统总线连接的处理器、存储器、网络接口、显示屏、温度传感器。其中，该智能终端的处理器用于提供计算和控制能力。该智能终端的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该智能终端的网络接口用于与外部的智能终端通过网络连接通信。该计算机程序被处理器执行时以实现一种基于知识蒸馏的肺音分析方法。该智能终端的显示屏可以是液晶显示屏或者电子墨水显示屏，该智能终端的温度传感器是预先在智能终端内部设置，用于检测内部设备的运行温度。

本领域技术人员可以理解，图3中示出的原理框图，仅仅是与本发明方案相关的部分结构的框图，并不构成对本发明方案所应用于其上的智能终端的限定，具体的智能终端可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

在一个实施例中，提供了一种智能终端，包括有存储器，以及一个或者一个以上的程序，其中一个或者一个以上程序存储于存储器中，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行以下操作的指令：

获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集；

获取神经网络模型，所述神经网络模型包括多个教师模型与学生模型，基于所述肺音数据集对多个教师模型进行训练，得到已训练的多个教师模型，所述已训练的多个教师模型为二分类模型，用于对所述肺音音频进行分类，且每一个已训练的教师模型对应两种肺音类别的分类；

根据所述已训练的多个教师模型分别计算所述肺音数据集的软标签，并将其转换后，得到多分类软标签；

基于所述多分类软标签和肺音数据集训练所述学生模型，得到已训练的学生模型，并利用所述已训练的学生模型对所述肺音音频进行分类，所述已训练的学生模型为多分类模型，所述已训练的学生模型对应于多种肺音类别的分类。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一非易失性计算机可读存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本发明所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、

电可编程 ROM (EPROM)、电可擦除可编程 ROM (EEPROM) 或闪存。易失性存储器可包括随机存取存储器 (RAM) 或者外部高速缓冲存储器。作为说明而非局限, RAM 以多种形式可得, 诸如静态 RAM (SRAM)、动态 RAM (DRAM)、同步 DRAM (SDRAM)、双数据率 SDRAM (DDRSDRAM)、增强型 SDRAM (ESDRAM)、同步链路 (Synchlink) DRAM (SLDRAM)、存储器总线 (Rambus) 直接 RAM (RDRAM)、直接存储器总线动态 RAM (DRDRAM)、以及存储器总线动态 RAM (RDRAM) 等。

综上, 本发明公开了一种基于知识蒸馏的肺音分类方法、系统、终端及存储介质, 所述方法包括: 获取肺音音频, 并基于所述肺音音频提取所述肺音音频的梅尔频谱特征, 得到肺音数据集; 基于所述肺音数据集对多个教师模型进行训练, 得到已训练的多个教师模型; 根据所述已训练的多个教师模型分别计算所述肺音数据集的软标签, 并将其转换后, 得到多分类软标签; 基于所述多分类软标签和肺音数据集训练所述学生模型, 得到已训练的多个学生模型, 并利用所述已训练的学生模型对所述肺音音频进行分类。本发明可以基于知识蒸馏, 通过利用多个教师模型的知识, 指导学生模型学习, 从而充分利用多方肺音音频, 并准确对所述肺音音频进行分类。

最后应说明的是: 以上实施例仅用以说明本发明的技术方案, 而非对其限制; 尽管参照前述实施例对本发明进行了详细的说明, 本领域的普通技术人员应当理解: 其依然可以对前述各实施例所记载的技术方案进行修改, 或者对其中部分技术特征进行等同替换; 而这些修改或者替换, 并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

1. 一种基于知识蒸馏的肺音分类方法，其特征在于，所述方法包括：

获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集；

获取神经网络模型，所述神经网络模型包括多个教师模型与学生模型，基于所述肺音数据集对多个教师模型进行训练，得到已训练的多个教师模型，所述已训练的多个教师模型为二分类模型，用于对所述肺音音频进行分类，且每一个已训练的教师模型对应两种肺音类别的分类；

根据所述已训练的多个教师模型分别计算所述肺音数据集的软标签，并将其转换后，得到多分类软标签；

基于所述多分类软标签和肺音数据集训练所述学生模型，得到已训练的学生模型，并利用所述已训练的学生模型对所述肺音音频进行分类，所述已训练的学生模型为多分类模型，所述已训练的学生模型对应于多种肺音类别的分类。

2. 根据权利要求1所述的基于知识蒸馏的肺音分类方法，其特征在于，所述获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集，包括：

采集所述肺音音频，并对所述肺音音频进行音频数据增强处理，得到增强后的肺音音频；
对所述增强后的肺音音频进行切片，得到多个音频片段；

分别从每一个所述音频片段中提取所述梅尔频谱特征，得到所述肺音数据集。

3. 根据权利要求2所述的基于知识蒸馏的肺音分类方法，其特征在于，所述音频数据增强处理包括：时间变换处理、音高变换处理和 μ 律压缩处理。

4. 根据权利要求2所述的基于知识蒸馏的肺音分类方法，其特征在于，所述获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集之后，包括：
对每一个所述音频片段进行零均值归一化处理。

5. 根据权利要求1所述的基于知识蒸馏的肺音分类方法，其特征在于，所述基于所述肺音数据集对多个教师模型进行训练，得到已训练的多个教师模型，包括：

获取所述肺音数据集中的预测标签，所述预测标签为所述肺音数据集中的梅尔频谱特征与预测音频类别的对应关系；

基于所述预测标签与真实标签，对多个教师模型进行训练，得到所述已训练的多个教师模型，所述真实标签为肺音类别。

6. 根据权利要求1所述的基于知识蒸馏的肺音分类方法，其特征在于，所述基于所述多分类软标签和肺音数据集训练所述学生模型，得到已训练的学生模型，并利用所述已训练的

学生模型对所述肺音音频进行分类，包括：

获取所述多分类软标签；

基于所述多分类软标签与肺音数据集，对所述学生模型进行训练，得到所述已训练的学生模型，所述已训练的学生模型的软标签接近所述已训练的多个教师模型的软标签；

利用所述已训练的学生模型对所述肺音音频进行分类，得到所述肺音音频的类别。

7. 根据权利要求1所述的基于知识蒸馏的肺音分类方法，其特征在于，所述教师模型与所述学生模型均为一个6层的卷积神经网络模型。

8. 一种基于知识蒸馏的肺音分析系统，其特征在于，所述系统包括：

音频特征获取单元，用于获取肺音音频，并基于所述肺音音频提取所述肺音音频的梅尔频谱特征，得到肺音数据集；

教师模型训练单元，用于获取神经网络模型，所述神经网络模型包括多个教师模型与学生模型，基于所述肺音数据集对多个教师模型进行训练，得到已训练的多个教师模型，所述已训练的多个教师模型为二分类模型，用于对所述肺音音频进行分类，且每一个已训练的教师模型对应两种肺音类别的分类；

软标签拟合单元，用于根据所述已训练的多个教师模型分别计算所述肺音数据集的软标签，并将其转换后，得到多分类软标签；

学生模型训练单元，用于基于所述多分类软标签和肺音数据集训练所述学生模型，得到已训练的学生模型，并利用所述已训练的学生模型对所述肺音音频进行分类，所述学生模型为多分类模型，所述已训练的学生模型对应于多种肺音类别的分类。

9. 一种智能终端，其特征在于，包括有存储器，以及一个或者一个以上的程序，其中一个或者一个以上程序存储于存储器中，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于执行如权利要求1-7中任意一项所述的方法。

10. 一种非临时性计算机可读存储介质，其特征在于，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行如权利要求1-7中任意一项所述的方法。

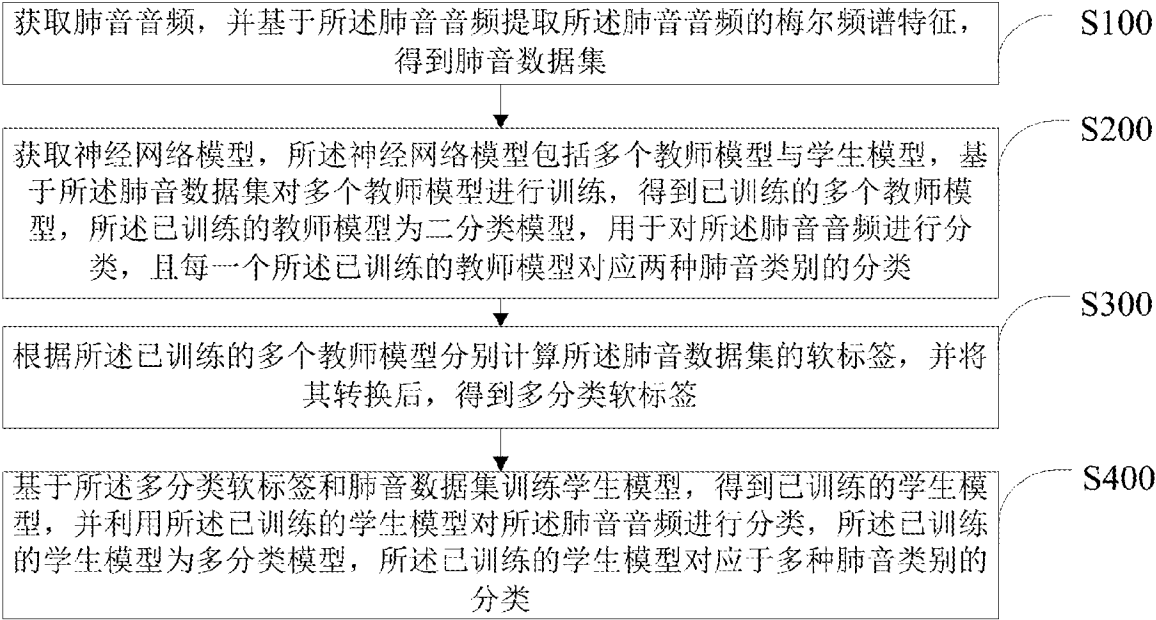


图 1

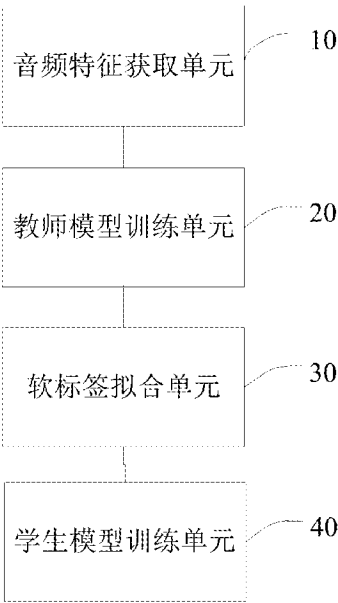


图 2

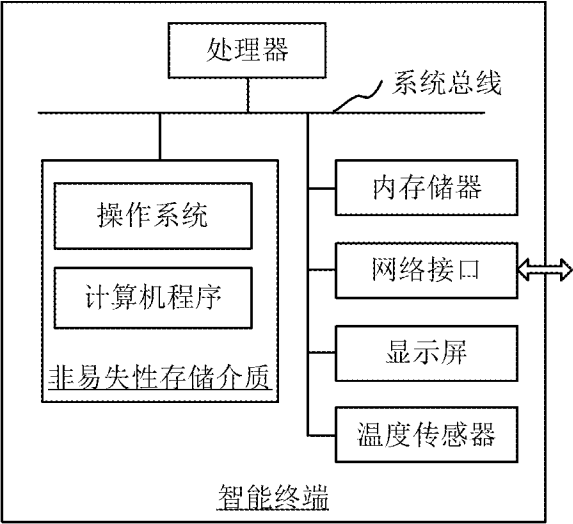


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/130136

A. CLASSIFICATION OF SUBJECT MATTER

G10L 25/51(2013.01)i; G10L 25/66(2013.01)i; G10L 25/30(2013.01)i; G10L 25/24(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

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, CNKI: 肺音, 肺病, 分类, 分析, 集成知识蒸馏, 神经网络, CNN, DNN, RNN, LSTM, lung, sound, disease, Artificial Neural Networks

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	李坚强 等 (LI, Jianqiang et al.). "基于集成知识蒸馏的肺病分析 (Lung Disease Analysis Using Ensemble Knowledge Distillation)" 控制理论与应用 (Control Theory & Applications), 27 September 2020 (2020-09-27), ISSN: 1000-8152, pages 1-7 abstract, sections 2-4, 5.3	1-10
PX	CN 111933185 A (SHENZHEN UNIVERSITY) 13 November 2020 (2020-11-13) claims 1-10, description paragraphs [0001]-[0051], figures 1-3	1-10
A	CN 110970042 A (SUZHOU MEINUOAI MEDICAL TECHNOLOGY CO., LTD.) 07 April 2020 (2020-04-07) entire document	1-10
A	CN 110705624 A (GUANGDONG UNIVERSITY OF TECHNOLOGY) 17 January 2020 (2020-01-17) entire document	1-10
A	CN 107818366 A (CHENGDU LICHUANG KUNLUN NETWORK TECHNOLOGY CO., LTD.) 20 March 2018 (2018-03-20) 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

05 July 2021

Date of mailing of the international search report

14 July 2021

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/CN2020/130136**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106022258 A (CHENGDU JISEN TECHNOLOGY CO., LTD.) 12 October 2016 (2016-10-12) entire document	1-10
A	US 2011098591 A1 (TECHNION RES. & DEV. FOUNDATION) 28 April 2011 (2011-04-28) entire document	1-10

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/130136

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111933185	A	13 November 2020	None			
CN	110970042	A	07 April 2020	None			
CN	110705624	A	17 January 2020	CN	110705624	B	16 March 2021
CN	107818366	A	20 March 2018	None			
CN	106022258	A	12 October 2016	None			
US	2011098591	A1	28 April 2011	WO	2009144725	A1	03 December 2009
				EP	2281193	A1	09 February 2011
				EP	2281193	B1	15 August 2012
				US	8366630	B2	05 February 2013

国际检索报告

国际申请号

PCT/CN2020/130136

A. 主题的分类

G10L 25/51(2013.01)i; G10L 25/66(2013.01)i; G10L 25/30(2013.01)i; G10L 25/24(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

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, CNKI: 肺音, 肺病, 分类, 分析, 集成知识蒸馏, 神经网络, CNN, DNN, RNN, LSTM, lung, sound, disease, Artificial Neural Networks

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	李坚强等. “基于集成知识蒸馏的肺病分析” 控制理论与应用, 2020年 9月 27日 (2020 - 09 - 27), ISSN: 1000-8152, 第1-7页 摘要、第2-4、5.3节	1-10
PX	CN 111933185 A (深圳大学) 2020年 11月 13日 (2020 - 11 - 13) 权利要求1-10、说明书第[0001]-[0051]段, 附图1-3	1-10
A	CN 110970042 A (苏州美糯爱医疗科技有限公司) 2020年 4月 7日 (2020 - 04 - 07) 全文	1-10
A	CN 110705624 A (广东工业大学) 2020年 1月 17日 (2020 - 01 - 17) 全文	1-10
A	CN 107818366 A (成都力创昆仑网络科技有限公司) 2018年 3月 20日 (2018 - 03 - 20) 全文	1-10
A	CN 106022258 A (成都济森科技有限公司) 2016年 10月 12日 (2016 - 10 - 12) 全文	1-10
A	US 2011098591 A1 (TECHNION RES. & DEV. FOUNDATION) 2011年 4月 28日 (2011 - 04 - 28) 全文	1-10

☐ 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

2021年 7月 5日

国际检索报告邮寄日期

2021年 7月 14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

授权官员

吴静

传真号 (86-10)62019451

电话号码 86-(010)-62085705

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/130136

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	111933185	A	2020年 11月 13日	无			
CN	110970042	A	2020年 4月 7日	无			
CN	110705624	A	2020年 1月 17日	CN	110705624	B	2021年 3月 16日
CN	107818366	A	2018年 3月 20日	无			
CN	106022258	A	2016年 10月 12日	无			
US	2011098591	A1	2011年 4月 28日	W0	2009144725	A1	2009年 12月 3日
				EP	2281193	A1	2011年 2月 9日
				EP	2281193	B1	2012年 8月 15日
				US	8366630	B2	2013年 2月 5日