



(12)发明专利申请

(10)申请公布号 CN 110890102 A

(43)申请公布日 2020.03.17

(21)申请号 201910844907.9

G06N 3/04(2006.01)

(22)申请日 2019.09.07

(71)申请人 创新奇智(重庆)科技有限公司

地址 400039 重庆市九龙坡区火炬大道69
号重庆启迪科技园9号楼201室

(72)发明人 张发恩 王忠强 唐永亮 黄家水

(74)专利代理机构 北京中南长风知识产权代理
事务所(普通合伙) 11674

代理人 郑海

(51)Int.Cl.

G10L 25/30(2013.01)

G10L 25/51(2013.01)

G10L 25/21(2013.01)

G10L 25/18(2013.01)

G06N 3/08(2006.01)

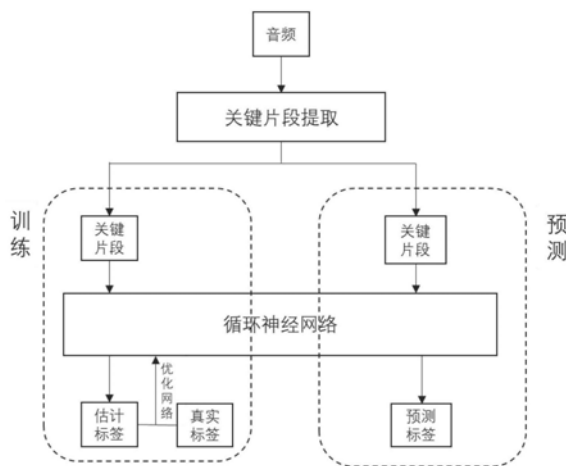
权利要求书1页 说明书5页 附图4页

(54)发明名称

一种基于RNN声纹识别的发动机缺陷检测算法

(57)摘要

本发明公开了智能质检领域的一种基于RNN声纹识别的发动机缺陷检测算法,包括以下具体步骤:S1:对所有录制的音频进行分割,从中提取出发电机加油高速运转时的关键片段;S2:对提取出的关键片段进行标注,以构建训练集;S3:搭建深度学习网络模型并进行训练;S4:使用训练好的网络模型对未知发电机的声音进行检测,结合发动机声音关键片段提取算法极大地减小了需要处理的数据量,提高了整个方法的准确度;深度学习的异常声音检测算法通过大量已知标签的样本即可自动学习出正常及各类异常声音的特点,而无需人工对声音的特点进行分析,可以处理变长序列输入的循环神经网络模型避免了传统使用卷积神经网络对音频进行处理的问题。



1. 一种基于RNN声纹识别的发动机缺陷检测算法,其特征在于:包括以下具体步骤:

S1:对所有录制的音频进行分割,从中提取出发电机加油高速运转时的关键片段;

S2:对提取出的关键片段进行标注,以构建训练集;

S3:搭建深度学习网络模型并进行训练;

S4:使用训练好的网络模型对未知发电机的声音进行检测。

2. 根据权利要求1所述的一种基于RNN声纹识别的发动机缺陷检测算法,其特征在于:所述步骤S1的具体步骤如下:

S1.1:将录制的原始音频生成二维声谱图;

S1.2:通过观察声音的二维声谱图,从中找出高频部分,如果某一时间的高频声音含量低于阈值,则将其分割并去除;剩余片段就是含高频声音的片段,即发动机高速运转时的声音关键片段。

3. 根据权利要求2所述的一种基于RNN声纹识别的发动机缺陷检测算法,其特征在于:所述步骤S1.1的具体步骤为:通过获取录制的原始音频中的一维振幅信息,使用长度 $N=2048$,滑动距离为512的窗口函数为 $w(n)=0.5\left(1-\cos\frac{2\pi n}{N-1}\right)$ 将其划分为若干帧;并对每个长

度为2048的帧执行离散傅立叶变换: $x_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} (k=0, \dots, N-1)$ 为每帧计算出长度为

2048维的特征向量,将所有帧的特征向量堆叠,得到大小为 $2048 \times n$ 的二维声谱图,其中, n 为帧的数。

4. 根据权利要求3所述的一种基于RNN声纹识别的发动机缺陷检测算法,其特征在于:所述步骤S1.2的具体步骤为:将2048维的频率特征向量的前800维作为低频信号,其余部分

作为高频信号,则每一帧的高频信号能量为: $E_{high} = \sum_{k=800}^{2048} X_k^2$ 将高频能量高于阈值的连续帧

组合,得到若干音频片段,去除其中持续时间较短的片段,剩余的片段被用于进行接下来的操作。

5. 根据权利要求1所述的一种基于RNN声纹识别的发动机缺陷检测算法,其特征在于:所述步骤S2中,标注方式采用逐一进行标注。

6. 根据权利要求4所述的一种基于RNN声纹识别的发动机缺陷检测算法,其特征在于:所述步骤S3中,深度学习网络模型采用长短期记忆网络LSTM,具体实现步骤如下:

S3.1:以长度不定、高度为2048的声谱图作为输入,寻找并学习出序列在时间上的关系;

S3.2:使用两层LSTM网络提取输入序列的特征,使用两个全连接层DENSE进行分类;

S3.3:输出该音频的分类,为一个长度为 $M+1$ 的向量,其中 M 为异常的种类,该向量中的值位于0到1之间,分别表示了该段音频为正常或具有某种异常的概率,即是否存在异常以及异常类型。

7. 根据权利要求6所述的一种基于RNN声纹识别的发动机缺陷检测算法,其特征在于:所述步骤S4中,检测后若所有片段均识别为正常,则认为该发动机不存在缺陷;如果其中有片段识别为异常,则认为该发动机存在缺陷。

一种基于RNN声纹识别的发动机缺陷检测算法

技术领域

[0001] 本发明涉及智能质检技术领域,具体为一种基于RNN声纹识别的发动机缺陷检测算法。

背景技术

[0002] 目前,基于深度神经网络的算法在智能质检领域得到越来越多的应用。但是,现有算法大多是对零件的外观进行检测,其处理对象为图像,使用的网络结构为卷积神经网络。然而,对于无法从外观上识别缺陷的产品,如发动机等,则不能使用上述基于图像的方法。在实际生产过程中,工人可以通过识别发动机运转时的异常声音以进行缺陷检测,但是在计算机领域,通过音频检测产品缺陷的技术仍然为空白。

[0003] 现有的大部分异常声音检测算法,都是通过人工识别出异常声音的特点,并总结出一套算法流程对未知声音进行判断。此类算法并不能自动学习出异常声音的特点,因此适用范围较小,且无法重复用于不同类型的声音检测。另外存在一种基于卷积神经网络的声音分类算法。它首先将每个音频规整为20帧长度相同的窗口;接着为每个窗口中提取出12维的梅尔倒谱系数及其一阶与二阶差分,共计36维的特征向量;之后将每个视频所有窗口的特征向量作为720*1或36*20的图像,分别使用一维与二维卷积神经网络对已知类别的声音进行训练;最后,对于未知类别的声音,使用训练完成的网络对其进行预测,即可得到该声音的类别。但是卷积神经网络对于音频来说并不完全适用,由于音频长度通常不固定,因此需要先对音频进行裁剪等预处理操作。

[0004] 总的来说,现有技术存在以下缺点:

[0005] 第一,现有算法使用的声音特征大多为梅尔倒谱系数,该特征根据人耳的感知特点对实际声音进行了处理,因此非常适用于处理语音,但是对于发动机等非语音类声音,该特征无法完全体现声音的特点;

[0006] 第二,大多异常声音检测算法通常由人工找出异常声音的特点,并设计一系列固定步骤进行判断,而不是自动学习出异常声音的特点,因此这些算法不具有通用性,对于不同的实际问题需要设计不同的算法;

[0007] 第三,一些算法采用卷积神经网络对音频进行处理,这样虽然可以自行学习出异常声音的特征并进行检测,但是卷积神经网络要求输入具有相同大小,而音频通常却无法达到这一要求;如果通过裁剪等方式对音频进行预处理,则可能丢失部分信息。

[0008] 基于此,本发明设计了一种基于RNN声纹识别的发动机缺陷检测算法,以针对组装完成的发动机中某些故障无法通过视觉观测进行识别这一问题进行解决。

发明内容

[0009] 本发明的目的在于提供一种基于RNN声纹识别的发动机缺陷检测算法,结合发动机声音关键片段提取算法、深度学习的异常声音检测算法和可以处理变长序列输入的循环神经网络模型,以解决上述背景技术中提出的问题。

[0010] 为实现上述目的,本发明提供如下技术方案:一种基于RNN声纹识别的发动机缺陷检测算法,包括以下具体步骤:

[0011] S1:对所有录制的音频进行分割,从中提取出发电机加油高速运转时的关键片段;

[0012] S2:对提取出的关键片段进行标注,以构建训练集;

[0013] S3:搭建深度学习网络模型并进行训练;

[0014] S4:使用训练好的网络模型对未知发电机的声音进行检测。

[0015] 优选的,所述步骤S1的具体步骤如下:

[0016] S1.1:将录制的原始音频生成二维声谱图;

[0017] S1.2:通过观察声音的二维声谱图,从中找出高频部分,如果某一时间的高频声音含量低于阈值,则将其分割并去除;剩余片段就是含高频声音的片段,即发动机高速运转时的声音关键片段。

[0018] 优选的,所述步骤S1.1的具体步骤为:通过获取录制的原始音频中的一维振幅信息,使用长度 $N=2048$,滑动距离为512的窗口函数为 $w(n)=0.5\left(1-\cos\frac{2\pi n}{N-1}\right)$ 将其划分为若

干帧;并对每个长度为2048的帧执行离散傅立叶变换: $x_k=\sum_{n=0}^{N-1}x_n e^{-\frac{2\pi i}{N}kn}$ ($k=0,\dots,N-1$) 为

每帧计算出长度为2048维的特征向量,将所有帧的特征向量堆叠,得到大小为 $2048\times n$ 的二维声谱图,其中, n 为帧的数。

[0019] 优选的,所述步骤S1.2的具体步骤为:将2048维的频率特征向量的前800维作为低频信号,其余部分作为高频信号,则每一帧的高频信号能量为: $E_{high}=\sum_{k=800}^{2048}X_k^2$ 将高频能量

高于阈值的连续帧组合,得到若干音频片段,去除其中持续时间较短的片段,剩余的片段被用于进行接下来的操作。

[0020] 优选的,所述步骤S2中,标注方式采用逐一进行标注。

[0021] 优选的,所述步骤S3中,深度学习网络模型采用长短期记忆网络LSTM,具体实现步骤如下:

[0022] S3.1:以长度不定、高度为2048的声谱图作为输入,寻找并学习出序列在时间上的关系;

[0023] S3.2:使用两层LSTM网络提取输入序列的特征,使用两个全连接层DENSE进行分类;

[0024] S3.3:输出该音频的分类,为一个长度为 $M+1$ 的向量,其中 M 为异常的种类,该向量中的值位于0到1之间,分别表示了该段音频为正常或具有某种异常的概率,即是否存在异常以及异常类型。

[0025] 优选的,所述步骤S4中,检测后若所有片段均识别为正常,则认为该发动机不存在缺陷;如果其中有片段识别为异常,则认为该发动机存在缺陷。

[0026] 与现有技术相比,本发明的有益效果是:

[0027] 1、采用发动机声音关键片段提取算法:利用发动机声音的特点,将高速运转时的关键片段分割出来,不仅极大地减小了需要处理的数据量,同时避免了大量无用片段的干

扰,提高了整个方法的准确度;

[0028] 2、采用基于深度学习的异常声音检测算法:采用了深度学习算法,通过大量已知标签的样本即可自动学习出正常及各类异常声音的特点,而无需人工对声音的特点进行分析,同时,本方法可以方便地用于处理其它类似问题,而无需进行大量修改;

[0029] 3、采用可以处理变长序列输入的循环神经网络模型:使用了可以接受变长序列作为输入的循环神经网络作为输入层,并在此基础上构建了一个深度神经网络模型,从而成功地避免了传统使用卷积神经网络对音频进行处理,需要所有的输入具有相同的规模,但是音频的长度通常不固定的问题。

附图说明

[0030] 为了更清楚地说明本发明实施例的技术方案,下面将对实施例描述所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0031] 图1为本发明原理框图;

[0032] 图2为本发明关键片段提取算法原理图;

[0033] 图3为本发明循环神经网络结构图;

[0034] 图4为本发明实施例检测算法原理图。

具体实施方式

[0035] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0036] 请参阅图1-3,本发明提供一种技术方案:一种基于RNN声纹识别的发动机缺陷检测算法,包括以下具体步骤:

[0037] S1:对所有录制的音频进行分割,从中提取出发电机加油高速运转时的关键片段;

[0038] 具体步骤如下:

[0039] S1.1:将录制的原始音频生成二维声谱图,通过获取录制的原始音频中的一维振幅信息,使用长度 $N=2048$,滑动距离为512的窗口函数为 $w(n)=0.5\left(1-\cos\frac{2\pi n}{N-1}\right)$ 将其划分为若干帧;并对每个长度为2048的帧执行离散傅立叶变换:

$$x_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad (k=0, \dots, N-1)$$
 为每帧计算出长度为2048维的特征向量,将所有帧的特征向量堆叠,得到大小为 $2048 \times n$ 的二维声谱图,其中, n 为帧的数;

[0040] S1.2:通过观察声音的二维声谱图,从中找出高频部分,如果某一时间的高频声音含量低于阈值,则将其分割并去除;剩余片段就是含高频声音的片段,即发动机高速运转时的声音关键片段,在图2的声谱图中,横轴为时间,纵轴为频率,对应点的值为该时间该频率

的强度，

[0041] 具体步骤为：将2048维的频率特征向量的前800维作为低频信号，其余部分作为高频信号，则每一帧的高频信号能量为： $E_{high} = \sum_{k=800}^{2048} X_k^2$ ，将高频能量高于阈值0.05的连续帧

组合，得到若干音频片段，去除其中持续时间较短小于20帧的片段，剩余的片段被用于进行接下来的操作。

[0042] S2：对提取出的关键片段进行标注，标注方式采用逐一进行标注，在训练集中需要保证有足够数量的正常与异常声音片段。如果一个音频为正常，则其所有片段可以标注为正常；但是如果一个音频为异常，并不是其中所有片段都为异常，因此需要逐一进行标注，以构建训练集；

[0043] S3：搭建深度学习网络模型并进行训练，采用深度学习网络模型采用长短期记忆网络LSTM，为循环神经网络RNN的一种，具体实现步骤如下：

[0044] S3.1：接受之前得到的以长度不定、高度为2048的声谱图作为输入，寻找并学习出序列在时间上的关系；

[0045] S3.2：使用两层LSTM网络提取输入序列的特征，使用两个全连接层DENSE进行分类；

[0046] S3.3：输出该音频的分类，输出该音频的分类，为一个长度为M+1的向量，其中M为异常的种类，该向量中的值位于0到1之间，分别表示了该段音频为正常或具有某种异常的概率，即是否存在异常以及异常类型；

[0047] S4：使用训练好的网络模型对未知发电机的声音进行检测，对于一段待检测的音频，同样需要先从中提取出关键片段，之后对这些片段使用训练好的模型进行异常检测，检测后若所有片段均识别为正常，则认为该发动机不存在缺陷；如果其中有片段识别为异常，则认为该发动机存在缺陷。

[0048] 具体工作原理如下所述：

[0049] 将所有音频首先通过关键片段提取算法提取出发动机高速运转时的片段；之后，在训练过程中，将训练样本送入神经网络，其估计结果与其真实标签共同用于对网络进行优化；在训练完成后，网络就可以用于预测一个音频是否存在异常，只需将片段输入训练好的网络即可得到预测标签。

[0050] 1、发动机声音关键片段提取算法的优点

[0051] 通常只有在发动机高速运转时才会出现明显的异常声音，因此一般仅对这些片段进行异常声音检测，而忽略其它片段。本算法利用发动机声音的特点，将高速运转时的关键片段分割出来，不仅极大地减小了需要处理的数据量，同时避免了大量无用片段的干扰，提高了整个方法的准确度。

[0052] 2、基于深度学习的异常声音检测算法

[0053] 在本方法中采用了深度学习算法，通过大量已知标签的样本即可自动学习出正常及各类异常声音的特点，而无需人工对声音的特点进行分析。同时，本方法可以方便地用于处理其它类似问题，而无需进行大量修改。

[0054] 3、可以处理变长序列输入的循环神经网络模型

[0055] 虽然使用卷积神经网络也可以对音频进行处理，但是它们需要所有的输入具有相

同的规模,但是音频的长度通常不固定,因此使用卷积神经网络十分不便。在本方法中,使用了可以接受变长序列作为输入的循环神经网络作为输入层,并在此基础上构建了一个深度神经网络模型,从而成功地避免了上述问题。

[0056] 实施例:

[0057] 如图4中的一段发动机音频,通过关键片段提取步骤可以从中提取出11段关键片段;这些片段经过已经训练好的神经网络进行预测,其中有一个片段判断为正时啸叫异常,其它片段判断为正常;因此最终判断该发动机具有正时啸叫异常。

[0058] 在本说明书的描述中,参考术语“一个实施例”、“示例”、“具体示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不一定指的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任何一个或多个实施例或示例中以合适的方式结合。

[0059] 以上公开的本发明优选实施例只是用于帮助阐述本发明。优选实施例并没有详尽叙述所有的细节,也不限制该发明仅为所述的具体实施方式。显然,根据本说明书的内容,可作很多的修改和变化。本说明书选取并具体描述这些实施例,是为了更好地解释本发明的原理和实际应用,从而使所属技术领域技术人员能很好地理解和利用本发明。本发明仅受权利要求书及其全部范围和等效物的限制。

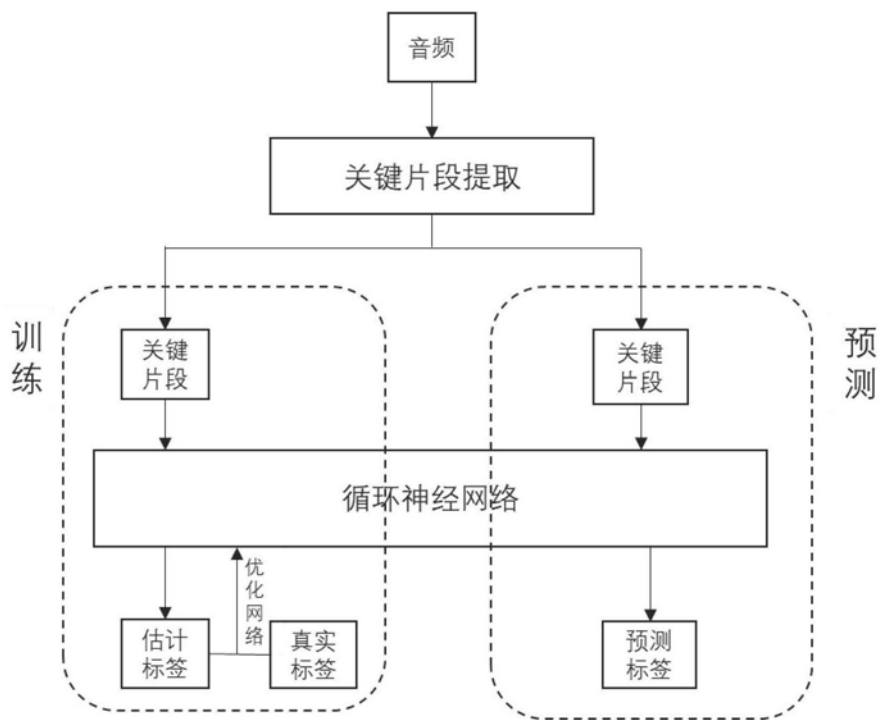


图1

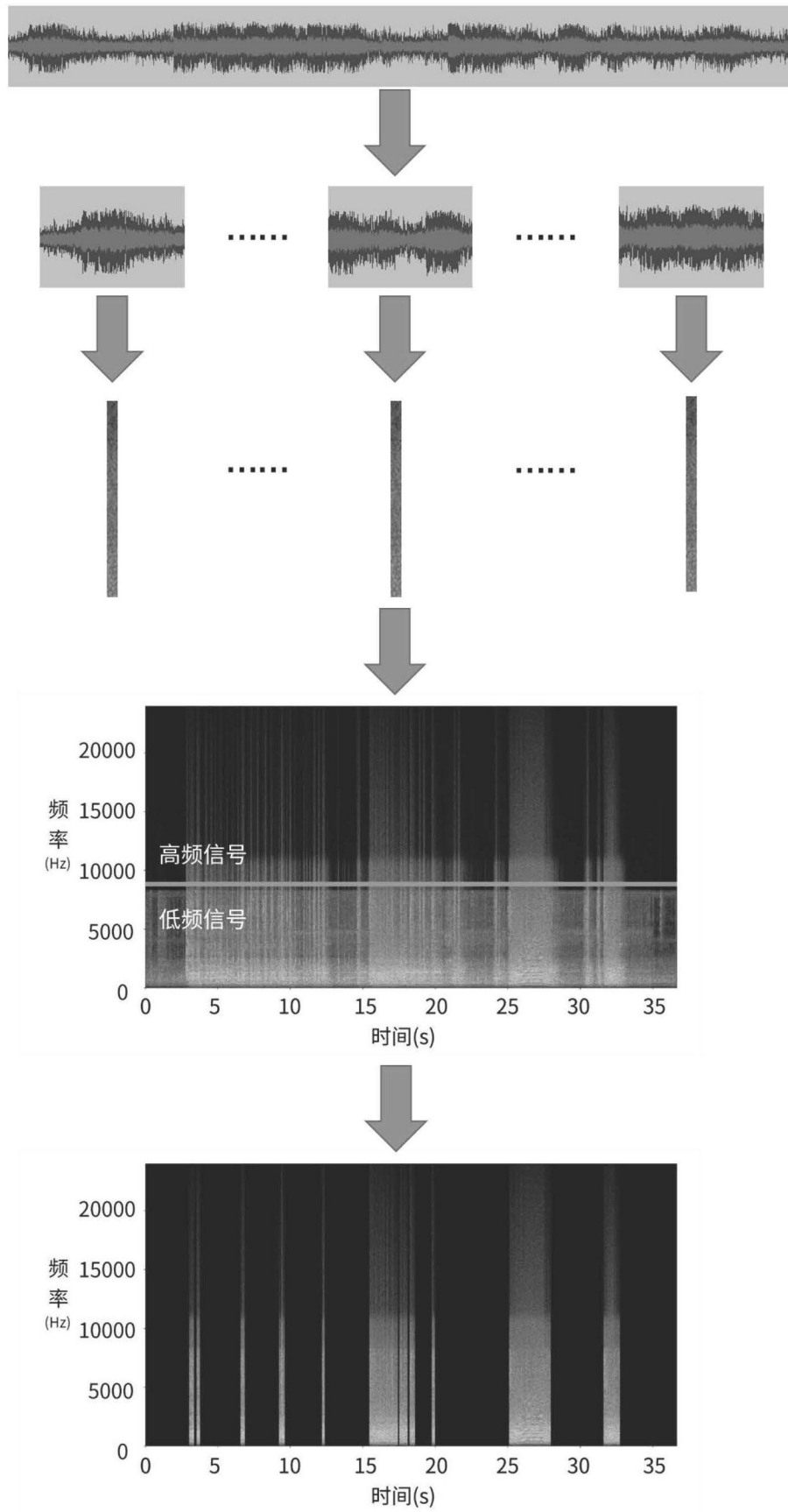


图2

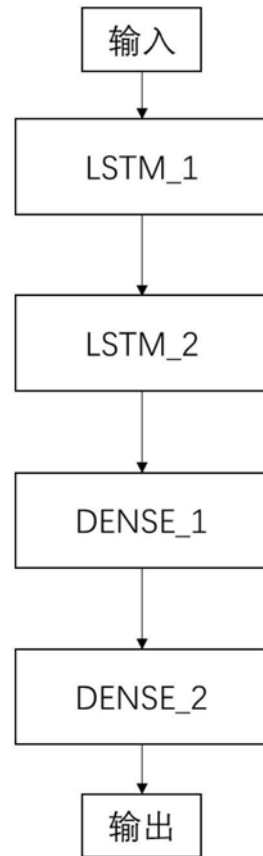


图3

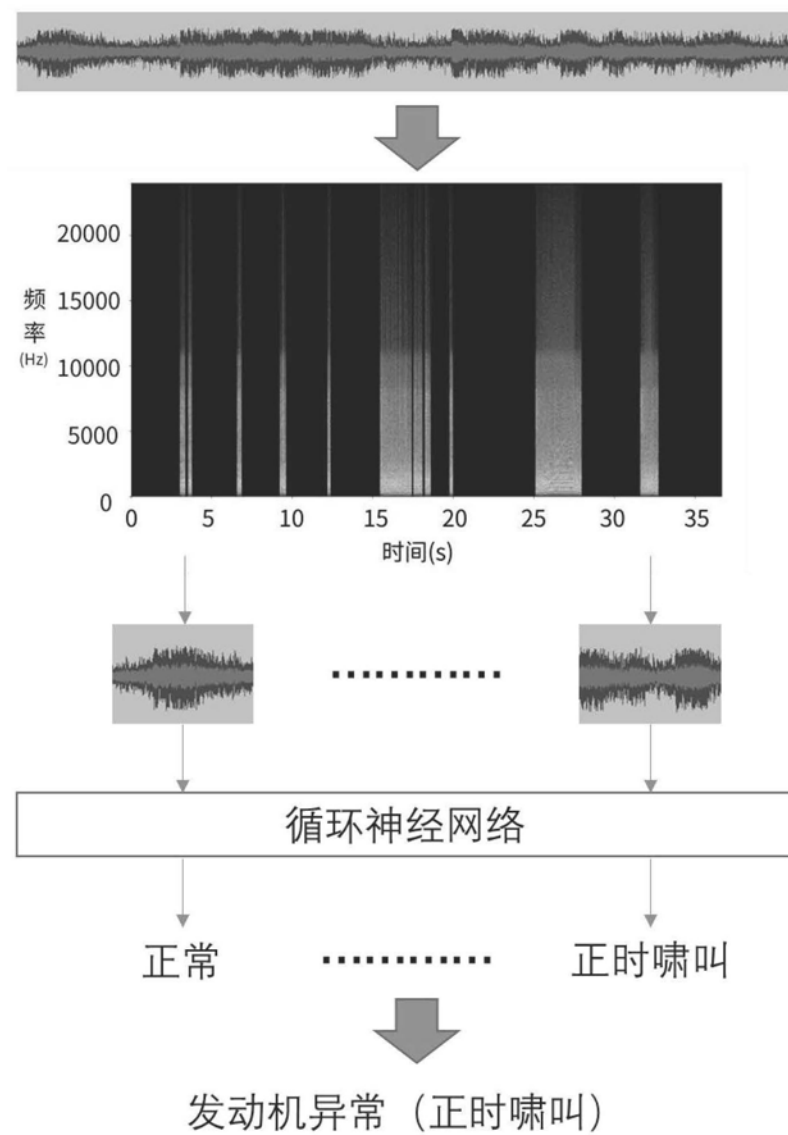


图4