(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 108564954 B
(45)授权公告日 2020.01.10

(21)申请号 201810225142.6

(22)申请日 2018.03.19

(65)同一申请的已公布的文献号
申请公布号 CN 108564954 A

(43)申请公布日 2018.09.21

(73)专利权人 平安科技(深圳)有限公司
地址 518000 广东省深圳市福田区八卦岭工业区平安大厦六楼

(72)发明人 赵峰 王健宗 肖京

(74)专利代理机构 深圳市沃德知识产权代理事务所(普通合伙) 44347
代理人 高杰 于志光

(51)Int.Cl.
G10L 17/00(2013.01)
G10L 17/02(2013.01)
G10L 17/04(2013.01)
G10L 25/24(2013.01)
G10L 25/30(2013.01)
G06N 3/04(2006.01)

(56)对比文件
CN 108564955 A,2018.09.21,
CN 107610707 A,2018.01.19,
CN 107808659 A,2018.03.16,
CN 107527620 A,2017.12.29,
CN 105261358 A,2016.01.20,
US 2006025995 A1,2006.02.02,

审查员 陈宬

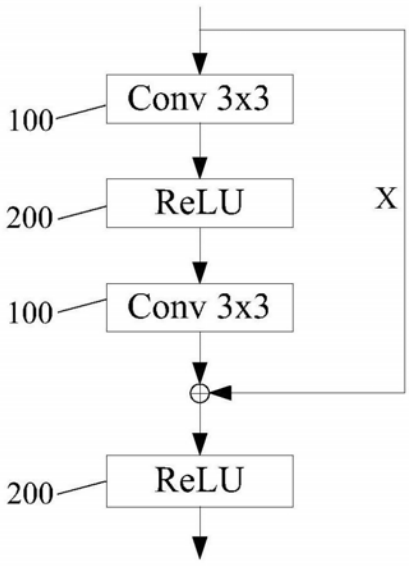
权利要求书3页 说明书10页 附图5页

(54)发明名称

深度神经网络模型、电子装置、身份验证方法和存储介质

(57)摘要

本发明公开一种深度神经网络模型、电子装置、身份验证方法和存储介质，该方法包括：在收到待进行身份验证的目标用户的当前语音数据后，获取待验证的身份对应的标准语音数据，将这两个标准语音数据分别分帧处理，以得到当前语音帧组和标准语音帧组；利用预设滤波器分别提取出两个语音帧组中的各个语音帧的预设类型声学特征；将提取出的预设类型声学特征输入预先训练好的预设结构深度神经网络模型，以得到当前语音数据和标准语音数据各自对应的预设长度的特征矢量；计算得到的两个特征矢量的余弦相似度，并根据计算出的余弦相似度大小确定身份验证结果。本发明技术方案提升了说话人身份验证的准确性。



CN 108564954 B

1.一种深度神经网络模型,其特征在于,该深度神经网络模型包括:

第一层结构:是由多层堆叠的有相同预设结构的神经网络层,每个预设结构的神经网络层包括:两个串联的CNN卷积层,两个修正线性单元ReLU,及一个将两个串联的CNN卷积层跨层直连的直连操作X,其中,各个ReLU与各个CNN卷积层一一对应,且各个ReLU分别串联在对应的CNN卷积层后,所述直连操作X将两个串联的CNN卷积层的第一个CNN卷积层的卷积操作的输入与第二个CNN卷积层的卷积操作的输出相加,并将结果送入到第二个CNN卷积层对应的ReLU操作中;

第二层结构:是平均层,此层的作用是沿时间轴向对矢量序列求平均值,它将第一层结构输出的二维矢量序列进行平均化;

第三层结构:是DNN全连接层;

第四层结构:是归一化层,此层将上一层的输入按照L2范数进行归一化,得到长度为1的归一化后的特征矢量;

第五层结构:是损失层,损失函数L的公式为:$L=\sum_{i=0}^{N}\max\left(S_i^{13}-S_i^{12}+\alpha,0\right)$,其中α是取值范围在0.05～0.2之间的常量,$S_i^{12}$代表属于同一说话人的两个特征矢量的余弦相似度,$S_i^{13}$代表不属于同一说话人的两个特征矢量的余弦相似度。

2.如权利要求1所述的深度神经网络模型,其特征在于,所述深度神经网络模型的训练过程为:

S1、获取预设数量语音数据样本,对各个语音数据样本分别标注代表对应的说话人身份的标签;

S2、分别对每个语音数据样本进行活动端点检测,将语音数据样本中非说话人的语音删除,得到预设数量的标准语音数据样本;

S3、将得到的标准语音数据样本的第一百分比作为训练集,第二百分比作为验证集,所述第一百分比与第二百分比的和小于等于100%;

S4、将所述训练集和验证集中的各个标准语音数据样本按照预设的分帧参数分别进行分帧处理,以获得每个标准语音数据样本对应的语音帧组,再利用预设滤波器分别提取出每个语音帧组中的各个语音帧的预设类型声学特征;

S5、将所述训练集中的各个语音帧组对应的预设类型声学特征划分成M批,分批输入所述深度神经网络模型中进行迭代训练,并在所述深度神经网络模型训练完成后,采用验证集对所述深度神经网络模型的准确率进行验证;

S6、若验证得到的准确率大于预设阈值,则模型训练结束;

S7、若验证得到的准确率小于或者等于预设阈值,则增加获取的语音数据样本的数量,并基于增加后的语音数据样本重新执行上述步骤S1-S5。

3.如权利要求2所述的深度神经网络模型,其特征在于,所述预设滤波器为梅尔滤波器,所述预设类型声学特征为梅尔频率倒谱系数MFCC。

4.如权利要求2所述的深度神经网络模型,其特征在于,所述深度神经网络模型迭代训练的过程包括:

根据模型的当前参数将当前输入每个语音帧组对应的预设类型声学特征转化为对应

的一个预设长度的特征矢量；

从各个特征矢量中进行随机选取以获得多个三元组，第i个三元组(x_{i1},x_{i2},x_{i3})由三个不同的特征矢量x_{i1}、x_{i2}和x_{i3}组成，其中，x_{i1}和x_{i2}对应同一个说话人，x_{i1}和x_{i3}对应不同的说话人，i为正整数；

采用预先确定的计算公式计算x_{i1}和x_{i2}之间的余弦相似度S_i^{12}，并计算x_{i1}和x_{i3}之间的余弦相似度S_i^{13}；

根据余弦相似度S_i^{12}、S_i^{13}及预先确定的损失函数L更新模型的参数，所述预先确定的损失函数L的公式为：$L=\sum_{i=0}^{N}\max\left(S_i^{13}-S_i^{12}+\alpha,0\right)$，其中α是取值范围在0.05～0.2之间常量，N是获得的三元组的个数。

5.一种电子装置，其特征在于，所述电子装置包括存储器和处理器，所述存储器上存储有可在所述处理器上运行的身份验证系统，所述身份验证系统被所述处理器执行时实现如下步骤：

在收到待进行身份验证的目标用户的当前语音数据后，从数据库中获取待验证的身份对应的标准语音数据，将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理，以得到所述当前语音数据对应的当前语音帧组和所述标准语音数据对应的标准语音帧组；

利用预设滤波器分别提取出当前语音帧组中各个语音帧的预设类型声学特征和标准语音帧组中各个语音帧的预设类型声学特征；

将提取出的当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征输入预先训练好的预设结构深度神经网络模型，以得到所述当前语音数据和所述标准语音数据各自对应的预设长度的特征矢量，其中，所述预设结构深度神经网络模型为权利要求1或3所述的深度神经网络模型；

计算得到的两个特征矢量的余弦相似度，并根据计算出的余弦相似度大小确定身份验证结果，所述身份验证结果包括验证通过结果和验证失败结果。

6.如权利要求5所述的电子装置，其特征在于，在将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理的步骤之前，该处理器还用于执行所述身份验证系统，以实现以下步骤：

分别对所述当前语音数据和标准语音数据进行活动端点检测，将所述当前语音数据和所述标准语音数据中的非说话人的语音删除。

7.一种身份验证方法，其特征在于，该身份验证方法包括：

在收到待进行身份验证的目标用户的当前语音数据后，从数据库中获取待验证的身份对应的标准语音数据，将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理，以得到所述当前语音数据对应的当前语音帧组和所述标准语音数据对应的标准语音帧组；

利用预设滤波器分别提取出当前语音帧组中各个语音帧的预设类型声学特征和标准语音帧组中各个语音帧的预设类型声学特征；

将提取出的当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型

声学特征输入预先训练好的预设结构深度神经网络模型，以得到所述当前语音数据和所述标准语音数据各自对应的预设长度的特征矢量，其中，所述预设结构深度神经网络模型为权利要求1、3、4中任意一项所述的深度神经网络模型；

计算得到的两个特征矢量的余弦相似度，并根据计算出的余弦相似度大小确定身份验证结果，所述身份验证结果包括验证通过结果和验证失败结果。

8.如权利要求7所述的身份验证方法，其特征在于，在将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理的步骤之前，所述身份验证方法还包括步骤：

分别对所述当前语音数据和标准语音数据进行活动端点检测，将所述当前语音数据和所述标准语音数据中的非说话人的语音删除。

9.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有身份验证系统，所述身份验证系统可被至少一个处理器执行，以使所述至少一个处理器执行如下步骤：

在收到待进行身份验证的目标用户的当前语音数据后，从数据库中获取待验证的身份对应的标准语音数据，将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理，以得到所述当前语音数据对应的当前语音帧组和所述标准语音数据对应的标准语音帧组；

利用预设滤波器分别提取出当前语音帧组中各个语音帧的预设类型声学特征和标准语音帧组中各个语音帧的预设类型声学特征；

将提取出的当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征输入预先训练好的预设结构深度神经网络模型，以得到所述当前语音数据和所述标准语音数据各自对应的预设长度的特征矢量，其中，所述预设结构深度神经网络模型为权利要求1、3、4中任意一项所述的深度神经网络模型；

计算得到的两个特征矢量的余弦相似度，并根据计算出的余弦相似度大小确定身份验证结果，所述身份验证结果包括验证通过结果和验证失败结果。

10.如权利要求9所述的计算机可读存储介质，其特征在于，在将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理的步骤之前，该处理器还用于执行所述身份验证系统，以实现以下步骤：

分别对所述当前语音数据和标准语音数据进行活动端点检测，将所述当前语音数据和所述标准语音数据中的非说话人的语音删除。

深度神经网络模型、电子装置、身份验证方法和存储介质

技术领域

[0001] 本发明涉及声纹识别技术领域，特别涉及一种深度神经网络模型、电子装置、身份验证方法和存储介质。

背景技术

[0002] 说话人识别通常称为声纹识别，是生物识别技术的一种，常被用来确认某段语音是否是指定的某个人所说，是“一对一判别”问题。说话人识别广泛应用于诸多领域，例如，在金融、证券、社保、公安、军队及其他民用安全认证等领域都有着广泛的应用需求。

[0003] 说话人识别包括文本相关识别和文本无关识别两种方式，近年来文本无关说话人识别技术不断突破，其准确性较之以往有了极大的提升。然而在某些受限情况下，比如采集到的说话人有效语音较短(时长小于5秒的语音)的情况下，现有的文本无关说话人识别技术的准确性不高，很容易出错。

发明内容

[0004] 本发明的主要目的是提供一种深度神经网络模型、电子装置、身份验证方法和存储介质，旨在旨在提升说话人身份验证的准确性。

[0005] 为实现上述目的，本发明提出的深度神经网络模型，包括：

[0006] 第一层结构：是由多层堆叠的有相同预设结构的神经网络层，每个预设结构的神经网络层包括：两个串联的CNN卷积层，两个修正线性单元ReLU，及一个将两个串联的CNN卷积层跨层直连的直连操作X，其中，各个ReLU与各个CNN卷积层一一对应，且各个ReLU分别串联在对应的CNN卷积层后，所述直连操作X将两个串联的CNN卷积层的第一个CNN卷积层的卷积操作的输入与第二个CNN卷积层的卷积操作的输出相加，并将结果送入到第二个CNN卷积层对应的ReLU操作中；

[0007] 第二层结构：是平均层，此层的作用是沿时间轴向对矢量序列求平均值，它将第一层结构输出的二维矢量序列进行平均化；

[0008] 第三层结构：是DNN全连接层；

[0009] 第四层结构：是归一化层，此层将上一层的输入按照L2范数进行归一化，得到长度为1的归一化后的特征矢量；

[0010] 第五层结构：是损失层，损失函数L的公式为：$L=\sum_{i=0}^{N}\max\left(S_i^{13}-S_i^{12}+\alpha,0\right)$，其中α是取值范围在0.05～0.2之间的常量，$S_i^{12}$代表属于同一说话人的两个特征矢量的余弦相似度，$S_i^{13}$代表不属于同一说话人的两个特征矢量的余弦相似度。

[0011] 优选地，所述深度神经网络模型的训练过程为：

[0012] S1、获取预设数量语音数据样本，对各个语音数据样本分别标注代表对应的说话人身份的标签；

[0013] S2、分别对每个语音数据样本进行活动端点检测，将语音数据样本中非说话人的语音删除，得到预设数量的标准语音数据样本；

[0014] S3、将得到的标准语音数据样本的第一百分比作为训练集，第二百分比作为验证集，所述第一百分比与第二百分比的和小于等于100％；

[0015] S4、将所述训练集和验证集中的各个标准语音数据样本按照预设的分帧参数分别进行分帧处理，以获得每个标准语音数据样本对应的语音帧组，再利用预设滤波器分别提取出每个语音帧组中的各个语音帧的预设类型声学特征；

[0016] S5、将所述训练集中的各个语音帧组对应的预设类型声学特征划分成M批，分批输入所述深度神经网络模型中进行迭代训练，并在所述深度神经网络模型训练完成后，采用验证集对所述深度神经网络模型的准确率进行验证；

[0017] S6、若验证得到的准确率大于预设阈值，则模型训练结束；

[0018] S7、若验证得到的准确率小于或者等于预设阈值，则增加获取的语音数据样本的数量，并基于增加后的语音数据样本重新执行上述步骤S1-S5。

[0019] 优选地，所述预设滤波器为梅尔滤波器，所述预设类型声学特征为梅尔频率倒谱系数MFCC。

[0020] 优选地，所述深度神经网络模型迭代训练的过程包括：

[0021] 根据模型的当前参数将当前输入每个语音帧组对应的预设类型声学特征转化为对应的一个预设长度的特征矢量；

[0022] 从各个特征矢量中进行随机选取以获得多个三元组，第i个三元组(x_{i1},x_{i2},x_{i3})由三个不同的特征矢量x_{i1}、x_{i2}和x_{i3}组成，其中，x_{i1}和x_{i2}对应同一个说话人，x_{i1}和x_{i3}对应不同的说话人，i为正整数；

[0023] 采用预先确定的计算公式计算x_{i1}和x_{i2}之间的余弦相似度S_i^{12}，并计算x_{i1}和x_{i3}之间的余弦相似度S_i^{13}；

[0024] 根据余弦相似度S_i^{12}、S_i^{13}及预先确定的损失函数L更新模型的参数，所述预先确定的损失函数L的公式为：$L=\sum_{i=0}^{N}\max\left(S_i^{13}-S_i^{12}+\alpha,0\right)$，其中α是取值范围在0.05～0.2之间常量，N是获得的三元组的个数。

[0025] 本发明还提出一种电子装置，所述电子装置包括存储器和处理器，所述存储器上存储有可在所述处理器上运行的身份验证系统，所述身份验证系统被所述处理器执行时实现如下步骤：

[0026] 在收到待进行身份验证的目标用户的当前语音数据后，从数据库中获取待验证的身份对应的标准语音数据，将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理，以得到所述当前语音数据对应的当前语音帧组和所述标准语音数据对应的标准语音帧组；

[0027] 利用预设滤波器分别提取出当前语音帧组中各个语音帧的预设类型声学特征和标准语音帧组中各个语音帧的预设类型声学特征；

[0028] 将提取出的当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征输入预先训练好的预设结构深度神经网络模型，以得到所述当前语音数据和

所述标准语音数据各自对应的预设长度的特征矢量，其中，所述预设结构深度神经网络模型为上述任一项所述的深度神经网络模型；

[0029] 计算得到的两个特征矢量的余弦相似度，并根据计算出的余弦相似度大小确定身份验证结果，所述身份验证结果包括验证通过结果和验证失败结果。

[0030] 优选地，在将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理的步骤之前，该处理器还用于执行所述身份验证系统，以实现以下步骤：

[0031] 分别对所述当前语音数据和标准语音数据进行活动端点检测，将所述当前语音数据和所述标准语音数据中的非说话人的语音删除。

[0032] 本发明还提出一种身份验证方法，该身份验证方法包括：

[0033] 在收到待进行身份验证的目标用户的当前语音数据后，从数据库中获取待验证的身份对应的标准语音数据，将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理，以得到所述当前语音数据对应的当前语音帧组和所述标准语音数据对应的标准语音帧组；

[0034] 利用预设滤波器分别提取出当前语音帧组中各个语音帧的预设类型声学特征和标准语音帧组中各个语音帧的预设类型声学特征；

[0035] 将提取出的当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征输入预先训练好的预设结构深度神经网络模型，以得到所述当前语音数据和所述标准语音数据各自对应的预设长度的特征矢量，其中，所述预设结构深度神经网络模型为上述任一项所述的深度神经网络模型；

[0036] 计算得到的两个特征矢量的余弦相似度，并根据计算出的余弦相似度大小确定身份验证结果，所述身份验证结果包括验证通过结果和验证失败结果。

[0037] 优选地，在将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理的步骤之前，所述身份验证方法还包括步骤：

[0038] 分别对所述当前语音数据和标准语音数据进行活动端点检测，将所述当前语音数据和所述标准语音数据中的非说话人的语音删除。

[0039] 本发明还提出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身份验证系统，所述身份验证系统可被至少一个处理器执行，以使所述至少一个处理器执行如下步骤：

[0040] 在收到待进行身份验证的目标用户的当前语音数据后，从数据库中获取待验证的身份对应的标准语音数据，将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理，以得到所述当前语音数据对应的当前语音帧组和所述标准语音数据对应的标准语音帧组；

[0041] 利用预设滤波器分别提取出当前语音帧组中各个语音帧的预设类型声学特征和标准语音帧组中各个语音帧的预设类型声学特征；

[0042] 将提取出的当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征输入预先训练好的预设结构深度神经网络模型，以得到所述当前语音数据和所述标准语音数据各自对应的预设长度的特征矢量，其中，所述预设结构深度神经网络模型为上述任一项所述的深度神经网络模型；

[0043] 计算得到的两个特征矢量的余弦相似度，并根据计算出的余弦相似度大小确定身

份验证结果，所述身份验证结果包括验证通过结果和验证失败结果。

[0044] 优选地，在将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理的步骤之前，该处理器还用于执行所述身份验证系统，以实现以下步骤：

[0045] 分别对所述当前语音数据和标准语音数据进行活动端点检测，将所述当前语音数据和所述标准语音数据中的非说话人的语音删除。

[0046] 本发明技术方案，通过将接收到待验证身份的目标用户的当前语音数据和待验证身份的标准语音数据先进行分帧处理，利用预设滤波器提取分帧处理得到的各个语音帧的提取出预设类型声学特征，再将提取出的预设类型声学特征输入到预先训练好的预设结构深度神经网络模型，预设结构深度神经网络模型分别将当前语音数据对应的预设类型声学特征和标准语音数据对应的预设类型声学特征转化为对应的特征向量后，计算两个特征向量的余弦相似度，根据余弦相似度大小确认验证结果。本实施例技术方案，通过将语音数据先分帧处理为多个语音帧并根据语音帧提取预设类型声学特征，使得即使在采集到的有效语音数据很短时，也能提取根据采集到的语音数据提取得到足够多的声学特征，再采用本发明的深度神经网络模型根据提取出得到声学特征进行处理，能够显著增强模型对输入数据的特征提取能力，减轻网络层次加深时性能降低的风险，提高输出验证结果的正确率。

附图说明

[0047] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中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示出的结构获得其他的附图。

[0048] 图1为本发明深度神经网络模型较佳实施例中第一层结构的一个预设结构的神经网络层的结构示意图；

[0049] 图2为本发明深度神经网络模型训练过程的流程示意图；

[0050] 图3为本发明身份验证方法一实施例的流程示意图；

[0051] 图4为本发明身份验证系统一实施例的运行环境示意图；

[0052] 图5为本发明身份验证系统一实施例的程序模块图；

[0053] 图6为本发明身份验证系统二实施例的程序模块图。

[0054] 本发明目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

具体实施方式

[0055] 以下结合附图对本发明的原理和特征进行描述，所举实例只用于解释本发明，并非用于限定本发明的范围。

[0056] 本发明提出一种深度神经网络模型，用于说话人身份识别验证。

[0057] 本实施例的深度神经网络模型的结构包括：

[0058] 第一层结构：是由多层堆叠(例如9～12层堆叠)的有相同预设结构的神经网络层，如图1所示，每个预设结构的神经网络层包括：两个串联的CNN卷积层100(例如，所述CNN卷积层100可以采用：3*3的卷积核，步长为1*1，通道数为64)，两个修正线性单元ReLU200，及一个将两个串联的CNN卷积层100跨层直连的直连操作X，其中，各个ReLU200与各个CNN卷积

层100一一对应，且各个ReLU200分别串联在对应的CNN卷积层100后，所述直连操作X将两个串联的CNN卷积层100的第一个CNN卷积层100的卷积操作的输入与第二个CNN卷积层100的卷积操作的输出相加，并将结果送入到第二个CNN卷积层100对应的ReLU200操作中；

[0059] 第二层结构：是平均层，此层的作用是沿时间轴向对矢量序列求平均值，它将第一层结构输出的二维矢量序列进行平均化；

[0060] 第三层结构：是DNN全连接层；

[0061] 第四层结构：是归一化层，此层将上一层的输入按照L2范数进行归一化，得到长度为1的归一化后的特征矢量；

[0062] 第五层结构：是损失层，损失函数L的公式为：$L=\sum_{i=0}^{N}\max\left(S_i^{13}-S_i^{12}+\alpha,0\right)$，其中α是取值范围在0.05～0.2之间的常量，$S_i^{12}$代表属于同一说话人的两个特征矢量的余弦相似度，$S_i^{13}$代表不属于同一说话人的两个特征矢量的余弦相似度。

[0063] 采用本实施例的深度神经网络模型，能够显著增强模型对输入数据的特征提取能力，减轻网络层次加深时性能降低的风险。

[0064] 本实施例中的深度神经网络模型的训练过程为：

[0065] S1、获取预设数量语音数据样本，对各个语音数据样本分别标注代表对应的说话人身份的标签；

[0066] 先准备好预设数量(例如，10000个)语音数据样本，各个语音数据样本都是已知说话人身份的语音数据；这些语音数据样本中，每一个说话人身份或部分的说话人身份对应有多个语音数据样本，将各个语音数据样本标注上代表对应的说话人身份的标签。

[0067] S2、分别对每个语音数据样本进行活动端点检测，将语音数据样本中非说话人的语音删除，得到预设数量的标准语音数据样本；

[0068] 对语音数据样本进行活动端点检测，以检测出每个语音数据样本中的非说话人的语音(例如，静音或噪音)并删除，避免语音数据样本中存在与对应的说话人身份的声纹特征无关的语音数据，而影响对模型的训练效果。

[0069] S3、将得到的标准语音数据样本的第一百分比作为训练集，第二百分比作为验证集，所述第一百分比与第二百分比的和小于等于100％；

[0070] 例如，将得到的标准语音数据样本的70％作为训练集，30％作为验证集。

[0071] S4、将所述训练集和验证集中的各个标准语音数据样本按照预设的分帧参数分别进行分帧处理，以获得每个标准语音数据样本对应的语音帧组，再利用预设滤波器分别提取出每个语音帧组中的各个语音帧的预设类型声学特征；

[0072] 其中，预设的分帧参数例如，每隔25毫秒分帧，帧移10毫秒；该预设滤波器例如为梅尔滤波器，通过梅尔滤波器提取出的预设类型声学特征为MFCC(Mel Frequency Cepstrum Coefficient，梅尔频率倒谱系数)频谱特征，例如，36维MFCC频谱特征。

[0073] S5、将所述训练集中的各个语音帧组对应的预设类型声学特征划分成M批，分批输入所述深度神经网络模型中进行迭代训练，并在所述深度神经网络模型训练完成后，采用验证集对所述深度神经网络模型的准确率进行验证；

[0074] 对训练集中的预设类型声学特征进行分批处理，划分成M(例如30)批，分批方式可

按照语音帧组为分配单位，每一批中分配等量或不等量的语音帧组对应的预设类型声学特征；将训练集中的各个语音帧组对应的预设类型声学特征按照分成的批次逐一的输入深度神经网络模型中进行迭代训练，每一批预设类型声学特征使所述预设结构胜读神经网络模型迭代一次，每次迭代都会更新得到新的模型参数，通过多次迭代训练完成后，该深度神经网络模型已经更新为较佳的模型参数；迭代训练完成后，则利用验证集对该深度神经网络模型的准确率进行验证，即将验证集中的标准语音数据两两分组，每次输入一个分组中的标准语音数据样本对应的预设类型声学特征到该深度神经网络模型，根据输入的两个标准语音数据的身份标签，确认输出的验证结构是否正确，在完成对各个分组的验证后，根据验证结果正确次数计算准确率，例如对100个分组进行验证，最终得到验证结果正确的有99组，则准确率就为99％。

[0075] S6、若验证得到的准确率大于预设阈值，则模型训练结束；

[0076] 系统中预先设置了准确率的验证阈值(即所述预设阈值，例如98.5％)，用于对所述深度神经网络模型的训练效果进行检验；若通过所述验证集对所述深度神经网络模型验证得到的准确率大于所述预设阈值，那么说明该深度神经网络模型的训练达到了标准，此时则结束模型训练。

[0077] S7、若验证得到的准确率小于或者等于预设阈值，则增加获取的语音数据样本的数量，并基于增加后的语音数据样本重新执行上述步骤S1-S5。

[0078] 若是通过所述验证集对所述深度神经网络模型验证得到的准确率小于或等于所述预设阈值，那么说明该深度神经网络模型的训练还没有达到了预期标准，可能是训练集数量不够或验证集数量不够，所以，在这种情况时，则增加获取的语音数据样本的数量(例如，每次增加固定数量或每次增加随机数量)，然后在这基础上，重新执行上述步骤S1-S5，如此循环执行，直至达到了步骤S6的要求，则结束模型训练。

[0079] 本实施例中，所述深度神经网络模型迭代训练的过程包括：

[0080] 根据模型的当前参数将当前输入每个语音帧组对应的预设类型声学特征转化为对应的一个预设长度的特征矢量；

[0081] 从各个特征矢量中进行随机选取以获得多个三元组，第i个三元组(x_{i1},x_{i2},x_{i3})由三个不同的特征矢量x_{i1}、x_{i2}和x_{i3}组成，其中，x_{i1}和x_{i2}对应同一个说话人，x_{i1}和x_{i3}对应不同的说话人，i为正整数；

[0082] 采用预先确定的计算公式计算x_{i1}和x_{i2}之间的余弦相似度S_i^{12}，并计算x_{i1}和x_{i3}之间的余弦相似度S_i^{13}；

[0083] 根据余弦相似度S_i^{12}、S_i^{13}及预先确定的损失函数L更新模型的参数，所述预先确定的损失函数L的公式为：$L=\sum_{i=0}^{N}\max\left(S_i^{13}-S_i^{12}+\alpha,0\right)$，其中α是取值范围在0.05～0.2之间常量，N是获得的三元组的个数。

[0084] 其中，模型参数更新步骤为：1.采用反向传播算法计算该深度神经网络的梯度；2.采用mini-batch-SGD(即小批量随机梯度下降)方法更新该深度神经网络的参数。

[0085] 本发明还提出一种身份验证方法，该身份验证方法基于上述实施例任一项所述的深度神经网络模型。

[0086] 如图3所示，图3为本发明身份验证方法一实施例的流程示意图。

[0087] 本实施例中，该身份验证方法包括：

[0088] 步骤S10，在收到待进行身份验证的目标用户的当前语音数据后，从数据库中获取待验证的身份对应的标准语音数据，将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理，以得到所述当前语音数据对应的当前语音帧组和所述标准语音数据对应的标准语音帧组；

[0089] 身份验证系统的数据库中预先存储有每个身份的标准语音数据，在收到待进行身份验证的目标用户的当前语音数据后，根据目标用户要求验证的身份(待验证的身份)，身份验证系统在数据库中获取该待验证的身份对应的标准语音数据，然后再分别对接收到的当前语音数据和获取到的标准语音数据按照预设的分帧参数进行分帧处理，以得到所述当前语音数据对应的当前语音帧组(包括当前语音数据经分帧得到的多个语音帧)和所述标准语音数据对应的标准语音帧组(包括标准语音数据经分帧得到的多个语音帧)。其中，所述预设的分帧参数例如，每隔25毫秒分帧，帧移10毫秒。

[0090] 步骤S20，利用预设滤波器分别提取出当前语音帧组中各个语音帧的预设类型声学特征和标准语音帧组中各个语音帧的预设类型声学特征；

[0091] 在得到当前语音帧组和标准语音帧组后，身份验证系统在利用预设滤波器分别对当前语音帧组和标准语音帧组中的各个语音帧进行特征提取，以提取出当前语音帧组中的各个语音帧对应的预设类型声学特征和标准语音帧组中的各个语音帧对应的预设类型声学特征。例如，该预设滤波器为梅尔(Mel)滤波器，提取出的预设类型声学特征为36维MFCC(Mel Frequency Cepstrum Coefficient，梅尔频率倒谱系数)频谱特征。

[0092] 步骤S30，将提取出的当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征输入预先训练好的预设结构深度神经网络模型，以得到所述当前语音数据和所述标准语音数据各自对应的预设长度的特征矢量，其中，该预设结构深度神经网络模型为上述实施例所述的深度神经网络模型；

[0093] 步骤S40，计算得到的两个特征矢量的余弦相似度，并根据计算出的余弦相似度大小确定身份验证结果，所述身份验证结果包括验证通过结果和验证失败结果。

[0094] 身份验证系统中具有预先训练好的预设结构深度神经网络模型，该模型为采用样本语音数据的对应的预设类型声学特征迭代训练的模型；身份验证系统在对当前语音帧组和标准语音帧组中的语音帧进行特征提取后，将当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征输入该预先训练好的预设结构深度神经网络模型中，模型将当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征分别转化为一个预设长度的特征矢量(例如，长度为1的特征矢量)，再计算得到的两个特征矢量的余弦相似度，根据计算出的余弦相似度的大小确定身份验证结果，即将该余弦相似度与预设阈值(例如0.95)比较，若该余弦相似度大于预设阈值，则确定身份验证通过，反之，则确定身份验证失败。其中，余弦相似度计算公式为：$\cos(x_i,x_j)=x_i^T x_j$，x_i和x_j代表两个特征矢量，T为预先确定值。

[0095] 本实施例技术方案，通过将接收到待验证身份的目标用户的当前语音数据和待验证身份的标准语音数据先进行分帧处理，利用预设滤波器提取分帧处理得到的各个语音帧的提取出预设类型声学特征，再将提取出的预设类型声学特征输入到预先训练好的预设结

构深度神经网络模型，预设结构深度神经网络模型分别将当前语音数据对应的预设类型声学特征和标准语音数据对应的预设类型声学特征转化为对应的特征向量后，计算两个特征向量的余弦相似度，根据余弦相似度大小确认验证结果。本实施例技术方案，通过将语音数据先分帧处理为多个语音帧并根据语音帧提取预设类型声学特征，使得即使在采集到的有效语音数据很短时，也能提取根据采集到的语音数据提取得到足够多的声学特征，再采用本发明的深度神经网络模型根据提取出得到声学特征进行处理，能够显著增强模型对输入数据的特征提取能力，减轻网络层次加深时性能降低的风险，提高输出验证结果的正确率。

[0096] 进一步地，本实施例在将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理的步骤之前，所述身份验证方法还包括步骤：

[0097] 分别对所述当前语音数据和标准语音数据进行活动端点检测，将所述当前语音数据和所述标准语音数据中的非说话人的语音删除。

[0098] 在采集的当前语音数据和预先存储的标准语音数据中都包含一些非说话人语音部分(例如，静音或噪音)，如果这些部分不删除掉，则对当前语音数据或标准对语音数据进行分帧处理后得到的语音帧组中，会出现包含非说话人语音部分的语音帧(甚至个别语音帧中全为非说话人语音)，这样，利用预设滤波器根据这些包含非说话人语音部分的语音帧提取出的预设类型声学特征属于杂质特征，会降低预设结构深度神经网络模型得出结果的准确性；故本实施例在对语音数据分帧处理之前，先检测当前语音数据和标准语音数据中的非说话人语音部分，并将检测到的非说话人语音部分删除，本实施例采用的非说话人语音部分的检测方式为活动端点检测(Voice Activity Detection，VAD)。

[0099] 此外，本发明还提出一种身份验证系统。

[0100] 请参阅图4，是本发明身份验证系统10较佳实施例的运行环境示意图。

[0101] 在本实施例中，身份验证系统10安装并运行于电子装置1中。电子装置1可以是桌上型计算机、笔记本、掌上电脑及服务器等计算设备。该电子装置1可包括，但不仅限于，存储器11、处理器12及显示器13。图4仅示出了具有组件11-13的电子装置1，但是应理解的是，并不要求实施所有示出的组件，可以替代的实施更多或者更少的组件。

[0102] 存储器11在一些实施例中可以是电子装置1的内部存储单元，例如该电子装置1的硬盘或内存。存储器11在另一些实施例中也可以是电子装置1的外部存储设备，例如电子装置1上配备的插接式硬盘，智能存储卡(Smart Media Card，SMC)，安全数字(Secure Digital，SD)卡，闪存卡(Flash Card)等。进一步地，存储器11还可以既包括电子装置1的内部存储单元也包括外部存储设备。存储器11用于存储安装于电子装置1的应用软件及各类数据，例如身份验证系统10的程序代码等。存储器11还可以用于暂时地存储已经输出或者将要输出的数据。

[0103] 处理器12在一些实施例中可以是一中央处理器(Central Processing Unit，CPU)，微处理器或其他数据处理芯片，用于运行存储器11中存储的程序代码或处理数据，例如执行身份验证系统10等。

[0104] 显示器13在一些实施例中可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED(Organic Light-Emitting Diode，有机发光二极管)触摸器等。显示器13用于显示在电子装置1中处理的信息以及用于显示可视化的用户界面。电子装置1的部件11-13通过系统总线相互通信。

[0105] 请参阅图5，是本发明身份验证系统10较佳实施例的程序模块图。在本实施例中，身份验证系统10可以被分割成一个或多个模块，一个或者多个模块被存储于存储器11中，并由一个或多个处理器(本实施例为处理器12)所执行，以完成本发明。例如，在图5中，身份验证系统10可以被分割成分帧模块101、提取模块102、计算模块103及结果确定模块104。本发明所称的模块是指能够完成特定功能的一系列计算机程序指令段，比程序更适合于描述身份验证系统10在电子装置1中的执行过程，其中：

[0106] 分帧模块101，用于在收到待进行身份验证的目标用户的当前语音数据后，从数据库中获取待验证的身份对应的标准语音数据，将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理，以得到所述当前语音数据对应的当前语音帧组和所述标准语音数据对应的标准语音帧组；

[0107] 身份验证系统的数据库中预先存储有每个身份的标准语音数据，在收到待进行身份验证的目标用户的当前语音数据后，根据目标用户要求验证的身份(待验证的身份)，身份验证系统在数据库中获取该待验证的身份对应的标准语音数据，然后再分别对接收到的当前语音数据和获取到的标准语音数据按照预设的分帧参数进行分帧处理，以得到所述当前语音数据对应的当前语音帧组(包括当前语音数据经分帧得到的多个语音帧)和所述标准语音数据对应的标准语音帧组(包括标准语音数据经分帧得到的多个语音帧)。其中，所述预设的分帧参数例如，每隔25毫秒分帧，帧移10毫秒。

[0108] 提取模块102，用于利用预设滤波器分别提取出当前语音帧组中各个语音帧的预设类型声学特征和标准语音帧组中各个语音帧的预设类型声学特征；

[0109] 在得到当前语音帧组和标准语音帧组后，身份验证系统在利用预设滤波器分别对当前语音帧组和标准语音帧组中的各个语音帧进行特征提取，以提取出当前语音帧组中的各个语音帧对应的预设类型声学特征和标准语音帧组中的各个语音帧对应的预设类型声学特征。例如，该预设滤波器为梅尔(Mel)滤波器，提取出的预设类型声学特征为36维MFCC(Mel Frequency Cepstrum Coefficient，梅尔频率倒谱系数)频谱特征。

[0110] 计算模块103，用于将提取出的当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征输入预先训练好的预设结构深度神经网络模型，以得到所述当前语音数据和所述标准语音数据各自对应的预设长度的特征矢量，其中，该预设结构深度神经网络模型为上述实施例所述的深度神经网络模型；

[0111] 结果确定模块104，用于计算得到的两个特征矢量的余弦相似度，并根据计算出的余弦相似度大小确定身份验证结果，所述身份验证结果包括验证通过结果和验证失败结果。

[0112] 身份验证系统中具有预先训练好的预设结构深度神经网络模型，该模型为采用样本语音数据的对应的预设类型声学特征迭代训练的模型；身份验证系统在对当前语音帧组和标准语音帧组中的语音帧进行特征提取后，将当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征输入该预先训练好的预设结构深度神经网络模型中，模型将当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征分别转化为一个预设长度的特征矢量(例如，长度为1的特征矢量)，再计算得到的两个特征矢量的余弦相似度，根据计算出的余弦相似度的大小确定身份验证结果，即将该余弦相似度与预设阈值(例如0.95)比较，若该余弦相似度大于预设阈值，则确定身份验证通过，反

之，则确定身份验证失败。其中，余弦相似度计算公式为：$\cos(x_i, x_j) = x_i^T x_j$，$x_i$和$x_j$代表两个特征矢量，T为预先确定值。

[0113] 本实施例技术方案，通过将接收到待验证身份的目标用户的当前语音数据和待验证身份的标准语音数据先进行分帧处理，利用预设滤波器提取分帧处理得到的各个语音帧的提取出预设类型声学特征，再将提取出的预设类型声学特征输入到预先训练好的预设结构深度神经网络模型，预设结构深度神经网络模型分别将当前语音数据对应的预设类型声学特征和标准语音数据对应的预设类型声学特征转化为对应的特征向量后，计算两个特征向量的余弦相似度，根据余弦相似度大小确认验证结果。本实施例技术方案，通过将语音数据先分帧处理为多个语音帧并根据语音帧提取预设类型声学特征，使得即使在采集到的有效语音数据很短时，也能提取根据采集到的语音数据提取得到足够多的声学特征，再采用本发明的深度神经网络模型根据提取出得到声学特征进行处理，能够显著增强模型对输入数据的特征提取能力，减轻网络层次加深时性能降低的风险，提高输出验证结果的正确率。

[0114] 如图6所示，图6为本发明身份验证系统二实施例的程序模块图。

[0115] 本实施例中，身份验证系统还包括：

[0116] 检测模块105，用于在将当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理之前，分别对所述当前语音数据和标准语音数据进行活动端点检测，将所述当前语音数据和所述标准语音数据中的非说话人的语音删除。

[0117] 在采集的当前语音数据和预先存储的标准语音数据中都包含一些非说话人语音部分(例如，静音或噪音)，如果这些部分不删除掉，则对当前语音数据或标准对语音数据进行分帧处理后得到的语音帧组中，会出现包含非说话人语音部分的语音帧(甚至个别语音帧中全为非说话人语音)，这样，利用预设滤波器根据这些包含非说话人语音部分的语音帧提取出的预设类型声学特征属于杂质特征，会降低预设结构深度神经网络模型得出结果的准确性；故本实施例在对语音数据分帧处理之前，先检测当前语音数据和标准语音数据中的非说话人语音部分，并将检测到的非说话人语音部分删除，本实施例采用的非说话人语音部分的检测方式为活动端点检测(Voice Activity Detection，VAD)。

[0118] 进一步地，本发明还提出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身份验证系统，所述身份验证系统可被至少一个处理器执行，以使所述至少一个处理器执行上述任一实施例中的身份验证方法。

[0119] 以上所述仅为本发明的优选实施例，并非因此限制本发明的专利范围，凡是在本发明的发明构思下，利用本发明说明书及附图内容所作的等效结构变换，或直接/间接运用在其他相关的技术领域均包括在本发明的专利保护范围内。

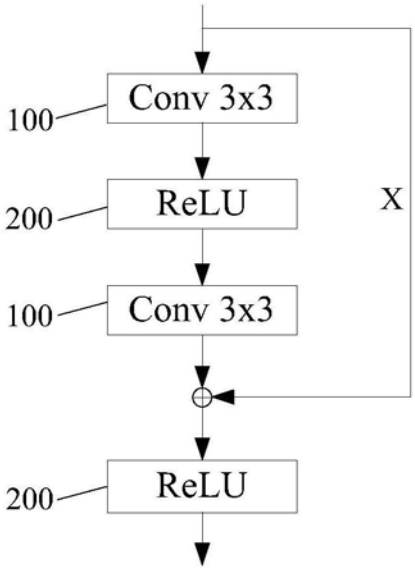


图1

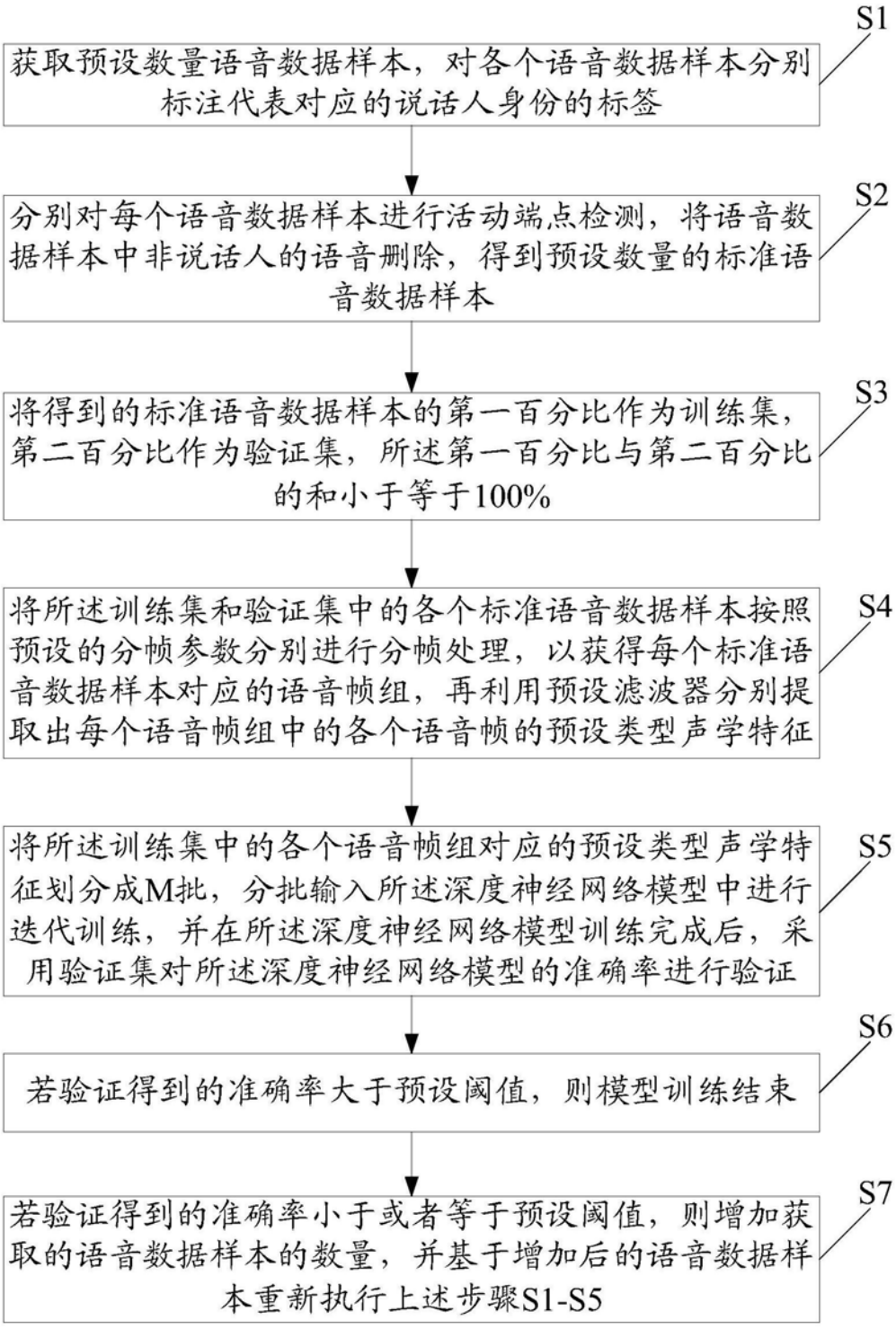

S1
获取预设数量语音数据样本，对各个语音数据样本分别标注代表对应的说话人身份的标签
S2
分别对每个语音数据样本进行活动端点检测，将语音数据样本中非说话人的语音删除，得到预设数量的标准语音数据样本
S3
将得到的标准语音数据样本的第一百分比作为训练集，第二百分比作为验证集，所述第一百分比与第二百分比的和小于等于100%
S4
将所述训练集和验证集中的各个标准语音数据样本按照预设的分帧参数分别进行分帧处理，以获得每个标准语音数据样本对应的语音帧组，再利用预设滤波器分别提取出每个语音帧组中的各个语音帧的预设类型声学特征
S5
将所述训练集中的各个语音帧组对应的预设类型声学特征划分成M批，分批输入所述深度神经网络模型中进行迭代训练，并在所述深度神经网络模型训练完成后，采用验证集对所述深度神经网络模型的准确率进行验证
S6
若验证得到的准确率大于预设阈值，则模型训练结束
S7
若验证得到的准确率小于或者等于预设阈值，则增加获取的语音数据样本的数量，并基于增加后的语音数据样本重新执行上述步骤S1-S5

图2

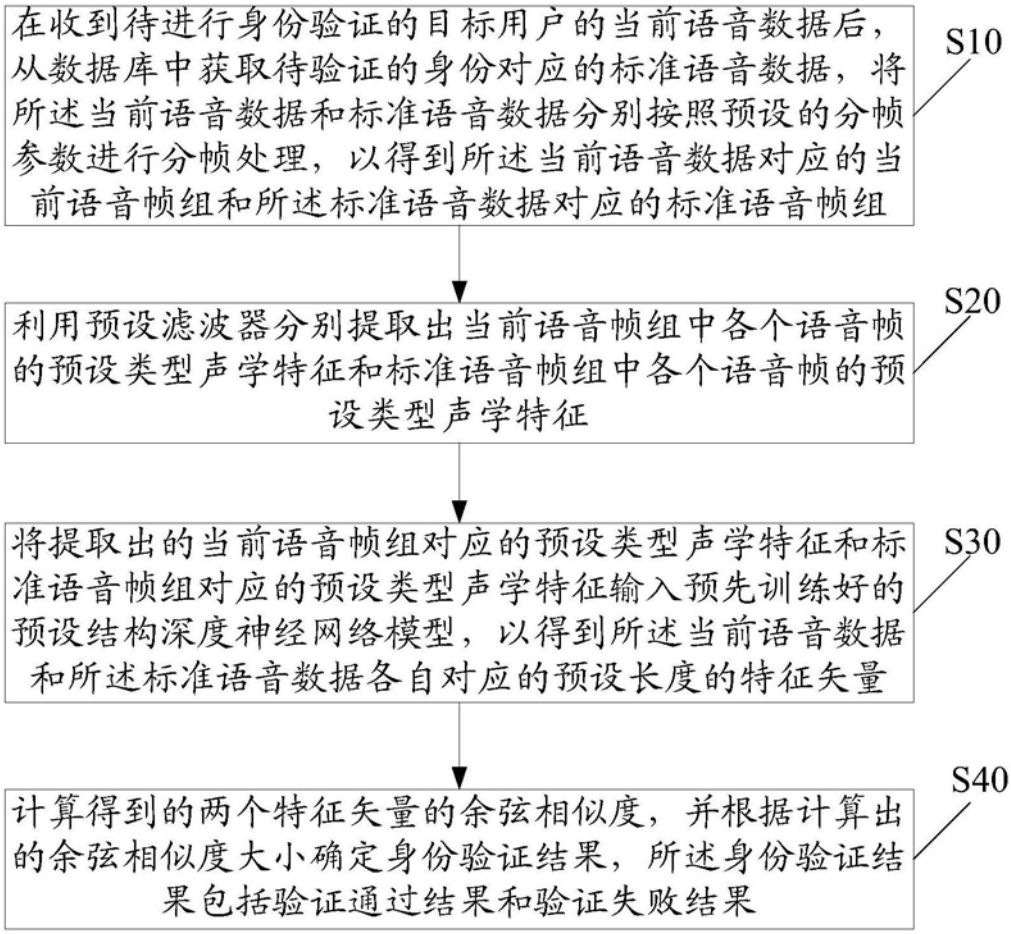

在收到待进行身份验证的目标用户的当前语音数据后，从数据库中获取待验证的身份对应的标准语音数据，将所述当前语音数据和标准语音数据分别按照预设的分帧参数进行分帧处理，以得到所述当前语音数据对应的当前语音帧组和所述标准语音数据对应的标准语音帧组
S10
利用预设滤波器分别提取出当前语音帧组中各个语音帧的预设类型声学特征和标准语音帧组中各个语音帧的预设类型声学特征
S20
将提取出的当前语音帧组对应的预设类型声学特征和标准语音帧组对应的预设类型声学特征输入预先训练好的预设结构深度神经网络模型，以得到所述当前语音数据和所述标准语音数据各自对应的预设长度的特征矢量
S30
计算得到的两个特征矢量的余弦相似度，并根据计算出的余弦相似度大小确定身份验证结果，所述身份验证结果包括验证通过结果和验证失败结果
S40

图3

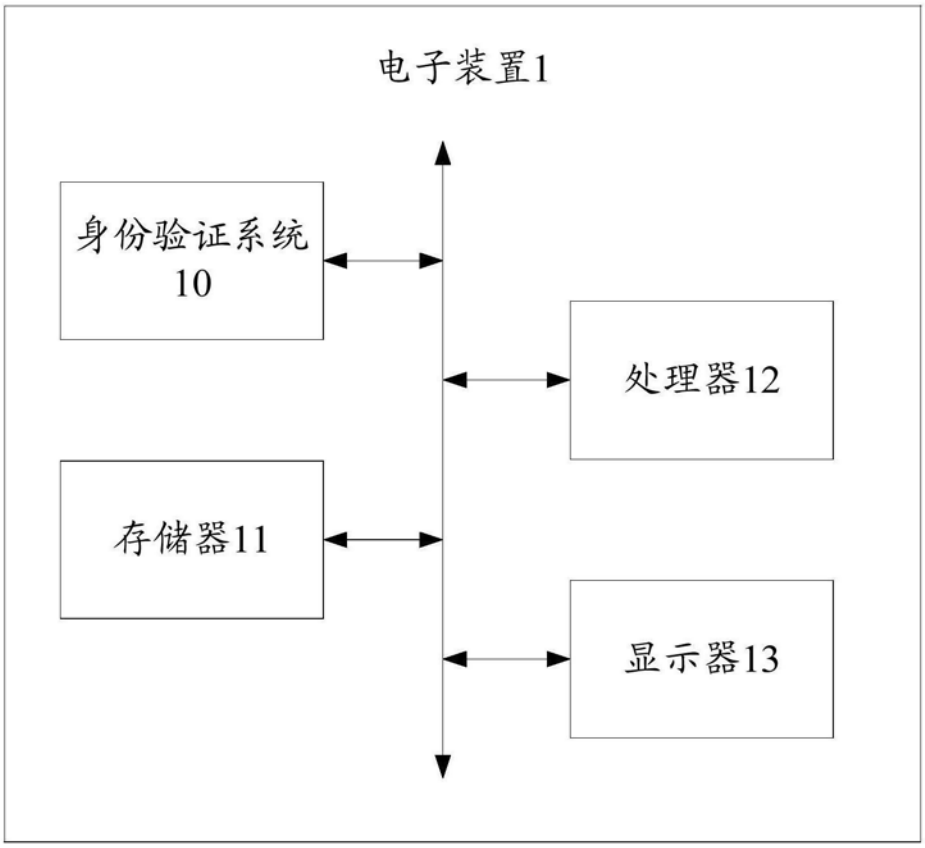

电子装置1
身份验证系统10
处理器12
存储器11
显示器13

图4

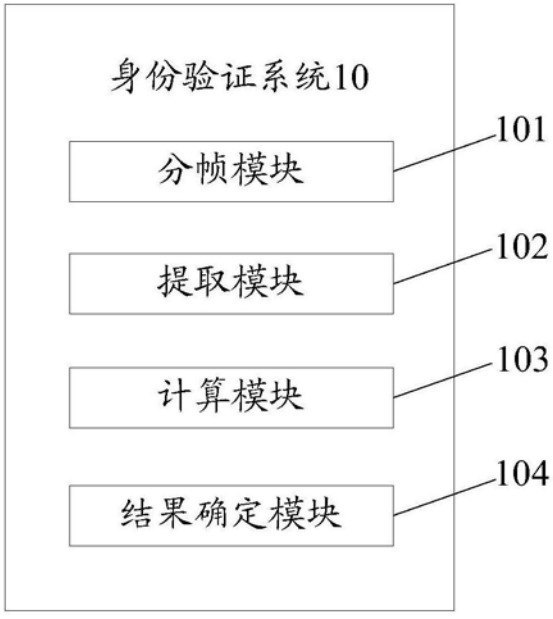

身份验证系统10
分帧模块
101
提取模块
102
计算模块
103
结果确定模块
104

图5

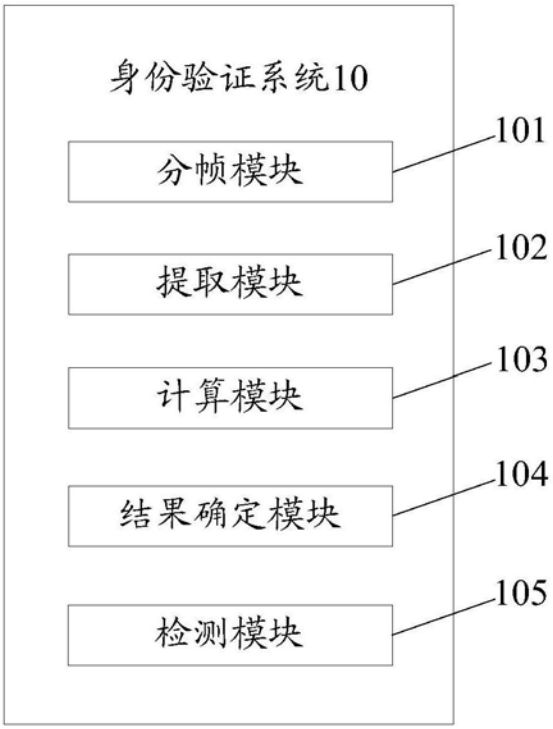


图6