深度残差网络专利

技术领域

本发明创造涉及深度学习以及残差网络，具体基于短时傅里叶变换以及神经网络中的深度残差网络一种时域信号分析系统。

背景技术

时域信号分析在当前社会各个领域都有涉及，具体包括：

机械领域（检测机械发出的音频信号与正常参数值进行比对，以此来检测故障），工业领域（工业上的超声探测信号处理），民生领域（乐器的音调识别），勘探领域，医疗领域等领域。

传统时域信号在分析时只是直接对时域信号进行分析，时域信号可以直观的观测到信号的形状，但不能用有限的参数对信号进行准确的描述。对信号进行时域分析时，有时一些信号的时域参数相同，但并不能说明信号就完全相同，因为信号不仅随时间变化，还与频率、相位等信息有关，这就需要进一步分析信号的频率结构，但传统的信号识别并不能做到这点，以至于识别系统的精度不足以及复杂场景无法精确使用。

发明内容

本发明创造涉及深度学习以及残差网络，具体基于短时傅里叶变换以及神经网络中的深度残差网络一种时域信号分析系统。

概括为：

1. 由音频采集设备从外界采集时域信号，且输入至主控设备。
2. 对采集数据集中进行预处理。
3. 将时域信号进行傅立叶快速变换得到频域信号。
4. 频域信号通过深度残差网络进行标准化处理。
5. 进入残差网络，开始训练。
6. 对音频进行一个编码。
7. 输出编码至所需设备，进行应用。

本发明创造的创新点为：

* 深度残差网络对音频时域信号的分析
* 时域信号的预处理
* 深度残差网络在时域信号分析中的优越性

以下为具体分析：

1. 深度残差网络对音频时域信号的分析

语音信号是一种典型的非平稳信号,其非平稳性是由声波震动过程产生的。在一段极短时间内，我们可以把它看作一段平稳的信号。因为傅里叶分析是分析线性系统和平稳信号稳态特性的强有力的手段,而短时傅里叶分析,也叫时间依赖傅里叶变换,就是在短时平稳的假设下,用稳态分析方法处理非平稳信号的一种方法。语谱图是语音信号短时频谱的时间一一强度表示。语音信号首先被分割为相互重叠的若干段帧,对每一段加窗处理,然后进行快速傅里叶变换,得到这个信号的短时频谱估计。获得声音语谱数据后,由于其维度较高直接进行模式识别，计算量大且识别精度高,因而需要对其进行特征的提取。采用深度神经网络,可以充分描述特征之间的相关关系。因此,我们把语音时频特征送入层叠自动编码器深度学习网络进行特征的自动提取。由于深度学习网络采用模拟人脑的多层结构,可以逐级地进行信息特征的抽取,最终形成适合模式分类的较理想特征。这种深层次结构与人脑处理信息有很大相似性,是分层处理的。

但随着层数的增加,深度神经网络会变得越来越难以训练,梯度消失或者梯度爆炸问题都会阻止网络达到收敛状态,并且网络性能并不一定随着网络层数的增加而提升。深度残差网络则可以解决深层网络的性能降低问题。相比于普通的深度神经网络,它更容易优化,随着层数的增多,它的性能表现逐渐提升,并且它具有足够的泛化能力。对于网络层数的选取来说，多少层都是可以的,但是对于信号识别这个问题,层数太少使深度残差学习网络的精度不够高,层数太多在训练过程中需要耗费大量的成本。关于此项发明创造,构建30层深度残差学习网络具有最优性能,所以本发明中构建的是30层深度残差学习网络。残差网络就解决了网络过深从而导致效果不好的问题，深度学习网络的建模技术,能够无缝地与传统的时频处理技术和分类器相结合,在不引起任何额外耗费的情况下,大幅度地提升诊断的精确度。

1. 时域信号的预处理

音频信号是一种非平稳的时变信号，它携带着各种信息。一般而言音频处理的目的有两种，一种是对音频信号进行分析，提取特征参数，用于后续处理；一种是加工音频信号，如在语音增强中对含噪语音进行背景噪声抑制，以获得相对“干净”的语音。根据分析参数不同，可分为时域分析和变换域（频域、倒谱域）分析，其中时域分析是最简单、最直观的方法，它直接对音频信号的时域波形进行分析，提取主要语音短时能量和平均幅度、短时平均过零率、短时自相关函数和短时平均幅度差函数等。实际的音频信号是模拟信号，因此在对音频信号进行数字处理之前，首先要将模拟语音信号S(t)以采样周期T采样，将其离散化为S(n)，采样周期的选择应根据模拟语音信号的带宽来确定，以避免信号的频域混叠失真。音频信号的预处理一般包括预加重处理和加窗处理。对输入的数字音频进行预加重，其目的是为了对音频的高频部分进行加重，去除外界环境辐射的影响，增加音频的高频分辨率。一般通过传递函数为H(z)=1-az^-1的高通数字滤波器来实现预加重，其中a为预加重系数，一般为0.9<a<1。设n时刻语音采样值为x(n),经过预加重处理后的结果为y(n)=x(n)-ax(n-1)

在进行预加重数字滤波处理后，接下来进行加窗分帧处理，音频信号是一种随时间而变化信号，主要分为浊音和清音两大类，由于发声物体的惯性运动，可以认为在一小段时间里（一般为10ms-30ms），音频信号近似不变，即音频信号具有短时平稳性。这样，可以把音频信号分为一些短段来进行处理，音频信号的分帧是采用可移动的有限长度窗口进行加权的办法来实现的，一般每秒的帧数为33-100帧，分帧虽然可以采用连续分段的方法，但一般要采用如图所示交叠分段的方法，为了使帧与帧之间平滑过渡，保持连续性，前一帧与后一帧的交叠部分为帧移，帧移和帧长的比值一般为0-1/2。

1. 深度残差网络在时域信号分析中的优越性

主要体现为随着层数的增加,深度神经网络会变得越来越难以训练,梯度消失或者梯度爆炸问题都会阻止网络达到收敛状态,并且网络性能并不一定随着网络层数的增加而提升。深度残差网络则可以解决深层网络的性能降低问题。相比于普通的深度神经网络,它更容易优化，在将时域信号分析过程中，帧与帧之间可以相对的平滑过渡。

具体为：音频信号首先被分割为相互重叠的若干段帧,对每一段加窗处理，但由于其维度较高直接进行模式识别，计算量大且识别精度高,因而需要对其进行特征的提取（目的为去除外界环境辐射的影响，增加音频的高频分辨率）。对此采用深度神经网络,可以充分描述特征之间的相关关系。但随着层数的增加,深度神经网络会变得越来越难以训练,梯度消失或者梯度爆炸问题都会阻止网络达到收敛状态,并且网络性能并不一定随着网络层数的增加而提升。深度残差网络则可以解决深层网络的性能降低问题。相比于普通的深度神经网络,它更容易优化,随着层数的增多,它的性能表现逐渐提升,并且它具有足够的泛化能力。这对于层数较多的深度神经网络有很大的优化。

由于时域信号在实际应用中存在各种问题，如：

* 传统时域信号在应用中精度不够。
* 在复杂场景中传统时域信号的精确度明显不足。

本发明创造即为致力于解决以上问题。

1. 本发明创造在时域信号输入时进行了一个预处理，且传统的时域信号分析是直接对时域信号进行一个处理，而本产品的创新点就是在传统的分析基础上，将时域信号转化为频域信号，时域可以直观的观测到信号的形状，但不能用有限的参数对信号进行准确的描述。对信号进行时域分析时，有时一些信号的时域参数相同，但并不能说明信号就完全相同，因为信号不仅随时间变化，还与频率、相位等信息有关，这就需要进一步分析信号的频率结构，并在频率域中对信号进行描述。频域分析可以将复杂信号分解为简单的信号（正弦信号）的叠加，可以更加精确的了解信号的“构造”。从而解决了精度不够的问题
2. 而在复杂场景中也是同样，当我们在系统中把时域信号转换为频域信号，频域信号就可忽略某些参数影响，从而解决在复杂场景中信号分析精度不够的问题。

具体实施方案：

本发明创造涉及深度学习以及残差网络，具体基于短时傅里叶变换以及神经网络中的深度残差网络一种时域信号分析系统。

该系统可应用于多个领域，如初学者的钢琴教学及打分，也可用于机器故障的识别，广泛应用于工业上的超声探测信号处理（工业），乐器的音调识别（民生，教育，艺术），地理地形震动探测等领域，在此选取汉语声音音频的实例。应当理解，本发明只解释于本发明，但不限于本发明。

实例分析过程如下：

本实例采用在汉语的条件下，对90%以上的拼音结构组成进行行辨识和分类，实际上是一个针对于一门复杂语言的多分类问题（非语义标识），并不结合语义语境做出解释。之所以选择汉语，是因为汉语是一种象形文字，有着较为复杂的读音结构，能发出较多的声音信号，且具有规律性，较为典型。而本项目的方向为多种信号的共同解决方案。

1. 由声音采集设备采集发生物体发出的时域信号（当前例子中为人或计算机发出的汉语语音）并输入处理设备。
2. 对采集数据集中进行预处理。
3. 将时域信号（汉语声音信号）进行傅立叶快速变换得到频域信号。
4. 频域信号通过深度残差网络进行标准化处理：

当输入频域信号之后（参数大致为3000），经过全连接层输出之后输出神经元参数为（6000），之后进入Basic normalization层进行标准化操作，进入BN层之后。

1. 首先进行（X - means）／variance此项操作 ，(其中means为平均值，其更新方式为 ：

rate \*本次的训练means + （1-rate）\*上一次的means

variance的更新方式和means相同，variance为标准差，即为：

rate \*本次的训练variance + （1-rate）\*上一次的variance)

1. 对之前求得的值进行线性映射，其中w和b随网络进行线性调整。

其中，BN保证了了神经网络单元保持活性。

当结束BN层的操作后，再一次经过全连接层输出之后，进入BN层进行标准化处理，共经过四次标准化处理后，将以上所述结构作为一个整体，将此结构循环运行七次，第八次运行时经过全连接层后的神经元参数为（12000），再重复上述循环结构七次，跳出循环后，运用softmax函数。

其中运用到shortcut链接与激活函数relu以及激活函数softmax，shortcut链接作用在残差网络中的某几层，而激活函数relu作用在每一层全连接层处理，激活函数softmax作用在最后的数据处理中。起分类作用。shortcut链接作用：构成了残差网络的一部分，对输入的参数进行全联接（Dense）处理以及线性映射（linear）。

1. 激活函数 relu作用：对于人工神经网络而言激活函数起到的作用可以说就是**进行非线性变换，增加系统的非线性表达能力**，没有其他的作用，但这些简单非线性激活函数的引入可以加快了网络的训练速度，在深层的网络中表现出了明显的优势。
2. 分类函数softmax：softmax用于多分类过程中，它将多个神经元的输出，映射到（0,1）区间内，可以看成概率来理解，从而来进行多分类！假设我们有一个数组，V，Vi表示V中的第i个元素，那么这个元素的softmax值就是Si=ei\ej
3. 进入残差网络，进行训练

深度残差学习网络构建完成之后,对采集信号数据集中进行预处理,将预处理后的图像作为深度残差学习网络的输入进行训练,同时对验证集进行过拟合检查,即判断深度残差学习网络对验证集图像处理的损失函数是否呈现先下降,后反而上升的现象,当验证集的损失函数不再下降时候,即可暂时停止训练。将预处理后的测试集图像作为训练终止后的深度残差学习网络的输入。

残差网络中的优化器：Adam

优化器是模型训练的指导教练，能告诉模型权值该怎么调整，调整多大量，最终目标是让模型权值 theta 调整到最优，使得代价函数 f(theta) 最小。Adam 是一种基于基于梯度的优化算法，该方法实现简洁，计算高效，内存占用少，适合非平稳目标函数，超参数有符合直觉的解释，无需复杂的调参过程。

该网络模型为：

网络的输⼊为常用声音信号在时域下的的特征矩阵（3000）

输入为已知时域下的汉语声音信号，输出为对应的所需结构。

网络模型的具体组成：

网络由传统的两层全联接网络组成 具体参数见图结构 。网络模型的输出为常用的1184种拼⾳音结构 如 a（一）a（二声）为两种 。训练量和检验量为八种音色，分别为：电子合成音男，电子合成音女，情感男声，情感女声，以及以上四种分别加 gauss白噪音之后的叠加音 其中七个音色为训练量，一个音色为检验量，且训练结果正确率在百分之95以上，收敛效果良好。

由此可知，本系统概括为：（如图一所示）

1. 由音频采集设备从外界采集时域信号，且输入至主控设备。
2. 对采集数据集中进行预处理。
3. 将时域信号进行傅立叶快速变换得到频域信号。
4. 频域信号通过深度残差网络进行标准化处理。
5. 进入残差网络，开始训练。
6. 对音频进行一个编码。
7. 输出编码至所需设备，进行应用。

将时域信号转换为频域信号（转换过程如图二所示）

特别的，本系统在进入残差网络训练前对时域信号进行一个处理，将时域信号转换为频域信号，此过程是傅立叶变换之后，对频域信号数据进行标准化、归一化处理，作为网络的输入信号。频域分析可以将复杂信号分解为简单的信号（正弦信号）的叠加，可以更加精确的了解信号的“构造”。以此来解决传统时域信号精确度不足的问题。

网络模型（如图三所示）

本发明使用的是较为传统的残差网络结构，并没有用到二维数据处理，全部为一维数据处理（这也是时域信号转为频域信号的一个特征），shortcut函数在一维的情况下，相当于没有激活函数的全联接层，其作用为线性映射保证在输出相加的时候，维度和矩阵形式是相同的，BN层在每一个残差里的具体设置同样重要，具体分析及具体参数如下。

当输入频域信号之后（参数大致为3000），经过全连接层输出之后输出神经元参数为（6000），之后进入Basic normalization层进行标准化操作，进入BN层之后。

1. 首先进行（X - means）／variance此项操作 ，(其中means为平均值，其更新方式为 ：

rate \*本次的训练means + （1-rate）\*上一次的means

variance的更新方式和means相同，variance为标准差，即为：

rate \*本次的训练variance + （1-rate）\*上一次的variance)

1. 对之前求得的值进行线性映射，其中w和b随网络进行线性调整。

其中，BN保证了了神经网络单元保持活性。

当结束BN层的操作后，再一次经过全连接层输出之后，进入BN层进行标准化处理，共经过四次标准化处理后，将以上所述结构作为一个整体，将此结构循环运行七次，第八次运行时经过全连接层后的神经元参数为（12000），再重复上述循环结构七次，跳出循环后，运用softmax函数。

其中运用到shortcut链接与激活函数relu以及激活函数softmax，shortcut链接作用在残差网络中的某几层，而激活函数relu作用在每一层全连接层处理，激活函数softmax作用在最后的数据处理中。起分类作用。shortcut链接作用：构成了残差网络的一部分，对输入的参数进行全联接（Dense）处理以及线性映射（linear）。

1. 激活函数 relu作用：对于人工神经网络而言激活函数起到的作用可以说就是**进行非线性变换，增加系统的非线性表达能力**，没有其他的作用，但这些简单非线性激活函数的引入可以加快了网络的训练速度，在深层的网络中表现出了明显的优势。
2. 分类函数softmax：softmax用于多分类过程中，它将多个神经元的输出，映射到（0,1）区间内，可以看成概率来理解，从而来进行多分类！假设我们有一个数组，V，Vi表示V中的第i个元素，那么这个元素的softmax值就是Si=ei\ej

实验结果

本发明创造进行多次试验后，得出一个平均值，正确率为96.83% ，在当前环境，传统方案得信号识别正确率为82.7%。该具体试验只是为了验证我们的方法的可行性，在一些其他时域方案中，本系统会在参数上做相应的调整，会有一部分网络上小范围的参数更改，但本系统的思路和处理框架不会变。

发明意义

本发明创造延伸出来的结果是在时域信号处理的过程中使用我们的方法将会有很优秀的表现，且足够克服传统方案中所无法攻克的精度和复杂环境问题。