文章编号：0427-7104(2021)03-0360-07

基于挤压和激励残差网络的歌声检测

桂文明[1,2]，吕家伟[1]，梁颖红[1]，敖志强[3]

(1. 金陵科技学院 软件工程学院，江苏 南京 211169；2. 南京邮电大学 宽带无线通信与传感网技术教育部重点实验室，江苏 南京 210003；3. 南昌航空大学 软件学院，江西 南昌 330063)

摘 要：本文提出一种基于挤压和激励残差网络的歌声检测算法，运用该算法，不需要对音乐信号进行复杂的特征工程处理，仅需对网络输入简单朴素的声学特征，便能通过多层次卷积以及挤压和激励操作，学习到更多的有效特征，从而达到比当前流行的检测算法更强的性能. 算法中，残差结构使得网络可以轻松扩展深度，挤压和激励模块能对深度残差网络中学习到的多个特征进行自动融合，进而使得学习到的歌声特征整体更有效. 为了验证算法的可行性和有效性，本文选择了2个公开的数据集进行实验，并以目前性能最好的歌声检测框架之一作为基线系统，实验结果证明了本算法的性能领先于基线系统.

关键词：歌声检测；音乐信息检索；挤压和激励网络；残差网络；卷积神经网络

中图分类号：TP391 **文献标志码**：A

DOI:10.15943/j.cnki.fdxb-jns.2021.03.015

歌声检测(Singing Voice Detection，SVD)是判断存在于数字音频形式的音乐中的每一小段音频是否含有人的歌声的过程，其检测精度一般在50～200 ms之间. 在每一小段音乐中，除了歌声，一般还含有演奏乐器的声音，要在混合乐器和人声的音乐片段中判断是否含有歌声，虽然对人来说是轻而易举的，但对机器来说却是颇具挑战性的工作. 歌声检测是音乐信息检索(Music Information Retrieval，MIR)领域重要的基础性工作，很多其他研究比如歌手识别、歌声分离、歌词对齐等都把歌声检测作为事前必备技术或者增强技术. 例如，在歌手识别过程中，首先对音乐进行歌声检测就是事前必备技术，只有检测到歌声后才能通过歌手鉴别过程进行歌手识别；在歌词对齐过程中，如果能准确地检测出歌声的位置，那么必然增强歌词对齐的准确性.

歌声检测的过程一般包括预处理、特征提取、分类和后处理等几部分，其中特征提取和分类是最重要的两大步骤. 输入的音频文件一般是物理样本级的，例如wav，mp3等文件. 特征提取是从音频信号中提取能表达含有歌声和不含歌声的音频之间区别的鉴别信息. 较简单的鉴别信息是短时傅里叶变换后的时频图，常用的特征还包括线性预测系数(Linear Predictive Coefficient，LPC)、感知线性预测系数(Perceptual Linear Predictive Coefficient，PLPC)、过零率(Zero Cross Rate，ZCR)、Mel频率倒谱系数(Mel Frequency Cepstral Coefficient，MFCC)、动谱特征(Fluctogram)、谱平坦因子(Spectral flatness)、谱收缩因子(Spectral contraction)等. 这里的大多数特征都是在时频图基础上提取的. 分类过程是采取机器学习等方法对特征信息进行分类，并根据特征分类来检测歌声. 主要的分类方法包括基于传统分类器的方法和基于深度神经网络(Deep Neural Network，DNN)的方法，前者包括支持向量机(Support Vector Machine，SVM)、隐马尔可夫模型(Hidden Markov Model，HMM)、随机森林(Random Forest，RF)等；后者包括采用卷积神经网络(Convolutional Neural Network，CNN)[1]和循环神经网络(Recurrent Neural Network，RNN)[2]的方法.

在现有的歌声检测算法中，研究者们总是试图通过精心设计某种特征，然后选择某种分类器进行分类. 当单一特征不能满足要求时，则采用组合多种特征[3-4]的方法，可以说歌声检测的发展历史就是研究者们寻找和设计特征的历史. 这种特征工程(Feature engineering)存在的弊端是人工设计特征周期长，以

收稿日期：2021-01-02

基金项目：国家自然科学基金(61872199)

作者简介：桂文明(1974—)，男，博士，副教授，E-mail：guiwenming@126.com

及设计的特征适应性不可靠.事实上,DNN 不仅可以充当歌声检测框架的分类器作用,还可以通过多层次的学习,对歌声进行多层次的特征提取[5].因此,一方面,采用适当的 DNN 框架,可以学习到歌声的特征,不需要进行复杂的特征工程;另一方面,DNN 框架既可充当特征提取器又可充当分类器,可减少环节,使算法框架更简单.在歌声检测的现有算法框架中,DNN 既作为特征提取器又作为分类器的框架并不多,大部分基于 DNN 框架的处理过程是先进行复杂的特征工程,然后再把特征输入 DNN 分类器.据我们所知,仅输入简单朴素的特征如对数 Mel 时频图的工作,只有 Schlüter 等的 CNN 方案[1,6].然而在该方案中,CNN 的深度有限,仅有 14 层,我们称之为浅层 CNN(Shallower Convolutional Neural Network, SCNN).受限于浅层深度,网络的学习能力有限,从而导致学习到的歌声特征有限.如果在 Schlüter 等的浅层方案中想进一步通过简单堆叠卷积层来达到提高深度的目的则是无法实现的,因为这会导致梯度问题,使得堆叠的网络无法训练或退化.本文提出一种基于挤压和激励残差网络的歌声检测算法,一方面,残差网络使得网络深度在可以避免梯度问题和退化问题的情况下对深度进行扩展;另一方面,挤压和激励网络可通过学习调整各层特征的重要性,并自动融合这些特征,送入到网络的下一层.

1　相关工作

1.1　残差网络

残差网络(Residual Network, ResNet)来源于图像分类领域,其在很大程度上解决了梯度爆炸和网络退化问题,使得网络可以构建得很深[7].残差网络由残差结构叠加组成,残差结构的一般构造如图 1(a)所示.残差网络在原有网络上堆叠身份映射(Identity),且能保持网络的性能不变.它不是直接学习堆叠网络的潜在映射 $H(x)$,而是通过增加身份映射后拟合一个残差映射(Residual mapping)$F(x)=H(x)-x$.残差映射相比潜在映射更容易优化,从而解决了深度增加后导致的梯度问题和网络退化问题.此外,残差结构是无侵入式结构,可以叠加到其他网络中,用以提升网络的深度和性能.

图 1(a)中的 $F(x)$可根据需要采取不同的结构.图 1(b)和图 1(c)是用于构建深度残差卷积神经网络的两种典型的 $F(x)$的结构:基本块(Basic block)和瓶颈块(Bottleneck block)的结构.图 1(b)和图 1(c)中,n,m 分别表示经过 1×1 或 3×3 卷积后的特征图数量,也就是通道数量.图 1(b)与图 1(c)的不同在于后者具有 3 个卷积层,且中间特征图数量也发生了变化,但二者的输出特征图数量都是相同的.

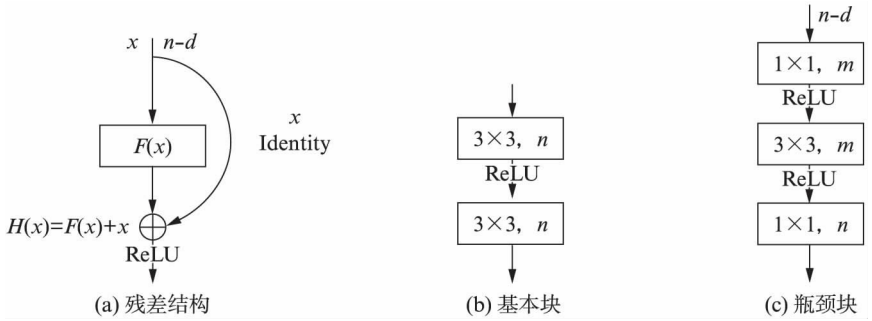

图 1　残差结构和两种典型的 $F(x)$块结构

Fig. 1　Residual structure and the two typical block for $F(x)$

1.2　挤压和激励操作

挤压和激励网络(Squeeze-and-Excitation Network, SENet)来源于图像分类,在 LSVRC 2017(ImageNet Large Scale Visual Recognition Challenge)图像分类竞赛中,基于该技术的网络获得了最好的成绩.SENet 的核心是利用挤压和激励操作挖掘卷积神经网络中通道间的关系,并调整通道的权重.SENet 模块的结构见图 2(见第 362 页).假定上一层卷积输出 F 是高和宽为 h 和 w 的图片,通道数量为 c,挤压操作是一个全局平局池化层,将 c 个通道压缩成 c 个描述符;激励操作的第 1 步是一个门机制,具体包括第 1 个全连接层将 c 个描述符以 r 倍降维,然后利用 ReLU 函数进行非线性化,接着是第 2 个全连接层以 r 倍增维;激励操作第 2 步首先利用 Sigmoid 激活函数对通道进行权重估值,然后通过 Scale 操作

对各通道按权重估值进行调整,最后调整后的通道 F' 进入下一层网络.挤压和激励操作使得各通道对下一层网络的作用发生变化,权重不再是相等的,而是通过学习得到的.本算法中,将 F 看成是学习到的特征,该特征先经过权重重估和调整,再送入下一层网络.特征权重重估和调整的过程就是特征自动融合的过程.

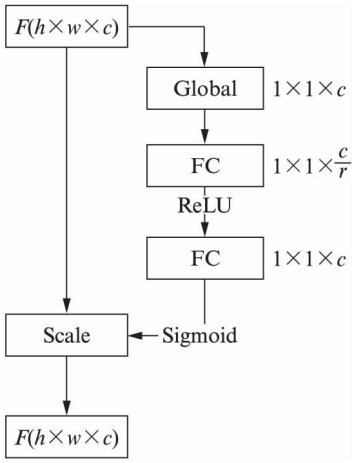

图 2 SENet 模块图

Fig. 2 SENet module

2 算法思路

本文基于挤压和激励残差网络的歌声检测算法是通过残差网络构建深度卷积神经网络,深度可到 200 甚至更深,从而增强对歌声特征的学习能力,产生不同层次的特征;而对于卷积神经网络学习到的不同层次的歌声特征,本算法通过挤压和激励模块来重估其对歌声分类的重要性.本文算法较浅层 CNN 的方法有两个方面的改进:一方面是在网络深度方面能避免出现退化现象的情况下得到扩展;另一方面是各层次的特征权重得到重估,使得分类效果得到提升.

2.1 网络输入

本算法的网络输入不是经过复杂特征工程处理后的特征,而是歌声检测最常用的、最简单朴素的对数 Mel 时频图(Log Mel-spectrogram).计算过程中采样率取为 22 050 Hz,帧长为 1 024,帧移为 315,Mel 频率数量取 80 个,频率区间为[27.5,8 000.0](单位 Hz).每个音频文件可得到一个行数为 80 的对数 Mel 时频图矩阵,我们从该矩阵的起始列位置开始逐个提取大小为 80×115 的图像,读取图像时每跳为 5 列,然后将图像输入到网络.

2.2 网络模块

结合挤压和激励操作的深度残差卷积模块(SEResNet)如图 3 所示.SEResNet 模块由 SENet 模块结合 ResNet 的基本块和瓶颈块构成.两种模块中,输入通道数量和输出通道数量都是一致的.我们根据网络的深度选择不同模块来构造全栈网络.其中 SRB 模块用于构造深度为 18 和 34 的网络,而 SRT 模块用来构造更深的网络,包括深度为 50,101,152 和 200 的网络.

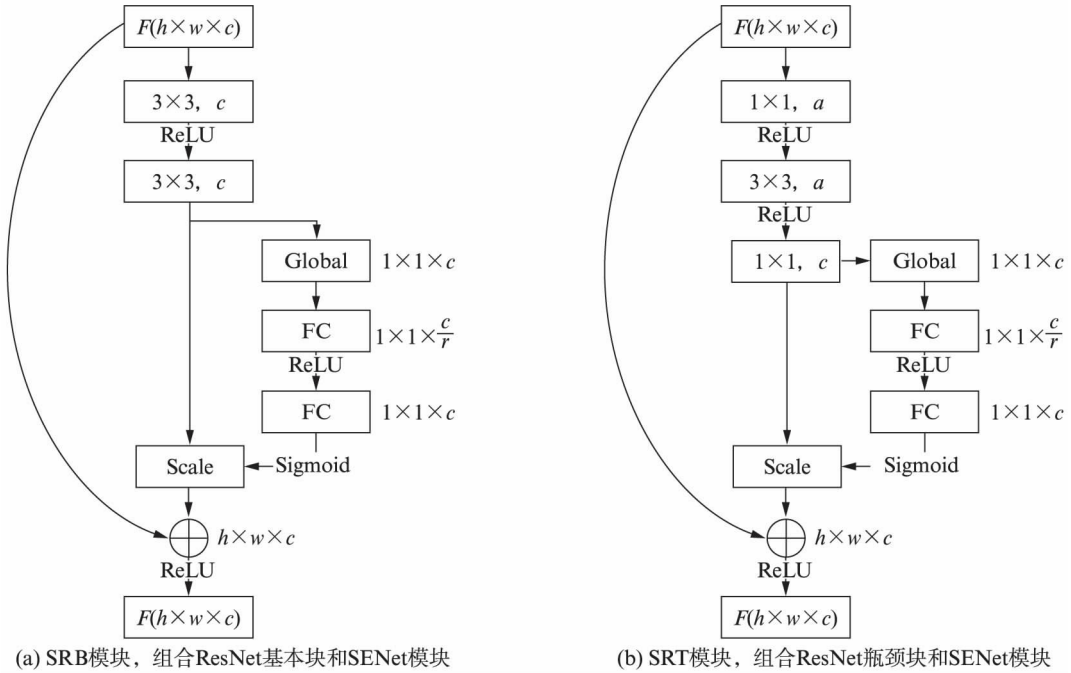

图 3 SEResNet 模块

Fig. 3 SEResNet module

2.3　全栈网络

如前所述，全栈网络是根据不同的 SEResNet 模块来构建的. 本文算法设计的 SEResNet 深度包括 18，34，50，101，152 和 200 共 6 种深度，下面以 SEResNet200（深度为 200 的 SEResNet）为例来说明全栈网络的结构，如表 1 所示. 大小为 80×115 的图像在进入残差网络前，先经过大小为 7×7，步数（Stride）为 2 的卷积层 conv1，此时，conv1 的输出将缩小为 40×58. 随后，输出的特征图进入挤压和激励残差结构中. 在 conv2，先经过大小为 3×3，步数为 2 的最大值池化层，特征图大小再次缩小至 20×29，随后，进入挤压和激励残差网络层，经过 1 个 SRT 残差块，再经过挤压和激励模块进行特征权重重估，图中 fc，[16，256] 表示挤压和激励网络结构中第 1 个和第 2 个全连接卷积层 FC（图 2）的输出维数，降维和增维倍数均为 16. 中括号外的×3 表示中括号内挤压和激励残差网络栈的深度为 3，即网络堆叠的深度. 对应 conv3，conv4，conv5 的堆叠深度分别为 12，48 和 3，我们称这 4 个堆叠深度序列为规模参数. 在 conv3，conv4，conv5 中，没有最大值池化层，但是由于它们的结构中存在步数为 2 的卷积层，因此，图像大小仍然和 conv2 层一样缩减为一半. 在最后的 FC 层，先经过一个 2 维的自适应平均池化层，其输出通道数为 1，再进入一个全连接卷积层，最终网络输出为 1 维向量 $\boldsymbol{o}$，含 2 个值 o_0，o_1，可以用来判断是否含有歌声.

表 1　深度为 200 的 SEResNet 的全栈网络结构

Tab. 1　The network structure of the SEResNet with the depth 200

卷积层名称	网络结构	输出图片大小
conv1	7×7，stride 2 3×3，max pool，stride 2	40×58
conv2	[1×1，64 3×3，64 1×1，256 fc，[16，256]] ×3	20×29
conv3	[1×1，128 3×3，128 1×1，512 fc，[32，512]] ×12	10×15
conv4	[1×1，256 3×3，256 1×1，1 024 fc，[64，1 024]] ×48	5×8
conv5	[1×1，512 3×3，512 1×1，2 048 fc，[128，2 048]] ×3	3×4
fc	average pool，fc，2-d	1×1

对于深度分别为 18，34，50，101 和 152 的网络，其规模参数分别为[2，2，2，2]，[3，4，6，3]，[3，4，6，3]，[3，4，23，3] 和 [3，8，36，3]. 输出图片大小和 SEResNet200 保持一致. 值得注意的是，规模参数并不是必须如上所述. 事实上，我们通过实验发现规模参数呈递增形，即形如[3，8，16，23] 的 SEResNet152 效果要更好，但是本文的重点在于“深度”和“特征的权重重估”，因此对于 SERenet 中各深度网络的具体组织形式没有进行研究.

2.4　加权交叉熵损失函数

由于歌声检测是二分类，所以本算法采用的是二分类交叉熵损失函数. 歌声检测的数据集中歌声和非歌声的样本数一般是不平衡的，通常是歌声样本数要多于非歌声样本数，因此，我们在损失函数中加入了权重，权重设为数据集中的样本数量比例. 上述挤压和激励残差网络的输出是 2 个值 o_0，o_1，先用 sigmoid 函数转换成概率值，再加入到下述损失函数中进行计算. 设 N 个样本预测为歌声的概率为 x_i，样本的标签为 y_i，权重为 w_i，其中 $i\in[1,N]$，则加权交叉熵损失函数为

$$l(x,y)=-\frac{1}{N}\sum_i w_i[y_i\log x_i+(1-y_i)\log(1-x_i)]. \tag{1}$$

这里的对数函数的底可以为 2，e，10.

3　实验和结果

3.1　数据集和基线系统

我们选择公开音乐数据集 RWC（Real Word Computing）中的流行歌曲[8] 和公开数据集 Jamendo[9]（简称 JMD）作为实验数据集. RWC 包含 100 首流行歌曲，时长共 407 min. 我们把 RWC 分成训练、验证和测试 3 个数据集，划分方式是将数据集文件结尾为 0—4 的文件划为训练集，将结尾为 5 和 6 的文件划为验证集，将结尾为 7—9 的文件划为测试集，这种划分是准随机的方法，以保证实验结果的公正性. JMD

包含 93 首歌曲,时长共 371 min. 我们保持 JMD 的训练、验证和测试集不变[10]. RWC 和 JMD 的歌声和非歌声的样本数量比分别为 1.12 和 1.55.

我们选择文献[10]中的系统作为比较的基线系统,该文献实现了目前国际上最先进的歌声检测算法,包括基于 SCNN 的模型,并公开了代码,可以认为是一个第三方的评估系统. 该 SCNN 包含 4 个卷积层和 3 个全连接层,是目前获得检测准确率最高的框架之一,网络输入正是对数 Mel 时频图. 我们将直接引用并比较该文献提供的 JMD 数据集上的实验结果. 对于 RWC,我们运行该系统的代码产生实验结果. 本算法采用 Pytorch,并借助 Homura 包(https://github.com/moskomule/homura.)进行开发和实现. 对于 JMD 数据集,SCNN 有最高准确率为 93.2%的报告[4],但因该系统实施了特征工程和数据增强,且没有提供实现细节,故没有作为基线系统.

3.2 结果比较

为了公正比较,在实验中我们没有通过调参选取最好的结果来进行比较,而是保持除深度之外所有的参数不变. 我们这里选取不同深度的挤压和激励残差网络进行实验是为了研究深度对检测结果的影响. 由于在歌声检测中,我们推荐深度在 50 以上的网络,因此,我们取深度为 50,101,152 和 200 的统计数据作为比较的数据,结果见表 2.

表 2 本文算法和 SCNN 在 JMD 和 RWC 下的结果比较

Tab. 2 Comparison between the proposed algorithm and the SCNN on JMD and RWC

数据集	算法模式	λ_A/%	F/%	λ_P/%	λ_R/%	k_{FP}	k_{FN}
JMD	SCNN	86.80	86.30	83.70	89.10	15.10	10.90
	SEResNet $\mu\pm\sigma$	88.25±0.30	87.85±0.41	84.73±1.18	91.25±1.95	14.35±1.61	8.75±1.95
	提升*	1.45	1.55	1.03	2.15	−0.75	−2.15
RWC	SCNN	87.94	89.73	91.46	88.07	12.25	11.93
	SEResNet $\mu\pm\sigma$	91.41±0.55	93.09±1.01	93.38±1.45	92.81±0.92	11.17±1.28	7.19±0.92
	提升*	3.47	3.36	1.92	4.74	−1.08	−4.74

注:其中 $\mu\pm\sigma$ 指的是相应指标的均值和方差. * 表示二者之差. k_{FP} 表示预测类别为歌声,真实类别为非歌声的样本数,即假正例的数量;k_{FN} 表示预测类别为非歌声,真实类别为歌声的样本数,即假负例的数量.

从表 2 中可看出,本算法所有指标均有不同程度的提升,这说明相对于 SCNN,本文算法的"深度"和"特征权重重估"的有效性. 在 JMD 和 RWC 上,本文算法的准确率(Accuracy)λ_A 均值分别较 SCNN 的有所提升. F 值(F-measure)是精确率(Precision)λ_P 和召回率(Recall)λ_R 的综合,本文算法的 F 值也较 SCNN 的有所提升. 假负例(FN)数量 k_{FN} 在 JMD 和 RWC 上都降低最多,这带来了召回率提升最多的效果.

3.3 不同深度对结果的影响

由于数据集的特征分布和数据集大小不一样,不同深度的 SEResnet 表现也不一样. 图 4 是在 JMD 和 RWC 上不同深度的准确率的变化情况. RWC 上的准确率随着深度的增加有一个显著的上升趋势,而 JMD 上的则曲折上升,二者的最高准确率均落在深度较大的网络上,这告诉我们在应用 SEResNet 时应首先考虑深度较大的网络. 需要说明的是为全面评估深度的影响,图中增加了深度为 18 和 34 的数据.

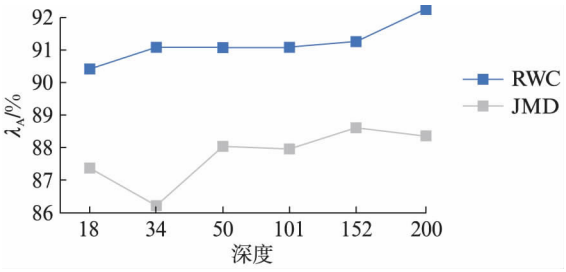

图 4 不同深度的 SEResNet 检测准确率的变化情况

Fig. 4 The accuracies of SEResNets change at the different depths

4 结 语

本文提出了一种基于挤压和激励残差网络的歌声检测算法,残差结构使得网络的深度可以扩张至 200 层甚至更多,挤压和激励嵌入在残差结构中,可对网络各层次学习到的特征进行权重重估,从而弱化对歌声检测权重小的特征,而强化权重大的特征. 通过实验证实,本算法的准确率等指标相对 SCNN 均有

提升. 进一步地，通过对深度为 18，34，50，101，152，200 的挤压和激励残差网络进行实验，检测算法的最佳性能均体现在较大的深度上，这说明在实际中应用本算法时，应该重视较大深度的网络. 值得注意的是，本算法和其他基于深度学习的算法一样，在一定程度上依赖于训练数据集的构造，其泛化效果有待验证. 比如本算法的训练数据主要是含有歌声和乐器的混合音乐，而其模型是否适用于说唱类型的音乐尚需进一步研究. 提升模型泛化性能的一种解决方案是在训练数据集中加入目标检测类型的音乐，使得模型能学习到该类型音乐中歌声的特征.

致谢：感谢南京邮电大学宽带无线通信与传感网技术教育部重点实验室开放研究基金资助；感谢金陵科技学院和澳大利亚昆士兰科技大学中外合作办学高水平示范性建设工程资助；感谢江苏省教育厅高校优秀中青年教师和校长境外研修项目资助.

参考文献：

[1] SCHLÜTER J. Learning to pinpoint singing voice from weakly labeled examples [C]// International Society for Music Information Retrieval. New York, USA: ISMIR, 2016: 44-50.

[2] LEHNER B, WIDMER G, BÖCK S. A low-latency real-time-capable singing voice detection method with LSTM recurrent neural networks [C]// 23rd European Signal Processing Conference. Nice, France: IEEE, 2015: 21-25.

[3] LEHNER B, WIDMER G, SONNLEITNER R. On the reduction of false positives in singing voice detection [C]// IEEE International Conference on Acoustics, Speech and Signal Processing. Florence, Italy: IEEE, 2014: 7480-7484.

[4] LEHNER B, SCHLÜTER J, WIDMER G. Online, loudness-invariant vocal detection in mixed music signals [J]. *IEEE/ACM Transactions On Audio, Speech, and Language Processing*, 2018, **26**(8): 1369-1380.

[5] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]// European conference on computer vision. Zurich, Switzerland: Springer, 2014: 818-833.

[6] SCHLÜTER J, GRILL T. Exploring data augmentation for improved singing voice detection with neural networks [C]// International Society for Music Information Retrieval. Malaga, Spain: ISMIR, 2015: 121-126.

[7] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 770-778.

[8] GOTO M, HASHIGUCHI H, NISHIMURA T, et al. RWC music database: Popular, classical and jazz music databases [C]// Proceedings of the 3rd International Conference on Music Information Retrieval. Paris, France: ISMIR, 2002, **2**: 287-288.

[9] RAMONA M, RICHARD G, DAVID B. Vocal detection in music with support vector machines [C]// 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. Nevada, USA: IEEE, 2008: 1885-1888.

[10] LEE K, CHOI K, NAM J. Revisiting singing voice detection: A quantitative review and the future outlook [C]// International Society for Music Information Retrieval Conference. Paris, France: ISMIR, 2018: 155-161.

Singing Voice Detection Algorithm Based on a Squeeze-and-Excitation Residual Network

GUI Wenming[1,2], LÜ Jiawei[1], LIANG Yinghong[1], AO Zhiqiang[3]

(1. *School of Software Engineering, Jinling Institute of Technology, Nanjing, Jiangsu* 211169, *China*;
2. *Key Lab of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu* 210003, *China*;
3. *School of Software, Nanchang Hangkong University, Nanchang, Jiangxi* 330063, *China*)

Abstract: In this paper, we proposed an algorithm based on the squeeze-and-excitation residual network. Other than those algorithms armed with complicated feature engineering, the proposed network could learn more effective features by the hierarchical convolution collaborated with the squeeze-and-excitation operation, only fed with the naive acoustic feature. In this algorithm, the residual structure can easily extend the depth of convolutional network, and the squeeze-and-excitation operation can fuse the learned multiple features by the adjusted weights, and furtherly can improve the overall performances. To prove the feasibility and effectiveness, we conducted the experiments on the two public datasets. Compared with one of the state of the art base line, the proposed algorithm produced the significantly better performance.

Keywords: singing voice detection; music information retrieval; squeeze-and-excitation network; residual network; convolutional neural network

通向统一存储器之路　微电子学院周鹏团队发现超快电荷存储原理

以“四个面向”为引领，复旦大学微电子学院教授周鹏团队针对主流电荷存储器技术，发现了制约硅基闪存技术的原理瓶颈，提供了可以应用于硅材料的器件模型，实现了匹敌易失内存技术的超快速度，为统一存储器的发展提供了技术途径. 北京时间 6 月 3 日，相关成果以“Ultrafast non-volatile flash memory based on van der Waals heterostructures”为题在 *Nature Nanotechnology* 在线发表.

闪存自 1967 年被发明以来，由于其高密度低成本的特性，已经占据了先进非易失存储技术 99％的市场. 然而自从东芝公司实现商业化技术后，工作在量子隧穿机制下的硅基闪存编程时间一直在百微秒量级，无法实现对速度有较高要求的内存级应用. 那么量子隧穿机制是注定不能实现更快的速度吗？

周鹏团队从源头出发，首次发现了双三角隧穿势垒超快电荷存储机理，突破了传统经验束缚，获得了内存 DRAM 技术级编程速度. 研究发现，在存储与擦除的工作过程中，势垒高度决定了电荷隧穿通过的难易程度，栅耦合比决定了栅极控制电压产生的电荷密度，良好界面保证了不会引入额外沾污或缺陷. 从以上三大方面看，现有的硅/氧化硅界面非常完美，周鹏团队发现并证明了栅耦合比、势垒高度是决定电荷存储器速度的根本因素.

周鹏团队根据此超快电荷存储原理建立了通用器件模型，设计并制备出同时具备三大要素的范德华异质结闪存，采用工业界标准阈值漂移测试和高温加速老化测试方案，验证了 20 ns 编程时间和 10 年数据保持能力；并对器件进行了理论模拟计算，实验数据和理论模拟结果吻合一致；同时探讨了三大要素的不同程度缺失导致器件速度衰退的物理机制，为在硅体系中开展应用指出了原则性的研发路径.

来源：微电子学院　发布时间：2021-06-07