

# 残差网络与注意力机制结合的啰音检测方法

杨淋坚 张宇

(广东工业大学, 广东 广州 510006)

**摘要:** 为解决啰音强度和性质易改变而导致的支持向量机人工参数选择困难、检测精度不高等问题, 提出一种残差网络与注意力机制结合的啰音检测方法。通过残差网络加深网络结构提取更多层次的信息, 同时加入注意力机制进一步挖掘通道层面与空间维度特征, 实现啰音检测。使用自主研发的数字听诊器记录的呼吸音进行实验。实验结果表明: 本文提出的方法相较于 SVM 和 ResNet50 啰音检测精度分别提高了 6.83% 和 1.58%。

**关键词:** 啰音检测; 信号处理; 残差网络; 注意力机制

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1674-2605(2021)01-0007-05

**DOI:** 10.3969/j.issn.1674-2605.2021.01.007

## 0 引言

据《国际呼吸学会论坛报告》<sup>[1-2]</sup>以及《2020 世界卫生统计报告》<sup>[3]</sup>显示: 全世界超过 10 亿人患有呼吸系统疾病, 其中每年有 400 万人死亡, 呼吸系统疾病已成为全球第三大死亡病因; 医患数量差距较大, 45% 成员国每 1000 人拥有不到 1 名医生。我国肺疾病的知晓率及肺功能检查普及率较低<sup>[4-5]</sup>。

肺音亦称呼吸音, 其变化可直接反映肺部生理和病理变化, 可通过听诊器听到<sup>[6]</sup>。呼吸音一般分为正常音和附加音。其中附加音是指异常声音叠加在正常呼吸音上, 啰音是常见的呼吸附加音<sup>[7-9]</sup>, 它是哮喘、肺炎、慢性阻塞性肺病等呼吸道疾病的早期征兆<sup>[10-13]</sup>。啰音通常由医生听诊辨析, 主观性强且易受外部因素影响, 而啰音自动检测更为客观稳定<sup>[14]</sup>。因此, 研究啰音检测方法, 通过数字听诊器自动分析呼吸音并及时发现啰音, 对肺疾病的预防和诊断起着关键作用<sup>[15]</sup>。

在呼吸音识别领域, 国内外学者提出了许多方法。ZHANG J 提出利用熵描述信号的频谱模式用以检测啰音是否存在, 并在经验阈值的基础上进行分类<sup>[16]</sup>。ZHANG K 等通过提取和分析数字化肺录音的频谱信息, 对肺信号的小波频谱进行分析, 提取数学形态学特征集, 实现呼吸音的自动识别<sup>[17]</sup>。JAKOVLJEVI 等利用谱减法去除噪声, 将梅尔倒谱系数 (Mel-Frequency Cepstral coefficients, MFCCs) 及其一阶导数

作为输入特征, 提出一种基于隐马尔可夫模型和高斯混合模型的呼吸音分类方法<sup>[18]</sup>。SERBES 等提出一种新的非线性谱特征提取算法, 采用非动态可调 Q 因子小波变换将信号分为高频、低频以及噪声 3 个通道; 利用短时傅里叶变换进行特征提取及特征融合, 从而分类呼吸音<sup>[19]</sup>。PERNA D 等提出基于 MFCCs 的数据预处理和循环神经网络 (recurrent neural network, RNN) 模型的学习框架, 长短时记忆网络和门控循环单元作为 RNN 的高级架构, 用于检测呼吸附加音<sup>[20]</sup>。

上述研究的数据均来自教科书光盘、在线教程以及少量自采音频, 出于教学目的, 声音较为干净且受试者少, 导致模型鲁棒性较差, 不利于临床应用<sup>[21]</sup>。本文以实验室自主研发的数字听诊器在医院临床采集的呼吸音信号为实验数据, 提出采用残差网络 (ResNet) 与卷积块注意力机制单元 (convolutional block attention module, CBAM) 结合的啰音检测方法, 以提高啰音检测的准确率。

## 1 数据预处理与特征提取

### 1.1 数据集

正常呼吸音采集于肺健康的志愿者, 异常呼吸音 (啰音) 采集于就诊患者, 包含男女患者, 年龄范围在 0.5 岁~70 岁, 且患有不同的肺部疾病以及严重程度也各不相同, 并由医生对采集的样本进行标注。听诊器信号采样频率为 8000 Hz, 每个样本时长 12 s。

共采集325名受试者合计2620个肺音样本,其中2000个为正常呼吸音,620个为啰音。根据受试者不同,把数据集分成5份,使用跨被试五折交叉验证对结果进行评估。

## 1.2 预处理和特征提取

呼吸音包括低频和低频成分,它们在时域和频域都有重叠。正常呼吸音频率在60 Hz~600 Hz; 啰音频率在100 Hz~2500 Hz<sup>[22-24]</sup>。为确保采集呼吸音的有效性,通过频谱分析验证信号的分布,如图1所示。

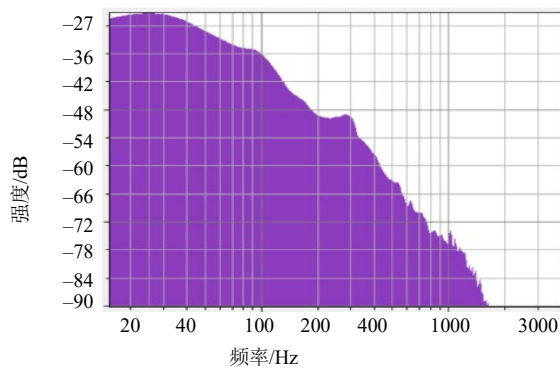


图1 呼吸音频谱分析

肺部声音微弱且容易受噪声(心跳声、听诊器接触摩擦声等)影响,使用8阶巴特沃斯带通滤波器<sup>[25]</sup>将目标频率保持在100 Hz~2500 Hz。为排除听诊器接触以及离开人体时可能产生的噪声,只截取信号中间的10 s。

在传统机器学习方法中,使用MFCCs作为SVM的输入特征;在深度残差网络中,使用梅尔谱图(Mel Spectrogram)<sup>[26-27]</sup>作为ResNet的输入特征。

## 2 ResNet与CBAM结合的啰音检测模型

### 2.1 模型结构

ResNet<sup>[28]</sup>结构容易修改和扩展,通过调整残差块内的通道数量以及堆叠的块数量,可调整网络的宽度和深度,得到不同表达能力的网络;只要训练数据足够,逐步加深网络,即可获得更好的性能表现,而不用担心网络退化问题。

CBAM<sup>[29]</sup>是一种结合了空间(Spatial)和通道(Channel)的卷积块注意力机制单元。通过在

ImageNet-1K上测试证明:增加CBAM后,大部分网络的分类错误率都有一定程度的降低。同时通过grad-CAM可视化分析发现:增加CBAM的网络模型将注意力更准确地放在正确待分类对象上<sup>[29]</sup>。

由于啰音的强度和性质易改变,部位易变换,瞬间数量可明显增减。为挖掘啰音的深层次特征,本文结合ResNet50与CBAM作为啰音检测模型CBAM-ResNet。残差块(ResBlock)加入CBAM如图2所示,啰音检测流程如图3所示。原始音频信号经过预处理和特征提取转变成梅尔谱图作为网络的输入特征,输出为信号是否包含啰音。

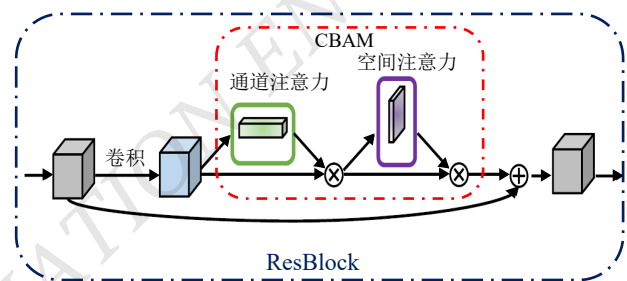


图2 在残差块中加入CBAM<sup>[29]</sup>

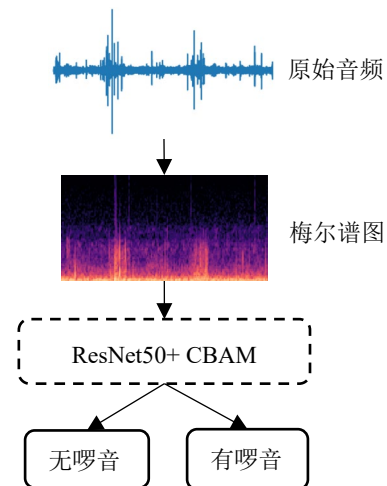


图3 ResNet50与CBAM结合的啰音检测流程

### 2.2 模型优化和损失函数

为提高模型的精度和避免模型过拟合问题,采用自适应学习率优化算法Adam<sup>[30]</sup>。由于正常呼吸音和啰音分布不平衡,使用 $\alpha$ -balanced focal loss<sup>[31]</sup>作为损失函数,公式为

$$FL(p_t) = -\alpha_t(I - p_t)^\gamma \log(p_t) \quad (1)$$

$$p_t = \begin{cases} p & \text{if } y = 1 \\ I - p & \text{otherwise} \end{cases} \quad (2)$$

式中,  $y \in \{\pm 1\}$  为真实类别;  $p \in [0, 1]$  为  $y = 1$  的类的模型估计概率;  $\gamma$  为难易样本调节参数;  $\alpha$  为类别权重。一般取  $\gamma = 2$ ,  $\alpha = 0.25^{[31]}$ 。

### 3 实验和结果分析

#### 3.1 评估指标

采用特异性 (specificity,  $S_p$ ) 和灵敏度 (sensitivity,  $S_e$ ) 以及两者的平均值作为得分 (score) 对模型进行评估, 公式为

$$S_p = \frac{N_n}{N} \quad (3)$$

$$S_e = \frac{R_r}{R} \quad (4)$$

$$score = \frac{S_e + S_p}{2} \quad (5)$$

式中,  $N$ ,  $R$  分别为正常呼吸音与啰音的样本总数;  $N_n$ ,  $R_r$  分别为被检测出来的正常呼吸音与啰音样本数。

#### 3.2 实验结果和分析

使用 SVM, ResNet50 和 CBAM-ResNet 对啰音检测的五折交叉验证的实验结果如表 1 所示。

表 1 实验结果

| Model       | $S_p/\%$ | $S_e/\%$ | score/% |
|-------------|----------|----------|---------|
| SVM         | 75.75    | 68.37    | 72.08   |
| ResNet50    | 79.68    | 74.98    | 77.33   |
| CBAM-ResNet | 82.39    | 75.42    | 78.91   |

由表 1 可以看出: 相较于 SVM 和 ResNet50, CBAM-ResNet 啰音检测精度分别提高了 6.83% 和 1.58%。表明在啰音检测中, ResNet50 具有比 SVM 更好的性能; CBAM 既考虑不同通道像素的重要性, 又考虑了同一通道不同位置像素的重要性, 表现出更优的分类检测性能。综上所述, CBAM-ResNet 的啰音检测方法取得较好的检测效果。

### 4 结语

针对啰音强度和性质易变的特性, 本文提出一种

残差网络 (ResNet) 与卷积块注意力机制单元 (CBAM) 结合的啰音检测方法 CBAM-ResNet, 并使用自主研发的数字听诊器记录临床呼吸音作为实验数据进行啰音检测实验。实验结果表明: 本文提出的方法相较于需要人工参数的支持向量机方法取得更好的啰音检测精度。

#### 参考文献

- [1] Forum of International Respiratory Societies. The Global Impact of Respiratory Disease—Second Edition. Sheffield[M]. European Respiratory Society, 2017.
- [2] “World Lung Day 2019,” Forum of International Respiratory Societies[DB/OL]. (2019-09-25) [2020-10-14]. <https://www.thoracic.org/about/newsroom/press-releases/journal/2019/world-lung-day-2019-respiratory-groups-unite-to-call-for-healthy-lungs-for-all.php>.
- [3] “World Health Statistics 2020: Monitoring health for the SDGs, sustainable development goals,” World Health Organization, Tech. Rep., 2020 [DB/OL]. (2020-05-23) [2020-10-14]. [http://www.who.int/gho/publications/world\\_health\\_statistics/2020/en/](http://www.who.int/gho/publications/world_health_statistics/2020/en/).
- [4] LIWEN F, PEI G, Heling B, et al. Chronic obstructive pulmonary disease in China: a nationwide prevalence study[J]. Lancet Respiratory Medicine, 2018, 6(6): 421-430.
- [5] BRUSSELLE G G, KO F W. Prevalence and burden of asthma in China: time to act[J]. The Lancet, 2019, 394(10196):364-366.
- [6] EARIS J E, CHEETHAM B M G. Current methods used for computerized respiratory sound analysis[J]. European Respiratory Review, 2000, 10(77):586-590.
- [7] SOVIJARVI, A R A, DALMASSO F, VANDERSCHOOT J, et al. Definition of terms for applications of respiratory sounds[J]. Nki Distance Education, 2000, 10(6):138-165.
- [8] SARKAR M, MADABHAVI I, NIRANJAN N, et al. Auscultation of the respiratory system[J]. Annals of Thoracic Medicine, 2015, 10(3): 158-168.
- [9] BOHADANA A, IZBICKI G, KRAMAN S S. Fundamentals of lung auscultation[J]. New England Journal of Medicine, 2014, 370(8): 744-751.
- [10] GURUNG A, SCRAFFORD C G, Tielsch J M, et al. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis[J]. Respiratory Medicine, 2011, 105(9):1396-1403.
- [11] PACIEJ R, VYSHEDSKIY A, BANA D, et al. Squawks in

- pneumonia[J]. Thorax, 2004, 59(2): 177-178.
- [12] MUNAKATA M, UKITA H, DOI I, et al. Spectral and waveform characteristics of fine and coarse crackles[J]. Thorax, 1991, 46(9): 651-657.
- [13] SOVIJARVI A R A. Characteristics of breath sounds and adventitious respiratory sounds[J]. Eur Respir Rev, 2000, 10: 591-596.
- [14] 李真真,吴效明.基于 S 变换的罗音信号检测算法[J].华南理工大学学报(自然科学版),2013,41(06):1-5.
- [15] 李真真,吴效明.附加性呼吸音信号处理的研究进展[J].生物医学工程学杂志,2013,30(05):1131-1135.
- [16] ZHANG J, SER W, YU J, et al. A novel wheeze detection method for wearable monitoring systems[C]//2009 International Symposium on Intelligent Ubiquitous Computing and Education. IEEE, 2009: 331-334.
- [17] ZHANG K, WANG X, HAN F, et al. The detection of crackles based on mathematical morphology in spectrogram analysis[J]. Technology and Health Care, 2015, 23(S2): S489-S494.
- [18] JAKOVLJEVIN, LONAR-TURUKALO T. Hidden Markov Model Based Respiratory Sound Classification[M]// Precision Medicine Powered by pHealth and Connected Health. 2017.
- [19] SERBES G, ULUKAYA S, KAHYA Y P. An automated lung sound preprocessing and classification system based on spectral analysis methods[M]//Precision Medicine Powered by pHealth and Connected Health. Springer, Singapore, 2018: 45-49.
- [20] PERNA D, TAGARELLI A. Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks[C]//2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). IEEE, 2019: 50-55.
- [21] ADHI P R X, STUART B, ESTHER R V, et al. Automatic adventitious respiratory sound analysis: a systematic review[J]. PLOS ONE, 2017, 12(5):e0177926.
- [22] EARLE B. WEISS, C. JEFFREY CARLSON. Recording of breath sounds[J]. American Review of Respiratory Disease, 1972, 105(5):835-9
- [23] GAVRIELY N, PALTI Y, ALROY G. Spectral characteristics of normal breath sounds[J]. Journal of applied physiology, 1981, 50(2): 307-314.
- [24] FORGACS P, NATHOO A R, Richardson H D. Breath sounds[J]. Thorax, 1971, 26(3): 288-295.
- [25] SELESNICK I W, BURRUS C S. Generalized digital butterworth filter design[J]. IEEE Trans. Signal Process, 1998, 46(6):1688-1694.
- [26] ROCHA B M, FILOS D, MENDES LUÍS, et al. An open access database for the evaluation of respiratory sound classification algorithms[J]. Physiological Measurement, 2019, 40(3): 035001.
- [27] SHI L, DU K, ZHANG C, et al. Lung Sound Recognition Algorithm Based on VGGish-BiGRU[J]. IEEE Access, 2019, (99):1.
- [28] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.
- [29] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [30] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [31] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2): 318-327.

## Rale Detection Method Based on Residual Network and Attention Mechanism

Yang Linjian ZhangYu

(Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** In order to solve the problems caused by the easy change of rales intensity and properties, such as difficulty in selecting artificial parameters of support vector machine, poor detection accuracy and so on. A rales detection method based on residual network and attention mechanism is proposed. Through the residual network to deepen the network structure to extract more levels of information, while adding the attention mechanism to further mine the channel level and spatial dimension features to achieve rales detection. We used a self-developed digital stethoscope to record a total of 2620 breath sounds in 325 subjects. The experimental results show that compared with SVM and ResNet50, the proposed method improves the accuracy of rale detection by 6.83% and 1.58% respectively.

(下转第 40 页)