

报告题目：毕业设计调研（二）

任永文

2023 年 3 月 17 日

概述

训练篇

测试篇

数据加载

1. 从.wav格式音频中随机选取2s（如果不够需要补全）
2. 写Dataloader加载数据

特征提取

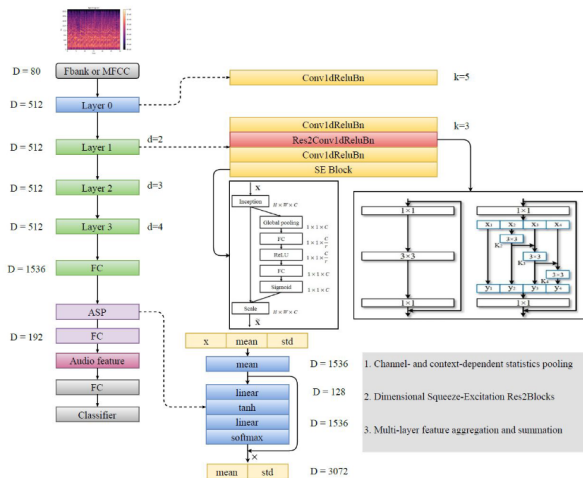
1. 采样频率：16kHz
2. 预加重：丰富高频信号 $y(t) = x(t) - \alpha x(t - 1)$
3. 分帧加窗：信号分成25ms一帧后偏移，加一个窗方便傅里叶
4. 傅里叶变换计算能量谱
5. 过滤取得FBank或Mel频谱
6. 过DCT得到MFCCs
7. 归一化

数据加强

RIR数据集和Musan数据集

1. RIR混响
2. Music: 1段音乐
3. Noise: 1段噪音
4. Babble: 3-8段演讲
5. TVnoise: 演讲+音乐
6. Spectrum Augmentation: 频谱图上随机遮盖一行一列数据

网络模型



损失函数

1. Softmax
2. Triple loss
3. L-softmax
4. SphereFace
5. Center Loss
6. Feature Normalization
7. AMsoftmax&Cosface
8. AAM softmax

概述

训练篇

测试篇

最终分数

1. 读取测试集的数据
2. 将测试集的数据放入网络提取特征Embedding，归一化
3. 对提取到的Embedding两两计算cos相似度（欧式距离）得到最终分数
4. 可以计算score1（整段音频）+score2（分段提取多段音频）

性能指标

1. EER: 弃真=存伪
2. minDCF: 存伪比弃真更重要 $C_{Det}(\theta) = C_{Miss} \cdot P_{Target} \cdot P_{(Miss)}(\theta) + C_{FalseAlarm} \cdot (1 - P_{Target}) \cdot P_{(FalseAlarm)}(\theta)$

分数后端

1. Z-norm
2. T-norm
3. T-norm
4. ZT-norm
5. S-norm
6. Adaptive S-norm

For Further Reading I



A. Author.

Handbook of Everything.

Some Press, 1990.