文件编号：从容应队-SWC2020-从容应队

受控状态：■受控 □非受控

保密级别：□公司级 □部门级 ■项目级 □普通级

采纳标准：CMMI DEV V1.2



****

素音

**Divoice**

项目开发文档

**Version 1**

2019.12.19

**Written by 从容应队**



**All Rights Reserved**

目录

[1 问题聚焦 1](#_Toc22848191)

[1.1 问题描述 1](#_Toc22848192)

[1.2 问题抽象 1](#_Toc22848193)

[1.3 问题定位 1](#_Toc22848194)

[1.4 问题评估 1](#_Toc22848195)

[1.5 问题分解 1](#_Toc22848196)

[2 相关工作 1](#_Toc22848197)

[3 技术方案 1](#_Toc22848198)

[3.1 技术方向 1](#_Toc22848199)

[3.2 技术选择 1](#_Toc22848200)

[3.3 结果期望 1](#_Toc22848201)

[4 技术实践 1](#_Toc22848202)

[4.1 使用的开发框架及依赖的Library 1](#_Toc22848203)

[4.2 技术实践过程 1](#_Toc22848204)

[5 结果验证 1](#_Toc22848205)

记录更改历史

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **更改原因** | **版本** | **作者** | **更改日期** | **备 注** |
| 1 | 创建 | 1.0.0 | 队员A | 2019/12/19 | 创建 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

# 问题聚焦

## 问题描述

Divoice有三个主要问题待解决：

1. 如何利用用户提供的采样音频优化音频分离效果
2. 对于用户提供的混音素材，如何使用训练好的模型去比对识别、分离
3. 对用户音频编辑内容提供相关素材推荐

## 问题抽象

基于问题一：Divoice需要将用户的采样音频中提取声纹特征，然后将此声纹特征写入用户常用采样音的数据集中。

基于问题二：根据用户选定的音频内含有的声纹标签，在相应的类中的分别进行声纹特征比对并拣选出匹配程度高的音频波段，形成新的音频文件，从而实现音频分离。对于比对度不高的剩余音频波段，再选用树表查找算法进行进一步比对分离，并返回比对度较高的声纹特征类别标签。

基于问题三：基于训练模型，对于不同场景中的常出现的声音特征写成集合，根据识别用户提供音频中的声纹特征，以及用户使用历史，猜测用户音频编辑风格，并对用户提供推荐素材。

## 问题定位

## 问题评估

## 问题分解

# 相关工作

1. ResNet（Residual Neural Network）由微软研究院的Kaiming He等四名华人提出，通过使用ResNet Unit成功训练出了152层的神经网络，并在ILSVRC2015比赛中取得冠军，在top5上的错误率为3.57%，同时参数量比VGGNet低，效果非常突出。ResNet的结构可以极快的加速神经网络的训练，模型的准确率也有比较大的提升。同时ResNet的推广性非常好，甚至可以直接用到InceptionNet网络中。
2. 2017年，北京联合大学北京市信息服务工程重点实验室发表了一篇论文《噪声环境下的FastICA算法研究》，介绍了独立成分分析(ICA)算法的改进算法——FastICA，并引入牛顿下山法来降低初值选择的敏感性，提出差商法来减少计算量来针对FastICA算法的缺陷进行改良，使得收敛更加稳定。
3. 2019年王敏发表《基于深度神经网络和多元损失的说话人识别》提出一种基于深度神经网络和beyond triplet loss相结合的说话人识别方法,模型通过梅尔频率倒谱系数(MFCC)提取MFCC声学特征,对MFCC声学特征提取说话人声纹特征,然后进行多元损失的模型训练。实验结果表明,DNN-BTL算法在说话人识别领域比高斯混合-隐马尔可夫模型(GMM-HMM)具有更好的识别效果。

# 技术方案

## 技术方向

本项目中对音频文件进行处理，需使用深度学习相关知识。

## 技术选择

所用模型为DNN（ResNet，即残差网络）

## 

图3.2.1.DNN加feed-forward architecture结构图

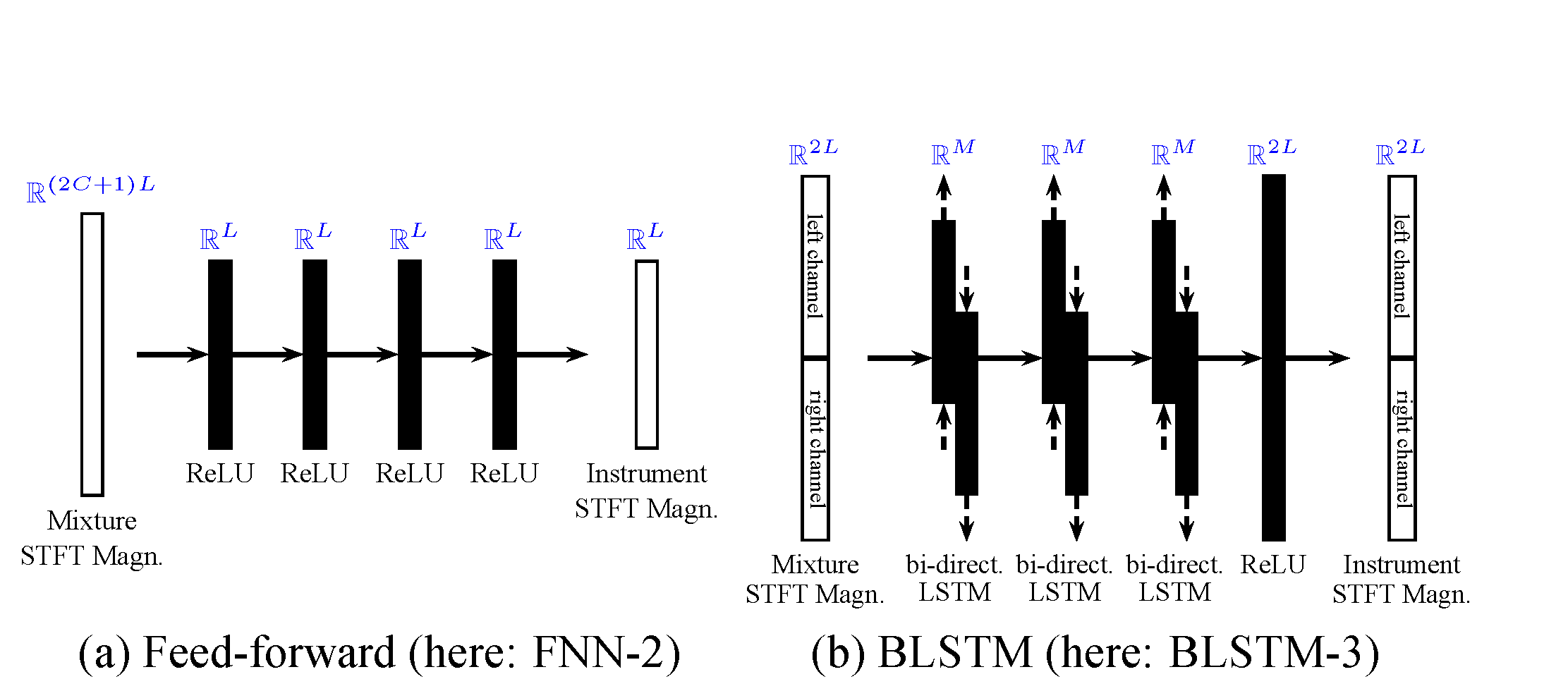


图3.2.2 DNN提取声纹结构示意图

## 结果期望

# 技术实践

## 使用的开发框架及依赖的Library

## 技术实践过程

# 结果验证