**Chapter 01 基于DNN模型的声纹识别实战**

**实验目的**

1. 理解语音特征的原理和应用
2. 掌握常用声学软件包的基本使用方法
3. 了解语音识别系统的基本架构和原理

**实验内容**

基于LibriSpeech数据集进行声纹识别。首先进行数据预处理，将语音特征抽取至数据文件中，然利用深度学习神经网络进行模型训练及测试。

**实验原理**

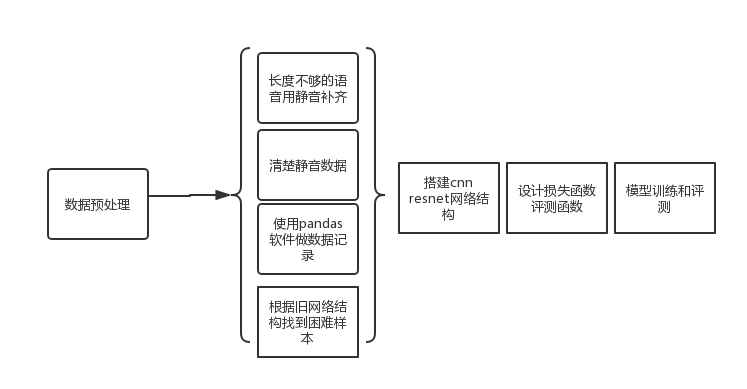
利用librosa软件包的短时傅里叶变换抽取语音特征，得到fbank的声音特征。将特征送入由多层卷积、池化和全连接层组成的深度学习网络中进行训练。借鉴人脸识别里面经典的triple loss，将语音特征类比为人脸特征作为特征比对任务。最终得到百分之99以上精度的模型。

**实验环境**

硬件：x86\_64 软件：Win10、Python 3.6.10、numpy 1.19.1、tensorﬂow 1.5等详细查看requirements

**实验步骤**

**整体流程图**

****

第一种方式 jupyter notebook

Cmd进入项目根目录,然后输入jupyter notebook

推荐按照顺序学习:

1. 准备数据.ipynb
2. 训练.ipynb
3. 模型结构.ipynb
4. 测试.ipynb



第二种方式

如果电脑中没有安装jupyter可以进入html版本教程目录

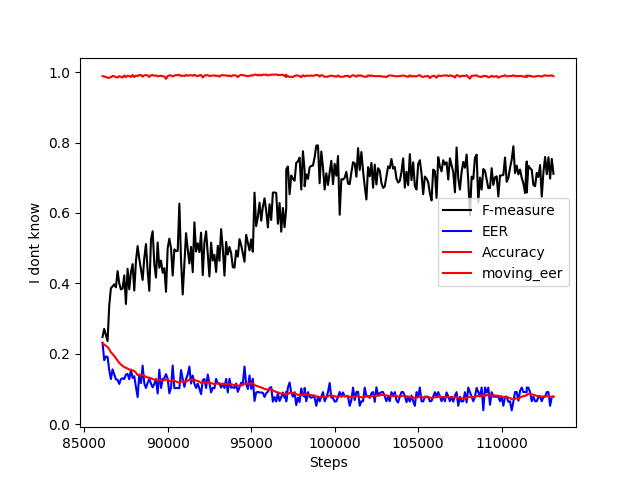
直接观看html版本的jupyter实验记录进行学习，注意浏览器打开文件较大会比较慢。大概30秒左右可以打开。

第三种方式

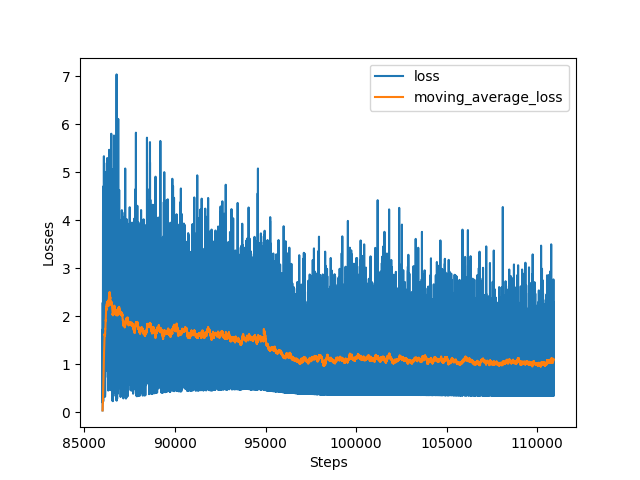
代码方式打开。推荐用pycharm的debug模式运行代码。因为函数调用结构比较复杂，debug的调用栈会让学习清晰很多。

train.py 模型会进行训练和评测  
 test\_model.py 进行测试.

官方效果展示：



红线的准确率接近于百分之99



训练过程的loss下降情况。

**实验细节讲解**

pre\_process.py 会进行数据的预处理工作从原始数据集的flac格式最终转化为npy格式。

里面包含数据清洗和补全等工作。

train.py 模型会进行训练和评测 cpu速度大概5分钟一次网络bp算法迭代。

test\_model.py cpu速度几乎秒算。