基于LSTM和pix2pix-Gans的语音驱动人脸动画系统

主要流程

1. 数据集制作
2. 数据预处理
3. 模型训练
4. 正向计算

数据集制作

要求：录制视频，人脸无表情变化，发音标准，头部尽量不要晃动

数据集预处理：

用ffmpeg将视频中的音频分离出来，用于提取mfcc特征,并将视频按照每秒二十帧的速度拆分成图片，用于pix2pix网络的训练

数据预处理：

1. 音频数据的处理(音频数据处理.py)

作用：提取音频特征

主要参数：窗宽：0.05s对应每秒20帧的图片处理速度

滑动窗宽：对应于LSTM的网络长度num\_unit

输出：csv文件

1. 视频图片处理过程：

图片人脸截取：按照眼角距离等比例截取人脸图像

人脸特征提取：在截取好的人脸图像的基础上利用Dilb工具提取人脸68个特征点，利用opencv将68个点连线画成简笔画,对应人脸图像进行输出。

进行pix2pix的resize处理，将人脸68个点坐标信息保存为csv文件并输出。（脸部特征提取。）

模型训练 :

1. 利用人脸特征图片和人脸图片训练pix2pix,将训练好的模型生成器部分进行固化。
2. 利用mfcc.csv和landmark(嘴部19个点，脸部轮廓16个点)训练双层的lstm模型，将训练好的模型进行固化并保存。（LSTM.py）

正向计算：

加载两个模型做正向计算

1. 加载LSTM模型预测语音对应的嘴部和人脸下巴的动作
2. 根据预测合成数据的68个点。进行pix2pix的正向计算，输出人脸图片。（run.py）