

STIS-课程编号-01510202



TensorFlow实践

声控智能

智能系统实验室

清华大学基础工业训练中心

目录

- 声控智能设计
- 网络模型训练
- 安卓应用开发

声控智能设计

- 实验步骤:

1. 语音指令设计

输出： 24句语音指令；

2. 语音指令录音

输出： 统一格式的录音文件集*.wav文件；

3. 基于TensorFlow的深度学习训练

输出： 网络模型文件*.pb；

4. 基于AndriodStudio的安卓应用开发

输出： 安卓应用的apk文件。

语音指令设计

- 24条语音指令

a蓝牙开机

b蓝牙拨打电话 / bb蓝牙打电话

c蓝牙接听电话 / cc蓝牙接电话

d蓝牙拒接

e蓝牙播放音乐 / ee蓝牙开始音乐

f蓝牙暂停音乐 / ff蓝牙停止音乐

g蓝牙上一首 / gg蓝牙上一曲

h蓝牙下一首 / hh蓝牙下一曲

i蓝牙音量增大 / ii蓝牙声音增大 / iii蓝牙音量增加 / iiii蓝牙声音增加

j蓝牙音量减小 / jj蓝牙声音减小

























k蓝牙关机

l蓝牙电量提醒 / ll蓝牙还剩多少电 / lll蓝牙还剩多少电量

语音指令录音

- 24个语音文件
- *.wav等格式

网络 > 166.111.134.69 > team-saturn > Temp_Audio > 2015620190

	名称	修改日期	类型	大小
✦	 a	2016/5/10 0:41	AMR 文件	2 KB
✦	 b	2016/5/10 0:41	AMR 文件	3 KB
✦	 bb	2016/5/10 0:41	AMR 文件	3 KB
✦	 c	2016/5/10 0:41	AMR 文件	3 KB
	 cc	2016/5/10 0:41	AMR 文件	3 KB
	 d	2016/5/10 0:41	AMR 文件	3 KB
	 e	2016/5/10 0:41	AMR 文件	3 KB
	 ee	2016/5/10 0:41	AMR 文件	3 KB
	 f	2016/5/10 0:41	AMR 文件	3 KB
	 ff	2016/5/10 0:41	AMR 文件	4 KB
	 g	2016/5/10 0:41	AMR 文件	3 KB
	 gg	2016/5/10 0:41	AMR 文件	4 KB
	 h	2016/5/10 0:41	AMR 文件	3 KB
	 hh	2016/5/10 0:41	AMR 文件	3 KB
	 i	2016/5/10 0:41	AMR 文件	3 KB
	 ii	2016/5/10 0:41	AMR 文件	3 KB
	 iii	2016/5/10 0:41	AMR 文件	4 KB
	 iii	2016/5/10 0:41	AMR 文件	3 KB
	 j	2016/5/10 0:41	AMR 文件	3 KB
	 jj	2016/5/10 0:41	AMR 文件	3 KB
	 k	2016/5/10 0:41	AMR 文件	3 KB
	 l	2016/5/10 0:41	AMR 文件	3 KB
	 ll	2016/5/10 0:41	AMR 文件	3 KB
	 ll	2016/5/10 0:41	AMR 文件	3 KB

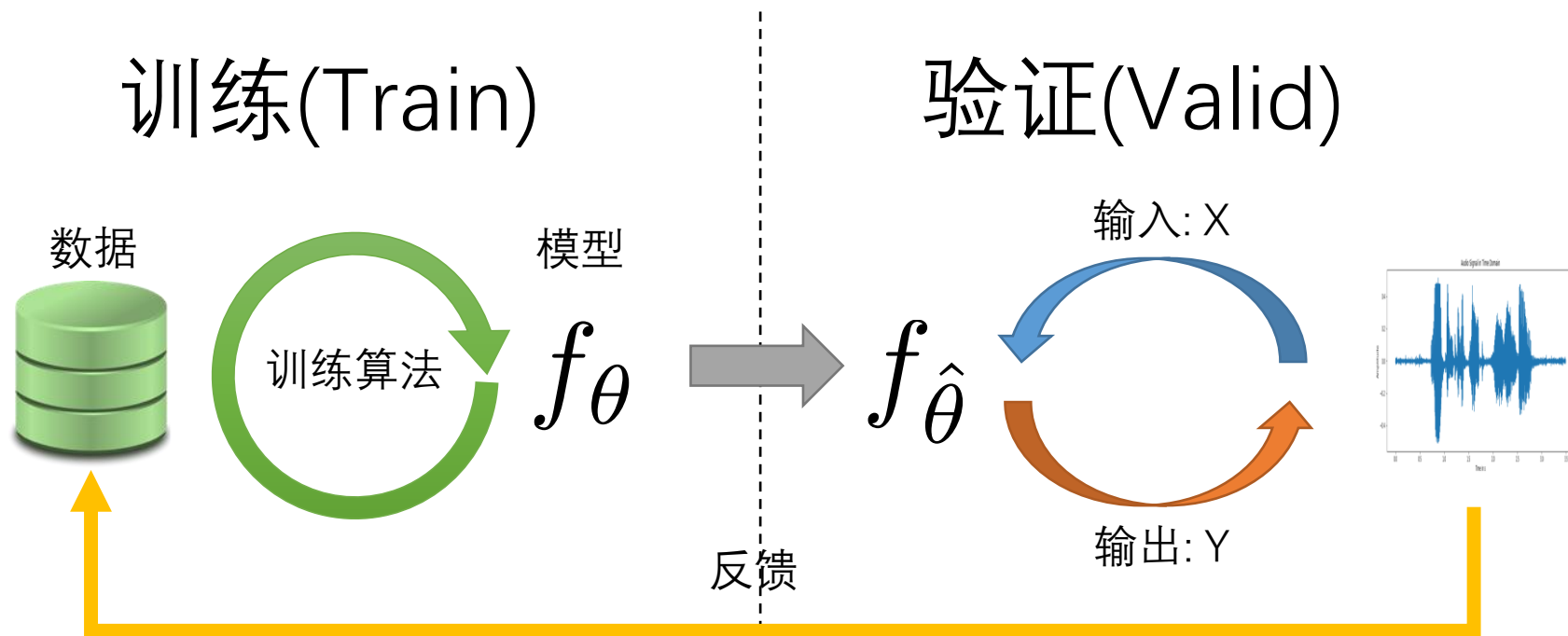
声控智能-实验过程

- (1) 用笔记本或者手机录下以上语音指令文件，生成对应指令的.wav文件，文件名即为指令名字编号。
- (2) 对应文件生成频谱图 (spectrogram) *.spec文件 [1]
- (3) 使用TensorFlow搭建3层全连接的神经网络 (24个softmax输出) 。
- (4) 使用采集的数据集，训练神经网络。
- (5) 对神经网络进行验证准确率。
- (6) 将训练好的神经网络部署在树莓派智能硬件平台上。

[1] Spectrogram, <https://en.wikipedia.org/wiki/Spectrogram>

网络模型训练

神经网络训练与验证



- 训练过程
 - 相对耗时
 - 迭代批处理过程

- 进行快速和鲁棒的预测
 - 快速推断
 - 检测与追踪反馈

计算平台使用

- (1) 录音采集（用手机设备）
- (2) 录音数据预处理（用工作站）
- (3) 神经网络训练（用GPU超级计算平台）
- (4) 神经网络部署（智能硬件）

audioPlot

- 录音文件预处理
- 统一化格式为*.wav文件
- 展示为波形文件和频谱图文件
- 删除不正确的数据，防止污染训练结果

深度网络的实际训练过程（1）

- 网络的训练将样本数据“分批训练”。
 - 先将整个训练集分成多个大小相同的子集，每个子集称为一个迷你批次（batch）。
 - 子集的大小由参数迷你批次大小（minibatch_size）控制。
 - 每个批次的数据被依次送入网络进行训练。
 - 训练完一个迷你批次，被称为一次迭代（iteration）。
 - 一个时代（epoch）是指训练集中所有训练样本都被送入网络，完成一次训练的过程。

深度网络的实际训练过程 (2)

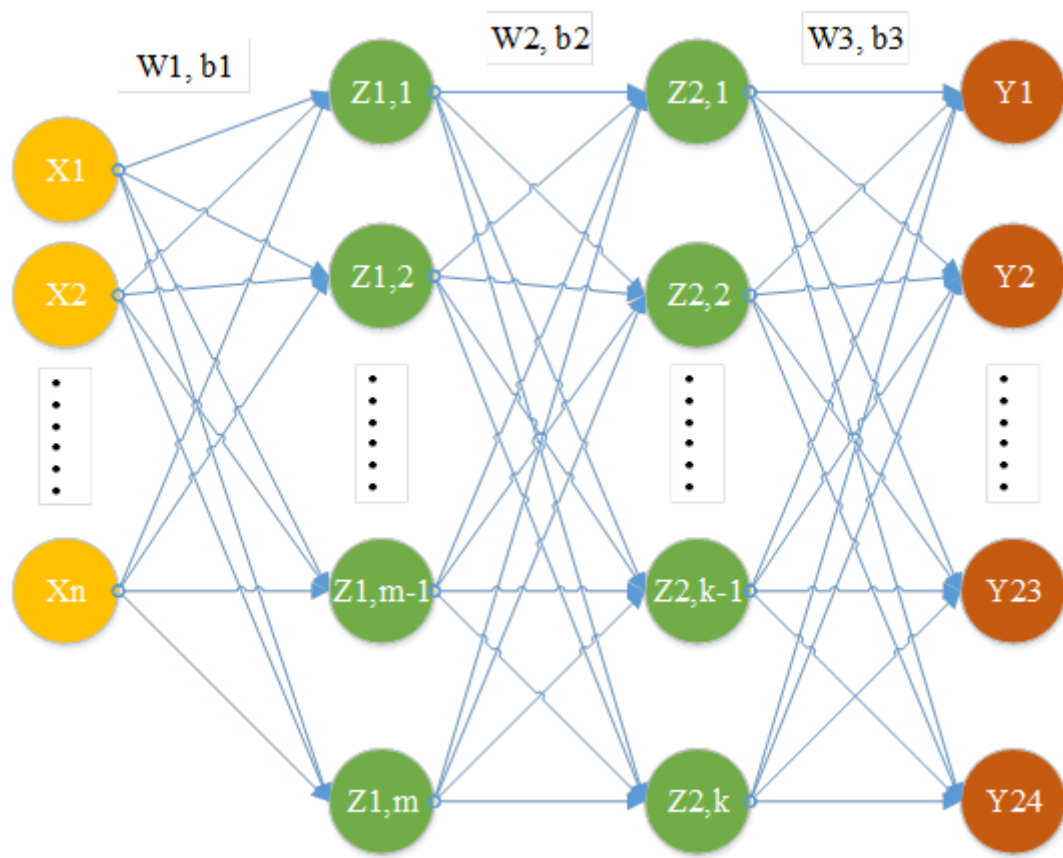
- 时代 (epoch) 数目记录了整个训练集被反复训练的次数
- 迭代 (iteration) 数目记录网络参数的调整次数。
- 时代 (epoch) 和迭代 (iteration) 之间的数量关系由批次大小 (batch_size) 和训练集大小决定。
- 比如, 训练集有100张照片, 分为50个子集, 每个子集2张, 那么 Batch_size =2。 如果epoch =10, iteration =500。

GPU工作站

- CPU:
 - Intel Xeon E5-2630 v3(8核16线程)
- GPU:
 - 8块NVIDIA GeForce TitanX 12G GDDR5
- 128GB内存
- 12TB存储



3层前馈网络模型



audioNet

- TensorFlow/Keras实现
- 谷歌CLDNN模型
- Python 语音
- 输入带标签的*.wav文件
- 输出*.h5文件, 转*.pb文件

实验结果：数据增强

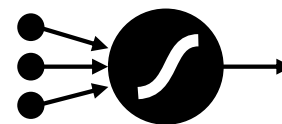
小数据



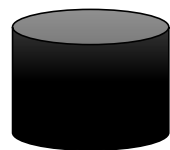
训练



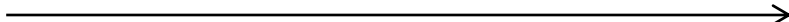
小模型



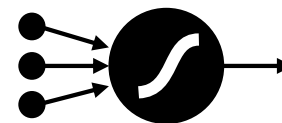
大数据



训练



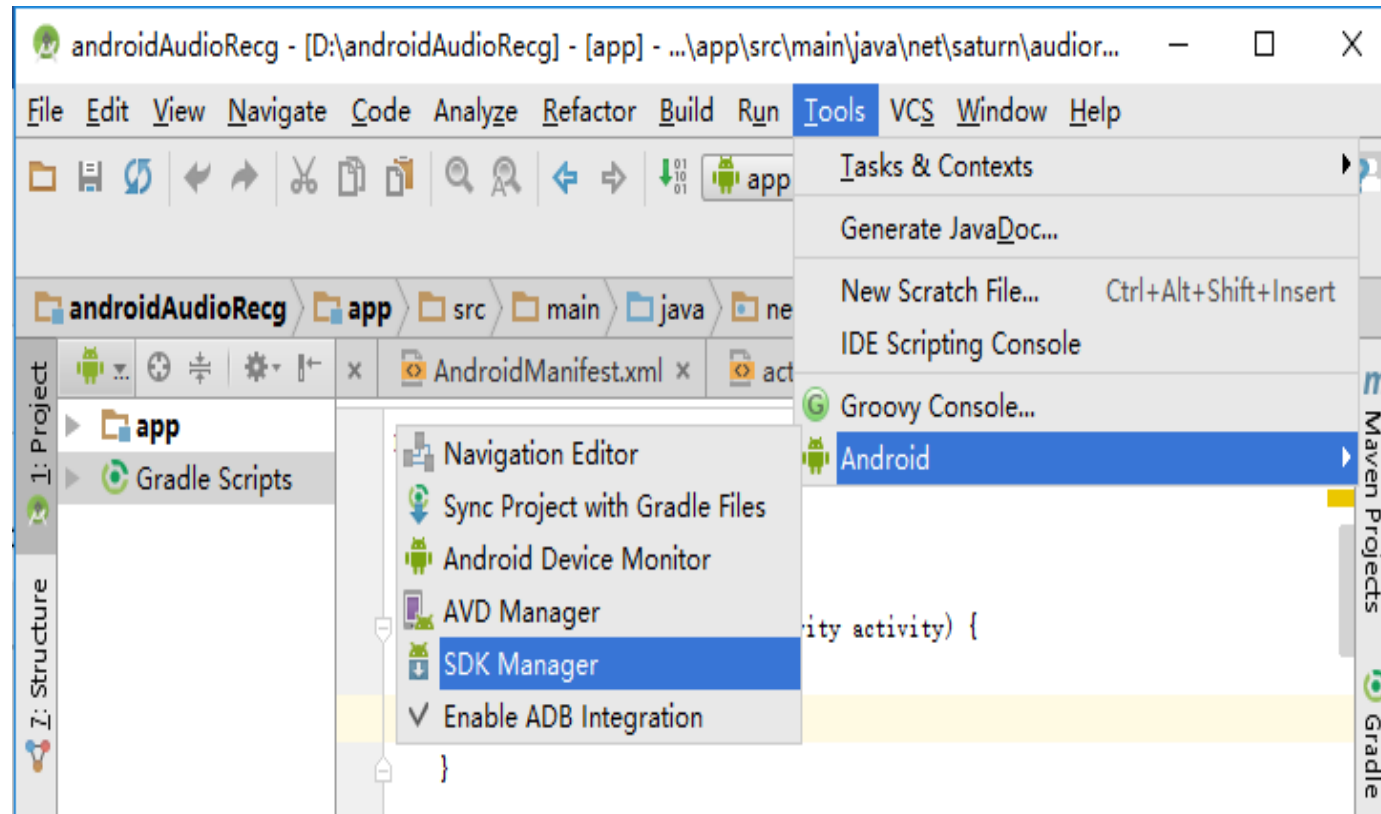
大模型



安卓应用开发

androidAudioRecg

- 谷歌android开发环境Android Studio



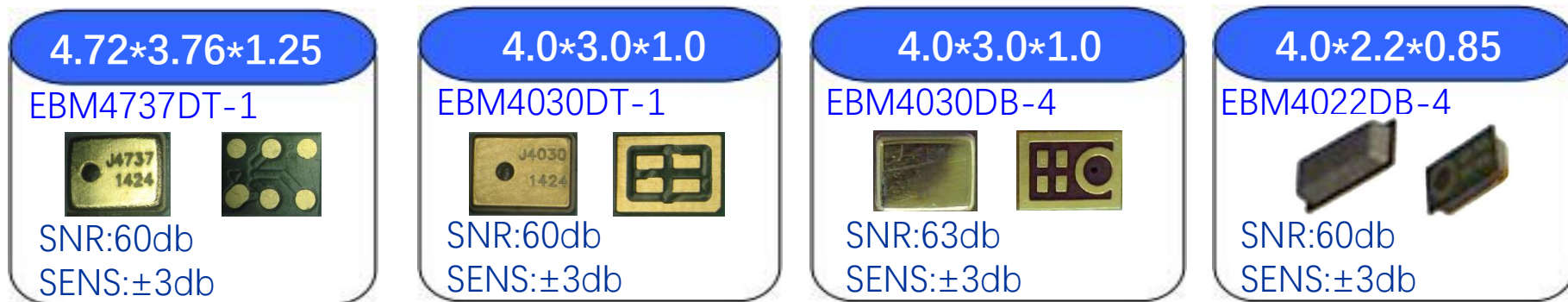
实现原理

- 在安卓应用中调用手机的录音（Record）功能
 - 需要获得系统的录音权限，否则会出现各种错误提示。
- 使用AudioRecord类来进行录音。好处是录音结果就是内存中的数组，以浮点数、整数的数组形式给出，方便算法处理。
- 载入推断模型文件，调用TensorFlow库的推断接口：
TensorflowInferenceInterface.

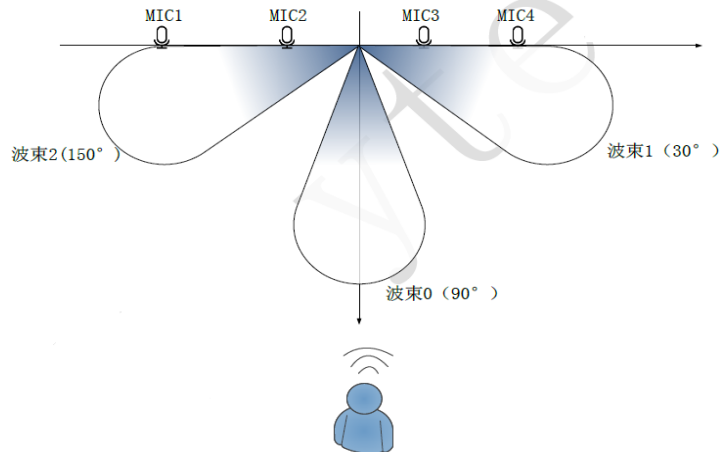
声控智能硬件

产业未来

硅麦克风模组



4 麦克风阵列形成 3 个波束，每个波束对应 60° 的录音范围。



- 基于4麦克风阵列的语音前端解决方案：
- 模块利用麦克风阵列的空域滤波特性，通过对唤醒人的角度定位，形成定向拾音波束，并对波束以外的噪声进行抑制，提升远场拾音质量、保证识别效果。
- 实现180度声源定位、5米拾音、降噪、回声消除、语音唤醒 等功能。

声控智能硬件应用

- AIoT物联网
- 语音唤醒万物
- 语音交互



展望

- 全中文语音识别系统开发
 - 中文语料准备
 - 模型选择, 如DeepSpeech2/3等
 - 模型训练与部署

谢谢指正！