

音频鉴伪挑战赛

团队：起飞

成员：白吴斌（毕业于加州大学戴维斯分校）

涂修建（上海优也信息科技有限公司）

基本思路

- Stage 1: Train on real/fake, test on real/fake
- Stage 2: Train on speakers(52), test on speakers
- Features: melspectrograms -> pictures(saves storage, much faster)



- Models: CNN

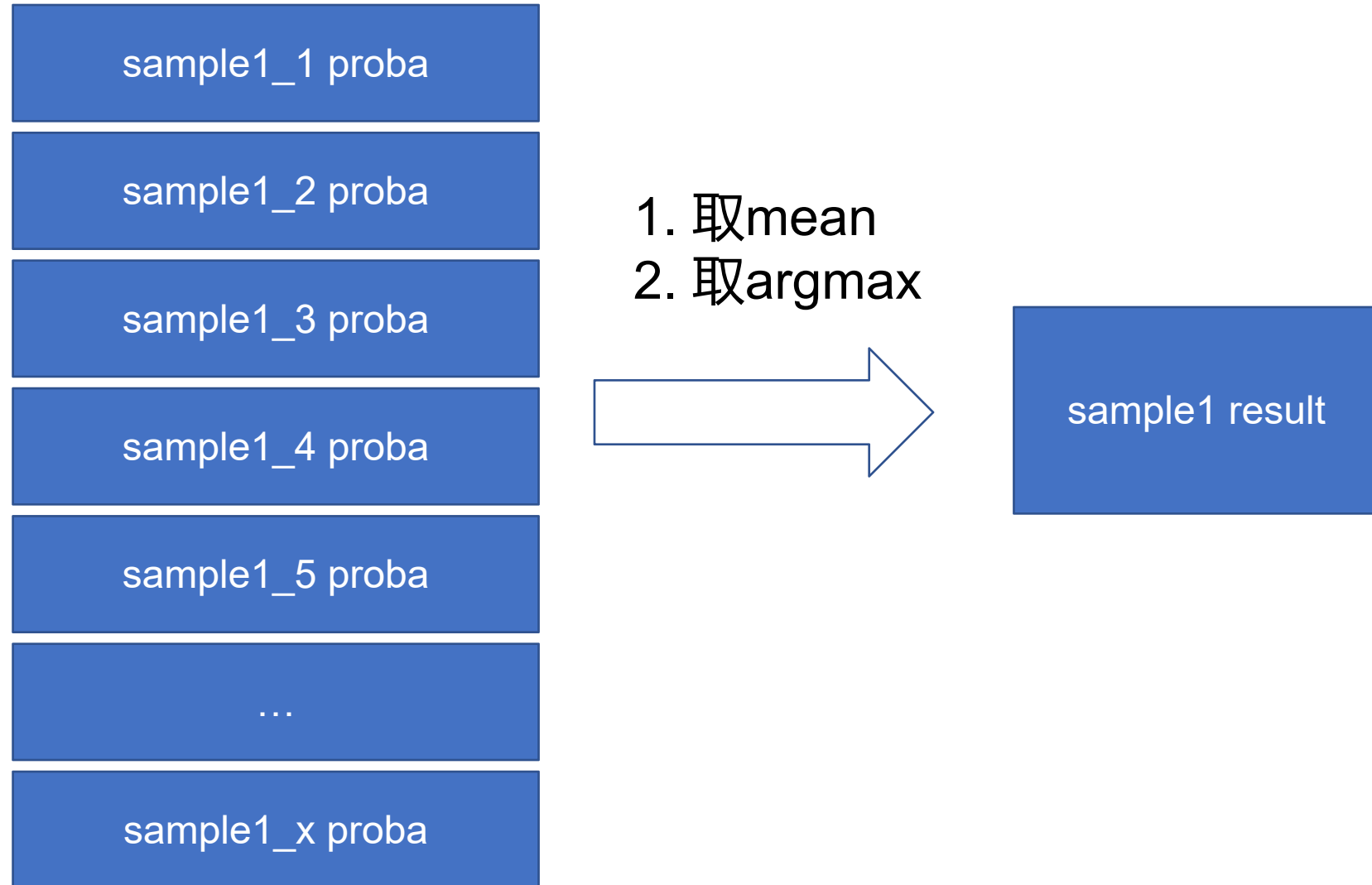
Features: melspectrograms

- 使用 `librosa.feature.melspectrogram` 方法从音频中提取得到melspectrogram特征，划窗长度选择2s。
- 考虑到提取的特征矩阵中数值为float类型，存储和训练时占用很大的硬盘和内存空间，将之转换为uint8的图片格式，图片大小是 (47,128)，空间占用大大降低，也便于使用resnet等图片处理的模型训练。

Model

- Model :
- Basic Building Block: Conv2D -> BatchNormalization -> Activation -> MaxPool2D
- Stack of these: model 2: around 7-8 layers of stacks of CNN
- model 1: simpler, 2 layers less.
- Conv2D: differ in stride and kernel size
- Accuracy/Validation Accuracy: 99%+

Prediction



谢谢大家 Q&A

- 白吴斌 wubinbai@yahoo.com