

报告题目：毕业设计调研（三）

任永文

2023 年 3 月 17 日

Data Preparation

Dataset

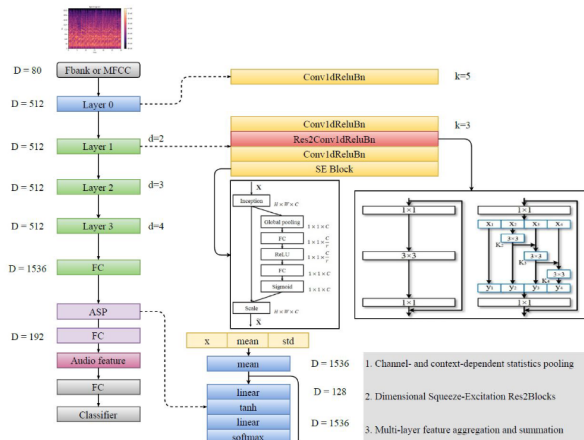
1. common_language_kpd
2. musan
3. rir_noises

Data Augmentation

1. Mixture of non-speech noises(music noise babble), which are from MUSAN
2. Reverbation injection, using simulated impulse responses from RIR_NOISES
3. SpecAugmentation

System Description

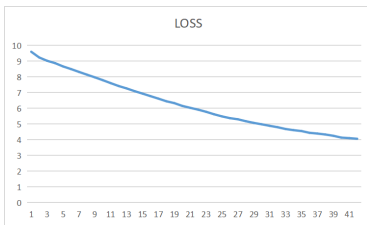
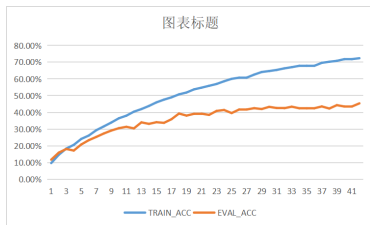
1. features: Fbank
2. baseline: ecapa-tdnn
3. classifier: 192 to 45
4. finetune: adam + ce-loss/aam-softmax



Experiment settings

1. input features:resample to 16000Hz, 80-dim fbank
2. model settings:(channels:[1024,1024,1024,1024,3072],
attention_channels:128, output:192)
3. train strategies: lr:1e-4, lr_decay:0.97, batch_size:64

Experiment results



| loss_function | lr | batch | loss | acc |
|----------------------|-------------|-----------|--------|--------------------------|
| crossentropy | 1e-3 | 64 | | 65% |
| aam+crossentropy | 1e-3 | 64 | slowly | |
| aam+kldiv | * | 4 | slowly | |
| aam+kldiv | 1e-3 | 64 | | rise to 20% then decline |
| aam+kldivloss | 1e-4 | 64 | | 45% |
| aam+kldiv+pretrain | 1e-4 | 64 | well | 65%+ |

For Further Reading I



TaoRuijie,speechbrain

ECAPA-TDNN.

<https://github.com/TaoRuijie/ECAPA-TDNN>

<https://github.com/speechbrain/speechbrain>