Speech analysis and contrast System

Defense for speech recognition course design

1953348 叶栩冰

1953066 刘昕宇

1953196 张铃沛

2021.12.30

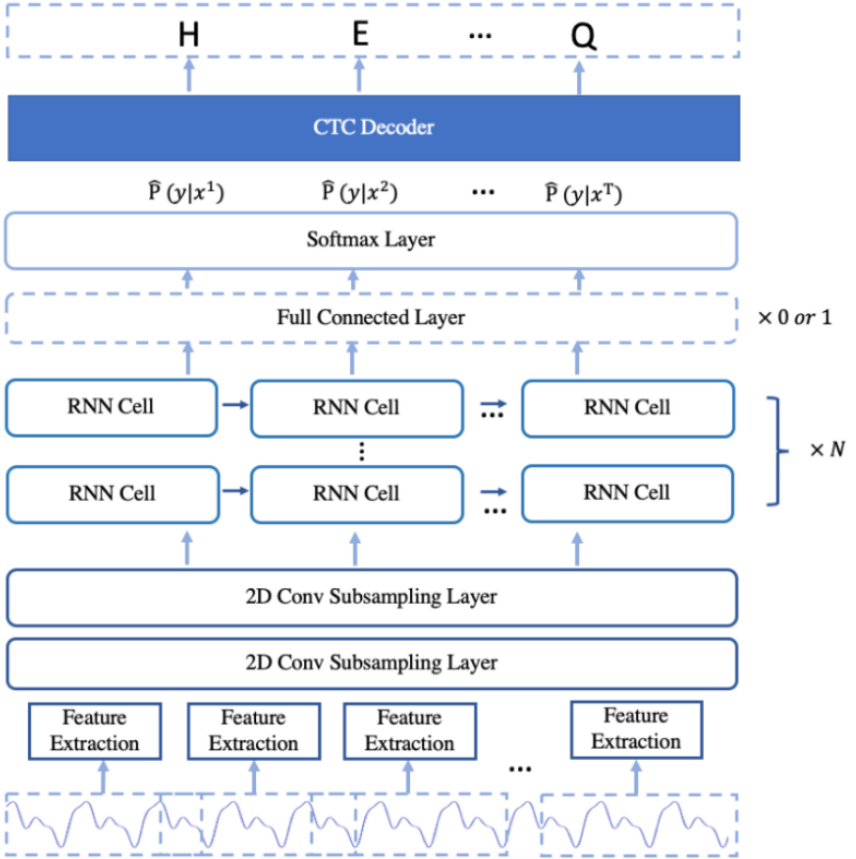
CONTENTS

Speech to text (Chinese)

Paddle Deep Speech

- AIShell
- SpenAugment
- 2 * Conv + 5 * LSTM
- CTC
- CER – 0.08452
- Score

Deep Speech 2 : End-to-End Speech Recognition in English and Mandarin



Emotion & gender analysis

- RAVDESS
- Label
- mfcc 216*13 -> 216*1
- CNN1D

0 – female angry

1 – female calm

2 – female fearful

3 – female happy

4 – female sad

5 – male angry

6 – male calm

7 – male fearful

8 – male happy

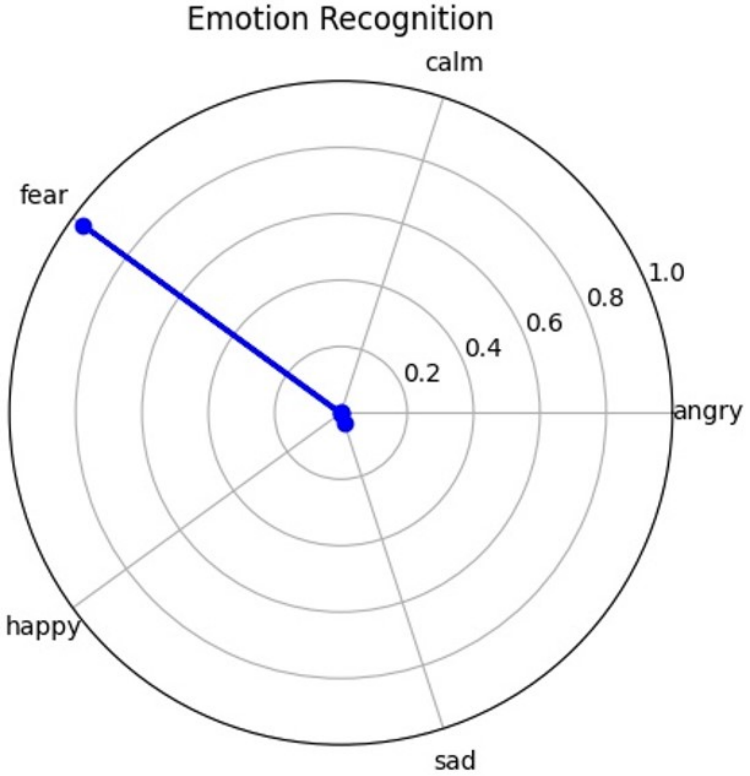
9 – male sad

Emotion & gender analysis

Micro precision 0.9103232533889468
Micro recall 0.9375327527527535
Micro f1-score 0.9291981861612104

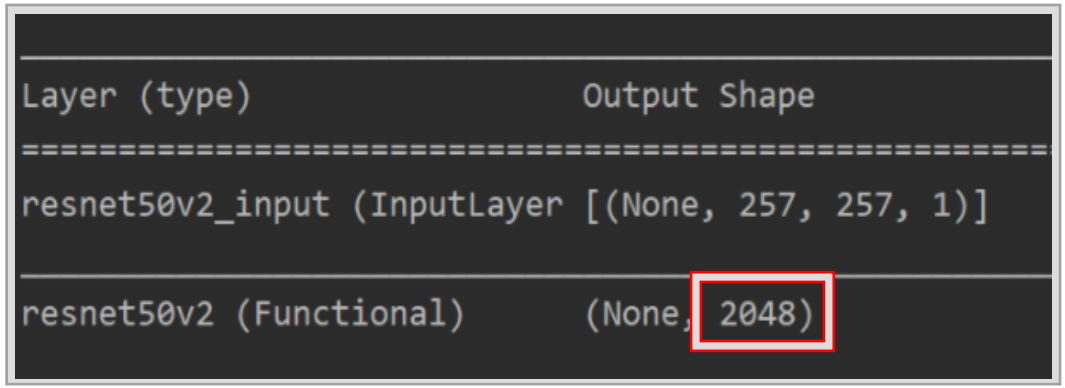
Each class ROC
0.9892936461954422
0.9964271919660099
0.9850086906141368
0.9914662997296253
0.9901264967168791
0.9944838740826574
0.9922392787524366
0.9900661452298185
0.9872175550405563
0.9778872151409811

AUC: 0.989140679092956



Voice print contrast

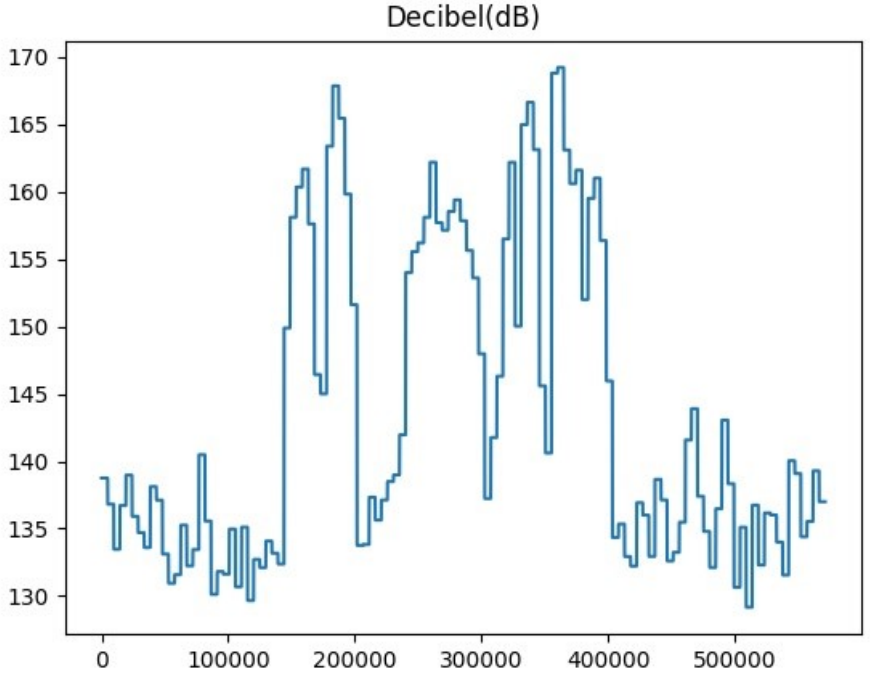
- zhvoice
- Short-Time Fourier Transform
- Resnet-50
- ArcFace



$$dist = \frac{np.dot(feature1, feature2)}{np.linalg.norm(feature1) * np.linalg.norm(feature2)}$$

Volume & else

$$volume = 10 * log_{10} \sum_{i=1}^{n} s_i^2$$



Demonstration

Thanks a lot !

2021.12.30