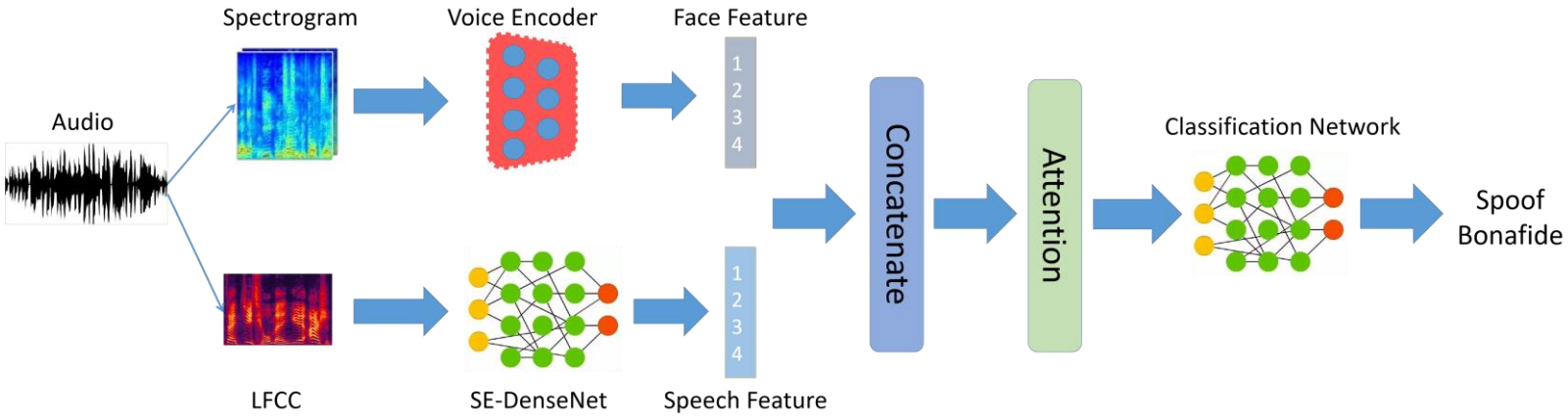
Physiological-Physical Feature Fusion for Automatic Voice Spoofing Detection

Junxiao XUE, Hao ZHOU

Frontiers of Computer Science, DOI: 10.1007/s11704-022-2121-6

Problems & Ideas

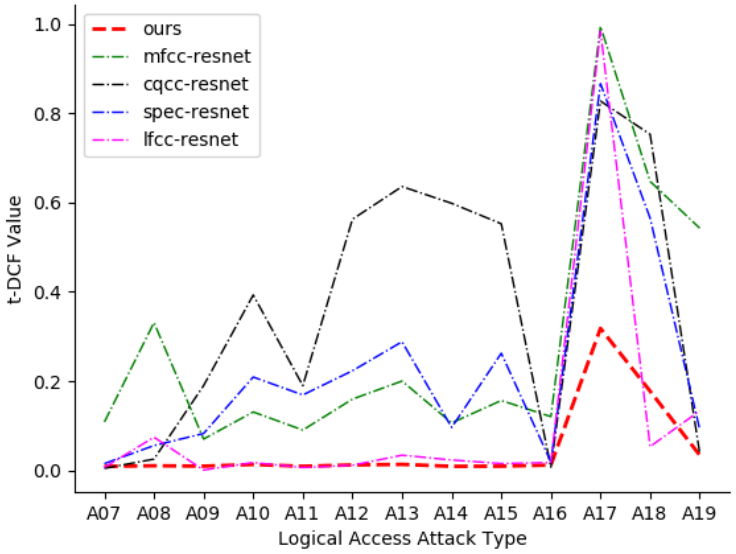
- Problems of automatic voice spoofing detection:
  - It is difficult to design a robust anti-spoofing detection strategy.
  - The existing methods only consider the physical features of speech resulting in poor performance in detection.
- Ideas: An automatic voice spoofing detection method based on combining physiological and physical features.



Our model structure figure. The structure of our model mainly includes five modules. The first module is the speech processing module. We can get the spectrogram and linear cepstral representation of speech through it. The second module is the voice encoder module, which can turn the spectrogram into face features. The third module is the SE-DenseNet module. The input of this module is LFCC, and the output is the speech feature. The fourth module is the feature fusion module, in which we fuse speech feature and face feature by an attention-based feature fusion method. The fifth module is the classification network. This module feeds the fusion features into the classification network to complete the spoofing attack detection.

Main Contributions

- Contributions:
  - A densely connected convolutional neural network with squeeze-and-excitation block is used to extract physical feature.
  - A new form of feature, physiological feature, is used to improve the performance of voice spoofing detection.
  - An attention-based feature fusion strategy is used to fuse physical feature and physiological feature.



t-DCF scores of our model and three baseline models for different attack types (including TTS and VC) in the logical access.