

# Recognition of Spoofed Voice Using Convolutional Neural Networks



Huixin Liang, Xiaodan Lin, Qiong Zhang, Xiangui Kang  
Sun Yat-sen University

## Motivation

- Spoofed voice, including voice conversion and voice transformation can easily ruin the automatic speaker verification (ASV) system.
- The detection for less distorted voices, especially for the disguising factors of  $\pm 4$  semitones that are more popularly employed in the methods of disguise, is not satisfactory.

## Proposed method

There are three steps in the proposed method. First, we should use the equation (1) to disguise the original voice. Second, the short-time Fourier transform (STFT) are taken to preprocess the speeches. Finally, all the data will be sent to the proposed network.

$$f = 2^{\alpha/12} \cdot f_0 \quad (1)$$

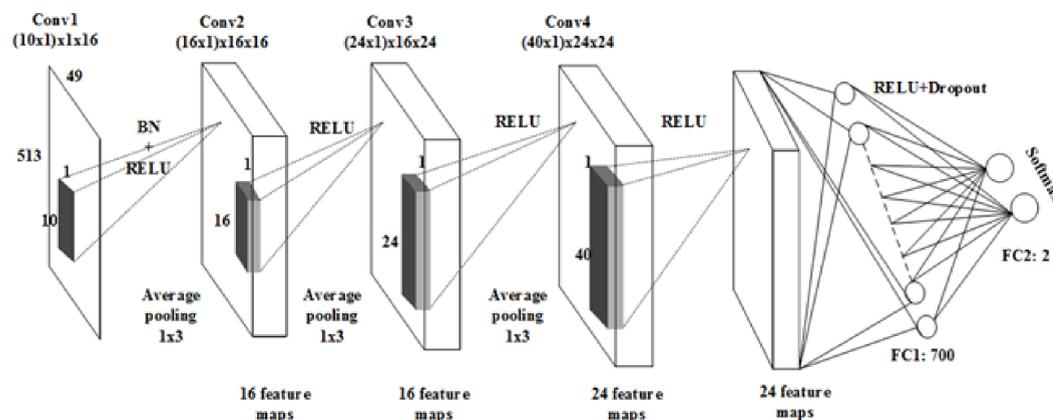


Fig. 2 The architecture of the proposed network

## Results

- Data: the proposed network is trained and tested on TIMIT, NIST and UME corpora. The audio format is WAV, with an 8KHz sampling rate, 16-bit quantization and mono.
- Intra-database evaluation, cross-database evaluation and the Evaluation of the pitch of  $\pm 4$  semitones are reported in Table 1, Table 2 and Fig 3.

Table 1. the detection accuracy of intra-database evaluation

Training database	Testing database	The proposed method	Wu's method [14]
TIMIT_1	TIMIT_2	96.52%	95.87%
NIST_1	NIST_2	95.93%	94.56%
UME_1	UME_2	94.85%	93.63%
Average	Average	95.77%	94.69%

Table 2. the detection accuracy of cross-database evaluation

Training database	Testing database	The proposed method	Wu's method [14]
TIMIT_1/ NIST_1	UME_2	94.37%	93.35%

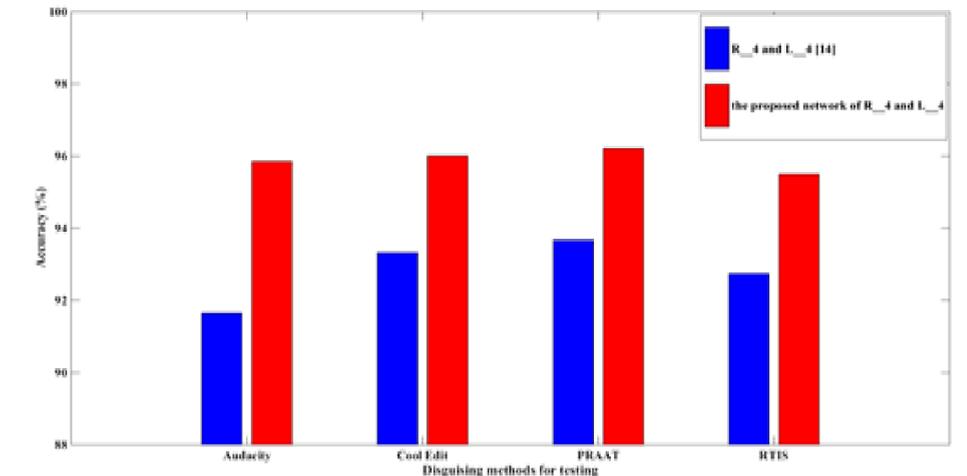


Fig 3. The detection accuracy for spoofed voice by the semitone of  $\pm 4$  from NIST [23], it shows an advantage Gap of 4.2% over Wu's method [14] in Audacity.

## Conclusion

- The proposed method is superior to the state-of-art methods, especially in recognizing less distorted voices.
- The performance of the proposed approach is excellent, even for different databases.

## References

- [14] Wu H, Wang Y, Huang J. "Identification of Electronic Disguised Voices," *IEEE Transactions on Information Forensics & Security*, vol. 9, no.3, pp. 489-500, Mar. 2014.
- [18] Y. LeCun, "Generalization and network design strategies," *In Proceedings of Connectionism in Perspective*, 1989.