Chen, Xuanjun

r09922165@ntu.edu.tw | xjchen.tech | xjchenGit | +86 130-4898-8547 (*WeChat*)

Education

National Taiwan University (NTU) Sep 2020 - Jan 2023
Master of Science in Computer Science and Information Engineering (GPA: 4.20/4.30) *Taipei Taiwan*
Advised by Prof. Hung-Yi Lee and Prof. Jyh-Shing Roger Jang.

National Taiwan University of Science and Technology (Taiwan Tech) Sep 2018 - Jun 2020
Bachelor of Science in Computer Science and Information Engineering (GPA: 4.11/4.30) *Taipei Taiwan*

Shenzhen Institute of Information Technology (SZIIT) Sep 2015 - Jul 2018
Junior College in Electronic Information Technology Engineering *Shenzhen Guangdong*

Experience

National Taiwan University - Multimedia Information Retrieval Lab Oct. 2020 - Present
Research Assistant, supervised by Prof. Jyh-Shing Roger Jang *Taipei Taiwan*

Audio-Visual Neural Network Model

- Proposed multimodal Transformer distillation for the audio-visual synchronization model, reducing teacher parameters by 83.52%, achieving competitive performance and providing comprehensive analysis. [1].
- Revealed the vulnerability of the audio-visual active speaker detection model in many ways, such as single- and multi-modal attacks, three-attack algorithms, white- and black-box attackers, and training-aware and inference-aware scenarios. [2]
- Proposed audio-visual interaction loss enables inter-class dispersion and intra-class compactness, which outperforms the adversarial training by 33.14 mAP (%) under multi-modal attacks. [2]

Automatic Speaker Verification and Spoofing Countermeasures (ASVspoof)

- Ranked 3rd out of 42 teams in the logical access track of the ASVspoof 2021 challenge. [Ranking]
- Proposed ASD-ResNetSE model combines generalized end-to-end pre-training and adversarial fine-tuning, achieves competitive performance with only 22.5% teacher parameters and 19.4% teacher MACs [3].

Singer Separation for Karaoke Content Generation

- Combined the singing voice separation model and the sound source separation model to construct a singer separation dataset and application [4].

Publication & Preprint (*Equal Contribution, †Equal Correspondence)

[1] **Mutlimodal Transformer Distillation for Audio-Visual Synchronisation**
Xuanjun Chen, Haibin Wu, Chung-Che Wang, Hung-Yi Lee†, and Jyh-Shing Roger Jang†
Submitted to ICASSP 2023. [*arXiv*]

[2] **Push-Pull: Characterizing the Adversarial Robustness for Audio-Visual Active Speaker Detection**
Xuanjun Chen*, Haibin Wu*, Helen Meng†, Hung-Yi Lee†, and Jyh-Shing Roger Jang†
In 2022 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2022. [*arXiv*][*demo*]

[3] **Adversarial Speaker Distillation for Countermeasure Model on Automatic Speaker Verification**
Yen-Lun Liao*, **Xuanjun Chen***, Chung-Che Wang, and Jyh-Shing Roger Jang
In Proc. 2nd Symposium on Security and Privacy in Speech Communication (pp. 30-34). [*ISCA*][*arXiv*]

[4] **Singer Separation for Karaoke Content Generation**
Hsuan-Yu Chen, **Xuanjun Chen**, and Jyh-Shing Roger Jang
arXiv preprint arXiv:2110.06707 (2021). [*arXiv*][*demo*]

Award & Scholarship

Distinguished Academic Record Award (4 years), Taipei Kwong Tong Community Associations 2020 - 2023
Certificate of Achievement (3 times), Department of CSIE, Taiwan Tech (top 5% Student) 2019 - 2020
National Bronze Award, 3rd China College Students' "Internet +" Innovation and Entrepreneurship Competition 2017
Guangdong Provincial Gold Award, 3rd China College Students' "Internet +" Innovation and Entrepreneurship Competition 2017
3rd Place of Academic Award, Department of EITE, SZIIT (top 20% Student) 2017
National Encouragement Scholarship, SZIIT (only 3% student) 2016

Activity

Reviewer, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2023
Volunteer, the Fourth Cross-Strait Youth Maker Competition, Tongji University, Shanghai, China 2019

Skill

Programming languages: Python, C++, C
ML/AI: Pytorch, Numpy, Pandas, Matplotlib
Miscellaneous: MySQL, Git, Shell, Latex, Django
Languages: Mandarin, Cantonese, English