

CHEN, XUANJUN

✉ d12942018@ntu.edu.tw | 🏠 xjchen.tech | 🌐 xjchenGit | 📞 +86 130-4898-8547 (WeChat)

EDUCATION

National Taiwan University (NTU)

Ph.D. in Graduate Institute of Communication Engineering

M.S. in Computer Science and Information Engineering, GPA: 4.19/4.30

Advisor: Prof. Hung-Yi Lee and Prof. Jyh-Shing Roger Jang

Taipei Taiwan

Sep. 2023 - Present

Jan. 2023

National Taiwan University of Science and Technology (Taiwan Tech)

B.S. in Computer Science and Information Engineering, GPA: 4.11/4.30

Taipei Taiwan

Jun. 2020

EXPERIENCE

National Taiwan University - Speech Processing and Machine Learning Lab

Research Assistant, supervised by Prof. Hung-Yi Lee

Taipei Taiwan

Sep. 2023 - Present

National Taiwan University - Multimedia Information Retrieval Lab

Research Assistant, supervised by Prof. Jyh-Shing Roger Jang

Taipei Taiwan

Oct. 2020 - Present

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].

PUBLICATIONS & PREPRINTS

(*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[†]

In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024.

[arXiv][Code]

[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. [IEEE][arXiv][Demos][Poster]

[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)., 2022. [ISCA][arXiv]

AWARDS & HONORS

Distinguished Academic Record Award (4 years), Taipei Kwong Tong Community Associations

2020 - 2023

Certificate of Achievement (3 semesters), Dept. of CSIE, Taiwan Tech (top 5% of students)

2019 - 2020

3rd Prize of Academic Award, Dept. of EIET, SZIIT (top 20% of students)

2017

National Encouragement Scholarship, SZIIT (only 3% of students)

2016

ACTIVITIES

Reviewer, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

2023-2024

Reviewer, the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)

2024

Volunteer, the Fourth Cross-Strait Youth Maker Competition, Tongji University, Shanghai, China

2019

SKILLS

Programming languages: Python, C++, C

ML/AI: Pytorch, Numpy, Pandas, Matplotlib

Miscellaneous: MySQL, Git, Shell, Latex, Django

Languages: Mandarin, Cantonese, English