LEYING ZHANG

✉zhangleying@sjtu.edu.cn ✆(+86) 18621098717
3-225 SEIEE Building, Shanghai Jiao Tong University
800 Dongchuan Road, Minhang District, Shanghai, China 200240

RESEARCH INTERESTS

Text-to-speech, Audio generation, Speaker Verification, Multi-modality

EDUCATION

Shanghai Jiao Tong University Sep. 2023 - Present
PhD, Computer Science and Engineering Supervisor: Prof. Yanmin Qian
Shanghai Jiao Tong University Sep. 2021 - 2023
Master, Electronic Information Supervisor: Prof. Yanmin Qian
Télécom Paris (Institut polytechnique de Paris) Sep. 2021 - Feb. 2022
Exchange Student, Data science and Image processing
Shanghai Jiao Tong University Sep. 2017 - Jun. 2021
Bachelor of Information Engineering and French (double degree)

PUBLICATIONS

[C1] **Leying Zhang**, Zhengyang Chen and Yanmin Qian. "Adaptive Large Margin Fine-tuning for Speaker Verification". *2023 IEEE International Conference on Acoustics, Speech and Signal Processing* (**ICASSP**), *June. 2023*

[C2] **Leying Zhang***, Zhengyang Chen* and Yanmin Qian. "Enroll-Aware Attentive Statistics Pooling for Target Speaker Verification ". *23rd Annual Conference of the International Speech Communication Association* (**InterSpeech**), *Sep. 2022*

[C3] **Leying Zhang**, Zhengyang Chen and Yanmin Qian. "Knowledge Distillation from Multi-Modality to Single-Modality for Person Verification". *22nd Annual Conference of the International Speech Communication Association* (**InterSpeech**), *Sep. 2021*

RESEARCH EXPERIENCE

Robust speaker verification Shanghai Jiao Tong University
Supervised by Prof. Yanmin Qian *Oct. 2020 - Dec.2022*

[R1] Adaptive Large Margin Fine-tuning Apr. 2022 - Dec. 2022
Proposed duration-based and similarity-based adaptive large margin fine-tuning strategies. Achieved performance improvement under duration mismatch scenarios and overcomes the performance degradation problems of previous fine-tuning methods.

[R2] Target Speaker Verification May. 2021 - Apr. 2022
Proposed enroll-aware attentive statistic pooling (EA-ASP) layer to help the speaker verification system extract specific speaker's information from multi-speaker speech. Achieved ∼50% relative Equal Error Rate reduction on multi-speaker simulation dataset based on VoxCeleb1.

[R3] Knowledge Distillation from Multi-Modality to Single-Modality Oct. 2020 - Apr. 2021
Applied knowledge distillation from multi-modal system to single-modal system at label, embedding and distribution levels. Achieved 10% Equal Error Rate improvement for the visual system and analyzed the audio system performance and knowledge distillation quality on the VoxCeleb1 evaluation set.

INDUSTRY EXPERIENCE

Research Intern Microsoft, USA
Supervised by Yao Qian *Apr. 2022 - present*

Target speech extraction: Investigated diffusion-based model for target speech extraction. Proposed an efficient generative approach for handling multi- and singlespeaker scenarios in both noisy and clean conditions. Improved inference process of diffusion model in order to regenerate and optimize speech quality based on pre-processed speech from a discriminative model.

Research Intern Microsoft Research Asia
Supervised by Xu Tan *Nov. 2022 - Mar. 2023*

Audio generation: Implemented vector-quantized diffusion model with classifier-free guidance. Achieved 10% improvement over baseline. Investigated latent diffusion model's effects by fine-tuning Stable diffusion.
Text-to-speech: Utilized vector-quantized diffusion model for text-to-speech on large-scale dataset with different neural audio codecs. Generated high-quality speech and get improvements on zero-shot text-to-speech.

TEACHING EXPERIENCE

Teaching Assistant - Machine Learning Fall, 2022
Teaching Assistant - Mobile Communication Systems Fall, 2022

HONORS AND AWARDS

National Scholarship 2022
First place in CN-Celeb Speaker Recognition Challenge 2022 2022
ISCA and Interspeech Travel Grant 2021
Outstanding Graduates of Shanghai 2021
Outstanding student leader of SJTU 2021
Guanghua Scholarship 2020
SJTU Class B Scholarship 2019

SKILLS

Programming skills: Python, Pytorch, C/C++, Matlab
Languages: Chinese(native), English(IELTS 7.5), French(DELF B2), Spanish(Beginner)
Extracurriculars: Piano, Yoga, Badminton