Ruchao Fan

SPAPL 63-134 Engineering IV Building, 420 Westwood Plaza, Los Angeles, CA 90095

Homepage Email Github Phone

Research Interests

Speech Processing, Self-supervised and Unsupervised Learning, Domain Adaptation, End-to-End Speech Recognition (Non-autoregressive transformer, CTC, Transducer), Children's ASR

Education

University of California, Los Angeles (UCLA) **Los Angeles, U.S.A.**
Ph.D. in Electrical and Computer Engineering *Sept. 2019 - Present*

Beijing University of Posts and Telecommunications (BUPT) **Beijing, China**
M.S. in Information and Communication Engineering *Sept. 2016 - Jun. 2019*
B.Eng. in Communication Engineering *Sept. 2012 - Jun. 2016*

Professional Experience

UCLA - Speech Processing and Auditory Perception Lab **Los Angeles, U.S.A.**
Graduate Student Researcher - Advisor: Prof. Abeer Alwan *Sept. 2019 - Present*
Research on children's ASR as a low-resource task
- Develop data augmentation and model pre-training methods to improve children's ASR
- Investigate and develop non-autoregressive models for children's ASR to improve decoding efficiency

Microsoft Corporation. - MSR **Redmond, U.S.A.**
Research Intern - Mentors: Dr. Guoli Ye and Dr. Jinyu Li *June 2021 - Sept. 2021*
Research on spell correction for end-to-end speech recognition
- Investigated methods of non-autoregressive spell correction for a transformer-transducer

PAII Inc. - US Research Lab **Palo Alto, U.S.A.**
Research Scientist Intern - Mentors: Dr. Wei Chu and Dr. Peng Chang *June 2020 - Sept. 2020*
Research on non-autoregressive transformers for end-to-end speech recognition
- Used CTC alignment as extra information for token-level acoustic embedding extraction
- Proposed an error-based sampling method during inference to improve performance

Sogou Inc. - Voice Interaction Technology Center **Beijing, China**
Research Intern - Mentors: Dr. Pan Zhou and Dr. Wei Chen *Apr. 2018 - Aug. 2019*
Research on attention-based encoder-decoder (AED) end-to-end speech recognition
- Proposed an online AED with 3.5% relative WER degradation compared to an offline AED
- Improved transformer's performance with parallel-schedule sampling and relative positional encoding

BUPT - Pattern Recognition and Intelligent System Lab **Beijing, China**
Research Assistant - Advisor: Prof. Gang Liu *Sept. 2016 - Jan. 2019*
Research on automatic speech recognition
- Experimented with connectionist temporal classifiers (CTC) for phoneme recognition [URL]
- Explored methods of frequency warping (normalization) for speaker adaptation

Publications

[9] Gary Yeung, **Ruchao Fan**, and Abeer Alwan, “Fundamental frequency feature warping for frequency normalization and data augmentation in child automatic speech recognition, ”Speech Communication, 2021, doi: https://doi.org/10.1016/j.specom.2021.08.002.

[8] **Ruchao Fan**, Wei Chu, Peng Chang, Jing Xiao and Abeer Alwan, "An Improved Single Step Non-autoregressive Transformer for Automatic Speech Recognition," Proc. Interspeech 2021, pp. 3715-3719.

[7] Jinhan Wang, Yunzheng Zhu, **Ruchao Fan**, Wei Chu and Abeer Alwan, "Low Resource German ASR with Untranscribed Data Spoken by Non-native Children–INTERSPEECH 2021 Shared Task SPAPL System," Proc. of Interspeech 2021, pp. 1279-1283.

[6] **Ruchao Fan**, Wei Chu, Peng Chang, and Jing Xiao, “CASS-NAT: CTC Alignment-based Single Step Non-autoregressive Transformer for Speech Recognition,” in ICASSP 2021, IEEE, pp. 5889–5893.

[5] **Ruchao Fan**, Amber Afshan and Abeer Alwan, "BI-APC: Bidirectional Autoregressive Predictive Coding for Unsupervised Pre-training and its Application to Children’s ASR," in ICASSP 2021, IEEE, pp. 7023–7027.

[4] Gary Yeung, **Ruchao Fan**, and Abeer Alwan, "Fundamental Frequency Feature Nomalization and Data Augmentation for Child Speech Recognition," in ICASSP 2021, IEEE, pp. 6993-6997.

[3] Vijay Ravi, **Ruchao Fan**, Amber Afshan, Huanhua Lu, and Abeer Alwan, “Exploring the use of an unsupervised autoregressive model as a sharedencoder for text-dependent speaker verification,” Proc. Interspeech 2020, pp. 766–770, 2020.

[2] **Ruchao Fan**, Pan Zhou, Wei Chen, Jia Jia, and Gang Liu, “An Online Attention-Based Model for Speech Recognition,” in Proc. Interspeech 2019, 2019, pp. 4390–4394.

[1] **Ruchao Fan** and Gang Liu, “CNN-based audio front end processing on speech recognition,” in 2018 International Conference on Audio, Language and Image Processing (ICALIP). IEEE, 2018, pp. 349–354.

Arxiv Papers

[1] Pan Zhou, **Ruchao Fan**, Wei Chen, and Jia Jia, “Improving generalization of transformer for speech recognition with parallel schedule sampling and relative positional embedding,” arXiv preprint, arXiv:1911.00203.

Skills and Coursework

Computer Languages and Open-source Framework

- Python, C/C++, Shell and Matlab, Pytorch, Tensorflow, Kaldi, Espnet, Fairseq;

Coursework

- ECE: Matrix Analysis, Digital Speech Processing, Advanced Topics in Speech Processing, Linear

Programming; CS: Natural Language Generation, Algorithmic Machine Learning
- Overal GPA: 3.93/4.0

Honors & Awards

Academic Honors
- One-year research funding from PAII Inc. (2020-2021)
- A first-year graduate fellowship for the ECE Ph.D. program at UCLA (2019-2020)
- Huawei Enterprise Scholarship (2017)
- Graduate Academic Scholarship (2016-2018)
- Undergraduate National Inspirational Scholarship (2013-2015)