Jinhan Wang

612-598-4012 wang7875@g.ucla.edu linkedin.com/in/jinhan-wang-607393175

Education

University of California, Los Angeles **Present – Sept 2024 (Expected)**
Ph.D, Electrical and Computer Engineering *Los Angeles, CA*

University of California, Los Angeles **June 2021**
Master of Science, Electrical and Computer Engineering (GPA: 4.00/4.00) *Los Angeles, CA*

University of Minnesota, Twin Cities **May 2019**
Bachelor of Electrical Engineering (Minor: Computer Science) (GPA: 3.96/4.00) *Minneapolis, MN*

Beijing Jiaotong University **July 2017**
Bachelor of Engineering, Automation (GPA: 91.2/100, Top 1 out of 28 students) *Beijing, China*

Technical Skills

Languages: Python, MATLAB, R
Technologies/Frameworks: Linux, Tensorflow, Keras, PyTorch, Hugging face, PEFT, Scikit-learn, NLTK, Kaldi

Research Projects

Depression Detection with Acoustic Cue and Linguistic Cue from LLM **Nov 2023 – Present**
Instructor: Prof. Abeer Alwan, University of California, Los Angeles

- Significance of depression from linguistics and acoustics varies across individuals. Some tend to have more salient patterns in terms of monotonic pitch, and hesitation, but others tend to choose different words.
- The first approach tries to exploit the potential of large language models in depression detection, in addressing the challenge of inter-correlation between acoustic and linguistic cues.
- Asynchronous alignment between significant acoustic cues and linguistic cues.
- Experiments will be conducted over English and Mandarin datasets to verify robustness across various domains.

Speechformer-CTC: Sequential Modeling of Depression from Speech Signals **June 2023 – Present**
Instructor: Prof. Abeer Alwan, University of California, Los Angeles

- Capture fine-grained non-uniform depression patterns through temporal classification for better depression detection.
- Propose one-hot policy and Hubert policy to generate pseudo CTC-label for alignment over multiple hierarchical stages of speech signals.
- The dynamic pattern of non-uniform depression states is verified through differential decoded CTC-label distribution.
- Word-level sequential modeling is shown to be the most promising configuration for depression detection and has great compatibility with ASR features on English and Mandarin datasets.
- The generic methods are applicable for paralinguistic speech processing tasks, such as emotion recognition, and Alzheimer's disease detection.

Turn-taking and Backchannel Prediction with Acoustic and LLM Fusion **June 2023 – Sept 2023**
Instructor: Long Chen, Amazon

- Utilize the language understanding capability of large language models in solving turn-taking and backchannel location prediction tasks in conversational dialog, to build a more natural voice assistant system.
- Propose a novel instruction-multitask fine-tuning reformulation by providing the model with explicit instructions to improve backchannel location detection sensitivity.
- Fuse acoustic and linguistic modalities to obtain complementary information.
- Explore the effect of including content history in instruction.

Privacy-preserving Depression Detection with Speaker Disentanglement **Sept 2022 – Oct 2023**
Instructor: Prof. Abeer Alwan, University of California, Los Angeles

- Propose an adversarial minimization-maximization technique to attenuate speaker-related information and preserve depression characteristics. Address the privacy concern when applying automatic systems in clinic use cases.
- Propose a novel Non-uniform Speaker Disentanglement framework to leverage differential behaviors between model layers, in terms of extracting depression/speaker information with varying quality and quantity.
- Three non-adversarial approaches based on loss equalization across speakers with KL-divergence/Variance/Cross-entropy are proposed to resolve the instability issue in adversarial methods.

EdgeDD: Device Directedness Personalization on the Edge **June 2022 – Sept 2022**

Instructor: Tobias Menne, Amazon

- A lightweight late fusion method, centroid distance fusion (CDF), is proposed to combine cloud-end model prediction and user-end personalized model prediction to achieve a better device-directedness utterance detection performance.
- Propose the method, environment variable estimation (EVE), in estimating environment attribute as a quantitative measurement of domain variation between universal set and target set.
- Incorporate adapter to further mitigate computation burden on user-end for faster personalization.
- Personalized detection performance can be improved with as few as 2 available sentences from the target speaker.

Unsupervised IDL Pretraining for Depression Detection from Speech **April 2021 – April 2022**

Instructor: Prof. Abeer Alwan, University of California, Los Angeles

- A generic unsupervised pretraining framework, Instance Discriminative Learning (IDL), is proposed to train a high-level, low-dimensional feature extractor for downstream task initialization.
- Investigate various augmentation methods in the pretraining stage and analyze their effects w.r.t depression state.
- Explore the correlation between speaker-identity information and depression status by setting sampling strategies in the pretraining stage.
- Propose a novel sampling technique, Pseudo-instance Sampling (PIS), to utilize clustering algorithms to reveal a deeper correlation between IDL embeddings and underlying acoustic units (depression status, specifically).
- With the help of the ablation speaker classification study, we verify that speaker-related information might help with depression classification, i.e. widely-used speaker-related features can also be used to distinguish depression status.

VADOI: Voice-Activity-Detection Overlapping Inference for Long-form ASR **June 2021 – Sept 2021**

Instructor: Xiaosu Tong, Amazon

- Incorporating VAD into the previous proposed Partial Overlapping Inference (POI) method for better ASR long-form decoding, resolving the challenge of word boundary cropping.
- A comprehensive comparison of OI and POI with different levels of OI modeling.
- Propose a novel Soft-Match mechanism to mitigate misaligned but similar words challenge.
- Achieve equivalent performance in terms of WER with 20% computation cost reduction.

Low Resource German ASR with Untranscribed Data of Non-native Children **March 2021 – April 2021**

Instructor: Prof. Abeer Alwan, University of California, Los Angeles

- Propose novel Non-speech Discriminative Loss (NSDL) to handle the majority of long-term non-speech segments within utterances, including laugh, hesitation, and noise.
- Utilize novel Bi-APC unsupervised pretraining strategy to learn common knowledge from the untranscribed dataset.
- Apply Incremental Semi-Supervised Learning (ISSL) to generate pseudo transcriptions for the untranscribed dataset in multiple folds. (Filter the pseudo transcription by log-likelihood)
- Augment the data using VTLP, Speed Perturbation, Pitch Perturbation, Volume Perturbation, and Noise Perturbation (background and foreground).
- Improve the results through language model re-scoring.

Study Data-driven based Comment Generation in Restaurant Domain **Oct 2020 – Jan 2021**

Instructor: Prof. Nanyun Peng, University of California, Los Angeles

- Investigate different input representations including Padding and Filling (PAF), Structured Data Embedding (SDE), and Boundary Based Embedding (BBE).
- Design multiple encoder-decoder models with an attention mechanism. Implement Transformer and compare.
- Regularize the model with the proposed NOS method, to control the complexity of the generated sentences in a soft manner. By applying NOS, the generated sentences are forced to have more subordinate sentences rather than the concatenation of simple sentences.

Experience

Speech Processing and Auditory Perception Laboratory **June 2021 – Present**

Graduate Student Researcher *Los Angeles, CA*

- Sequential modeling of depression detection from speech with Connectionist Temporal Classification, in capturing dynamic/non-uniform depression patterns over the temporal domain for an individual.
- Unsupervised pretraining techniques in depression detection from speech signals.
- Implement research on data augmentation method in ASR for Children's Mandarin dataset with style-variation.
- JIBO Children and Kindergarten dataset preparation for education purposes.

Amazon **June 2021 – Sept 2021 & June 2022 – Sept 2022 & June 2023 – Sept 2023**

Applied Scientist Intern *Sunnyvale, CA*

- 2021: Improve long-form ASR performance through novel decoding operation, named VADOI, as an extension of POI in mitigating boundary distortion.
- 2022: Propose a light-weight fusion method, CDF, to improve Device-directedness utterance detection on the user end in low-resource scenario.
- 2023: Utilize large language model and fuse it with acoustic modality for speaker activity detection in conversational dialog, to get a more natural voice assistant system.

Publications

Wang, J., Zhu, Y., Fan, R., Chu, W., & Alwan, A. (2021). Low Resource German ASR with Untranscribed Data Spoken by Non-Native Children: INTERSPEECH 2021 Shared Task SPAPL System. Interspeech 2021

Ravi, V., **Wang, J.**, Flint, J., & Alwan, A. (2022). Fraug: A Frame Rate Based Data Augmentation Method for Depression Detection from Speech Signals. In ICASSP 2022 2022 (pp. 6267 6271). IEEE.

Wang, J., Tong, X., Guo, J., He, D., & Maas, R. (2022). VADOI: Voice Activity Detection Overlapping Inference for End-To-End Long Form Speech Recognition. In ICASSP 2022 2022 (pp. 6977 6981). IEEE.

Ravi, V., **Wang, J.**, Flint, J., & Alwan, A. (2022). A Step Towards Preserving Speakers' Identity While Detecting Depression Via Speaker Disentanglement. Interspeech 2022

Wang, J., Ravi, V., Flint, J., & Alwan, A. (2022). Unsupervised Instance Discriminative Learning for Depression Detection from Speech Signals. Interspeech 2022

Fan, R., Zhu, Y., **Wang, J.**., & Alwan, A. (2022). Towards Better Domain Adaptation for Self-supervised Models: A Case Study of Child ASR. IEEE Journal of Selected Topics in Signal Processing

Wang, J., Ravi, V., Flint, J., & Alwan, A. (2023). Non-uniform Speaker Disentanglement for Depression Detection from Speech Signals. Interspeech 2023

Ravi, V., **Wang, J.**, Flint, J., & Alwan, A. (2023). Enhancing accuracy and privacy in speech-based depression detection through speaker disentanglement. Computer Speech & Language (2023): 101605

Wang, J., et al. (2023). Turn-taking and Backchannel Prediction with Acoustic and Large Language Model Fusion. ICASSP 2024

Awards

Dean List student for academic achievement in University of Minnesota	Sept 2017 – May 2019
High distinction graduation student awarded by University of Minnesota	May 2019
Dean List student for academic achievement in University of Minnesota	Sept 2017 – May 2019
First-class Social Work Scholarship awarded by Beijing Jiaotong University	Sept 2016
Second-class Scholarship for Academic Excellence awarded by Beijing Jiaotong University	Sept 2016
Third-class Scholarship for Social Practice awarded by Beijing Jiaotong University	Sept 2016
Outstanding Volunteer for the 120th anniversary of Beijing Jiaotong University	Sept 2016

Relevant Coursework

- Matrix Analysis for Scientists and Engineers
- Linear Programming
- Stochastic Processes
- Digital Image Processing
- Digital Speech Processing
- Speech and Image Processing Systems Design
- Neural Networks and Deep Learning
- Large-Scale Data Mining
- Large-Scale Social and Complex Networks
- Signal and Image Processing for Biomedicine
- Algorithmic Machine Learning
- Security in Circuits and Embedded Systems
- Advanced Topics in Natural Language
- etc.