Siyuan Shan

📞+1-919-537-6515 | 📧 shansiliu@outlook.com | 🖥️ Personal Website | LinkedIn | Google Scholar

🪪 Work Authorization: EB1 Green Card recipient. Authorized to work in the U.S. **without** sponsorship.

🏠 San Francisco Bay Area, CA | Open to **local opportunities only**

Objective

👩‍💻 Seeking opportunities for Applied Scientist / Research Scientist / MLE roles in areas of LLM / Speech / Audio GenAI.
📄 Publications at top ML conferences such as NeurIPS / AAAI / ICASSP on topics of Generative AI / Audio Synthesis / Transformer / Attention Mechanism / Voice Conversion / Meta-learning / Memory-augmented Neural Nets.
⚙️ Hands-on experience for LLM fine-tuning / Direct Preference Optimization / Parallel Training / LLM evaluation / synthetic data generation / model deployment / prompt engineering.
⚙️ Knowledge of signal processing techniques (e.g., STFT, FFT, MFCC extraction), audio datasets (e.g., LibriSpeech, NSynth), SOTA deep learning models (e.g., fastspeech, diffusion models, differentiable digital signal processing, WavLM), vocoders, and evaluation metrics (e.g., MOS, intelligibility, speaker similarity).

Experience

- **Walmart Global Tech [✳]** *Feb 2023 - Now*
  Senior Data Scientist - Conversational AI Sunnyvale, CA
  - Fine-tune a Llama-3-3B model with DPO, Knowledge Distillation and SFT to make it a judge model that can automatically evaluate answers generated by Walmart Chatbot. Training data consists of synthetic data generated by a large Llama-3.3-70B model as well as human-labeled data from real traffic. The fine-tuned model achieves 20% increase in F1 scores compared to the model using prompt engineering.
  - Create training data & fine-tune & evaluate BERT model for Customer Intent Classification (e.g. search, add to cart, request agents, cancel orders, etc.) and Named Entity Recognition (e.g. brand, product, quantity, etc.). The model is deployed using Triton Inference Server to support a wide range of Walmart applications, including voice ordering, text-to-shop, customer care chatbots, and AskSam.
  - Develop prompts to improve the performance of GPT-4(o) and Llama-3-8b on generating product search keywords from a conversation between customers and an AI chatbot. Prepare a dataset using Labelbox for this task with 10k turns of conversations and human annotations of search keywords, which could be used for evaluation and fine-tuning.

- **Tiktok [TikTok]** *Summer 2020 - 2022*
  Summer Research Intern @ Speech, Audio & Music Intelligence (SAMI) Team Mountain View, CA
  - 2022 Summer: Develop a TTS model based on FastSpeech2 that can edit speech by editing input texts. The model is trained in a way similar to BERT to predict masked spectrogram based on the bidirectional spectrogram and phoneme context. Our model can generate realistic and natural speech with only 5 seconds of speech from an unseen speaker.
  - 2021 Summer: Propose a neural audio synthesis technique for high-fidelity, real-time, and interactive instrument sound generation in the time domain. Our model learns a dictionary of one-period waveforms (i.e. wavetables) end-to-end with a GRU autoencoder. Compared to DDSP, our model runs 10x faster with the same generation quality. Write a paper accepted by *ICASSP 2022* as well as a patent.
  - 2020 Summer: Extend GANSynth for flexible instrument sound generation by interpreting the latent space of GAN. Experiments show that our model can find a direction in the latent space to control a specific aspect of the generated sounds, such as velocity, duration, distortion, and reverb.

- **University of North Carolina at Chapel Hill [UNC]** *Aug 2018 - May 2023*
  PhD Researcher in Machine Learning at LUPA Lab Chapel Hill, NC
  - Develop a variational set expansion model based on Transformer for voice conversion (altering a persons voice while preserving linguistic content). Our method does not require any text annotation, support any-to-any speaker conversion and only need a very short target speaker's voice prompt (3s). Our method achieves better speaker similarity and similar speech quality compared to SOTA. Write a paper accepted by *AAAI 2024*.
  - Develop a hierarchical time series imputation model based on Transformer. The model is trained in a way similar to BERT to predict missing time series data based on observed context. The proposed method achieves lower imputation error compared to recurrent models (RNNs, ODEs). Write a paper accepted by *ICASSP 2023*.
  - Develop a meta-learning model that learns memory-augmented neural networks for improving image classification performance. Our method improves classification accuracy by 1%-3% on CIFAR-100 against baselines like ResNets. Write a paper accepted by *NeurIPS 2020*.
  - Develop an interpretable model based on kernel mean embeddings for single cell set classification. Write a paper accepted by *ACM-BCB 2022*.

- **Kaggle Freesound General-Purpose Audio Tagging Challenge[kaggle]** *May 2018 - Aug 2018*
  Beijing, China
  - Combine 1D ConvNets to process raw audio and 2D ConvNets to process Mel Spectrogram for audio classification. Our team ranked 16/558 (top 3%) among all participants.

- **Microsoft Research Asia [ ]** *May 2016 - Dec 2017*
  Undergraduate Researcher at Microsoft Research Asia Beijing, China
  - Develop a framework to detect and recognize subtitles in East Asian languages from videos. Write a paper accepted by *Signal Processing: Image Communication*.
  - Develop an unsupervised medical image registration method based on CNNs and Spatial Transformer Networks.

Education

- **University of North Carolina at Chapel Hill** *Aug 2018 - May 2023*
  PhD in Computer Science Chapel Hill, NC
  - Thesis: Leveraging Related Instances for Better Prediction.
- **Beihang University** *Sep 2013 - May 2017*
  Bachelor in Biomedical Engineering Beijing, China
  - GAP: 3.9/4.0, Ranking: 1/54
- **Czech Technical University in Prague** *Aug 2015 - Jan 2016*
  Exchange Student in Electrical Engineering Prague, Czech Republic

Publications

C=Conference, J=Journal, P=Patent, S=In Submission, T=Thesis

[C] **Siyuan Shan**, Yang Li, Amartya Banerjee, Junier Oliva (2024). **Phoneme Hallucinator: One-shot Voice Conversion via Set Expansion**. In *AAAI Conference on Artificial Intelligence (AAAI) 2024*.

[C] **Siyuan Shan**, Yang Li, Junier Oliva (2023). **NRTSI: Non-Recurrent Time Series Imputation**. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2023*.

[T] **Siyuan Shan** (2023). **Leveraging Related Instances for Better Prediction**. In *PhD Thesis 2023*.

[C] **Siyuan Shan**, Lamtharn Hantrakul, Jitong Chen, Matt Avent, David Trevelyan (2022). **Differentiable Wavetable Synthesis**. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2022*.

[C] **Siyuan Shan**, Vishal Baskaran, Haidong Yi, Jolene Ranek, Natalie Stanley and Junier Oliva (2022). **Transparent Single-Cell Set Classification with Kernel Mean Embeddings**. In *ACM-BCB 2022*.

[C] Vishal Baskaran, Jolene Ranek, **Siyuan Shan**, Natalie Stanley and Junier Oliva (2022). **Distribution-based Sketching of Single-Cell Samples**. In *ACM-BCB 2022*.

[S] Yang Li, **Siyuan Shan**, Qin Liu, Junier Oliva (2021). **Towards Robust Active Feature Acquisition**. In *Arxiv 2021*.

[P] Lamtharn Hantrakul, **Siyuan Shan**, Jitong Chen, Matthew David Avent, David Trevelyan (2021). **Differentiable Wavetable Synthesis - An efficient method of neural audio synthesis using data-driven waveform dictionaries**. In *US Patent 17/525,814*.

[C] **Siyuan Shan**, Yang Li, Junier Oliva (2020). **Meta-Neighborhoods**. In *Neural Information Processing Systems (NeurIPS) 2020*.

[C] Yang Li, Haidong Yi, Christopher M Bender, **Siyuan Shan**, Junier Oliva (2020). **Exchangeable Neural ODE for Set Modeling**. In *Neural Information Processing Systems (NeurIPS) 2020*.

[J] Meijun Liu, Jicong Zhang, Wenxiao Jia, Qi Chang, **Siyuan Shan**, Yegang Hu, Dangxiao Wang (2019). **Enhanced executive attention efficiency after adaptive force control training: behavioural and physiological results**. In *Behavioural Brain Research*.

[J] Yan Xu, **Siyuan Shan**, Ziming Qiu, Zhipeng Jia, Zhengyang Shen, Yipei Wang, Mengfei Shi, I Eric, Chao Chang (2018). **End-to-end subtitle detection and recognition for videos in East Asian languages via CNN ensemble**. In *Signal Processing: Image Communication*.

[S] **Siyuan Shan**, Wen Yan, Xiaoqing Guo, Eric I-Chao Chang, Yubo Fan, Yan Xu (2018). **Unsupervised end-to-end learning for deformable medical image registration**. In *Arxiv 2017*.

Skills

- **Programming Languages:** Python, Latex, SQL
- **Deep Learning Frameworks:** Pytorch, TensorFlow, Keras, Caffe

Activities

- **Reviewer for** ICML 2021-2024, NeurIPS 2021-2023, ICLR 2022-2024, ACM-MM 2022, TCBB, Interspeech 2023-2024, WASPAA 2023, ICASSP 2024.