

# Siyuan Shan

919-537-6515 | [shansiliu@outlook.com](mailto:shansiliu@outlook.com) | [shansiliu95.github.io](https://github.com/shansiliu95) | [Google Scholar](#)

Research Interests: Machine Learning, Artificial Intelligence, Audio Processing & Synthesis, Text Speech Multimodal Learning, Transformer, Few-shot/Meta Learning, Self-supervised Learning, Natural Language Processing

## ABOUT

---

Fast and self-motivated learner, great team player, efficient communicator. Solid machine learning and coding foundations.

I have more than 6 years of experience in developing cutting-edge machine learning algorithms in several areas, such as neural music synthesis, speech synthesis, meta-learning, interpretable machine learning, set modeling, time series imputation, bio-informatics, image registration, and biomedical signal processing, video caption recognition.

My efforts in these areas have led to 11 academic papers, 2 commercialized products, and 1 patent. My papers are published in top machine learning conference NeurIPS and top audio processing conference ICASSP. I also won the best paper award in ACM BCB 2022.

## EDUCATION

---

### University of North Carolina at Chapel Hill

*PhD in Computer Science*

Chapel Hill, NC

*Aug. 2018 – Feb. 2023*

### Beihang University

*Bachelor in Biomedical Engineering, GAP: 3.9/4.0, Ranking: 1/54*

Beijing, China

*Aug. 2013 – May 2017*

### Czech Technical University in Prague

*Exchange Student in Electrical Engineering*

Prague, Czech Republic

*Aug. 2015 – Jan 2016*

## EXPERIENCE

---

### Senior Software Engineer at Walmart

*Advisor: Philippe Hanrigou*

Feb 2023 – now

*Sunnyvale, CA*

- Conversational AI. Building intelligent Chatbots, Chatbot Frameworks for various requirements, and automating the chatbot building process using AI/ML.

### Research Intern at ByteDance AI Lab

*Advisor: Mingbo Ma*

May 2022 – Aug 2022

*Mountain View, CA*

- We propose a model to edit speech content by only editing the input text. Operations supported are word replacement, insertion and deletion. Supported languages are English and Chinese.
- Our model can generate realistic and natural speech with only 5 seconds of speech from an unseen speaker.
- Our model is based on a SOTA TTS model FastSpeech2 and is trained on large-scale with 24 V-100 GPUs.
- This model will be officially launched at TikTok soon.

### Research Intern at ByteDance AI Lab

*Advisor: Jitong Chen, Hanoi Hantrakul*

May 2021 – Aug 2021

*Mountain View, CA*

- We propose a neural audio synthesis technique that learns a dictionary of one-period waveforms (i.e. wavetables).
- The waveforms are learned end-to-end with a neural network that uses an attention mechanism to combine the waveforms.
- We achieve high-fidelity audio synthesis with as little as 10 to 20 wavetables and demonstrate how a data-driven dictionary of waveforms opens up unprecedented one-shot learning paradigms on short audio clips. Our method is also fast enough for realtime and interactive audio synthesis.
- This project empowers the music creation app Mawf.

### Research Intern at ByteDance AI Lab

*Advisor: Jitong Chen*

May 2020 – Aug 2020

*Mountain View, CA*

- Extending GANSynth for flexible instrument sound generation by interpreting the latent space of GAN

- Our model can control several key aspects of the generated sounds, such as velocity, duration, distortion and reverb

## Graduate Researcher

Aug 2018 – Nov 2022

Advisor: Junier Oliva

Chapel Hill, NC

- Developing a instance-wise feature selection algorithm based on Proximal Policy Optimization. We aim at improving the interpretability of machine learning models.
- Develop a multi-resolution time series imputation model NRTSI that is based on the attention mechanisms. Thanks to the non-recurrent nature, NRTSI does not suffer from the error compounding problem of previous works (e.g. NAOMI, LatentODE, NeuralCDE). NRTSI is inspired by the permutation-equivariant modeling of sets and achieves SOTA performance across 8 time series imputation benchmarks
- Develop a meta learning approach called Meta-Neighborhoods that learns a set of neighbors along with the model parameters and flexibly adapt the model using these neighbors at inference time. Better discriminative learning performance is achieved across a wide range of image classification datasets (e.g. CIFAR, CINIC, ImageNet, MNIST, SVHN) and regression datasets from the UCI machine learning repository
- Work on a set modeling model called ExNODE that adds beneficial inductive bias (e.g. permutation invariance and equivariance) to Neural Ordinary Differential Equation (NODE) to model sets. SOTA performance is achieved for point cloud classification, generation and temporal point cloud modeling
- Developing an interpretable single cell classification model based on random Fourier features, kernel mean embedding and kernel herding. Our model achieves better performance than STOA deep learning methods and provides a better interpretability.
- Develop a single cell clustering model that is build on Gaussian Mixture Model. The model has  $n$  cluster centers and each center only has a subset of foreground features selected.  $L_0$  regularization is applied to encourage the selected features to be sparse. There is another cluster center that handles background features.

## Kaggle Freesound General-Purpose Audio Tagging Challenge

Mar 2018 – May 2018

Advisor: Yi Ren

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

## Undergraduate Research Assistant

May 2016 – July 2017

Advisor: Yan Xu

Beijing, China

- Work on unsupervised medical image registration using U-Net and Spatial Transformer Network
- Work on video subtitle detection based on color histogram and recognition using CNNs

## Undergraduate Research Assistant

Feb 2016 – Apr 2016

Advisor: Jicong Zhang

Beijing, China

- Work on a project of exploiting patterns that are indicative of sustained attention from EEG data

## PUBLICATION

### Audio/Time Series/Sequence Data Processing & Synthesis

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

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

Meijun Liu, Jicong Zhang, Wenxiao Jia, Qi Chang, **Siyuan Shan**, Yegang Hu, Dangxiao Wang. “Enhanced executive attention efficiency after adaptive force control training: behavioural and physiological results,” in *Behavioural Brain Research* 2019.

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

**Siyuan Shan**, Yi Ren. “Automatic Audio Tagging with 1D and 2D Convolutional Neural Networks” in *Detection and Classification of Acoustic Scenes and Events* 2018.

## Meta Learning/Reinforcement Learning for Discriminative Learning

**Siyuan Shan**, Yang Li, Junier Oliva. “Meta-Neighborhoods,” in *Neural Information Processing Systems (NeurIPS)* 2020.

Yang Li, **Siyuan Shan**, Qin Liu, Junier Oliva. “Towards Robust Active Feature Acquisition,” in *Arxiv*, 2021.

## Set Modeling

**Siyuan Shan**, Vishal Baskaran, Haidong Yi, Jolene Ranek, Natalie Stanley and Junier Oliva. “Transparent Single-Cell Set Classification with Kernel Mean Embeddings,” in *ACM-BCB 2022 (Best Paper Award)*.

Yang Li, Haidong Yi, Christopher M Bender, **Siyuan Shan**, Junier Oliva. “Exchangeable Neural ODE for Set Modeling,” in *Neural Information Processing Systems (NeurIPS)* 2020.

Vishal Baskaran, Jolene Ranek, **Siyuan Shan**, Natalie Stanley and Junier Oliva. “Distribution-based Sketching of Single-Cell Samples,” in *ACM-BCB 2022*.

## Image Registration

**Siyuan Shan**, Wen Yan, Xiaoqing Guo, Eric I-Chao Chang, Yubo Fan, Yan Xu. “Unsupervised end-to-end learning for deformable medical image registration,” in *Arxiv* 2017.

## PATENT

---

Lamtharn Hantrakul, **Siyuan Shan**, Jitong Chen, Matthew David Avent, David Trevelyan. “Differentiable Wavetable Synthesis - An efficient method of neural audio synthesis using data-driven waveform dictionaries,” *US 17/525,814*, 2021.

## TECHNICAL SKILLS

---

**Languages:** Python, C/C++, MATLAB, Latex

**Deep Learning Frameworks:** Pytorch, TensorFlow, Keras, Caffe

## ACTIVITIES

---

Reviewer for ICML 2021, NeurIPS 2021 2022, ICLR 2022, ICML 2022, IEEE/ACM Transactions on Computational Biology and Bioinformatics, ACMMM 2022

Ranked 16/558 (top 3%) Kaggle Freesound General-Purpose Audio Tagging Challenge

## AWARDS

---

Best Paper Award, ACM BCB 2022

Outstanding Graduates of Beihang University, 2017

National Scholarship, 2015