# Xinru Shan

+86 18845797625 | libchanz@gmail.com | Homepage | Suzhou, China

## Research Interests

I am always passionate about systems, particularly in building innovation-driven high-performance system. My current work involves the development of large-scale cloud service system, providing me with a deeper insight of industrial cloud service architecture. Additionally, I have engaged in AI research, such as speech processing, deep learning, LLM application. I am eager to dive into the intersection of system and AI, such as machine learning system, distributed system and performance optimization.

**Keywords**: Systems, Machine Learning System, Distributed System, Cloud Architecture.

## Education

**Harbin Institute of Technology** Harbin, China

*M.Sc. in Computer Science, Faculty of Computing* *Sep 2020 – Jun 2022*

- GPA: top 10%
- Advisor: Jie Liu

**Harbin Institute of Technology** Harbin, China

*B.Sc. in Internet of Things, Faculty of Computing* *Aug 2016 – Jun 2020*

- GPA: 87.13/100(top 3%)
- Advisor: Jie Liu

## Research Experience

**Room Impulse Response Generator Based on VAE for Data Augmentation** Harbin, China

*Master's thesis, Advisor: Jie Liu, Harbin Institute of Technology* *Oct 2021 – May 2022*

- In comparison to the limitations of mainstream geometric methods, we propose a shallow, stable, and robust Room Impulse Response (RIR) generator based on Variational Autoencoder (VAE).
- Extend unsupervised RIR-VAE to RIR-CVAE, and the synthesis of Room Impulse Response involves generating varied acoustic scenes based on these acoustic parameters.
- In the evaluation experiment of Auto Speech Recognition, RIR-VAE demonstrates superior robustness compared to geometric methods, achieving a noteworthy 6.8% higher accuracy and an 8.52% lower Word Error Rate (WER).

**Cough Automatic Localization and Detection System** Harbin, China

*Bachelor's thesis, Advisor: Jie Liu, Harbin Institute of Technology* *Jan 2020 – May 2020*

- Amidst the COVID-19 pandemic, we implemented a real-time system for the detection and localization of cough sounds using a microphone array. This system seamlessly integrates sound source location, beamforming, and cough detection techniques.
- We propose to detect cough sound by SVM RBF-based classification model, which trained on AudioSet and YouTube audio, demonstrated remarkable performance metrics, including 95.3% F1 index, 91% sensitivity, 83% specificity, and a 0.9 AUC value.

**Robustness and Survey of Segment Anything Model** Suzhou, China

*Independent research, Advisor: Chaoning Zhang, Remote* *May 2023 – July 2023*

- Conduct survey on application and research work related to SAM.
- Investigate the application of SAM in autonomous driving and specifically explore its robustness under adverse weather conditions.

**Independent Research in DeepSpeed Team** Suzhou, China

*Independent research, Advisor: Zhen Zheng, Microsoft* *Oct 2023 – present*

- To gain a better understanding of DeepSpeed, delve into research papers and familiarize with GPU profiling.
- Explore the KV cache optimization in inference.

## Work Experience

**Microsoft** Suzhou, China

*Software Development Engineer* *Jul 2022 – Present*

- M365 Core team, work on the compliant microservice platform named COSMIC, which supports build and deploy service via Azure Kubernetes Service(AKS).
- Build and maintain Namespace Placement System(NPS) to end-to-end onboard namespace to cluster, achieve fast provisioning and 99.99% availability.
- Design and build reliable data flow based on OLAP log platform for monitoring, provide perspective for service owner and customer.
- Design and build a new NPS permission solution to improve the security in COSMIC.
- [Hack Project] To leverage GPT with troubleshooting scenarios, we propose SmartTSG, a solution that combines documentation, execution, and AI capabilities. This integration significantly improves troubleshooting efficiency while promoting knowledge reusability.

**Alibaba Group** Hangzhou, China

*Software Development Engineer Intern* *Jun 2021 – Aug 2021*

- Work on Idle Fish Application, the largest C2C second-hand trading platform in China.
- Contribute to log module for enhancing maintaining platform security, trace problems and data analysis.
- Optimize 5x query latency by improving concurrency and batching.

## Publications & Preprints

**Xinru Shan**, Chaoning Zhang. Robustness of Segment Anything Model (SAM) for Autonomous Driving in Adverse Weather Conditions. (arXiv preprint) 2023.

Chaoning Zhang, Sheng Zheng, Chenghao Li, Yu Qiao, Taegoo Kang, **Xinru Shan**, Chenshuang Zhang et al. A Survey on Segment Anything Model (SAM): Vision Foundation Model Meets Prompt Engineering. (arXiv preprint) 2023.

## Awards & Achievements

**Postgraduate Scholarship**

**The national college Internet of Things (IoT) competition. (The Third Prize)**

**People Scholarship (The First Prize)**

**Huawei ICT Competition (World Final)**

## Skills & Interests

**Programming:** Python, C#, Go, C/C++, Java, Cuda

**Tools:** Kubernetes, Docker, Git, Vim, GPU profiling

**Teaching Assistant:** C Programming Language (Spring 2021), Data and Structure (Fall 2021)

**Languages:** Mandarin (Native), English (Professional)

**Interests:** Badminton, Running, Guitar/Bass/Drum in band.