

# Susan Liang

(+1) 585-410-9586 | [sliang22@ur.rochester.edu](mailto:sliang22@ur.rochester.edu) | [liangsusan-git.github.io](https://liangsusan-git.github.io)

Second year Ph.D. candidate, Computer Science

*Looking for Summer 2024 Internship from June to August (3 Months)*

## EDUCATION

---

### University of Rochester

NY, USA

*Ph.D. of Computer Science*

*Sept. 2022 – May 2027 (Expected)*

- Advisor: Prof. Chenliang Xu
- GPA: 4/4

### University of Chinese Academy of Sciences

Beijing, China

*Bachelor of Computer Science*

*Sept. 2018 – Jul. 2022*

- Advisor: Prof. Shiguang Shan
- GPA: 3.90/4, Rank: 4/104

## RESEARCH INTERESTS

---

- Audio-Visual Understanding and Generation
- Implicit Neural Fields
- Multi-Modal Learning
- Trustworthy AI

## PUBLICATION

---

### **Dream Your Sounding Object Thousand Ways: Language-Guided Audio-Visual Personalization**

**Susan Liang**, Chao Huang, Yapeng Tian, Anurag Kumar, Chenliang Xu

(In submission)

### **DAVIS: High-Quality Audio-Visual Separation with Generative Diffusion Models**

Chao Huang, **Susan Liang**, Yapeng Tian, Anurag Kumar, Chenliang Xu

(In submission)

### **Revisit Audio-Visual Adversarial Robustness from Temporal and Modality Correlation Perspectives**

Zeliang Zhang\*, **Susan Liang**\*, Daiki Shimada\*, Chenliang Xu

(In submission)

### **AV-NeRF: Learning Neural Fields for Real-World Audio-Visual Scene Synthesis**

**Susan Liang**, Chao Huang, Yapeng Tian, Anurag Kumar, Chenliang Xu

Conference on Neural Information Processing Systems (NeurIPS), Dec. 2023

### **Neural Acoustic Context Field: Rendering Realistic Room Impulse Response With Neural Fields**

**Susan Liang**, Chao Huang, Yapeng Tian, Anurag Kumar, Chenliang Xu

International Conference on Computer Vision Workshops (ICCVW), Oct. 2023

### **UNICON+: ICTCAS-UCAS Submission to the AVA-ActiveSpeaker Task at ActivityNet Challenge 2022**

Yuanhang Zhang\*, **Susan Liang**\*, Shuang Yang, Xiao Liu, Zhongqin Wu, Shiguang Shan

IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2022

### **UniCon: Unified Context Network for Robust Active Speaker Detection**

Yuanhang Zhang\*, **Susan Liang**\*, Shuang Yang, Xiao Liu, Zhongqin Wu, Shiguang Shan, Xilin Chen

ACM International Conference on Multimedia (ACM MM), Oct. 2021 (Oral)

### **ICTCAS-UCAS-TAL Submission to the AVA-ActiveSpeaker Task at ActivityNet Challenge 2021**

Yuanhang Zhang\*, **Susan Liang**\*, Shuang Yang, Xiao Liu, Zhongqin Wu, Shiguang Shan

IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2021

\* indicates equal contribution.

## RESEARCH EXPERIENCE

---

### Semantic Correspondence

Sept. 2021 – Mar. 2022

*Vision and Learning Lab, **University of California, Merced***

*Advisors: Prof. Ming-Hsuan Yang and Dr. Taihong Xiao*

- Proposed a self-supervised deep learning approach for semantic correspondence.
  - Exploited contrastive learning and cycle consistency to learn discriminative and consistent features.
- 

### Vector Graphics Learning and Generation

Jun. 2021 – Aug. 2021

*Institute for AI Industry Research, **Tsinghua University***

*Advisors: Dr. Yizhi Wang and Dr. Hao Xu*

- Developed an encoder-decoder model to convert raster images to vector graphics.
  - Used a differentiable rasterization pipeline to enable supervision by raster images.
- 

### Active Speaker Detection

Oct. 2020 – Apr. 2021

*VIPL, Institute of Computing Technology, **Chinese Academy of Sciences***

*Advisors: Prof. Shiguang Shan and Dr. Shuang Yang*

- Developed an audio-visual multi-modal fusion scheme to detect when each visible speaker in the video is speaking.
  - Proposed an permutation-equivariant layer with the capability of processing all speakers in the scene simultaneously.
  - Exploited the skew-symmetry of inter-speaker relations which not only has reasonable interpretation but also reduces memory usage and FLOPS.
  - Conducted extensive experiments on multiple datasets (AVA-ActiveSpeaker, Columbia and RealVAD) with outstanding performance.
- 

### Face Deformation Field Generation and Lip Reading

Feb. 2020 – Sept. 2020

*VIPL, Institute of Computing Technology, **Chinese Academy of Sciences***

*Advisors: Prof. Shiguang Shan and Dr. Shuang Yang*

- Developed an encoder-decoder model to generate face deformation field (face-specific optical flow) which features the face motion.
  - Trained the deformation field in a self-supervised manner with no annotations.
  - Combined deformation field and gray-scale face images to recognize visual speech.
- 

## AWARDS

---

UCAS Outstanding Thesis Award

Jun. 2022

ActivityNet CVPR 2022 Workshop AVA Active Speaker Detection Challenge First Place

Jun. 2022

ActivityNet CVPR 2021 Workshop AVA Active Speaker Detection Challenge First Place

Jun. 2021

UCAS Overseas Graduate Studies Fellowship

Aug. 2021

UCAS Academy Fellowship

Sept. 2019 - Jun. 2021

## TECHNICAL SKILLS

---

**Languages:** Proficient in Python and C; Familiar with LaTeX

**Frameworks:** Proficient in PyTorch; Familiar with TensorFlow, PyTorch-Lightning, Diffusers, and nerfstudio

**Developer Tools:** Git, Docker, and Vim

## ACADEMIC SERVE

---

**Reviewer:** AAAI 2024, CVPR 2024