

梁苏叁

✉ sliang22@ur.rochester.edu · ☎ (+86) 138-6129-2824 · 🏠 <https://liangsusan-git.github.io/> ·

🎓 教育背景

- 罗彻斯特大学, 纽约, 美国 2022.08 – 至今
- 在读博士研究生 计算机视觉, 计算机专业, 预计 2027 年 5 月毕业
 - 导师: *Prof. Chenliang Xu*
- 中国科学院大学, 北京, 中国 2018.09 – 2022.06
- 学士 计算机科学与技术专业
 - 导师: 山世光研究员
 - 绩点: 3.90/4, 排名: 4/104

♡ 研究兴趣

多模态学习, 视觉语言大模型, 语音视觉重建与生成, 空间声音渲染

📄 论文

■ VIDCOMPOSITION: Can MLLMs Analyze Compositions in Compiled Videos?

Yunlong Tang, Junjia Guo, Hang Hua, **Susan Liang**, Mingqian Feng, Xinyang Li, Rui Mao, Chao Huang, Jing Bi, Zeliang Zhang, Pooyan Fazli, Chenliang Xu
Accepted by **CVPR'25**, 2025. (CCF-A)

■ Rethinking Audio-Visual Adversarial Vulnerability from Temporal and Modality Perspectives

Zeliang Zhang*, **Susan Liang***, Daiki Shimada, Chenliang Xu
Accepted by **ICLR'25**, 2025.

■ Language-Guided Joint Audio-Visual Editing Via One-Shot Adaptation

Susan Liang, Chao Huang, Yapeng Tian, Anurag Kumar, Chenliang Xu
Accepted by **ACCV'24**, 2024. (CCF-C)

■ High-Quality Visually-Guided Sound Separation from Diverse Categories

Chao Huang, **Susan Liang**, Yapeng Tian, Anurag Kumar, Chenliang Xu
Accepted by **ACCV'24**, 2024. 最佳论文提名. (CCF-C)

■ Learning to Transform Dynamically for Better Adversarial Transferability

Rongyi Zhu*, Zeliang Zhang*, **Susan Liang**, Zhuo Liu, Chenliang Xu
Accepted by **CVPR'24**, 2024. (CCF-A)

■ Random Smooth-based Certified Defense against Text Adversarial Attack

Zeliang Zhang, Wei Yao, **Susan Liang**, Chenliang Xu
Accepted by **EACL'24**, 2024. (CCF-C)

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

Susan Liang, Chao Huang, Yapeng Tian, Anurag Kumar, Chenliang Xu
Accepted by **NeurIPS'23**, 2023. (CCF-A)

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

Susan Liang, Chao Huang, Yapeng Tian, Anurag Kumar, Chenliang Xu
Accepted by **ICCVW'23**, 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
Accepted by **CVPRW'22**, 2022.

■ UniCon: Unified Context Network for Robust Active Speaker Detection

Yuanhang Zhang*, **Susan Liang***, Shuang Yang, Xiao Liu, Zhongqin Wu, Shiguang Shan, Xilin Chen
Accepted by **ACM MM'21**, 2021. Oral. (CCF-A)

■ 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
Accepted by **CVPRW'21**, 2021.

■ Scaling Concept With Text-Guided Diffusion Models

Chao Huang, **Susan Liang**, Yapeng Tian, Anurag Kumar, Chenliang Xu
arXiv preprint, 2024.

■ Video Understanding with Large Language Models: A Survey

Yunlong Tang*, Jing Bi*, Siting Xu*, Luchuan Song, **Susan Liang**, Teng Wang, Daoan Zhang, Jie An, Jingyang Lin, Rongyi Zhu, Ali Vosoughi, Chao Huang, Zeliang Zhang, Feng Zheng, Jianguo Zhang, Ping Luo, Jiebo Luo, Chenliang Xu
arXiv preprint, 2023.

在投论文

■ BinauralFlow: A Causal and Streamable Approach for High-Quality Binaural Speech Synthesis with Flow Matching Models

Susan Liang, Dejan Markovic, Israel D. Gebru, Steven Krenn, Todd Keebler, Jacob Sandakly, Frank Yu, Samuel Hassel, Chenliang Xu, Alexander Richard

Submitted to **ICML'25**, 2025. (CCF-A, In submission).

■ π -AVAS: Can Physics-Integrated Audio-Visual Modeling Boost Neural Acoustic Synthesis?

Susan Liang, Chao Huang, Yunlong Tang, Zeliang Zhang, Chenliang Xu

Submitted to **ICCV'25**, 2025. (CCF-A, In submission).

* 共同一作

👥 科研项目

📁 空间音频生成

2024.05 – 2024.08

Reality Labs Research, Meta

指导老师: Dr. Dejan Markovic, Dr. Alexander Richard

- 设计了一种新颖的 flow matching 模型，用于生成高质量的双耳音频。
- 提出了 causal 和流式 U-Net 架构，以支持近实时推理。

📁 语义匹配

2021.09 – 2022.03

加州大学默塞德分校视觉与学习实验室

指导老师: Prof. Ming-Hsuan Yang, Dr. Taihong Xiao

- 提出了一种用于语义匹配的自监督深度学习方法。
- 利用对比学习和循环一致性学习判别性和一致性的特征。

📁 矢量图学习与生成

2021.06 – 2021.08

清华大学智能产业研究院

指导老师: Dr. Yizhi Wang, Dr. Hao Xu

- 开发了一种编码器-解码器模型，将光栅图像转换为矢量图。
- 使用可微分光栅化框架，实现对光栅图像的监督训练。

📁 说话人检测

2020.10 – 2021.04

中国科学院计算技术研究所 VIPL 实验室

指导老师: Prof. Shiguang Shan, Dr. Shuang Yang

- 设计了一种音视频多模态融合方案，检测视频中可见说话人何时在说话。
- 提出了一个置换等变层，能够同时处理场景中的所有说话人。
- 利用说话人之间关系的反对称性，既有合理的解释，又减少了内存使用和计算量。
- 在多个数据集 (AVA-ActiveSpeaker, Columbia, RealVAD) 上进行了广泛实验，取得了优异性能。

中国科学院计算技术研究所 VIPL 实验室

指导老师: Prof. Shiguang Shan, Dr. Shuang Yang

- 开发了一种编码器-解码器模型，用于生成人脸变形场 (面部特征光流)，以捕捉面部运动。
- 采用自监督方式训练变形场，无需人工标注。
- 结合变形场与灰度人脸图像进行视觉语音识别。

🏆 获奖情况

北京市本科优秀毕业论文	2022 年 6 月
中国科学院大学 本科优秀毕业论文	2022 年 6 月
ActivityNet CVPR 2022 AVA 说话人检测挑战赛第一名	2022 年 6 月
ActivityNet CVPR 2021 AVA 说话人检测挑战赛第一名	2021 年 6 月
中国科学院大学 海外留学奖学金	2021 年 8 月
中国科学院大学 本科生学业奖学金	2020 年 10 月