

黎吉国

15652370966 · jgli@fudan.edu.cn/jiguo_li@qq.com · 工学博士 · [GitHub](#) | [Google Scholar](#)

个人简介

2016 年于华中科技大学取得学士学位，2022 年于中国科学院计算技术研究所取得博士学位，同年加入复旦大学信息科学与工程学院，担任博士后研究员，并获得上海市“超级博士后”激励计划资助。研究兴趣包括多模态生成，基于生成的编码方法等。

教育经历

- 复旦大学, 信息科学与工程学院, 跨模态生成, 博士后研究员 2022.9 - 至今
- 合作导师为[郑立荣教授](#)(国家特聘专家).
 - 研究方向为跨模态生成。研究项目获得上海市“超级博士后”激励计划资助。
- 北京大学, 视频与视觉技术国家工程研究中心, 多媒体技术方向, 联合培养 2016.9 - 2022.6
- 指导老师为北京大学[马思伟教授](#)(国家杰青).
 - 研究方向为图像处理与生成, 跨媒体分析, 虚拟现实等.
- 中国科学院计算技术研究所, 计算机应用技术, 多媒体技术方向, 工学博士 2016.9 - 2022.6
- 师从中国工程院院士[高文教授](#)(北京大学“博雅特聘教授”, 鹏城实验室主任).
 - 研究方向为图像处理与生成, 跨媒体分析, 虚拟现实等.
- 华中科技大学, 自动化, 模式识别与智能系统方向, 工学学士 2012.9 - 2016.6
- 研究方向为全景视频拼接, 导师为[曹治国教授](#)(原自动化学院院长).

工作经历

- 京东集团 | JD.Inc, 搜索推荐实习生 2022.03-2022.06
- 对商品销量建模, 预测商品在未来的销量, 为集团大促、仓储备货提供支撑, 申请发明专利一项。
- 字节跳动 | Bytedance, 多媒体算法实习生 2018.06-2020.09
- 在[许继征博士](#)指导下发表/接收期刊论文两篇, 会议论文两篇, 申请发明专利三项。
 - 在[施澍博士](#)指导下设计开发用于自动测量云游戏延时的算法和系统, 申请发明专利一项。
- 维境视讯 | VSCENE, 图像算法实习生 2016.09-2018.06
- 兼职图像算法实习生, 在[王悦博士](#)的指导下, 负责全景直播系统中的视频实时拼接算法的研究和实现, 相关技术和实现应用于维境视讯的全景直播系统中。
 - 相关技术发表期刊论文一篇, 申请发明专利一项, 相关系统获得 2017 年 AVS 产业创新奖。

项目经历

主要围绕跨模态分析与生成, 声纹识别与对抗攻击, 虚拟现实等三个方面, 加 ☆ 表示重点项目。

1 跨模态分析与生成

- 1.1 跨模态压缩: 一种面向语义保真的压缩框架 | Pytorch | Paper | Patent | ☆ 2020.01-2020.09
- 首次提出跨模态压缩: 通过设计一个人可理解的, 语义紧凑的压缩域, 将高语义冗余的视觉数据压缩至低语义冗余的压缩域, 同时保持语义的保真和可恢复性. 实现了跨模态压缩的一个实例: 图像-文本-图像的跨模态压缩框架. 图像-文本-图像的跨模态压缩框架在 IS, FID 等语义保真指标下的 RD 性能超过了 JPEG 和 JPEG2000.
 - 相关成果整理为论文, 已被多媒体领域顶级国际会议 ACM MM 2021 (CCF A) 接收. 相关技术已申请[发明专利](#).
- 1.2 基于语音描述的图像生成 | Pytorch | Paper | Patent | ☆ 2018.09-2019.06
- 探索直接从语音描述中生成图像, 而不借助中间文本表示的系统和方法. 基于迁移学习训练音频特征提取器, 利用 GAN 从语音特征中生成语义一致的图像. 在合成的语音数据上, 与基于文本的图像生成, 在 FID, IS 等指标上接近.
 - 相关成果整理为期刊[论文](#), 发表于信号处理领域顶级期刊 IEEE J-STSP (IF 6.865). 相关技术已申请[发明专利](#).

1.3 跨模态生成式图像压缩 | Pytorch | Paper | ☆ 2019.06-至今

- 研究基于跨模态生成模型 (GAN, Transformer, Diffusion model) 的图像编码方法, 设计新的跨模态图像编码框架.
- 相关成果投稿/发表在 IEEE TCSVT(CCF B), IJCV(CCF A), IEEE TMM(CCF B) 上.

2 声纹识别与对抗攻击

2.1 空时域融合的声纹识别方法 | Pytorch | Patent 2019.06-2019.12

- 传统的声纹识别先将声音数据变换为频谱, 基于频谱进行特征提取和声纹识别. 本项目尝试将声音数据的频谱处理和时域的特征提取相结合, 通过融合时域和频域的信息, 得到了比单一域特征表示更好的声纹识别结果.
- 相关技术与系统已申请发明专利.

2.2 面向声纹识别的对抗攻击方法 | Pytorch | Paper | Patent | ☆ 2019.06-2019.12

- 本文首次利用人耳对声音的感知特性: 人耳会重点关注有实际含义的声音而忽略无意义的噪声, 设计一种可以更好保持语音中音素信息的对抗攻击方法, 明显提升了针对声纹识别对抗攻击的效率.
- 相关成果已整理为论文, 被语音领域顶级国际会议 IEEE ICASSP2020 (CCF B) 接收. 相关技术已申请发明专利.

2.3 声纹识别对抗扰动的频谱分析 | Pytorch | Paper | Patent | ☆ 2019.06-2019.12

- 首次通过分析不同频带限制下的对抗扰动对声纹识别系统的攻击效率, 发现高频段的对抗扰动可以实现更高的攻击效率.
- 相关成果整理成论文, 已被 ACM 多媒体及应用汇刊 ACM TOMM (CCF B) 接收.

2.4 通用声纹识别对抗扰动 | Pytorch | Paper | Patent 2019.06-2019.12

- 首次利用对抗训练的方式学习面向声纹识别的通用对抗扰动, 提出一种面向通用对抗扰动的优化方法, 学到的通用对抗样本的对抗攻击效果明显高于随机噪声.
- 相关成果整理成论文, 被国际多媒体旗舰会议 IEEE ICME2020 (CCF B) 接收.

3 虚拟现实与云游戏

3.1 4K 全景视频直播系统 | Cuda | Paper | Patent | ☆ 2016.09-2018.06

- 863 项目: 真三维视频关键技术研究及先导验证, 演示系统之一, 负责其中的视频实时拼接部分. 利用视频采集卡实时获取 6-8 路视频流, 设计拼接算法将多路视频流逐帧实时拼接后显示或发送至推流服务器进行分发. 利用 GPU 进行像素级并行投影 (Cuda 实现), 利用 GPU 流水线进行拼接-亮度补偿-融合的加速, 从而实现多路高清视频在带 GPU 的移动工作站 (笔记本) 的实时处理.
- 部分技术整理成论文, 获得第十七届中国虚拟现实大会 (ChinaVR 2017) 最佳论文奖, 发表于 CCF A 类中文期刊《中国科学: 信息科学》. 相关技术已申请发明专利, 并应用于维境视讯有限公司全景视频直播系统中.

3.2 云游戏时延自动测量系统 | OpenCV | Patent 2020.01-2020.09

- 云游戏的延时是影响用户体验的最重要因素之一, 但是当前延时的测量依赖大量的人工, 费时费力. 本项目首次使用自动化的方式测量云游戏的延时. 通过后台实时收集鼠键控制输入和屏幕的变化, 分析控制输入和屏幕内容变化之间的对应关系, 筛选出响应显著的控制输入, 从而实现自动化的延时测量.
- 相关技术与系统已申请发明专利.

获奖荣誉

上海市“超级博士后”	2022
AVS 产业创新奖	2017
第十七届中国虚拟现实大会最佳论文奖	2017
华中科技大学 国家奖学金 & 国家励志奖学金	2015/2016
第九届 & 第十届全国大学生智能汽车大赛华南赛区一等奖	2014/2015
华中科技大学 校级三好学生 & 启明学院特优生 & 优秀学生干部 & 优秀党员 & 优秀毕业生	2014/2016
华中科技大学自动化学院 C 语言课程设计大赛一等奖	2014

论文发表

共发表/接收期刊论文 4 篇 (3 英文, 1 中文), 共发表/接收会议论文 4 篇 (2 CCF A, 2 CCF B).

1. **Li Jiguo**, Jia Chuanmin, Zhang Xinfeng, Ma Siwei, Gao Wen. "Cross modal compression: Towards human-comprehensible semantic compression." In Proceedings of the 29th ACM International Conference on Multimedia, pp. 4230-4238. 2021.
 - *Published, ACM MM 2021, CCF A, EI, First Author*
2. **Li Jiguo**, Zhang Xinfeng, Jia Chuanmin, Xu Jizheng, Zhang Li, Wang Yue, Ma Siwei, Gao Wen. Direct speech-to-image translation[J]. IEEE Journal of Selected Topics in Signal Processing(J-STSP), 2020, 14(3): 517-529.
 - *Published, IEEE J-STSP, 2020 IF 6.856, SCI, First Author, Homepage*
3. **Li Jiguo**, Zhang Xinfeng, Xu Jizheng, Ma Siwei, Gao Wen. Learning to Fool the Speaker Recognition[J]. ACM Transactions on Multimedia Computing, Communications, and Applications(TOMM), 2021, 17(3s): 1-21.
 - *Published, ACM TOMM, CCF B, IF 3.275, SCI, First Author, Homepage*
4. **Li Jiguo**, Zhang Xinfeng, Jia Chuanmin, Xu Jizheng, Zhang Li, Wang Yue, Ma Siwei, Gao Wen. Universal adversarial perturbations generative network for speaker recognition[C]. 2020 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2020: 1-6.
 - *Published, IEEE ICME 2020, CCF B, EI, First Author, Homepage*
5. **Li Jiguo**, Zhang Xinfeng, Xu Jizheng, Zhang Li, Wang Yue, Ma Siwei, Gao Wen. Learning to fool the speaker recognition[C]. 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 2937-2941.
 - *Published, IEEE ICASSP 2020, CCF B, EI, First Author, Homepage*
6. **Jiguo Li**, Liu Xiaobin, Tianzi Zhang, Lirong Zheng. Time-frequency Network for Robust Speaker Recognition[J]. IEEE Multimedia.
 - *Under Review, IEEE MM 2023, IF 3.491, SCI*
7. 黎吉国, 王悦, 张新峰, 马思伟, 高文. 一种鱼眼视频全景拼接中的亮度补偿算法 [J]. 中国科学: 信息科学, 2018.
 - *Published, CCF A, First Author*
8. Chang Jianhui, Zhang Jian, Xu Youmin, **Li Jiguo**, Ma Siwei, Gao Wen. "Consistency-Contrast Learning for Conceptual Coding." In Proceedings of the 30th ACM International Conference on Multimedia, pp. 2681-2690. 2022.
 - *Published, ACM MM 2022, CCF A, EI*
9. Zhang Pingping, Wang Shiqi, Wang Meng, **Li Jiguo**, Wang Xu, Kwong Sam. Rethinking Semantic Image Compression: Scalable Representation with Cross-modality Transfer[J]. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), 2023.
 - *Accepted, IEEE TCSVT 2023, CCF B, IF 5.859, SCI*
10. Chang Jianhui, Zhang Jian, **Li Jiguo**, Wang Shiqi, Mao Qi, Jia Chuanmin, Ma Siwei and Gao Wen. Semantic-aware Visual Decomposition for Image Coding[J]. International Journal of Computer Vision (IJCV).
 - *Minor Revision, IJCV 2023, CCF A, IF 13.369, SCI*
11. Gao Junlong, **Li Jiguo**, Jia Chuanmin, Ma Siwei and Gao Wen. Cross Modal Compression with Variable Rate Prompt[J]. IEEE Transactions on Multimedia(TMM).
 - *Under Review, TMM 2023, CCF B, IF 8.182, SCI*

专利申请

1. 黎吉国, 许继征, 马思伟, 王悦. 基于语音的图像生成方法、装置和电子设备. 公开号: 201910042614.9
2. 马思伟, 黎吉国, 贾川民, 赵政辉, 王苕社. 数据压缩的方法、装置、设备及存储介质. 公开号: 202010604773.6
3. 黎吉国, 许继征, 张莉, 王悦, 马思伟. 音频对抗扰动的测试方法、设备及存储介质. 公开号: 202011024815.5
4. 王悦, 姜晓龙, 魏建超, 黎吉国, 孟德国, 刘鑫. 一种全景视频处理制作系统及方法. 公开号: 201711100069.1
5. 施澍, 黎吉国, 张莉, 马思伟, 王悦. 一种系统延时确定方法、装置、设备和存储介质. 公开号: 202110172567.7
6. 黎吉国, 许继征, 张莉, 王悦, 马思伟. 声音识别方法、装置以及存储指令的计算机可读存储介质. 公开号: 202010694750.9

其它

- 技术能力: Python, C/C++, Pytorch, OpenCV, Linux, Cuda, Git, L^AT_EX, CET 6
- 学术服务: 长期担任国际期刊 IEEE TCSVT, IEEE MM 审稿人, 受邀担任国际会议 ICME2020-2023& ICASSP2021-2023& CVPR2021-2023 审稿人.