

胡世宇

Research Fellow, 新加坡南洋理工大学

📅 1995.10.29 ✉️ shiyu.hu@ntu.edu.sg 🔗 hushiyu1995 🌐 huuuuusy
🏠 <https://huuuuusy.github.io/> 🌐 <http://viig.aitestunion.com/>

个人概述

- 教育背景优秀** ■ 本硕博均毕业于顶尖大学/科研机构，本硕博学位论文答辩成绩均为优秀。
- 荣获多项奖励及荣誉，博士阶段荣获国家奖学金（获奖率：1%）和北京市优秀毕业生（获奖率：5%）。
- 科研基础扎实** ■ 博士阶段共计发表论文14篇。其中第一作者/通讯作者论文5篇，分别为3篇国际顶级期刊IEEE TPAMI和IJCV、1篇国际顶级会议NeurIPS和1篇《中国图象图形学报》。受Springer出版社邀请，将于2024年12月完成一本专著。
- 负责搭建并维护的科研平台获得全球130+国家和地区38.2万次访问（截止2024年6月）。
- 交流合作广泛** ■ 担任NeurIPS, CVPR, ECCV, AAAI, ACMMM, SCIENCE CHINA Information Sciences等顶级会议和期刊的审稿人，将在第31届ICIP、第27届ICPR和第17届ACCV会议上举办讲习班。
- 自2022年9月起，发起并定期组织基于计算机视觉的跨领域研讨会（40+次，10+学校，20+人），讨论主题涉及计算机视觉、认知神经科学、人智交互等。
- 协助指导本硕博生近10人开展科研工作，和学生一起成立视觉智能兴趣小组 (VIIG) 并在多个方向开展研究。

工作经历

- 2024.08 - 至今 ■ **Research Fellow**，南洋理工大学（NTU），物理与数学科学学院
- 方向：AI4Science，计算机视觉
 - **PI**: Assoc Prof. Kanghao Cheong (IEEE Senior Member)
- 2018.03 - 2018.11 ■ **Research Assistant**，香港大学（HKU），工程学院，计算机科学系
- 方向：高性能计算，异构计算
 - **PI**: Prof. Choli Wang
- 2016.08 - 2016.09 ■ **Research Intern**，中国科学院，电子学研究所（CASIE）

教育经历

- 2019.09 - 2024.01 ■ **博士**，中国科学院，自动化研究所（CASIA）
- 专业：计算机应用技术
 - 导师：黄凯奇研究员（IAPR会士，万人领军）
 - 合作导师：赵鑫教授（北京市杰青）
 - 论文题目：单目标跟踪中的智能评估技术研究
 - 答辩委员会：焦建彬教授，彭宇新教授（国家杰青），赵耀教授（IEEE会士，IET会士，国家杰青），王蕴红教授（IEEE会士，IAPR会士，CCF会士），唐明研究员
 - 论文答辩成绩：全优
- 2017.09 - 2019.07 ■ **硕士**，香港大学（HKU），工程学院，计算机科学系
- 专业：计算机科学
 - 导师：Prof. Choli Wang
 - 论文题目：NightRunner: Deep Learning for Autonomous Driving Cars after Dark
 - 论文答辩成绩：A+

教育经历 (continued)

2013.09 - 2017.07 ■ 学士, 北京理工大学 (BIT), 信息与电子学院, 信息工程菁英班

- 专业: 电子信息工程
- 论文题目: 基于深度神经网络的文本情感分类研究
- 论文答辩成绩: 全优

2015.07 - 2015.08 ■ 暑期学期, 加州大学伯克利分校 (UCB)

- 专业: 新媒体
- 课程成绩: A

研究基础 & 研究兴趣

视觉目标跟踪 ■ 通用场景、专用场景 (例如, 无人机) 下的单目标跟踪算法研究。

■ 面向单目标跟踪算法的鲁棒性、泛化性、安全性等研究。

视觉语言跟踪 ■ 基于长视频序列的多模态跟踪、视频理解和视觉推理研究。

■ 结合大语言模型 (LLMs) 和基础视觉模型 (LVMs) 的长视频理解研究。

■ 以视觉语言跟踪为代理任务的长视频序列下人智交互模式探究。

视觉基准建模 ■ 融合人类知识结构的大规模单模态、多模态数据集构建策略研究。

■ 面向视觉鲁棒性、泛化性、安全性的评估机制研究和评价指标设计。

智能评估技术 ■ 人机通用的视觉智能评估框架设计。

■ 以人为基准度量算法在感知、认知、推理等能力维度上的表现, 并深入分析算法和人类被试的瓶颈, 为类人建模、人机协作、人机融合等研究提供引导。

AI4Science ■ 认知神经科学: 基于类人建模的视觉任务设计、环境构建和人机能力分析。

■ 医学: 结合人工智能的医学图像处理技术研究 (例如, 细胞分割与跟踪, 低温电子显微镜图像去噪)。

■ 心理学: 面向焦虑、抑郁、强迫等心理维度的游戏化评测系统开发及智能心理评估技术研究。结合大语言模型 (LLMs) 和基础视觉模型 (LVMs) 的心理场景下视觉理解任务研究。

■ 教育学: 面向智能教育场景的人机交互技术研究, 包括多学科交叉视角下的智能教育框架设计、人机交互技术研究以及定性和定量分析。

研究经历

主要研究 (负责: 思路制定 & 方法设计 & 实验分析 & 论文撰写 & 平台开发)

2018.03 - 2018.11 ■ Darknet-Cross: 适配于异构计算的轻量化深度学习框架

✎ 高性能计算 | 异构计算 | 深度学习框架

✎ Darknet-Cross是一个轻量化的深度学习框架, 主要基于开源的深度学习算法库Darknet和yolov2_light进行二次开发, 并通过交叉编译成功移植到手机端, 实现利用移动GPU开展高效的算法推理。

✎ 该框架支持多种平台 (例如, Android和Ubuntu) 和多种型号GPU (例如, Nvidia GTX1070和Adreno 630) 下的算法加速处理。

✓ 相关工作是我在香港大学硕士毕业论文的一部分 (论文答辩成绩: A+)。

研究经历 (continued)

2019.11 - 2022.02 ■ VideoCube: 大规模多维度全局实例跟踪任务智能评估平台

✎ 视觉目标跟踪 | 大规模基准构建 | 智能评估技术

✎ 从类人建模出发对单目标跟踪进行拓展，并提出全局实例跟踪（global instance tracking, GIT）任务，使其应用范围拓展至对抗场景中。

✎ 提出基于影视学理论的视频叙事内容解耦框架，并基于此构建了一个包含746万视频帧的大规模多维度的全局实例跟踪任务智能评估平台VideoCube，该平台是目前规模最大的单目标跟踪基准

✎ 从人机对抗出发，首次在单目标跟踪任务中引入人类被试，旨在以人类为基准度量算法的视觉智能。

✓ 相关工作于2022年2月被CCF-A期刊TPAMI接收 [A1]，截止2024年6月，平台获得全球130+国家和地区38.2万次访问，1100余次下载，400余次算法测试。

2023.03 - 2023.09 ■ MGIT: 基于层级化语义框架的多模态全局实例跟踪任务基准

✎ 视觉语言跟踪 | 长视频理解与推理 | 层级化语义信息

✎ 在GIT任务和VideoCube基准的基础上，构建一个具有复杂视频叙事关系的多模态基准MGIT，旨在充分体现长视频所刻画的复杂时空关系和因果关系。

✎ 结合人类认知的层级化结构，创新性地提出一种多粒度的语义信息标注策略，旨在为算法提供高质量的语义信息，并通过实验验证该策略的有效性。

✎ 提出一套适配于多模态单目标跟踪任务的评测机制，并基于该评测机制对不同结构的算法进行详细的实验分析，从而深入挖掘算法的性能瓶颈。

✓ 相关工作于2023年9月被CCF-A会议NeurIPS接收 [A2]。

2021.07 - 2023.10 ■ SOTVerse: 用户自定义的单目标跟踪任务空间

✎ 视觉目标跟踪 | 动态开放环境建模 | 视觉评估技术

✎ 从高效利用研究资源开展评估的视角出发，提出3E范式作为任务分析框架，并从实验环境、评估系统和任务对象三个维度对单目标跟踪进行梳理。

✎ 对代表性单目标跟踪数据集进行整合，将原本静态封闭的数据空间融合为一个包含1256万帧的动态开放数据空间SOTVerse，并提出子空间构建算法，旨在帮助研究者依据研究目标快速定位高挑战性序列形成专用的实验环境。

✎ 提出全新的评测系统，使其能够适配于不同的评测需求和任务对象，并对23个代表性算法开展细粒度分析，有效挖掘出不同架构算法的性能瓶颈。

✓ 相关工作于2023年9月被CCF-A期刊IJCV接收 [A3]。

2022.05 - 2023.10 ■ BioDrone: 基于仿生无人机的单目标跟踪基准和鲁棒视觉研究平台

✎ 视觉目标跟踪 | 无人机视角跟踪 | 视觉鲁棒性研究

✎ 构建了首个基于仿生扑翼无人机的单目标跟踪基准数据集BioDrone，扑翼无人机特有的空气动力学结构会导致镜头出现剧烈抖动，为鲁棒视觉研究提供一个高质量的环境。

✎ 在BioDrone上复现并测试了20种代表性单目标跟踪算法，并针对算法的鲁棒性瓶颈开展细致分析。

✎ 设计了一个名为UAV-KT的基线算法，通过实验分析验证了该方法可以有效提取视觉特征，并在无人机场景的挑战因素下依旧保持鲁棒性。

✓ 相关工作于2022年5月-10月作为竞赛数据支撑举办第三届高速低功耗视觉理解挑战赛（在第五届中国模式识别与计算机视觉大会上举办），并于2023年10月被CCF-A期刊IJCV接收 [A4]。

2022.04 - 至今 ■ 基于视觉图灵的单目标跟踪智能评估技术

✎ 视觉目标跟踪 | 评估技术 | 视觉图灵

✎ 结合图灵测试思想，提出“视觉图灵”评估范式——以人类的视觉能力为基准，全面度量算法的视觉智能。

✎ 构建了一个公平对比人机动态视觉能力的实验环境，该环境可以涵盖任务对象在执行动态视觉任务时需要调用的感知和认知能力。

✎ 挑选合适的任务对象开展人机动态视觉能力测试，包括了20个代表性算法和15名受试者，并基于严格的流程开展实验。

✎ 设计出一套人机通用的多粒度动态视觉任务评估框架，基于中心点距离从三个粒度（帧级、序列级和组级）开展评估和分析。

✓ 相关工作于2021年和2023年发表中文综述2篇 [A5], [A6]，实验内容及主要结论正在完善 [P1]，准备投稿至Cell Patterns期刊 [O1]。此外，受Springer出版社邀请，将于2024年12月完成一本专著 [O2]。

2024.03 - 至今 ■ FIOVA：人类群体和大型视觉语言模型的视频理解能力对比基准

✎ 大型视觉语言模型 | 评估技术 | 视觉图灵

✎ 提出FIOVA基准，用于评估LVLM与人类在复杂视频描述任务中的理解差异。FIOVA包含3,002个长视频序列，五位标注员为每个视频提供注释，使得注释长度比现有基准长4至15倍，首次全面建立了人类视频理解的基线。

✎ 基于FIOVA基准，对六种最先进的LVLM（VideoLLaMA2、LLaVA-NEXT-Video、Video-LLaVA、VideoChat2、Tarsier、ShareGPT4Video）进行了深入评估，有效挖掘LVLM与人类的细粒度差异性。

✎ 实验结果强调了使用单一标注员作为评价基准的局限性，提出了新的评价视角，以捕捉语义理解、描述深度及一致性的差异，为未来实现人类水平的视频理解提供指导。

✓ 相关工作于2024年10月投稿至CAAI-A会议至ICLR [R1]，处于审稿中。

■ 基于大语言模型的多样化和个性化虚拟学生智能体

✎ AI4Education | 大语言模型 | AI智能体

✎ 提出了一种AI4Education框架SOE（场景 - 对象 - 评估），系统地构建了基于大语言模型的虚拟学生智能体，通过个性化的师生交互数据集进行微调，并开展多维度评估实验。

✎ 开发了一套用于生成虚拟学生智能体的理论框架，并将人类主观评价指标整合至GPT-4的评估中，结果显示人类评估者与GPT-4在虚拟学生智能体的真实性判断上具有高度一致性。

✎ 验证了大语言模型在教育情境中生成具有人类特征的个性化虚拟学生代理的能力，为AI4Education的应用奠定了基础。

✓ 相关工作于2024年10月投稿至CAAI-A会议至ICLR [R2]，处于审稿中。

独立开发（负责：平台维护 & 升级）

2020.07 - 2024.01 ■ GOT-10k: 大规模广覆盖单目标视觉跟踪评估平台

✎ 视觉目标跟踪 | 智能评估技术 | 平台开发维护

✎ GOT-10k是用来度量算法在未知目标类别和运动模式上泛化性的评估平台。该平台提供了1万段视频序列，包含563类目标类别和87种运动模式，总计150万个矩形框标注。

✎ 和已有的单目标跟踪基准相比，GOT-10k在目标类别的丰富程度上要远超过现有工作，其提出的训练-测试类别不重合的评测思路也成为单目标跟踪评估评测领域的研究规范。

✓ 相关工作于2019年11月被CCF-A期刊TPAMI接收；截止2024年2月，本平台获得全球150+国家和地区324万次访问，6千余次下载，1.7万余次算法测试。和维护初期相比，有18倍的用户增长。

研究经历 (continued)

合作研究（负责：思路讨论 & 实验分析 & 论文修改）

2019.05 - 2019.10 ■ 无色卡场景下的肤色检测系统

✎ 颜色恒常性 | 肤色检测算法 | 环境光照估计

✎ 在18种环境光照条件和4部手机的参数组合下，对110位用户进行肤色数据采集，并以德国CK公司的MPA9肤色检测仪的测试结果作为用户肤色的真值，构建包含7920张图片的肤色数据集。

✎ 基于椭圆皮肤模型得到图像中的关键皮肤区域，并利用开源颜色恒常性模型FC⁴对环境光照进行恢复，最后利用SVR回归计算用户的肤色检测结果。

✓ 相关工作已成功部署在华为手机官方应用《镜子》的AI测肤功能中。

2020.11 - 2021.03 ■ 基于深度学习方法的细胞跟踪项目

✎ 医学图像处理 | AI4Science | 细胞分割与跟踪

✎ 参考基于检测的跟踪范式，将每一帧的细胞分割结果进行整合，并利用孪生神经网络进行帧间细胞位置匹配，从而完成跟踪过程。

✓ 相关工作于2021年3月提交至ISBI细胞跟踪挑战赛，在2023年10月，依旧保持Fluo-C2FL-MS-C+数据集第二名和Fluo-C2FL-Huh7数据集第三名。

2023.08 - 2023.10 ■ 面向单粒子低温电磁场图像的去噪与恢复算法

✎ 医学图像处理 | AI4Science | 扩散模型

✎ 采样自低温电子显微镜的信息通过图像重建可以展示接近原子尺度的生物分子信息，但重建过程需要克服低信噪比和复杂的噪声结构。本工作提出一种带有后处理模块的扩散模型，以有效地去噪和恢复单粒子低温电镜图像。

✎ 在模拟数据集和真实数据集上的实验结果验证了方法的有效性。

✓ 相关工作于2023年12月被CCF-B会议ICASSP接收 [A7]。

2023.07 - 2024.09 ■ 基于类人记忆机制建模的鲁棒视觉语言跟踪算法

✎ 视觉语言跟踪 | 类人记忆建模 | 自适应提示

✎ 针对视觉语言跟踪算法在长视频中鲁棒性不足的情况，从类人记忆建模角度出发，设计了一个鲁棒的多模态跟踪算法MemVLT。

✎ MemVLT受认知领域的互补学习理论的启发，包含记忆存储和交互模块，旨在模拟人类记忆在海马体和新皮层之间的复杂调制过程。

✎ 在多个代表性视觉语言跟踪基准上的实验验证了MemVLT的有效性。

✓ 相关工作于2024年9月被CCF-A会议NeurIPS接收 [A8]。

2023.12 - 2024.09 ■ 基于视觉搜索机制建模的类人视觉目标跟踪算法

✎ 视觉目标跟踪 | 视觉搜索机制 | 视觉图灵测试

✎ 针对视觉目标跟踪算法难以应对时空不连续场景的难点，从类人视觉搜索机制建模角度出发，设计了一个鲁棒的视觉目标跟踪算法CPDTrack。CPDTrack受人类视觉系统通路的启发，为局部视野和全局视野设计了一种高效编码方式。

✎ 为验证CPDTrack的效果，构建了一个由包含时空不连续挑战性的序列构成的基准STDChallenge；并设计了一个视觉图灵测试，度量并量化人类被试的视觉搜索能力。

✎ 在STDChallenge基准和其他代表性基准验证了CPDTrack的有效性，并通过视觉图灵测试中人机的错误一致性分析，验证CPDTrack的决策与人类被试具有较高的相似性。

✓ 相关工作于2024年9月被CCF-A会议NeurIPS接收 [A9]。

研究经历 (continued)

2022.10 - 至今 ■ 面向真实应用场景的无约束空中手写技术

✎ 空中手写技术 | 视觉基准构建 | 人机交互

✎ 基于空中手写任务构建了一个大规模高质量的空中汉字手写数据集AWCV-100k, 旨在为人机交互研究提供一种更加自然通用的实验环境。

✎ AWCV-100k包含880万视频帧, 涉及多种环境、光照组合, 全面覆盖中文一级字符库GB2312-80的3755个字符, 是目前规模最大、覆盖范围最全面的中文空中手写视频数据集。

✎ 提出了基于视频的空中手写字符识别算法VCRec, 该基线算法可以从稀疏的视觉线索中提取指尖特征, 并使用一个时空序列模块进行分析。

✎ 在AWCV-100k数据集上复现并测试了代表性的空中手写算法和基线算法VCRec, 实验结果验证了VCRec的鲁棒性和有效性。

✓ 相关工作于2024年4月被CCF-B期刊TCSVT接收 [A10]; 后续工作在AWCV-100k的基础上引入了深度信息和手部关节点信息, 形成一个多模态基准MMAW-UCAS2024, 并提出相应的多模态基线算法MMRec, 将2024年11月投稿至CCF-A会议CVPR [O3]。

2023.01 - 至今 ■ 面向相似物体干扰挑战的单目标跟踪任务研究

✎ 视觉目标跟踪 | 相似物体干扰 | 高效数据利用

✎ 基于主流跟踪算法的运行原理及失败案例分析, 重新定义了算法视角下的相似物体干扰挑战, 并重点分析人机在面对相似物体干扰时的认知差异。

✎ 提出一种数据挖掘算法, 在无人工干预的情况下从代表性单目标跟踪数据集中挖掘出存在相似物体干扰的序列, 形成TrackingSOI数据集。

✎ 提出一种有效应对相似物体干扰的算法TransKT, 其可以将目标与表观信息相似的候选物进行区分, 从而获得鲁棒的视觉目标跟踪能力。

✓ 相关工作的初始版于2023年11月被EI会议CSAI接收为oral论文 [A11], 完整版于2024年1月投稿至CCF-B期刊TCSVT, 处于审稿中 [R3]。

2023.02 - 至今 ■ 基于电子沙盘的智能心理评测系统

✎ 心理评测系统 | 游戏化评测 | AI4Science

✎ 针对传统心理评测窥视感强、被试缺乏内在动机等缺陷, 构建了基于电子沙盘的智能心理评测系统, 可以对被试开展焦虑、抑郁、强迫等多个心理维度的评测和分析。

✎ 和传统的量表评测和游戏化评测相比, 基于电子沙盘的智能心理评测系统结合了人工智能技术, 可以挖掘游戏内动态过程数据中耦合的心理测量证据, 能够进一步提升测量真实性和可解释性。

✎ 在教育场景和公共安全场景招募被试, 并开展信效度验证和个案分析, 实验结果证实了电子沙盘的有效性。

✓ 相关工作所采用的沙盘主题识别模型于2023年9月被CCF-C会议PRCV接收 [A12]; 作为一个三学科交叉的系统性工作(心理学、游戏设计、人工智能), 其技术路线发展脉络和技术框架将分别以综述和技术论文的形式投稿至中国顶级心理学期刊 [A13], [R4]。后续工作构建了一个基于绘画投射的交互式智能心理评测系统VS-LLM, 于2024年6月被CCF-C会议PRCV接收 [A14]。

2024.01 - 至今 ■ 基于大语言模型的面向视觉语言跟踪任务的多样化文本生成方法

✎ 视觉语言跟踪 | 大语言模型 | 评估评测技术

✎ 基于LLM设计了一个名为DTLLM-VLT的框架, 该方法可以自动生成多粒度文本, 从而高效增强视觉语言跟踪环境的多样性。

✎ 将DTLLM-VLT框架部署在具有代表性的视觉语言跟踪基准中, 并对VLT算法进行了详细评估。实验结果表明, DTLLM-VLT框架与数据集相结合可以有效提升对视觉数据集理解能力。

✓ 相关工作于2024年4月被CCF-A会议CVPR附属的第三届视觉数据理解研讨会接收为oral论文, 并获得最佳论文提名奖 [A15]; 后续工作在DTLLM-VLT的基础上构建DTVLT视觉语言跟踪基准, 并开展细粒度的评估评测分析, 于2024年6月投稿至CCF-A会议NeurIPS [R5], 处于审稿中。

从协同走向融合：智能教育中人机交互的困境及对策研究

✎ 智能教育技术 | 人机交互 | AI4Science

✎ 采用跨学科的视角，结合教育学、认知心理学、计算机科学等领域的知识与方法，构建智能教育下人机交互形态演变的理论框架。

✎ 基于所构建的理论框架，深入分析从协同到融合阶段人机交互在教育实践中的转变，探讨智能教育下人机交互新变革中的核心问题。

✎ 基于所挖掘的核心问题，探索如何用理论指导和技术改进来改善智能教育中人机交互的效果，最终达到有效的人机融合。

✓ 相关项目获得上海智能教育研究院与华东师范大学2023年智能教育博士生科研基金资助，处于研究中 [O4]。

2024.05 - 至今 利用对齐的目标-上下文线索构建鲁棒的视觉语言跟踪器

✎ 视觉语言跟踪 | 多模态学习 | 上下文线索

✎ 针对视觉语言跟踪算法在长序列跟踪时的局限性，提出了一种新颖的跟踪器ATCTrack，该方法通过目标-上下文特征建模，获取与目标状态对齐的多模态线索，实现鲁棒的跟踪效果。

✎ 对于视觉模态，提出了一种高效的时序视觉目标-上下文建模方法，为跟踪器提供及时的视觉线索；对于文本模态，提出了基于文本内容的目标词识别方法，确保关注目标相关的词语，同时减少辅助上下文词语的干扰。

✎ 与现有基于视觉-文本相似度的文本处理方法相比，所提方法简化了任务，同时提高了准确性，在代表性基准上取得SOTA的性能。

✓ 相关工作于2024年8月投稿至CCF-A会议AAAI [R6]，处于审稿中。

2024.07 - 至今 基于多模态线索转换机制的视觉语言跟踪器

✎ 视觉语言跟踪 | 多模态学习 | Grounding模型

✎ 针对视觉语言跟踪面临视频和文本数据不平衡问题，提出一种即插即用的方法，将文本线索转化为视觉线索，从而提升跟踪效果。

✎ 利用基础模型的强大文本-图像对齐能力，将文本线索转换为可解释的视觉热图，并通过热图引导模块将这些热图与搜索图像融合，以更有效地指导跟踪。

✎ 该方法简化了VLT系统对文本提示的处理，且与现有跟踪器兼容。在多个主流基准上的实验结果证明了其在增强VLT性能方面的有效性，在代表性基准上取得SOTA的性能。

✓ 相关工作于2024年9月投稿至CCF-B会议ICASSP [R7]，处于审稿中。

基于多模态交互的视觉语言跟踪基准

✎ 视觉语言跟踪 | 多模态交互 | 评估评测技术

✎ 针对当前视觉语言跟踪任务在多轮交互机制上的缺失，提出了新的基准VLTm，首次引入多轮交互机制。

✎ 使用DTLLM-VLT生成多粒度文本，并通过文本更新和目标恢复实现多轮多模态交互。

✎ 设计了新的交互范式，在多次跟踪失败时提供更准确的文本和边界框，提高了跟踪器的鲁棒性和任务范围，扩展了VLT任务的下游应用。

✎ 通过对比传统VLT基准，验证了本工作所提出的交互范式下的跟踪器准确性和鲁棒性，为领域提供了更精细的多模态跟踪器评估方法，并为未来数据集扩展提供了新思路。

✓ 相关工作于2024年9月投稿至CCF-A会议NeurIPS附属的第一届视觉语言模型研讨会 [R8]，处于审稿中。

论文成果

已发表/接收 (Acceptance)

- 1 S. Hu, X. Zhao, L. Huang, and K. Huang, "Global instance tracking: Locating target more like humans", *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 45, no. 1, pp. 576–592, 2023.

DOI: 10.1109/TPAMI.2022.3153312.

- 2 **S. Hu**, D. Zhang, M. Wu, X. Feng, X. Li, X. Zhao, and K. Huang, "A multi-modal global instance tracking benchmark (mgit): Better locating target in complex spatio-temporal and causal relationship", in *The 37th Conference on Neural Information Processing Systems (NeurIPS, Poster)*, vol. 36, 2023, pp. 25 007–25 030.
- 3 **S. Hu**, X. Zhao, and K. Huang, "Sotverse: A user-defined task space of single object tracking", *International Journal of Computer Vision (IJCV)*, vol. 132, pp. 872–930, 2024.  DOI: 10.1007/s11263-023-01908-5.
- 4 X. Zhao , **S. Hu** , Y. Wang, J. Zhang, Y. Hu, R. Liu, H. Ling, Y. Li, R. Li, K. Liu, and J. Li, "Biodrone: A bionic drone-based single object tracking benchmark for robust vision", *International Journal of Computer Vision (IJCV)*, vol. 132, pp. 1659–1684, 2024.  DOI: 10.1007/s11263-023-01937-0.
- 5 **S. Hu**, X. Zhao, and K. Huang, "Visual intelligence evaluation techniques for single object tracking: A survey (单目标跟踪中的视觉智能评估技术综述)", *Journal of Images and Graphics (《中国图象图形学报》)*, 2023.
- 6 K. Huang, X. Zhao, Q. Li, and **S. Hu**, "Visual turing: The next development of computer vision in the view of human-computer gaming (视觉图灵：从人机对抗看计算机视觉下一步发展)", *Journal of Graphics (《图学学报》)*, vol. 42, no. 3, p. 339, 2021.  DOI: 10.11996/JG.j.2095-302X.2021030339.
- 7 J. Zhang, T. Zhao, **S. Hu**, and X. Zhao, "Robust single-particle cryo-em image denoising and restoration", in *The 49th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP, Poster)*, 2024.
- 8 X. Feng, X. Li, **S. Hu**, D. Zhang, M. Wu, J. Zhang, X. Chen, and K. Huang, "Memvlt: Visual-language tracking with adaptive memory-based prompts", *The 38th Conference on Neural Information Processing Systems (NeurIPS, Poster)*, 2024.
- 9 D. Zhang, **S. Hu**, X. Feng, X. Li, M. Wu, J. Zhang, and K. Huang, "Beyond accuracy: Tracking more like human via visual search", *The 38th Conference on Neural Information Processing Systems (NeurIPS, Poster)*, 2024.
- 10 M. Wu, K. Huang, Y. Cai, **S. Hu**, Y. Zhao, and W. Wang, "Finger in camera speaks everything: Unconstrained air-writing for real-world", *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 2024.
- 11 Y. Wang, **S. Hu**, and X. Zhao, "Rethinking similar object interference in single object tracking", in *The 7th International Conference on Computer Science and Artificial Intelligence (CSAI, Oral)*, 2023, pp. 251–258.
- 12 X. Feng, **S. Hu**, X. Chen, and K. Huang, "A hierarchical theme recognition model for sandplay therapy", in *The 6th Chinese Conference on Pattern Recognition and Computer Vision (PRCV, Poster)*, 2023, pp. 241–252.  DOI: 10.1007/978-981-99-8462-6_20.
- 13 K. Huang, Y. Kang, C. Yan, **S. Hu**, L. Wang, T. Tao, and W. Gao, "A review of intelligent psychological assessment based on interactive environment (基于交互环境的智能化心理测评)", *Chinese Mental Health Journal (《中国心理卫生杂志》)*, 2024.
- 14 M. Wu, Y. Kang, X. Li, **S. Hu**, X. Chen, Y. Kang, W. Wang, and K. Huang, "Vs-llm: Visual-semantic depression assessment based on llm for drawing projection test", *The 7th Chinese Conference on Pattern Recognition and Computer Vision (PRCV, Poster)*, 2024.
- 15 X. Li, X. Feng, **S. Hu**, M. Wu, D. Zhang, J. Zhang, and K. Huang, "Dtllm-vlt: Diverse text generation for visual language tracking based on llm", *3rd Workshop on Vision Datasets Understanding and DataCV Challenge in The IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024 (CVPRW, Oral, Best Paper Honorable Mention)*, 2024.
- 16 Y. Zhang, C. Liu, W. Chen, X. Xu, F. Wang, H. Li, **S. Hu**, and X. Zhao, "Revisiting instance search: A new benchmark using cycle self-training", *Neurocomputing (Neu)*, vol. 501, pp. 270–284, 2022.  DOI: 10.1016/j.neucom.2022.06.027.

预印本 (Preprint)

- 1 **S. Hu**, X. Zhao, Y. Wang, Y. Shan, and K. Huang, *Nearing or surpassing: Overall evaluation of human-machine dynamic vision ability*, 2023.  URL: https://openreview.net/forum?id=LGbZYw_pnsc.


在审稿 (Under Review)

- 1 **S. Hu***, X. Li*, X. Li, J. Zhang, Y. Wang, X. Zhao, and K. Cheong, "Can lvlms describe videos like humans? a five-in-one video annotations benchmark for better human-machine comparison", *The 13th International Conference on Learning Representations (ICLR, Under Review)*, 2024.
- 2 Y. Ma*, **S. Hu***, X. Li, Y. Wang, S. Liu, and K. Cheong, "Students rather than experts: A new ai for education pipeline to model more human-like and personalised early adolescences", *The 13th International Conference on Learning Representations (ICLR, Under Review)*, 2024.
- 3 Y. Wang, **S. Hu**, D. Zhang, M. Wu, T. Yao, Y. Wang, L. Chen, and X. Zhao, "Target or distractor? rethinking similar object interference in single object tracking", *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT, Under Review)*, 2024.
- 4 Y. Ren, X. Feng, **S. Hu**, Y. Kang, C. Yan, Y. Zeng, L. Wang, and K. Huang, "Intelligent psychological assessment with sandplay based on evidence-centered design theory (基于证据中心设计理论的智能心理沙盘测评系统)", *Acta Psychologica Sinica (《心理学报》, Under Review)*, 2024.
- 5 X. Li, **S. Hu**, X. Feng, D. Zhang, M. Wu, J. Zhang, and K. Huang, "Dtvlt: A multi-modal diverse text benchmark for visual language tracking based on llm", *The 13th International Conference on Learning Representations (ICLR, Under Review)*, 2024.
- 6 X. Feng, **S. Hu**, X. Li, D. Zhang, M. Wu, J. Zhang, X. Chen, and K. Huang, "Atctrack: Leveraging aligned target-context cues for robust vision-language tracking", *The 39th Annual AAAI Conference on Artificial Intelligence (AAAI, Under Review)*, 2024.
- 7 X. Feng, D. Zhang, **S. Hu**, X. Li, M. Wu, J. Zhang, X. Chen, and K. Huang, "Enhancing vision-language tracking by effectively converting textual cues into visual cues", *The 50th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP, Under Review)*, 2024.
- 8 X. Li, **S. Hu**, X. Feng, D. Zhang, M. Wu, J. Zhang, and K. Huang, "Visual language tracking with multi-modal interaction: A robust benchmark", *1st Workshop on Video-Language Models in The 39th Conference on Neural Information Processing Systems (NeurIPS, Under Review)*, 2024.

待投稿 (Ongoing Research)

- 1 **S. Hu**, J. Zhu, Y. Wang, X. Zhao, and K. Huang, "Vt³: A visual tracking turing test of human-machine dynamic vision ability", *Cell Patterns (In Preparation)*, 2024.
- 2 X. Zhao, **S. Hu**, and X. Yin, *Visual Object Tracking - An Evaluation Perspective*. Springer, 2024.
- 3 M. Wu, X. Li, **S. Hu**, Y. Cai, K. Huang, and W. Wang, "Unconstrained multimodal air-writing benchmark: Writing by moving your fingers in 3d", *The IEEE/CVF Conference on Computer Vision and Pattern Recognition 2025 (CVPR, In Preparation)*, 2024.
- 4 Y. Ma, Z. Yang, Y. Kang, X. Xue, and **S. Hu**, "From collaboration to integration: Research on the dilemma and countermeasures of human-computer interaction in intelligent education (从协同走向融合：智能教育中人机交互的困境及对策研究)", *Intelligent Education PhD Research Fund, supported by the Institute of AI Education Shanghai and East China Normal University (In Progress)*, 2024.

论文统计

期刊  **TPAMI**: IEEE Transactions on Pattern Analysis and Machine Intelligence (CCF-A 期刊, 计算机视觉领域Top-1期刊, IF=20.8) 已发表/接收×1 (一作×1)

论文统计 (continued)

- **IJCV**: International Journal of Computer Vision (CCF-A期刊, 计算机视觉领域Top-2期刊, IF=11.6) 已发表/接收×2 (一作×1, 通讯×1)
- **TCSVT**: IEEE Transactions on Circuits and Systems for Video Technology (CCF-B期刊, IF=8.3) 已发表/接收×1, 在审稿×1
- **JIG**: Journal of Images and Graphics (《中国图象图形学报》, CCF-B中文期刊)。已发表/接收×1 (一作×1)
- **JOG**: Journal of Graphics (《图学学报》, CCF-C中文期刊) 已发表/接收×1
- **Neu**: Neurocomputing (CCF-C期刊, IF=5.5) 已发表/接收×1
- **CMHJ**: Chinese Mental Health Journal (《中国心理卫生杂志》, CSSCI期刊, 中国顶尖心理学期刊) 已发表/接收×1
- **APS**: Acta Psychologica Sinica (《心理学报》, CSSCI期刊, 中国Top-1心理学期刊) 在审稿×1
- 会议 ■ **NeurIPS**: Conference on Neural Information Processing Systems (CCF-A会议) 已发表/接收×3 (一作×1)
- **ICLR**: International Conference on Learning Representations (CAAI-A会议) 在审稿×3 (一作×2)
- **AAAI**: Annual AAAI Conference on Artificial Intelligence (CCF-A会议) 在审稿×1
- **NeurIPSW**: Workshop in Conference on Neural Information Processing Systems (CCF-A会议研讨会) 在审稿×1
- **CVPRW**: Workshop in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CCF-A会议研讨会) 已发表/接收×1 (口头报告&最佳论文提名×1)
- **ICASSP**: IEEE International Conference on Acoustics, Speech, and Signal Processing (CCF-B会议) 已发表/接收×1, 在审稿×1
- **PRCV**: Chinese Conference on Pattern Recognition and Computer Vision (CCF-C会议) 已发表/接收×2
- **CSAI**: International Conference on Computer Science and Artificial Intelligence (EI会议) 已发表/接收×1 (口头报告×1)

技能

- 语言 ■ 汉语普通话(母语)和英语。
- 编程 ■ Python, Java, Matlab, C, \LaTeX .
- 开发 ■ Android, Flask, SQLite.
- Linux ■ Shell, OS虚拟化。
- 其他 ■ 学术研究, 团队协作能力, 领导力, 学术报告。

奖励及荣誉

- 2024 ■ **最佳论文提名**, CVPR第三届视觉数据理解研讨会。
- **北京市优秀毕业生**, 北京市教育委员会(获奖率: 5%)。
- 2023 ■ **国家奖学金**, 中华人民共和国教育部(获奖率: 1%)。
- **攀登一等奖学金**, 中国科学院自动化研究所。
- 2022 ■ **优秀学生**, 中国科学院大学。
- 2017 ■ **学业奖学金**, 北京理工大学。
- **优秀创新学生**, 北京理工大学。
- 2016 ■ **大学生奖学金**, 中国科学院。
- **学业奖学金**, 北京理工大学。

奖励及荣誉 (continued)

- 优秀团员，北京理工大学。
- 2015 ■ 国家一等奖，全国大学生数学建模竞赛（获奖率：1%）。
- 学业奖学金，北京理工大学。
- 数学建模竞赛校第一名，北京理工大学。
- 暑期社会实践优秀个人，北京理工大学。
- 暑期社会实践校二等奖，北京理工大学（队长）。
- 优秀学生干部，北京理工大学。
- 优秀团干部，北京理工大学。
- 优秀团支部，北京理工大学（团支书）。
- 十佳团日活动，北京理工大学（团支书）。
- 2014 ■ 学业奖学金，北京理工大学。
- 优秀学生，北京理工大学。
- 2013 ■ 学业奖学金，北京理工大学。

学术活动和服务

- 讲习班 ■ 31th IEEE International Conference on Image Processing (ICIP)
- 题目：An Evaluation Perspective in Visual Object Tracking: from Task Design to Benchmark Construction and Algorithm Analysis
 - 时间和地点：27-30 October, 2024, Abu Dhabi, United Arab Emirates
 - 时长：半天（3小时）
- 27th International Conference on Pattern Recognition (ICPR)
- 题目：Visual Turing Test in Visual Object Tracking: A New Vision Intelligence Evaluation Technique based on Human-Machine Comparison
 - 时间和地点：01-05 December, 2024, Kolkata, India
 - 时长：半天（3小时）
- 17th Asian Conference on Computer Vision (ACCV)
- 题目：From Machine-Machine Comparison to Human-Machine Comparison: Adapting Visual Turing Test in Visual Object Tracking
 - 时间和地点：08-12 December, 2024, Hanoi, Vietnam
 - 时长：半天（3小时）
- 副主编 ■ 期刊：Innovation and Emerging Technologies
- 审稿 ■ 会议：NeurIPS, ICLR, CVPR, ECCV, AAAI, ACMMM, AISTATS等
- 期刊：SCIENCE CHINA Information Sciences, IEEE Access, Journal of Computational Science, Journal of Electronic Imaging, Digital Signal Processing等

协助指导学生

- 博士生 ■ 武美奇，2022.08-至今，中国科学院大学（计算机视觉|人机交互技术）
- 丰效坤，2023.04-至今，中国科学院自动化研究所（视觉目标跟踪|视觉语言跟踪）
- 马一平，2023.08-至今，华东师范大学（智能教育技术|人机交互技术）
- 张岱凌，2023.08-至今，中国科学院自动化研究所（视觉目标跟踪|视觉图灵测试）
- 王怡沛，2024.08-至今，东南大学（多模态大模型|视觉目标跟踪）
- 李旭宸，2024.08-至今，中国科学院自动化研究所（视觉语言跟踪|多模态大模型|AI4Science）
- 硕士生 ■ 马一平，2022.05-2023.07，南京师范大学（智能教育技术|语音情感识别）

协助指导学生 (continued)

- 王怡沛, 2022.08-2024.07, 东南大学 (视觉目标跟踪 | 基于大语言模型的推荐系统)
- 本科生 ■ 朱峻佑, 2022.09-2023.07, 中国科学院大学 (视觉图灵测试)
- 胡力杭, 2022.09-2023.07, 中国科学院大学 (视觉目标跟踪)
- 张岱凌, 2022.09-2023.07, 东南大学 (视觉目标跟踪)
- 李旭宸, 2023.04-2024.07, 北京邮电大学 (视觉目标跟踪 | 视觉语言跟踪)

推荐人

黄凯奇研究员和赵鑫教授分别是我的博士导师和合作导师, 我非常荣幸和他们一起开展计算机视觉领域的研究。王卓立教授是我的硕士导师, 我非常荣幸在他的指导下开展高性能计算研究。目前, 我在南洋理工大学和Kanghao Cheng教授一起工作。

黄凯奇 研究员

研究员, IAPR会士, IEEE高级会员, 万人领军
中国科学院自动化研究所(CASIA)
中关村东路95号, 北京, 中国

✉ kqhuang@nlpr.ia.ac.cn

王卓立 教授

名誉教授
香港大学(HKU), 计算机科学系
薄扶林, 香港特别行政区, 中国

✉ clwang@cs.hku.hk

✉ choliwang@gmail.com

赵鑫 教授

教授, IEEE高级会员, 北京市杰青
北京科技大学(USTB), 计算机与通信工程学院
学院路30号, 北京, 中国

✉ xinzhaoustb.edu.cn

✉ xzhaopersonal@foxmail.com

Kanghao Cheng 教授

副教授, IEEE高级会员
物理与数学科学学院助理院长
科学学部助理院长

南洋理工大学(NTU)

南洋大道50号, 新加坡

✉ kanghao.cheong@ntu.edu.sg