

谢歆哲

实习时长: 大于三个月 | 实习意向: 视觉算法、多模态算法、大模型应用开发 (随时到岗)

电话: 18768078708 | 邮箱: xiexinzhe@zju.edu.cn | 意向城市: 浙江杭州

年龄: 26岁 | 性别: 男 | 政治面貌: 中共党员 | 个人网址: <https://github.com/Xinzhe99>



教育经历

2022.09 - 2027.06 浙江大学 研究方向: 人工智能 (国家重点实验室) 综测前1% 博士

研究领域: 多模态算法 (T2I, I2I, VLA)、计算机视觉与生成式图像算法, 涵盖多聚焦图像融合等

荣誉奖项: 博士生国家奖学金 (2025); 浙江大学五好研究生 (2024/2025)、优秀研究生 (2024/2025); 优秀学生干部 (2025)、优秀团员 (2025); 全国硕士统考专业第一、同级综素排名全系第一 (2025); 其余荣誉10+项; 硕博连读期间发表了8篇一作论文(远超毕业要求), 授权3项国家发明专利和4项软著

任职情况: 院研究生会 办公室成员; 班长; 作为核心成员参与多个重点科研项目。

2018.09 - 2022.06 温州大学 专业: 电子信息与工程 综测前5% 本科

荣誉奖项: 优秀毕业生 (2022)、“外教社·国才杯”全国英语阅读竞赛校赛 一等奖 (2021)、全国及浙江省大学生高等数学竞赛 二等奖 (2021)、“三好学生”与奖学金 多次 (2018-2022)、先进个人与优秀学生干部 多次 (2018-2022)、其它荣誉30+项

任职情况: 院团委·学生会 办公室主任; 班长; 学习委员

实习经历

2025.07 - 2025.10 杭州当虹科技股份有限公司 CTO Office - 算法实习生

- 三维体素化算法调研与落地:** 与宇数科技、云深处科技等机器人公司合作, 对雷达点云体素化算法进行系统调研与分析, 比较了OctoMap与voxblox框架在工程领域的可行性, 并独立实现当虹科技BlackEye Vision机器人超远距离远程操控系统的点云体素化功能, 助力机器人复杂环境盲操, 成果成功落地至 App 端并通过发布会完成产品演示与发布。
- 自主导航算法调研与验证:** 基于云深处X30工业机器狗的传感器配置, 系统调研合适的机器狗室内外自主导航与控制算法, 涵盖融合RTK、雷达里程计、IMU的定位算法及基于move_base框架的局部与全局路径规划, 为机器狗在开放环境中的精准循迹与稳定运动提供算法优化方案, 并完成初步验证。

2024.07 - 2024.09 共青团舟山市委 办公室 - 实习生(全市唯一)

- 会议统筹与管理:** 协助舟山市共青团 2024 年半年度工作会议的筹备工作, 负责议程规划、会务组织及核心材料准备, 通过精细化统筹保障了会议高效、顺利召开。
- 团务信息化与公文写作:** 负责团内日常信息化管理及数据统计, 建立规范的数据台账; 独立承担会议纪要、工作报告及信息汇总等重要文稿的撰写工作, 确保信息传递的准确性与时效性。

2022.10 - 2023.10 自然资源部 东海预报减灾中心 - 算法实习生

- 算法研发与部署:** 在校内外导师指导下, 参与多套水下成像传感系统的全生命周期开发。独立负责系统算法模块的设计与实现, 涵盖水下图像增强、深度估计、目标检测及超景深重建等方向, 完成算法从模型训练到部署优化的全流程工作。
- 产品落地与优化:** 参与项目落地与实际投入使用, 针对科研探测场景下复杂水下环境的设备稳定性与交互问题进行优化。通过多轮现场测试与算法迭代, 推动系统实现规模化部署, 显著提升运行稳定性与用户体验, 并顺利通过项目验收。

2020.10 - 2021.01 杭州海康威视责任有限公司 睿影 - 视觉算法实习生

- 算法优化与部署:** 参与杭绍线安检违禁品检测算法研发, 基于PyTorch对目标检测模型进行改进与微调, 在复杂场景下将违禁品检出率提升了5个百分点。
- 全流程验证体系:** 主导开发基于Python的自动化测试脚本, 集成了数据预处理、模型推理及后处理逻辑, 实现了对算法模型的全流程自动化验证与性能评估, 将单次回归测试时间缩短了200%。

项目经历

2022.9- 至今 海洋底栖生物原位实时显微监测系统研发及应用 主要负责人

中央引导地方科技发展资金 (No. 2025ZY01111)、省自然科学基金项目 (No. LQN26D060019)

- 硬件设计与实现:** 运用SOLIDWORKS完成系统整体结构设计, 确保其在水下环境的良好适应性 with 操作灵活性; 搭建光路并设计电子系统, 使其能稳定、高效地在水下运行。
- 软件开发与优化:** 基于NVIDIA ORIN NX平台, 进行嵌入式开发, 编写水下超景深显微镜核心工作代码; 基于PyQt开发服务端程序, 实现系统的高效运行与远程控制功能。
- 算法研发与创新:** 研发基于AI的多聚焦图像融合技术, 显著提升系统成像质量。针对多焦点图像栈融合问题, 创新性地提出多种融合框架, 在各个场景下均能实现高效高质量成像。

学术论文 (第一作者)

[1] Xie X, Guo B, He S, et al. Multi-focus image stack fusion with deterministic integration and generative restoration[J]. Engineering Applications of Artificial Intelligence. 【中科院1区Top在审】

Keywords: Generative Fusion; Latent Diffusion Model; ControlNet; Missing Focal Plane Restoration; Two-stage Framework

本文提出了 *StackMFF V4*，针对现有方法无法处理焦平面缺失及边缘伪影的问题，提出的两阶段生成式融合框架。该方法级联了确定性融合模块与基于 Latent Diffusion 的 IFControlNet，利用生成式先验重建缺失焦平面内容并消除边缘伪影。实验表明该方法有效解决了复杂场景下的细节丢失问题，实现了 SOTA 的融合性能。

[2] Xie X, Guo B, He S, et al. General multi focus image fusion [J]. Engineering Applications of Artificial Intelligence. 【中科院1区Top在审】

Keywords: Stack-based Fusion; Order-Agnostic; Cross-Layer Attention; Multi-class Classification

本文提出了 *StackMFF V3*，针对前代模型依赖有序输入的局限，提出的第三代通用型堆栈融合模型。创新性设计 Pyramid Fusion MLP 与像素级跨层注意力模块，在无需焦点顺序先验的情况下高效捕捉层内与层间依赖；通过将融合任务建模为像素级多分类问题，彻底消除了输入约束。实验表明该方法具有极强的通用性，在多基准上均达 SOTA。

[3] Xie X, Guo B, He S, et al. One-shot multi-focus image stack fusion via focal depth regression[J]. Engineering Applications of Artificial Intelligence, 2025, 162: 112667. 【中科院1区Top】

Keywords: Multi-focus Image Fusion; One-shot Framework; Focal-plane Depth Regression; Proxy Supervision; Real-time

本文提出了 *StackMFF V2*，针对现有方法处理图像堆栈时存在的误差累积与计算低效问题，我们将融合任务重构为焦平面深度回归的 One-shot 框架。该方法设计了三阶段网络架构，利用可微分软回归策略并以深度图作为代理监督信号，实现了无需人工标注的端到端训练；在 5 个公开数据集上的实验表明，该方法以极低的计算开销取得了 SOTA 性能。

[4] Xie, X., Qingyan, J., Chen, D., Guo, B., Li, P., & Zhou, S. (2025). StackMFF: end-to-end multi-focus image stack fusion network. Applied Intelligence, 55(6), 503. 【JCR: 2区】

Keywords: 3D CNN; Image Stack Fusion; Synthetic Data Generation; 3D Attention; End-to-End

本文提出了 *StackMFF*，针对传统堆栈融合的误差累积难题，提出基于 3D CNN 的端到端融合网络。我们构建了基于单目深度估计的数据合成流水线以解决训练数据短缺，并引入 3D 注意力机制与组合损失函数优化模型。实验表明该方法在速度与质量上均达 SOTA，建立了多焦点堆栈融合任务的强基准。

[5] Xie X, Guo B, Li P, et al. Multi-focus image fusion with visual state space model and dual adversarial learning[J]. Computers and Electrical Engineering, 2025, 123(Part D): 110238. 【JCR: 1区】

Keywords: Visual State Space Model; BridgeTune; Dual Adversarial Learning; Lightweight (0.05M); Real-time

本文提出了 *BridgeMFF*，针对现有网络在平坦区域难以判断聚焦状态的问题所提出的基于视觉状态空间模型的 BridgeMFF 网络。首创 BridgeTune 通用微调技术，利用双重对抗学习弥合语义与纹理差距，隐式学习局部空间分布。实验表明该方法在平坦区域表现优异，且模型仅 0.05M 参数量（单张 0.09s），实现了极高效的实时融合。

[6] Xie X, Guo B, Li P, et al. SwinMFF: toward high-fidelity end-to-end multi-focus image fusion via swin transformer-based network[J]. The Visual Computer, 2024: 1-24. 【JCR: 2区】

Keywords: Swin Transformer; Multi-focus Image Fusion; Long-range Dependencies; High Fidelity; End-to-End

本文提出了 *SwinMFF*，针对现有端到端融合方法因像素回归误差导致保真度不足的问题开发的基于 Swin Transformer 的多聚焦图像融合框架。通过捕捉图像间的长距离依赖关系显著降低回归误差，有效抑制边缘伪影并实现高保真融合。大规模实验证明，该方法在主客观指标上均优于 28 种 SOTA 方法。

[7] Xie X, Lin Z, Guo B, et al. LightMFF: A Simple and Efficient Ultra-Lightweight Multi-Focus Image Fusion Network[J]. Applied Sciences, 2025, 15(13): 7500. 【JCR 2区】

Keywords: Ultra-lightweight; Refinement Perspective; Edge Guidance; Real-time; Low FLOPs

本文提出了 *LightMFF*，针对现有方法计算资源需求高的问题所提出的超轻量级网络。我们创新性地将融合任务从分类视角重构为细化视角，利用边缘引导简化架构。模型仅 0.02M 参数与 0.06G FLOPs（计算量降低 98%），单张推理仅 0.02s，在维持 SOTA 性能的同时实现了极致的效率。

[8] Xie X, Guo B, Li P, et al. Underwater three-dimensional microscope for marine benthic organism monitoring[C]//OCEANS 2024-Singapore. IEEE, 2024: 1-4. 【领域内TOP国际会议论文集】

Keywords: Underwater 3D Microscope; Electrically Tunable Lens; In-situ Monitoring; Image Fusion; 3D Reconstruction

本文提出了一种水下显微成像系统，针对复杂海底环境下的微观监测难题，引入电调谐透镜，实现了工作距离动态调整；结合图像融合算法，从堆栈中重建高清晰度全焦图像。该系统具备微米级分辨率，实现了对底栖生物的长期原位实时监测。

技术栈与语言水平

- **技术栈**: Python (Advanced), C++ (Basic, mainly with ROS), PyTorch, TensorFlow (Optional), Hugging Face (Transformers, Diffusers), ROS & ROS2 (Noetic/Melodic), Git, Docker, Accelerate, NumPy, SciPy, OpenCV, Matplotlib, LaTeX, Markdown
- **语言水平**: 英语(CET-6, 分数达到浙江大学英语免修要求), 普通话(二甲)
- **其它**: [OpenFocus](#) 开源多聚焦图像融合系统项目作者; [StackMFF](#) 多聚焦图像融合系列算法作者; 长期深耕计算摄影与图像融合方向, 兼具对多个视觉领域研究进展的系统理解。拥有4年 AI 辅助编程经验, 具备出色的工程实现能力, 已在 GitHub 上开源并维护 10 余项视觉算法项目。持续关注前沿研究动态, 并能够将研究构想高效转化为可用系统与算法实现。