

吴世涵

电子科技大学 · 计算机科学与工程学院 · 计算机科学与技术

🏠 koorye.github.io · 🌐 koorye.github.io/blog · ✉ shihan.wu.koorye@outlook.com · ☎ +8615869663967

👤 个人介绍

我是来自电子科技大学计算机科学与技术专业的学术硕士 (学院排名 **Top 1.3%**)，研究方向聚焦：

- 视觉语言模型的高效与高性能迁移学习
- 机器人视觉-语言-动作模型 (VLA) 的训练与测试时适应
- 以一作/共一身份发表 **CVPR(CCF-A)** 论文 2 篇，另有 2 篇在投，获得国家奖学金、优秀毕业生等荣誉

🎓 教育背景

电子科技大学 计算机科学与技术 · 学术硕士 2023.9 ~ 2026.6

- 学业表现: GPA 3.94/4.0 | 学院排名: 6/454 (**Top 1.3%**)
- 核心荣誉: **国家奖学金**、优秀研究生、学业一等奖学金、新生一等奖学金等

电子科技大学 软件工程 · 工学学士 2019.9 ~ 2023.6

- 学业表现: GPA 3.94/4.0 | 专业排名: 18/181 (**Top 10%**) | 英语四六级: 579/467
- 荣誉奖项: **优秀毕业生**、“世强”专项奖学金、优秀学生奖学金等

🔬 科研论文

[ICLR 2025 投稿中 · 共同第一作者] Policy Contrastive Decoding for Robotic Foundation Models 2025.9

- 针对具身智能大模型 (VLA) 的虚假相关性问题，首次提出策略对比解码的解决方法
- 通过开放语义目标检测模型 Grounding DINO 和图像分割模型 SAM2 对图像观察中的目标实现实时分割和跟踪，并通过图像重绘框架 Inpaint Anything 抹除目标物体得到负样本
- 通过正负样本在 VLA 输出概率空间中的对比解码技术实现对虚假相关性的有效抑制
- 适用于基于自回归和扩散的多种 VLA 策略，无需训练实现多种主流 VLA+8%~41% 的性能提升
- 项目主页: <https://koorye.github.io/proj/PCD>, ArXiv: <https://arxiv.org/abs/2505.13255>

[ICRA 2025 投稿中 · 共同第一作者] InSpire: Vision-Language-Action Models with Intrinsic Spatial Reasoning 2025.9

- 针对具身智能大模型 (VLA) 的虚假相关性问题，提出基于空间推理和思维链技术的解决方法
- 引入空间推理问题作为多模态大模型 (VLM) 与具身智能大模型 (VLA) 的中间桥梁，引导模型的动作推理关注图像观察中的目标
- 实现 VLA 在可见 (+6.2%) 和未见 (+10%) 任务性能显著提升
- 项目主页: <https://koorye.github.io/proj/Inspire>, ArXiv: <https://arxiv.org/abs/2505.13888>

[CVPR 2025 (CCF-A) · 第一作者] Skip Tuning: Pre-trained Vision-Language Models are Effective and Efficient Adapters Themselves 2024.12

- 传统提示调优 (P-Tuning 等) 和适配器 (LoRA 等) 需要引入额外参数，且时间和内存效率提升有限
- 提出了针对视觉语言模型全新的高效微调方法，从长度和宽度两个维度修剪特征-梯度传播流
- 不改变结构的同时，实现 +1.04% 准确率以及 15x 时间效率和 6.4x 内存效率的提升
- 项目主页: <https://github.com/Koorye/SkipTuning>, ArXiv: <https://arxiv.org/abs/2412.11509>

[CVPR 2024 (CCF-A) · 共同第一作者] DePT: Decoupled prompt tuning 2023.11

- 针对现有提示调优方法在已见和未见类上的性能权衡问题，提出视觉语言模型的解耦学习方法
- 将视觉语言模型的输出特征空间分为任务特定和任务通用空间，在学习已见类知识的同时保持未见类上的通用泛化能力
- 作为即插即用的解耦方法，取得现有多种提示调优方法 +0.67%~2.65% 的性能普遍提升
- 项目主页: <https://github.com/Koorye/DePT>, ArXiv: <https://arxiv.org/abs/2309.07439>

💻 实习经历

北京智源人工智能研究院 (BAAI) 研究实习生 · 具身智能大模型 2025.6 - 2025.10

- 设计大型双臂交互式机器人数据集，包含 6 款双臂机器人本体和 350K+ 轨迹数据
- 探索具身智能大模型 (VLA) 的能力金字塔，设计双臂复杂任务的层次化技能标注
- 开发 Realman RM75, Agilix Piper, Pika 等双臂机器人的部署系统，以及基于 LeRobot 的数据平台

📁 项目经历

雷达信号智能检测系统 · 算法研发	2024.3 - 2024.7
‣ 设计雷达信号的实时检测系统，利用频域图实时检测并提取其中的雷达信号片段	
‣ 改进一维残差网络作为信号特征提取网络和分类器，通过数据增强以实现更好的分类性能	
‣ 设计开集检测算法，实现对已知和未知雷达信号的有效区分	
面向自动驾驶场景的复杂目标检测系统 · 算法研发	2023.1 - 2023.6
‣ 改进两阶段目标检测模型 <i>Faster R-CNN</i> 实现未知物体的小样本适应	
‣ 针对华为 <i>CODA</i> 数据集中的极端场景，利用开放语义视觉语言模型 <i>CLIP</i> 生成语义密集的软标签	
‣ 利用知识蒸馏技术对 <i>Faster R-CNN</i> 的 <i>RPN Head</i> 与分类器提供引导，实现未知物体的有效检测	
水面目标智能检测系统 · 算法研发	2022.7 - 2022.9
‣ 选用单阶段目标检测模型 <i>YOLOX-l</i> ，实现水面目标的实时检测	
‣ 针对长尾数据分布设计了 <i>Mosaic</i> 与 <i>Copy-Paste</i> 数据增强策略，提升困难目标的检测性能	
智能安全驾驶系统 · 后端开发与运维	2022.4 - 2022.8
‣ 面向安全驾驶监测需求，设计了一套嵌入式设备、服务端与检测算法结合的智能安全驾驶系统	
‣ 通过摄像头和可穿戴式设备实时收集用户图像和心跳、血压等信息，与服务端实时同步	
‣ 调用检测算法对用户的疲劳驾驶、吸烟等危险行为进行监控和反馈	
“FACE ME” 人脸美学评分系统与智能美妆推荐系统 · 算法研发	2021.10 - 2022.4
‣ 基于华南理工 <i>SCUT-FBP5500</i> 人脸美学数据集，利用残差网络 <i>ResNet-18</i> 回归预测对人脸进行评分	
‣ 设计基于个性化 <i>PageRank</i> 算法智能产品推荐系统，分析用户购买记录网络，为用户推荐感兴趣产品	

🔧 专利成果

[申请中 · 学生第二发明人]	视觉语言模型的微调方法	2025.2
[申请中 · 学生第一发明人]	基于低频增强的小样本图像分类迁移方法	2024.8

🏆 竞赛奖项

国际级 Meritorious Winner · 美国大学生数学建模大赛 (MCM)	2021.4
国家级铜奖 · 中国国际“互联网+”大学生创新创业大赛	2021.12
省部级一等奖 · 中国成都国际软件设计和应用竞赛	2021.10
省部级二等奖 · 全国大学生数学建模竞赛	2021.9
省部级银奖 · “挑战杯”四川省大学生创业计划竞赛	2022.7

</> 技术能力

核心能力: 视觉语言模型 (CLIP、LLaVa) 及微调 (Prompt Tuning、LoRA)、VLA 模型 (OpenVLA、 π_0 等)
编程语言: Python、Java、C、C#、JavaScript、SQL 等
算法框架: PyTorch、Keras、MMDetection、Scikit-learn 等
工程能力: 前后端 (Vue、SpringBoot)、数据库 (MySQL、Redis)、运维 (Docker)、游戏 (Unity3D)、移动端