马俊程

ucasmjc | majuncheng21@mails.ucas.ac.cn | 15628896318

教育背景

中国科学院大学 人工智能专业 2021.9 至今

- **GPA**: 3.91/4.00 • **排名**:7/70 • **专业课 GPA**:4.00/4.00
- **相关课程**: 模式识别与机器学习 (96, **Top 1/70**), 多媒体信息处理 (95, **Top 1/36**), 模式识别研讨课 (100, **Top 1/36**), 人工智能的数学基础 [矩阵论与优化理论](95), 人工智能原理 (95), 信号与系统 (94), 程序设计基础与实验 (97)
- **英语成绩**: CET-4：543, CET-6：525

实习经历

人大高瓴，GeWu-Lab 2023.10 - 2024.3

在胡迪教授的指导下，熟悉多模态领域的相关工作，具体探索了 *Audio-Visual Segmentation*(AVS) 任务，希望预测视频中声源的二值 *mask* (S4, MS3) 和语义 *mask* (AVSS)。基于任务分解的思想提出了一种双阶段渐进训练策略，并提出一个新的 AVS 框架，在三个子任务上均实现 SOTA，**该成果已被 ECCV 2024 接收（第一作者）**。

中国科学院自动化所，模式识别国家重点实验室 2023.3 - 2023.7

在张兆翔老师课题组进行科研实习，阅读目标检测、语义分割、多目深度估计等领域论文，对语义分割论文进行复现；熟悉 *Pytorch* 和 *MMSegmentation* 的使用，对鱼眼相机在地下车库采集的畸变视觉数据进行矫正和分割；学习 SLAM 和三维视觉相关知识，阅读 ORB-SLAM 源代码。

科研项目

Stepping Stones: A Progressive Training Strategy for Audio-Visual Semantic Segmentation 2023.10 - 2024.3

[Audio-Visual Alignment] **ECCV2024 Accepted.** 第一作者。 Stepping-Stones

- 观察到过去工作在视听语义分割 (AVSS) 任务上学习不充分而达到次优化，语义监督信息的引入反而降低了模型的 audio visual 模态间对齐能力。进一步将此现象归因于 AVSS 任务的固有复杂性和监督信号在端对端训练时的歧义性，并通过分解任务目标进行分阶段充分训练以实现 *step-by-step* 的全局优化。
- 提出了一种简单而有效的双阶段训练策略 *Stepping Stones*，实现由视听定位到语义理解的由粗到细的声源语义分割，在 AVSS 任务上将 mIoU 提升到 48.50%(+11.84%)，将 F-score 提升到 53.20%(+11.20%)。该训练策略具有很好的泛化性，可以直接应用到过去工作中，并实现效果提升。
- 提出了一个新的视听分割模型 Adaptive Audio Visual Segmentation (AAVS)，设计了一个 parameter-free 的 audio query 生成器来自适应融合语音特征，并引入 masked-attention 动态调整 audio query 的注意力范围，在单声源分割 (S4) 和多声源分割 (MS3) 任务上分别将 mIoU 提升到 83.18%(+1.12%) 和 67.30%(+2.30%).

A Survey on Diffusion-based Motion Generation 2024.4 - 2024.6

[AIGC] 科研实践第一阶段成果，在中国科学院自动化所张兆翔老师组完成。 Diff-Mogen

- 回顾了文生图扩散模型的代表性工作，并概述了 DDPM, ScoreSDE, 条件生成（Guidance）的数学原理。
- 将动作生成领域中，基于扩散模型的顶会文章根据其科研贡献进行归类和回顾，整理出当前的主流研究方向：动作扩散模型，可控性增强（文本控制/空间约束）、数据可得性。
- 总结并提出未来的研究方向：更先进的生成模型，数据受限下的生成，更灵活有效的控制，提高生成效率，更符合人类标准的评价指标。

When Segment Anything Model Meet Sound Source Localization. 2023.10 - 2023.12

[Audio-Visual Alignment] 《多媒体信息处理》大作业，课程最终得分 95。 GSAVL

- 声源定位任务的目标为在无监督设置下定位视频中的发声物体，过去方法只能捕获粗略的物体轮廓。受声源分割领域工作的启发，将视觉基础模型 SAM 应用到声源定位任务，提出了一个新的框架 Generalizable Shape-aware Audio-Visual Localization(GSAVL)，实现了更高的定位精度和泛化能力。
- 广泛阅读声源定位和声源分割的论文，梳理方法的发展脉络，并进行汇总和归类。

基于人像分割的实时虚拟背景替换系统 2023.3 - 2023.7

[Computer Vision] 《模式识别与机器学习》大作业，课程最终得分 96。 PRML

- 提出一个实时人像分割模型，在人像数据集上实现了最佳的精度-效率平衡，以 0.248M 参数达到 95.95% mIoU。
- 设计了多尺度上下文模块 SAPPM 和流对齐-注意力融合模块，并进行了充分实验验证有效性。

Robomaster 机器人竞赛 2021.9 - 2023.6

[Robotics] 担任电控组组长，负责机器人的电路控制；推进项目进展，书写 C 语言和单片机讲义培训新人。

Paper Reading 2023.3 至今

汇总了过去阅读论文时整理的笔记，包括对每篇文章的总结和理解，涵盖多个领域。 Paper-Reading

- AIGC: Diffusion Models, Text2Image, Motion generation, Controllable Generation.
- Computer Vision: (Real-time/Semi-Supervised/Few-shot) Semantic Segmentation, Vision Backbone, Object Detection.
- MutilModal Learning: Sound Source Localization, Audio Visual Segmentation.

研究兴趣

我对**图像/视频处理与分析**抱有浓厚的兴趣，也对**多模态大模型**、**AIGC** 等学术与工业界最前沿和最受关注的方向抱有浓厚的兴趣，希望未来可以与世界上其它最优秀的科研人员一起探索通往 AGI 之路，实现真正的智能！

专业技能

编程语言： Python, C, C++, Matlab
框架： Pytorch, MMSegmenation
工具： Latex, Git, Markdown……
性格： *Keep Self-Motivated, Passionate and Positive*