

# 崔若晨

年龄：21 岁 · 意向申请：博士 / 硕士  
手机：13522181392 · 邮箱：ruochencui421@gmail.com



## 教育背景

### 电子科技大学

2021.09 - 至今

- 计算机科学与工程学院（网络空间安全学院），人工智能专业
- 成绩：GPA：3.68 / 4.00 排名：25 / 69
- 语言：CET4：578 CET6：532
- 编程：熟悉 Python / C，熟悉 Pytorch 框架、DeepSpeed 框架，了解 Linux、Git

## 科研经历

### A Novel Window-Interaction Module Based on W-MSA

2023.03 - 2023.12

- 内容：Swin Transformer 实现了窗口局部自注意力，但是窗口间的信息交互有所不足。基于此，我们提出了一个基于分窗操作的双分支模块：窗口池化生成深度可分离卷积核，计算卷积核间的自注意力和通道间注意力；窗口卷积核聚合为全局卷积核；使用所有卷积核对原图作深度可分离卷积，最后使用  $1\times 1$  卷积聚合为一张图像。模块通过卷积核的信息交互实现了窗口间的信息交互，在去雨任务上实验，在减小计算量的同时取得了与 baseline 相当的效果。
- 贡献：作为论文第一作者，独立提出论文思路和编写论文代码，参与论文写作
- 成果：ICLR 2024 Tiny Paper Track “Invite to present”
- 链接：<https://openreview.net/pdf?id=ki4R0z0C4K>

### 针对多模态大模型幻觉问题提出 Post Interaction Block

2024.02 - 至今

- 内容：在当前的多模态大模型中，图像 tokens 被映射到文本空间中以实现跨模态信息交互，但是图像信息经过 LLM 后存在被丢失或被忽略的可能，从而增加出现幻觉的概率。基于此，我们提出了 Post Interaction Block (PIB)，这一模块用于 LLM backbone 与 lm\_head 中间，用以二次增强图像与文本的信息交互，从而缓解多模态大模型幻觉问题。此外，我们计算 PIB 模块前后的 KL 散度，用以评估多模态大模型生成的答案的可信度。在 POPE 验证集上使用 LLaVA-1.5，准确率实现了 0.7-2.1% 的提升。
- 贡献：独立调研文献，提出研究思路，编写主要代码

### 基于 Prompt Tuning 方法的医学高维小样本数据分类研究

2023.10 - 2023.12

- 内容：调用 HuggingFace 开源的 BERT 类大模型，使用 Prompt Tuning、P-Tuning、Prefix Tuning 等技术，将数据输入到 LLM 中进行分类。最终调用 MacBERT、使用 Prefix Tuning 在五分类的 ADNI 数据集上做到了 76% 的准确率，远超各种传统方法。
- 贡献：主要负责调研文献与算法讨论，参与代码编写工作

### 基于 Restormer、Swin Transformer 的图像复原研究

2022.12 - 2023.12

- 基于 Restormer、SwinIR、Uformer 等 Transformer 模型结构，自主进行模型结构上的探索。利用 SENet 增强 Attention 模块在捕捉通道信息上的能力，观察不同 Attention 模块下 GELU、LeFF、MSFN 等模块作为 FeedForward Network 的效果。在图像去雨与高光谱图像融合等任务上进行实验。

## 获奖情况

- 2024 年美国大学生数学建模比赛 M 奖
- 2024 年蓝桥杯 AI 办公赛道省级一等奖
- 2023 年大学生创新创业实践项目“深度学习图像超分辨率方法研究” 结题优秀

## 学生工作

- 电子科技大学青年志愿者协会外事部成员

## 自我评价

我对科研工作具有浓厚兴趣，从大一开始我主动参与各种课外项目，广泛探索；同时我具有读博的想法，这样可以保证科研的连续性，做出更多有影响力的工作，以便有足够的时间做出满意的工作。