苏彤睿

手机：(+86) 15264823092 · 邮箱：molarsu18@gmail.com

研究领域

- 具有较强代码能力，擅长 python，C 等语言，熟练使用 verl，mmlab 等框架。
- 感兴趣的方向包括**大模型的后训练**，**RL** 等。

教育背景

北京理工大学，本科 2021.08-2025.06

- **本科专业排名**：6/162 **荣誉称号**：北京理工大学一等奖学金（连续两年）等
- **获奖经历**：ICPC,CCPC 区域赛银牌

中科院计算所，硕士，**导师：陈薇研究员** 2025.09 至今

大模型后训练与自进化 RL 研究

外呼对话系统的后训练–美团履约实习 2025.08-2026.02

- **项目简介：**在复杂的销售、招聘等外呼场景下对模型进行 CPT、SFT、RLVR 训练。为解决传统单轮对话难以达成复杂场景最终目标的问题，构建了一套多轮场景下的数据构建、模型继续预训练以及分角色评测与训练的完整框架。
- **数据构建与预训练：**构建 CPT 和 SFT 单轮数据，分别训练目标模型与用户模拟器（如销售员与客户模型）以供后续流程使用。并与业务方紧密合作，针对不同角色需求构建了分角色评测 benchmark。
- **多轮 RLVR 探索：**在多轮场景下，引导目标模型和用户模拟器进行交互。结合轮次与对话级别评分，深入探索了目标模型与用户模拟器的协同进化机制。
- **技术栈:** GRPO, DAPO, GSPO 等常用 Policy 更新技术，LLM-as-Judge 给出细粒度混合 reward。
- **项目成果：**沉淀的 RLVR 多轮训练方案成功推广至多个业务场景。在核心业务场景下，GSB 评估相较于多轮训练前的 baseline 提升 60%。

零数据和零监督下的 RL 自进化 2025.10-至今

- **研究背景：**R-zero 框架在无标注、无监督环境下，通过“出题人-答题人”角色扮演及 Majority Voting 生成伪标签进行迭代进化。但在实验中发现，该机制在三次迭代后会出现明显的模式崩塌。
- **效果提升**：深入剖析发现，模式崩塌源于出题与答题同源，导致生成数据过度拟合模型自身分布并产生 Overconfidence。为此，我们创新性地引入 PPL（困惑度）作为数据筛选标准。初步验证表明该方法能有效打破这一恶性循环，显著缓解模式崩塌并提升模型性能，相关深度验证正在推进中。
- **性能加速**：R-zero 两模型隔离训练，有重复低效 rollout，导致训练缓慢，且导致 off policy，会有训推差距。为此，深入修改 Verl，形成 Co-verl 框架，两模型同时交互推理，实时 on policy 训练，训练收敛速度提升 3 到 4 倍。
- **技术栈:** Verl 框架，verifier free RL，课程学习

模型在线自蒸馏 2026.02-至今

- **研究背景：**自蒸馏指的是模型同时扮演受专家演示引导的教师角色与仅接收任务输入的学生角色，在学生自身生成的轨迹上进行自蒸馏，最小化二者间的反向 KL 散度以实现在线策略学习
- **技术思路:**
  - **探索过渡提示词（Transition Prompt）对 Teacher 分布的影响**在 OPSD 在线蒸馏中，过渡提示词的功能是在教师的前向传播过程中，确保模型能够自然地从“理解参考答案”过渡到“评价学生采样轨迹”的状态，事实上影响甚至塑造了了 Teacher 分布，单独固定的 Teacher 分布可能会导致 Reward bias, 因此可以探索多样性过度提示词带来的分布方差和偏差的变化。
  - **探索更大模型在线蒸馏**目前学界对在线自蒸馏的研究停留在较小模型上，其有效防止灾难性遗忘，细粒度 reward 等优势十分有前景，且其在更大模型上的性能在理论上应当更优，但尚缺乏探索。
- **技术栈：**OPSD，SDFT，SDPO 等在线自蒸馏技术

视觉泛化与安全研究

RaPA: Enhancing Transferable Targeted Attacks via Random Parameter Pruning
Tongrui Su, Qingbin Li, Shengyu Zhu, Wei Chen, Xueqi Cheng **(CVPR 2026)**

- **研究简介：**针对神经网络迁移攻击中的过拟合痛点，深入剖析了对抗样本过度依赖代理模型特定参数的现象。
- **解决方案：**提出基于随机剪枝的多版本代理模型生成方法进行 self-ensemble，大幅提升了迁移攻击的成功率与系统的整体鲁棒性。
- **开源贡献：**代码已开源于 Github: https://github.com/molarsu/RaPA

CausalSeg: Causality Guided Latent Disentanglement for Corruption Robust Segmentation
Likun Wang, Jianing Li, Wei Chen, **Tongrui Su**, Mingkun Zhang, Qingbin Li, Xueqi Cheng **(ECCV 2026 在投)**

- **研究简介：**聚焦语义分割模型在现实复杂场景下对噪声、模糊等 15 种干扰的脆弱性，提出 CausalSeg 框架。
- **解决方案：**利用结构因果模型 (SCM)，并创新引入多粒度调节 (MGC) 重建技术，实现全局语义表示与局部纹理因子的有效解耦，显著提升了 OOD(分布外) 图像分割性能。

Exploring Structured Semantic Priors Underlying Diffusion Score for Test-time Adaptation
Mingjia Li, Shuang Li, **Tongrui Su**, Longhui Yuan, Jian Liang, Wei Li **(NeurIPS 2024)**

- **研究简介：**基于深度理论挖掘，利用扩散分数中的结构化语义先验，赋能图像分类器和密集预测器的测试时适应 (TTA)。
- **解决方案：**提出的 DUSA 方法仅需从 Diffusion 的单一时间步提取知识，彻底摆脱了传统基于蒙特卡洛的时间步似然估计的计算负担，在各类 TTA 任务上均达到 SOTA 水平。
- **开源贡献：**代码已开源于 Github: https://github.com/BIT-DA/DUSA

复杂环境下的视觉感知落地项目 (轻舟智航实习 & ITSAC 竞赛) 2023.10-2024.12

- **远距离红绿灯检测与深度估计（轻舟智航）：**负责自动驾驶感知模块，在少量精准深度标注数据限制下，通过引入多焦距相机深度一致性 loss，有效突破了远距离红绿灯检测的准确度瓶颈。
- **轨道交通复杂场景语义分割（ITSAC 夺冠）：**应对极端天气、城乡混合等复杂条件下的 railsem19 数据集，通过 class weight、logits-constraint 等策略进行极限优化，最终斩获中国智能交通大会 (ITSAC) 语义分割赛题第一名。

ACM 经历

北京理工大学 epsilon 队 2023.10-2024.01

- The 2023 ICPC Asia-East Continent Final Contest 入围 2024.01
- The 2023 ICPC Asia HeFei Regional Contest 银奖 2023.11
- The 2023 ICPC Asia XiAn Regional Contest 银奖 2023.10
- 第九届中国大学生程序设计竞赛（深圳）银奖 2023.11
- 第十四届蓝桥杯大赛软件赛省赛北京市一等奖