

曹柏源

电话: +86 137-8910-4361 邮箱: boyuancao@126.com

个人网站: <https://kmittle.github.io/>



教育背景

复旦大学 - 985/QS 30 | 生物医学工程 博士 | GPA 3.5/4.0

2023.09 - 2028.06

南京农业大学 - 211/双一流 | 自动化 本科

2018.09 - 2023.06

最具影响力学子(全校仅十人), 国家奖学金(0.2%), 优秀毕业生 | GPA: 92.3/100 | 排名: 1/65

项目经历

★符合人类审美偏好的高分辨率图像生成 (关键词: AIGC/扩散模型/强化学习)

2024.03 - 至今

- 背景: 生成高质量的高分辨率图像具有重要价值, 然而主流的扩散模型受限于其训练分辨率而难以实现。
- 技术栈: Pytorch, 扩散模型 (Diffusion、Flow Matching), 强化学习 (PPO、GRPO)。
- 个人贡献: (1)基于提出的无需训练的注意力机制引导扩散模型去噪过程, 提升表征的空间结构一致性并使图像整体结构提前浮现, 为高频生成提供更好先验从而提升图像的色彩和细节。 (2)证明了隐空间插值上采样导致伪影, 提出像素空间上采样+扩散重采样优化图像细节。 (3)通过大量实验证明了所提出注意力引导机制与主流预训练扩散模型兼容 (e.g., StableDiffusion 1.5、2.1、XL、3.0, ControlNet, etc)。 (4)提出了一种高效的去噪渲染器, 显著提升图像视觉体验。 (5)提出了一个符合人类审美偏好的大规模高分辨率数据集。 (6)提出了一个适用于扩散模型的强化学习训练框架以解决强化学习训练过程中优势不稳定的问题。
- 成果: 基于上述贡献(1)(2)(3), 以第一作者身份于 CCF-A AI 顶会 NeurIPS 2025 发表 [Spotlight ★ TOP 3% 论文](#)[paper, project_page], 相比于 SOTA 方法, 客观指标提升 9%, 用户主观评测指标提升 1.5-2.5 倍, 生成速度提升 4-5 倍。基于上述贡献(4)(5)(6)撰写论文, 待以第一作者身份投稿至 CCF-A AI 顶会 ICML。

★高效 DiT 架构设计 (关键词: AIGC/DiT/线性注意力机制/混合专家模型)

2024.12 - 至今

- 背景: DiT 架构在生成领域取得了瞩目的效果, 但是受限于 $O(N^2)$ 的时间/空间复杂度, 在生成视频、高分辨率图像时面临挑战。
- 技术栈: Pytorch, 扩散模型 (Diffusion、Flow Matching), 线性注意力机制, 混合专家模型 (MoE)。
- 个人贡献: (1)提出了动态差分注意力机制, 通过动态 token 投影、动态度量核映射、动态 token 差分分别实现 token 表征解耦、token 相似度度量优化、query-to-key 查询校准, 提升了线性注意力机制在生成任务上的建模能力。 (2)证明了 token-wise 差分和 attention map-wise 差分的内在一致性。 (3)提出了 DyDi-LiT 架构, 相比 SOTA 模型, 在多项指标上都达到了最优性能。
- 成果: 已以第一作者身份投稿至 CCF-A AI 顶会 CVPR, 相比 DiT 架构, 将复杂度降至 $O(N)$ 的同时获得了 14.4%-22.3% 的生成性能指标提升。

★电-气混合式农业机器人抓手设计及自适应主动柔顺抓取控制研究 (关键词: 位姿检测/序列预测)

2021.03 - 2023.06

- 背景: 果蔬产品由于其形状多变、柔嫩易损的特性, 实现其自动化无损采摘和精细分拣存在挑战。
- 技术栈: Pytorch/TensorFlow/Keras, YOLO (X、V3-V7), DETR, UNet (nnUNet、TransUNet), Mask2Former, 结构重参数化, RCNN 系列, RetinaNet, SSD, Sklearn, OpenCV, PLC 控制, STM32 开发
- 个人贡献: (1)提出了第一个针对果蔬抓取位姿检测的数据集。 (2)提出了抓取位姿检测的实时、高效检测算法。 (3)设计了一种吞咽式采摘机械手及机器人系统。 (4)设计触觉传感手套, 根据人类真实抓取时序数据, 设计物体感知的抓取控制序列预测模型。
- 成果: 基于贡献(1)(2)以第一作者身份发表期刊论文一篇(SCI 一区 TOP, IF=8.3)[paper], 投稿期刊论文一篇(SCI 一区 TOP, IF=12.4, Minor & Major Revision)[code], 获得发明专利一项 (ZL202210144228.2)。参与发表 SCI 一区 TOP 论文两篇。

★农业非结构化复杂场景重建 (关键词: 三维重建/NeRF)

2022.08 - 2023.06

- 背景: 精确重建农业场景对于实现农业自动化作业具有重要意义, 然而由于农业场景的非结构化复杂特性, 精确重建面临挑战。
- 技术栈: Pytorch, NeRF
- 个人贡献: 引入视角解耦的颜色预测, 通过可学习的混合因子得到最终颜色, 根据最终颜色预测视角耦合的密度, 通过可学习混合因子得到最终密度。
- 成果: 重建性能超越当时 SOTA 模型 Tensorf, 撰写技术报告。

其他荣誉