

# 曹柏源

+86 137-8910-4361 · boyuancao@126.com  
https://kmiddle.github.io/



## 教育背景

### 复旦大学

2023.09 - 2028.06

类脑智能科学与技术研究院 生物医学工程 博士 GPA: 3.5/4.0

### 南京农业大学

2018.09 - 2023.06

人工智能学院 自动化 本科 GPA: 92.3/100 排名: 1/65  
最具影响力学子 (全校仅十人) 国家奖学金 (1.8%) 优秀毕业生

## 论文 (第一作者)

- [CCF-A, NeurIPS 2025 Spotlight ★ TOP 3%] RepLDM: Reprogramming Pretrained Latent Diffusion Models for High-Quality, High-Efficiency, High-Resolution Image Generation (paper, project)
- [AI 顶会, ECCV 2026, Under Review] Dynamic Differential Linear Attention: Enhancing Linear Diffusion Transformer for High-Quality Image Generation (paper)
- [SCI 一区 TOP, IF=8.3] Real-time, highly accurate robotic grasp detection utilizing transfer learning for robots manipulating fragile fruits with widely variable sizes and shapes
- [SCI 一区 TOP, IF=12.4, Minor & Major Revision] Real-Time, Robust and Highly Accurate Robotic Grasp Detection of Fruits with Variable Sizes and Shapes (project)

## 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 / Diffusion / 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 (DDPM、Flow Matching), 强化学习 (DDPO、GRPO)。
- 个人贡献: (1) 设计了一种无参数的注意力机制, 基于该机制提出了一种与网络架构和去噪模型无关的采样算法, 极大提升了生成图像的色彩、细节等视觉体验 (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) 通过大量实验验证了所提出的采样算法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) 提出了一个符合人类审美偏好的大规模高分辨率数据集。 (5) 提出了一种高效的去噪渲染器, 显著提升图像视觉体验。 (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, 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 AI 顶会 ECCV。

### 基于动态差分线性注意力的 DiT 架构优化 | 关键词: AIGC / DiT / Diffusion / MoE

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 (DDPM、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 and attention map-wise 差分的内在一致性。 (3) 提出了 DyDi-LiT 架构, 相比 SOTA 模型, 在多项指标上都实现了显著提升。
- 成果: 以第一作者身份投稿至 AI 顶会 ECCV, 相比 DiT 架构, 将复杂度降至  $O(N)$  的同时获得了 14.4%-22.3% 的性能提升。

### 电-气混合式机器人抓手设计及自适应主动抓取控制 | 关键词: 具身智能 / 位姿检测

2021.03 - 2023.06

- 背景: 机器人抓取是具身智能的核心问题之一, 然而机器人在抓取形状多变、柔嫩易损的物体时面临挑战。
- 技术栈: Pytorch/TensorFlow/Keras, YOLO 系列, DETR, UNet, 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) 提出了第一个针对易损物体 (果蔬等) 的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**), 获得发明专利一项 (ZL202210144228.2)。参与发表 **SCI 一区 TOP** 论文两篇。

## 其他论文

---

- [IMCL 2025] Jiaxin Ye, **Boyuan Cao**, Hongming Shan. Emotional face-to-speech.
- [CVPR 2026] Jiaxin Ye, Gaoxiang Cong, Chenhui Wang, Xin-Cheng Wen, Zhaoyang Li, **Boyuan Cao**, Hongming Shan. Hierarchical Codec Diffusion for Video-to-Speech Generation.
- [SCI 一区 TOP] Yuhao Bai, Yunxiang Guo, Qian Zhang, **Boyuan Cao**, Baohua Zhang. End-to-End lightweight Transformer-Based neural network for grasp detection towards fruit robotic handling.
- [SCI 一区 TOP] Congmin Guo, Chenhao Zhu, Yuchen Liu, Renjun Huang, **Boyuan Cao**, Qingzhen Zhu, Ranxin Zhang, Baohua Zhang. Multi-network fusion algorithm with transfer learning for green cucumber segmentation and recognition under complex natural environment

## 获奖荣誉

---

- 优秀本科毕业论文校级特等奖、江苏省二等奖 | 三好学生一等奖学金 | 全国三维数字化创新设计大赛总决赛一等奖; 全国大学生数学竞赛二等奖; 等 10 余项国省级竞赛荣誉