

AI算法

业务判断/算法实现/部署优化

个人信息

姓名：周干恭

生日：2002.4.24

电话：13537775065

性别：男

户籍：深圳

邮箱：2021252439@qq.com

个人主页：<https://goldenfishes.github.io/zhouqiangong/>

教育经历

北京理工大学珠海学院

数据科学与大数据技术

理学学士学位

2020.9 - 2024.6

工作经历

深圳小铁信息技术

AI算法

2024.1 - 至今

Multi-Agent System项目：以步骤划分的多智能体协作框架 25.3-25.8

实现多智能体协同框架（MAS）的其中一种构型。重新定义MAS内最小执行单元，通过Agent自主决策使用这些单元形成不同的工作模式。以此，我们实现了前所未有的决策自由度，无需为每个任务手动实现不同Agent的工作逻辑。与此同时，我们构建了包含 Task、Stage、Agent、Step 四个层级的状态架构，从任务和执行两个视角协同约束系统行为，从而实现拓扑结构优化与进度可控的统一。

技术报告：<http://arxiv.org/abs/2508.11294>；开源地址：<https://github.com/motern88/Allen>

Det-SAM2项目：检测模型自动给予SAM2提示修正的长视频分割框架 24.8-24.11

该项目基于SAM2实现了一个实时推理视频流并由检测模型自动提供条件提示的长视频分割pipeline，是第一个功能完备的长视频自提示SAM2框架。其支持推理过程中在线添加新类别、预加载记忆库等新功能。同时，我们实现pipeline中恒定的显存与内存开销，从而支持一次性推理无限长的视频。

技术报告：<https://arxiv.org/abs/2411.18977>；开源地址：<https://github.com/motern88/Det-SAM2>

Robot Action项目：机器人动作生成 24.6-24.11

首先对Aloha机械臂上ACT算法相关参数调试，随后在该机械臂上移植DiffusionPolicy算法进行验证和探索，最后在Aloha机械臂上部署RDT-170m。积累机器人领域的初步技术储备。

基于瑞尔曼机械臂构建自己的机器人（可移动底座、双臂、头部可动摄像头）。

HiLight一期项目：视频语言模型（VLM） 24.1-24.7

我们首先实现一个更细粒度模态对齐的VideoEncoder：在CLIP-ViP基础上引入Patch与Token之间的对比学习损失。其次使用改进后的CLIP-ViP和Long-CLIP组成视频-图像双塔视觉提取模块，经过特征融合后作为Gemma-2B语言模型的视觉输入，实现基于视频理解的对话能力。

其中从0-1积累相关技术储备包括但不限于：预训练对比学习Loss改进，模型特征可视化监控，大规模多模态数据集加载，DDP和DeepSpeed单机多卡分布式训练，多模态语言模型多轮对话推理，lora微调等。

技术报告：<https://arxiv.org/abs/2407.0735>；开源地址：<https://github.com/motern88/HiLight>

其他业务项目：

小铁台球智能大师：使用检测与双摄像头追踪算法支持"AI教练"与"AI裁判"子功能，并实现AI教练中摄像头与投影仪的自动校准功能。

数能觉醒（深圳）科技

AI研究员（实习）

2023.12 - 2023.12

Text2SQL项目：电商数据分析AI

面向亚马逊电商的数据管理平台。负责该Text2SQL的LLM-Agent的数据制备，利用LLM自动扩展数据集与自动评估/修正数据组建数据生成引擎，重点关注如何让模型将已有知识形成稳定输出。

深圳竞控科技

深度学习算法工程师（实习）

2023.7 - 2023.8

竞赛机器人：视觉抓取小车

从零独立负责视觉模型在公司竞赛机器人设备上的部署。完成多个候选主流目标检测模型的效果验证和训练微调，并以NCNN/TFLite将模型部署在其安卓设备上。

其他论文

《FreeV : Free Lunch in MultiModal Diffusion U-ViT》 第一作者 23.10-23.12

本文揭示了U-ViT架构在扩散模型中未被挖掘的潜力，初步研究了U-ViT架构在多模态扩散模型的视觉生成任务中的贡献，并提出了一个针对U-ViT架构下的改进方案“FreeV”——首次将基于U-Net的FreeU改进框架应用在了Transformer架构中。该FreeV框架不需要额外的训练或微调就能够大大提高生成质量，此文的关键见解是在U-ViT中同时权衡来自骨干网络、跳跃连接和特征融合后的特征图的贡献，以充分利用U-ViT两个组件的优势的同时也能避开U-ViT在特征融合中的局限性。

DOI: 10.36227/techrxiv.24633840.v1 项目地址：<https://github.com/GoldenFishes/FreeV>

《Optimal Use of Attention Mechanisms: Comparative Study in U-Net for Image Segmentation Tasks》 第一作者 23.5-23.10

本文侧重于两种不同的经典注意力机制，并研究了它们在添加到U-Net进行图像分割任务时的实现方法和效果。得出结论，注意力机制虽然实现了加权通道信息，但它们可能对计算机视觉图像分割中传统卷积特征提取的骨干网络产生负面影响。同时，它们可以有效地弥补编码器-解码器模型结构与编码高维语义信息相关的缺点。

接受于ICCVDM。