Curriculum Vitae

Personal details

Surname / Family Name: Luo

First name(s):   Yaxin

Mobile Phone: +971 585699266 (UAE); +86 17882057622 (China)

Email address: Yaxin.Luo@mbzuai.ac.ae

Personal Website: https://yaxin9luo.github.io/

Google Scholar: https://scholar.google.com/citations?user=tEaSCzYAAAAJ&hl=en

---

Research Interests

My long-term goal is to develop intelligent machines capable of perceiving, understanding, and creating multimodal content, such as videos. This includes multimodal machine learning, vision foundation models, and also efficient algorithms for foundational models. Recently, I am focusing on MLLM Agent, Especially, Tool Integrated Reasoning Training MLLM using Reinforcement Learning to equip the model with reasoning and tool calling abilities to solve more complex and long trajectory tasks.

Education

MBZUAI

PhD of Machine Learning   Expected graduation date: December,2029

Advisor: Dr. Zhiqiang Shen

Technical University of Denmark   2800 Kongens Lyngby
Bachelor of General Engineering (specialized in Machine Learning)   Expected graduation date: 21/12/2024

Bachelor thesis advisor: Prof. Dimitrios Papadopoulos

University of Edinburgh

Bachelor of Mathematics (Overall grade of taken courses: UK first-class)

Start date: 21/09/2020   Date of withdrawing: 19/03/2021

Reason for withdrawing: change major and country

Working Experience

MBZUAI

Research Assistant   Jan 2025 – present
Conducting a project of analyzing LLM's generalization ability on pure vision tasks using only image data. Also, explore reasoning in MLLMs.

First Author Conference/Journal Publications

ICLR 2025 γ-MoD: Exploring Mixture-of-Depth Adaptation for Multimodal Large Language Models
γ-MOD is a novel approach to enhance computational efficiency in Multimodal Large Language Models (MLLMs) by incorporating Mixture-of-Depth (MoD) layers. This plug-and-play strategy seamlessly replaces redundant dense layers, significantly reducing computational costs while maintaining performance.

ECCV 2024 APL: Anchor-based Prompt Learning for One-stage Weakly Supervised Referring Expression Comprehension.
The project focuses on developing an Anchor-based Prompt Learning (APL) method for one-stage weakly supervised Referring Expression Comprehension (REC). The key innovation lies in the Anchor-based Prompt Encoder (APE), which generates and fuses prompts related to position, color, and category into anchor features. This improves the visual description power, enhancing the weakly supervised alignment between vision and language. Two auxiliary losses, the text reconstruction loss and visual alignment loss, further promote better vision language alignment. APL demonstrated state-of-the-art performance on multiple benchmarks, including RefCOCO and ReferIt, with significant improvements in both accuracy and efficiency compared to existing methods.

Other Experience

IEEE Cybermatics Congress 2024   Aug 2024
Conference Local Team Member
Acting as a conference helper and the session chair of the Smart Data workshop

SciSec2024   Aug 2024
Conference helper

Institute of Social Science Survey, Peking University
Summer Internship   July 2019 – Sep 2019

Internship in a public psychological healthcare project led by the Ministry of Civil Affairs, China.

Other Skills

- Good practical skills in HPC training(Slurm), DeepSpeed for LLMs, and some theoretical knowledge in Distributed Computing
- Experience in 3 Course Projects in Embedded System Programming and Circuits
- Signal Processing and Knowledge in Acoustics
- Some Basic Knowledge in Bioinformatics from DTU bachelor's biology courses