Siyu Li

✉ wufu@bupt.edu.cn · ☎ 189-5935-2187 ·

🎓 Education

Beijing University of Posts and Telecommunications, Beijing, China   2021 - Present
B.Eng. in Artificial Intelligence
Rank: **65 / 201**, Grade: 87.92, GPA: 3.63/4.0.

👥 Internship Experience

Beijing University of Posts and Telecommunications - Guoshun Nan - Research Intern	2023.7 - 2024.2
Tsinghua University - THUNLP, Zhiyuan Liu - Research Intern	2023.10 - 2024.7
Tiamat AI - Research Intern	2024.5 - Present

♡ Research Publication

ContextBLIP: Doubly Contextual Alignment for Contrastive Image Retrieval 2023.11 - 2024.5

- ACL 2024 Findings，Advisors: Guoshun Nan
- We have discovered that even the most advanced visual capability models, such as GPT-4V, struggle with the challenging task of Image Retrieval from Contextual Descriptions (IRCD) . In response to this, we proposed a Plug-and-Play adapter based on BLIP, which employs a dual-context alignment strategy to tackle the difficulties inherent in IRCD. Remarkably, our adapter surpasses GPT-4V in IRCD performance with a minimal number of parameters.
- As the **co-first author**, I completed the idea development, experiments, coding, and paper writing.

Trans4D: Realistic Geometry-Aware Transition for Compositional Text-to-4D Synthesis 2024.7 - 2024.10

- Under Review，Advisors: Jiaming Liu
- TRANS4D, a novel framework for text-to-4D synthesis that focuses on generating high-quality 4D scenes with realistic geometry-aware transitions and complex object deformations. It utilizes MLLMs for physics-aware scene planning and introduces a geometry-aware transition network, allowing for more accurate scene transitions compared to previous methods.
- As **the second student author** of this paper, I independently handled the pipeline for MLLM's decomposition of 4D data, while also learning and completing some experiments and code related to the 4D generation process.

MIO: A Foundation Model on Multimodal Tokens   2024.1 - 2024.5

- Under Review，Advisors: Jie Fu
- MIO is a foundation model designed to align inputs from different sources at the token level like GPT-4O.
- As a member of the authors, my primary work involved handling data from different sources. I also contributed to the code development and paper writing.

⚙ Research Experience

XAgent: An Autonomous Agent for Complex Task Solving   2023.10 - 2024.1

- XAgent is an open-source experimental Large Language Model (LLM) driven autonomous agent that can automatically solve various tasks. It is designed to be a general-purpose agent that can be applied to a wide range of tasks

- As a member of the XAgent Team, I actively participated in the code writing, debugging, and contributed to the part of ideas development.

Encoder Fusion for MLLM 2024.5 - Present

- Still in Experimental Stage.
- Similar to experiments with the MoE Vision Encoder, this paper explores the impact and enhancement on MLLM capabilities by using Cross Attention to connect different ViT models as Vision Encoders and investigating the effects of their fusion.
- As the **first author**, I completed the idea development, data process, experiments and coding.

MLLM in Diffusion Model 2024.5 - Present

- Still in Experimental Stage. Advisors: Jiaming Liu
- Exploring the use of MLLM as the text encoder for text-to-image diffusion models to enhance prompt-following image generation. This experiment is complete, but due to similarities with other work, we continue to explore the downstream task about MLLM in Diffusion Model.
- As the **first author**, I completed the idea development, data process, experiments and coding.