Huanyu WANG

Email: whyisverysmart@sjtu.edu.cn Tel: (86) 176-1218-9497 GitHub: https://whyisverysmart.github.io

Education

Shanghai Jiao Tong University (SJTU), Shanghai, China   Sept. 2022 - Jun. 2026 (Expected)

Bachelor of Computer Science, Double Major in Applied Mathematics

GPA: 90.2 / 100 (Overall), **91.1 / 100** (Core)

Courses: Game Theory (97), Operating System (95), Programming Languages and Compilers (95), Data Structure (95), Principles and Methods of Program Design (95), Design and Analysis of Algorithms (95), Program Design Practice (94.5), Machine Learning (93), Computer Vision (93), Computer Networks (93).

Publication

- Fourier-LLaVA: Compressing Vision Tokens in the Frequency Domain for Large Vision-Language Models
  Huanyu Wang, Jushi Kai, Haoli Bai, Bo Jiang, Zhouhan Lin
- LoMo: Longer and More Videos Benchmark for Understanding and Temporal Grounding Tasks
  Chengyang Hu, Xinyu Zhou, **Huanyu Wang**, Danyu Shen, Ran Yi, Mengtian Li, Lizhuang Ma

Research Experiences

SJTU, LUMIA Lab (Language Understanding and Machine Intelligence Algorithms)   Oct. 2024 - Present

Research Assistant, Advisor: Prof. **Zhouhan Lin**

- Addressed the challenge of computational redundancy in Large Vision-Language Models (LVLMs), significantly improving the efficiency and accuracy of vision-language data processing.
- Developed an innovative approach by integrating the **2-dimentional Discrete Cosine Transform** with **LLaVA** and **Qwen** architectures, achieving a remarkable **93.75%** reduction in vision token redundancy.
- Utilized the LCS-558K dataset for further pretraining, optimizing feature alignment for better model performance.
- Conducted comprehensive evaluations across multiple benchmarks, delivering state-of-the-art results while reducing Floating Point Operations (FLOPs) by an impressive **6.18x** and improving inference speed by **31.2%**.

SJTU, DMCV Lab (Digital Media & Computer Vision Laboratory)   Sept. 2024 - Jan. 2025

Research Assistant, Advisor: Prof. **Lizhuang Ma**

- Designed a large-scale long-video dataset to evaluate the temporal reasoning capabilities of existing models, comprising **14,047+** videos ranging from **1 minute** to **4 hours**, and covering **6 distinct tasks**.
- Developed a high-efficiency automatic annotation pipeline leveraging large language models, enabling scalable and cost-effective labeling on only **2 Nvidia RTX 4090 GPUs**.
- Conducted extensive evaluations of open-source Vision-Language Models on the benchmark, providing insights and recommendations for improving long-video understanding capabilities.

Work Experiences

Gauss Lab, Huawei   Jul. 2024 - Sept. 2024

Software Development Intern, Advisor: Jinming Liao

- Collaborated with the Standby-Read team on the development of the **Gauss database kernel**.
- Designed and implemented novel C++ methods to address issues related to the database kernel state being unreadable, enhancing debugging and development efficiency by **60%**.
- Elevated functions to system-level and tested their effectiveness utilizing custom SQL scripts on a database cluster.

Skills

Computer: Python, PyTorch, Transformers, TensorFlow, Hugging Face, C++, LaTeX.

Model: LLaVA, Qwen, CLIP, GPT, BERT, BLIP, LLaMA, VILA, mPlugOwl3.

Language: English - TOEFL: **107** (Speaking 24)