

Shu Pu

pushuabc@gmail.com — Homepage — Scholar — Github — Last Update: Nov 1 2024

EDUCATION

Sep 2022- Jun 2026 (Expected) Huazhong University of Science and Technology Computer Science and Technology
Grade: 4.14/5, Rank: 12/35
Degree: Bachelor of Science
Major: Excellent engineer in Computer Science and Technology

In Submission

- **ICLR 2025** in Submission, Dongping Chen *, Ruoxi Chen *, **Shu Pu** *, ..., Ranjay Krishna, Yao Wan. ***ISG-Bench: Interleaved Scene Graph for Text-to-Image Generation Assessment*** [PDF]
- **CVPR 2025** in Submission, Haojie Zheng *, Tianyang Xu *, Hanchi Sun, **Shu Pu**, Ruoxi Chen, Lichao Sun. ***Thinking Before Looking: Is Visual Element the Core of Multimodal Reasoning?*** [PDF]

Future Plan

- *MLLM as a judge++* [Proposal], aiming on ICML
- *Swin Transformer Controll Auto-Regressive Image Generation* [Proposal], aiming on IJCAI
- *Image-free Training Method for LVLM* [Proposal], aiming on ICML
- *Enabling Generating anything ability of MLLMs*, aiming on ACL.

Experience

Research Internship in HUST, Supervised by Professor Yao Wan

Part-time, Research Intern

June 2024 – Present

- Researching on Unified generation models and their interleaved generation capabilities. We propose a benchmark **ISG-Bench** focusing on evaluating model's interleaved image-text generation capability and propose a brand-new evaluation metrics for interleaved context using scene graph.

Online Research Internship in Lehigh University, Supervised by Professor Lichao Sun

Part-time, Research Intern

June 2024 – Sep 2024

- Researching on MLLM-based Multi-Agents frameworks. We propose a framework to reduce the hallucination and enhance the reasoning ability of Multimodal Large Language Models. The **Thinking Before Looking** will be released after submission to CVPR.

Skills

Language: Mandarin (native), English (TOFEL 102).

Research Abilities: Proficient in coding and programming using python; Knowledge and experience in Multi-modal Perception and information retrieval, Text2Image Generation, LLM-based Agents and prompt engineering.

Interesting Field: Besides all the research I did, I'm also a motivated Computer Graphics learner, especially in 3D AIGC. I believe it is much more worthwhile to integrate AI into our daily use.

Others

Research is not all I need.

Hobbies: Football, R&B.