Leizhi Li
Shanghai, China — lileizhi@sjtu.edu.cn — (+86)13995621901

EDUCATION

Shanghai Jiaotong University, Shanghai, China — Sep. 2021 – Present
Bachelor of Computer Science, ACM Honor Class — Overall GPA: 3.56 — Major GPA 3.52

- Optimization — A
- Mathematical Analysis — 4.0/4.3
- Computer System Comprehensive Design — 4.3/4.3

RESEARCH INTEREST

I am broadly engaging with cutting-edge machine learning theories and technologies. My areas of focus include multi-modal generative models, guided diffusion, and vision-language foundation models. Additionally, I am interested in the pertinent stochastic explanations and mathematical principles behind elegant machine learning.

PROJECTS

LOSe *Project Link:* `https://github.com/TristoneLee/lOSe`

- LOSe is a Linux-like operating system designed and implemented on the RISC-V platform. LOSe supports Buddy memory allocation module, use/kernel space isolation, process scheduling, user mode I/O and applications, and interrupt handling. Lose can be run on QEMU, providing a command line interface.

GiFS *Project Link:* `https://github.com/TristoneLee/GiFS`

- GiFS is an independent implementation of the Google File System(GFS) which has been widely deployed in industry and inspired future studies in distributed file system. GiFS supports high-level concurrent file IO with weak consistency guarantee under diverse circumstances and fast and robust recovery confronting disastrous server failure.

Mx* Compiler *Project Link:* `https://github.com/TristoneLee/Compiler`

- This is an assignment of SJTU ACM class, a compiler for a language called Mx* into assembly code in RISC-V 32IMA. This compiler contains a full-stack procedure of compiling, including semantic parsing, building abstract syntax tree, IR representation, register allocation, and assembly code generation.

CURRENT WORK

Editable Human Motion Generation
As an extension of my project of the Computer Vision course lectured by Professor Cewu Lu, I am currently engaged in a project on editable human motion generation. By introducing explainable human motion representation, we aim to enable motion latent diffusion to generate motions based on human feedback and real-world requirements. This project allowed me to work with up-to-date generative models and accumulate a deeper comprehension of the ideas of VAE and diffusion through practice.

CLIP Driven Zero-shot Semantic Segmentation
Inspired by GroupViT, I have been exploring combining the grouping technique with vision-language foundation models like CLIP. As CLIP is trained on matching image-text pairs, its usage can be limited by its poor ability to process partial image tokens. By quantizing image tokens encoded by CLIP, we gain a more compact and structured representation of vision information. We hope this intrinsic property of discrete representation can enable better performance in zero-shot settings. I want to exploit these quantized representations further using graph techniques in data processing fields to extract deeper information that we can learn from CLIP and future foundation models for downstream tasks.