Shuai Liu

SHUAI005@e.ntu.edu.sg · (+86) 185-1185-5066 · https://github.com/choiszt

Education

Nanyang Technological University (NTU), Singapore 2024 – 2026 (Expected)
Master of Engineering (by Research) in Computer Science
MMLab, S-Lab@NTU, *Advised by Prof. Ziwei Liu*

Beijing University of Posts and Telecommunications (BUPT), Beijing, China 2020 – 2024
Bachelor in Artificial Intelligence
Average Mark: 90; **GPA:** 3.75 / 4.0

Publications

[1] Jingkang Yang, Jun CEN, Wenxuan Peng, **Shuai Liu**, Fangzhou Hong, Xiangtai Li, Kaiyang Zhou, Qifeng Chen, Ziwei Liu. 4D Panoptic Scene Graph Generation *(NeurIPS, 2023 Spotlight).*

[2] Jingkang Yang*, Yuhao Dong*, **Shuai Liu***, Bo Li*, Ziyue Wang, Jiamu Kang, ChenCheng Jiang, Haoran Tan, Yuanhan Zhang, Kaiyang Zhou, Ziwei Liu. Octopus: Embodied vision-language programmer from environmental feedback *(ECCV, 2024) (*equal contributions, 240+ stars on Github)*

[3] Kaichen Zhang, Bo Li, Peiyuan Zhang, Fanyi Pu, Joshua Adrian Cahyono, Kairui Hu, **Shuai Liu**, Yuanhan Zhang, Chunyuan Li, Ziwei Liu. LMMs-Eval: Reality Check on the Evaluation of Large Multimodal Models *(NeurIPS, 2024, Under Review)*

[4] Kairui Hu*, Fanyi Pu*, Kaichen Zhang*, **Shuai Liu***, Yuanhan Zhang* Bo Li*;†, Peiyuan Zhang, Ziwei Liu. Embracing Video Evaluations With LMMs-Eval *(Blog, 2024)*

Internship Experience

Shanghai AI Laboratory Aug. 2023 - July. 2024

Trainee Researcher. Advisor: Prof. Ziwei Liu

Focused on visual perception and reasoning in the open world, specifically in the areas of multimodality, scene graph generation and Embodied AI.

Visual Generalist Models

- Developed OctoGibson simulator based upon Omniverse and pipelines for VLM-powered embodied agent. Worked on enhancing VLMs' programming capabilities by designing high-level tasks and simulation feedback for RLHF.
- Provided mentorship in developing GTA-based environment using C++ and C# for task scripting and MineCraft-based environment.
- Explored the potential of scene graph-enhanced embodied task planner.
- Contributed new models for image and video evaluation in LMMs-Eval: Large Multimodal Model Evaluations.

4D Scene Graph Generation

- Took the lead in the development of the GTA-V environment and proposed a C++ framework for RGB-D sequence alignment using Script-Hook V.
- Contributed to expanding the SAIL-VOS 3D datasets and introduced the 'Segment-Any-RGBD' toolbox for panoptic annotations based on Segment-Anything.
- Conducted 3DSGG and PSG4DFormer baseline on Aria Digital Twin Dataset.

HAOMO.AI Nov. 2022 - Apr. 2023

Research Intern. Advisor: Prof. Yang Yang and Dr. Dong Cao

Focused on Autonomous Driving and BEV Perception.

3D Object Detection in Large-Scale Real-Road Scenario

- Trained and employed lightweight BEV Model with large-scale jointly labeled dataset.
- Conducted a survey of multisensor and multimodality alignment methods.

Multi-Sensor BEV Detection

- Developed novel BEV fusion schemes utilizing inverse transformation from BEV to 2D coordinates and ROI alignment to address feature loss from the lift-splat process in multi-view perception.
- Conducted a comprehensive Bird's Eye View (BEV) benchmark using data from various sources of multi-sensor datasets

Research Experience

Berkeley Artificial Intelligence Research (BAIR) Lab @ UC Berkeley July. 2024 - Present

Student Research Assistant. Adviser: Dr. Yutong Bai

MMLab, S-Lab @ Nanyang Technological University Apr. 2023 - Present

Student Research Assistant. Advisor: Prof. Ziwei Liu

National Laboratory of Pattern Recognition (NLPR) Chinese Academy of Sciences (CASIA) June 2022 - Nov. 2022

Student Research Assistant. Advisor: Prof. Yang Yang and Prof. Zhen Lei (IEEE Fellow)

Laboratory of Pattern Recognition and Intelligent Systems @ Beijing University of Posts and Telecommunications Sep. 2021 - June 2022

Student Research Assistant. Advisor: Prof. Ming Wu

- Processed pixel-level satellite remote-sensing road extraction images.
- Proposed COSIS, a novel weakly-segmentation based training pipeline to improve satellite image segmentation by leveraging collaborative voting, semantic image synthesis and consistency learning mechanisms.
- Proposed two by-produced weakly-supervised benchmark dataset for road extraction.

Project Experience

Accelerating the development of large multimodal models (LMMs) with lmms-eval
https://github.com/EvolvingLMMs-Lab/lmms-eval

- Contribute models for evaluation framework meticulously crafted for consistent and efficient evaluation of LMM.

SEDI: Segmentation-based Road Extraction via Weakly-Supervised Data
https://github.com/choiszt/spade_dlinknet

- Proposed a novel processing pipeline for training weakly-supervised segmentation dataset.
- Ranked 10% in DEEPGLOBE CVPR 2018 - Satellite Challenge

Large-Language Model in solving math problem
https://github.com/choiszt/all_project/tree/master/Professional_Practice

- Proposed supervised Fine-tuning and LoRA-tuning methods by leveraging the Aquila-chat 7b language model on the GSM8K dataset.

Developing a basic operating system kernel
https://github.com/choiszt/all_project/tree/master/Operating_System

- Took charge of writing the timers module and standardization, developing bandwidth monitor and implementing dynamic log file output.

Scene Graph Enhanced Embodied Task Planning with Large Language Models
https://github.com/choiszt/SG_VLM

- Explore the effect of scene graph in embodied task planning.

Self-distilled semantic segmentation for TGS Salt Identification Challenge
https://github.com/choiszt/all_project/tree/master/Pattern_Recognition_and_Machine_Learning

- Proposed a data-level self-distilled Unet for TGS Salt Identification.

Vision-Audio based bird calls identification *BirdCLEF23*
https://github.com/choiszt/all_project/tree/master/Experiments_on_Intelligent_Information_Network

- Took the lead in audio and Mel Spectrogram fusion for bird call identification based on the VGGish model.

Sentiment Analysis based Restaurant Rating System
https://github.com/choiszt/all_project/tree/master/Course_Project_NLP

- Took the lead in structure design and closed data loops.
- Implemented ChatGLM fine-tuning and model distillation.

Neural Network-Based Congestion Control Algorithm
https://github.com/choiszt/all_project/tree/master/Computer_Networking

- Implemented simulating WebRTC transmission in a simulator and used a Multilayer Perceptron (MLP) to predict bitrates for real-time audio-video communication (RTC).

A Survey: Multimodality in Embodied AI
https://github.com/choiszt/all_project/tree/master/multi_modal_processing

- Conducted a survey on the applications of Multimodality and Large Model in Embodied AI.