

Tong Zhang

✉ tongz27@uci.edu

☎ +86 18871162559

🌐 [the-star-sea](https://github.com/the-star-sea)

🌐 stonezhang.com

Education Background

University of California Irvine (UCI)

MS, Computer Engineering, GPA: 3.71/4

Irvine, USA

June 2023 - Sep. 2024

Southern University of Science and Technology (SUSTech)

Bachelor of Computer Science and Engineering, GPA: 3.65/4

Shenzhen, China

Sep. 2019 - June 2023

Research Interest

- **World Models for GUI Agents:** Modeling latent dynamics and planning strategies for software environments, enabling agents to simulate and manipulate interactive UI states through reasoning.
- **Controllable Image Generation with Visual Reasoning:** Designing energy-based and language-guided frameworks for image editing, incorporating semantic priors (e.g., SVGs, semantic region masks) to support multi-step, interpretable visual transformations.

Academic Experience

Human-Readable SVG Generation with Vision Language Models [PDF](#)

Assistant Prof. Haohan Wang

PyTorch

UIUC

Sep. 2023 - May 2024

- proposed S^2VG^2 , the first method combined with a vision language model for SVG generation
- introduced a specialized dataset named SVG-SHAPE, designed for evaluating SVG generation methods and reasoning of LLMs
- demonstrated state-of-the-art performance in SVG reasoning of LLMs and vision metrics

One-shot Controllable Head Avatar with Vertex-feature Transformer [PDF](#)

Prof. Xiaohui Xie

PyTorch

UCI

Apr. 2023 - June 2023

- proposed CVTHead, a one-shot controllable head avatar framework
- evaluated our method in comparison to other methods for cross-identity reenactment
- demonstrated state-of-the-art performance on VoxCeleb1 and VoxCeleb2

Trajectory Prediction and Driving Video Caption [PDF](#)

Assistant Prof. Hao Zhao

NumPy, PyTorch

AIR, Tsinghua University

May 2022 - Sep. 2022

- trained a novel end-to-end transformer generating descriptions and explanations of driving videos
- demonstrated state-of-the-art performance in driving video captioning

Professional Experience

Research Intern

vivo AI Research

PyTorch, LaTeX

Shenzhen, China

May 2025 - Present

- Designed an energy-based refinement algorithm for prompt-guided heatmaps, enabling high-fidelity and spatially coherent image edits.
- Proposed joint training of planner and world model in latent space for multi-round controllable editing via self-reflective reasoning.
- Explored VLM + CoT frameworks where language models iteratively generate, reflect, and revise

based on visual feedback from image generators.

Natural Language Processing Engineer [PDF](#)

Prof. Haizhou Li

PyTorch

- conducted a comprehensive analysis of 11 defensive mechanisms applied to 6 LLMs, evaluating their impact on model performance, over-refusal, and token overhead.
- proposed the 9 **meta-defenders** to systematically analyze the side effects of defense mechanisms, providing insight into the trade-offs between model safety and performance.
- implemented and evaluated token compression on audio language models

Lightweight OCR Models Support for OpenCV [Report](#)

OpenCV

PyTorch, ONNX, C++

- implemented the detection part of PP-OCRv2 model in OpenCV Zoo by ONNX
- implemented high level C++ API of PP-OCRv2 model in OpenCV
- implemented evaluation metrics of text detection (AP, Recall, Precision, Hmean) in OpenCV Zoo

Teaching Assistant for Introduction to Java Programming

Associate Prof. Yu Zhang

English, Java

- designed and graded a significant portion of the coursework, including assignments and projects.
- developed and managed an online judging platform for evaluating student code submissions

CUHK-Shenzhen

Nov. 2024 - April 2025

Google Summer of Code 2022

May 2022 - Sep. 2022

SUSTech

Mar. 2022 - June 2022

Publications

Tong Zhang, Yiming Chen, Simin Chen, Zexin Li, Xianghu Yue, Cong Liu, Wei Yang, Haizhou Li and Tao Xie, "Unintended Side Effects of Defense Mechanisms in Large Language Models: A Comprehensive Study", *Under Review*, 2025

Tong Zhang, Haoyang Liu, Peiyan Zhang, Yuxuan Cheng and Haohan Wang, "Beyond Pixels: Exploring Human-Readable SVG Generation for Simple Images with Vision Language Models", *Preprint*, 2024

Haoyu Ma, **Tong Zhang**, Shanlin Sun, Xiangyi Yan, Kun Han and Xiaohui Xie, "CVTHead: One-shot Controllable Head Avatar with Vertex-feature Transformer", *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024

Bu Jin, Xinyu Liu, Yupeng Zheng, Pengfei Li, Hao Zhao, **Tong Zhang**, Yuhang Zheng, Guyue Zhou and Jingjing Liu, "ADAPT: Action-aware Driving Caption Transformer", *IEEE International Conference on Robotics and Automation (ICRA)*, 2023