Muzi **Tao**

3650 McClintock Ave, Los Angeles, CA 90089

muzitao@usc.edu | github.com/muzi-tao | linkedin.com/in/muzi-tao | @tao_muzi

Education

University of Southern California — *Los Angeles, CA*

Ph.D. in Computer Science — *Aug 2024 - Now*

- Advisor: Prof. Xuezhe Ma
- GPA: 4.0
- Interests: Multimodal Large Language Model, Generative Model

Shanghai Jiao Tong University — *Shanghai, China*

B.E. in Computer Science — *Sept 2020 - Jun 2024*

- Member of **ACM Honors Class**, which is an elite CS program for top 1% talented students.

Interests

My major research interest lies in Computer Vision and Multimodality. Currently, I am working on the detailed visual processing capability of Multimodal LLMs and Multimodal Generation.

Research Experience

MLLM research for evaluation on low-level vision — *New York University*

Research intern advised by Prof. Saining Xie — *Apr. 2023 - Dec. 2023*

- Offered a vision-centric environment for evaluating multimodal systems. *(See in Selected Projects ArtQA for details)*

Research on image harmonization — *BCMI, Shanghai Jiao Tong University*

Undergraduate Researcher advised by Prof. Li Niu — *Jul. 2022 - Mar. 2023*

- Augment data for the task Image Harmonization by addressing a critical gap in the dataset iHarmony4 through the application of non-global color transformations.

Selected Projects

On the Idiosyncratic Gap between Image Captioning and Generation Models — *In Submission, 2025*

Multimodal Large Large Model, Idiosyncrasy, Image Captioning, Image Generation

- We propose attribution-as-evaluation as a complementary metric for prompt-following: instruction following should increase transfer of caption idiosyncrasies into images, narrowing the gap.

What Does a Visual Formal Analysis of the World's 500 Most Famous Paintings Tell Us About Multimodal LLMs? — ***ICLR 2024*** ***invite to present(notable)***

Multimodal Large Large Model, Detailed Visual Processing, Low-level Vision — ***tinypaper***

- Leverage the extensive knowledge base of a large language model and human-filtering as the pipeline of the high-quality benchmark to focus on low-level vision features.
- Most models exhibit gaps between detailed vision processing and high-level semantic understanding, highlighting significant potential for improvement.

DiffAnnot: Improved Neural Annotator with Denoising Diffusion Model — *Fall 2022*

3D Human body reconstruction, Diffusion Model (ICIPMC 2023) — *Python, PyTorch*

- Presented DiffAnnot, a new neural annotator applying denoising diffusion probabilistic models on existing 3D human reconstruction models.

Awards and Honors

SJTU	**Zhiyuan Honorary Scholarship,** Top 2% in SJTU each year	*2020, 2021, 2022, 2023*
SJTU	**Academic Excellence Scholarship,** For Outstanding Undergraduates	*2021, 2022, 2023*
SJTU	**Excellent Undergraduate Graduate Award,** in SJTU	*2024*

Skills

Programming Python, C/C++, JAVA, Rust, Verilog, MATLAB, Javascript.

Tools and Frameworks Linux, Shell, Git, LaTeX, Markdown, PyTorch, TensorFlow, Vivado, HTML.