WENTAO HU

Google Scholar

Phone: (+86) 18980399950 ⋄ Email: wentao002@e.ntu.edu.sg

EDUCATION

City University of Hong kong (QS 63)
Incoming Ph.D Student
Sep. 2026 – Jun. 2030 (expected)
Advisor: Prof. Xiangyu Zhao

Nanyang Technological University (QS 15)
M.Eng in Computer Science and Engineering (By Research)
GPA: 3.83/5.0
Aug. 2023 – Jan. 2026 (expected)
Advisor: Prof. Hanwang Zhang

Hunan University (Project 985)
B.Sc in Statistics
GPA: 3.55/4.0
Sep. 2019 – June. 2023

RESEARCH INTERESTS

Multimodal Large Language Model, AI + X

PUBLICATIONS

[1] **Wentao Hu***, Wang Lin*, Liyu Jia, Kaihang Pan, Zhang Majun, Zhou Zhao, Fei Wu, Jingyuan Chen, Hanwang Zhang
Vinci: Deep Thinking in Text-to-Image Generation using Unified Model with Reinforcement Learning
NeurIPS 2025 · (Co-first authors)

[2] Wang Lin, Feng Wang, Majun Zhang, **Wentao Hu**, Tao Jin, Zhou Zhao, Fei Wu, Jingyuan Chen, Sucheng Ren, Alan Yuille
WorldEdit: Towards Open-World Image Editing with a Knowledge-Informed Benchmark
ICLR 2026

[3] Wang Lin*, Liyu Jia*, **Wentao Hu***, Kaihang Pan, Zhongqi Yue, Wei Zhao, Jingyuan Chen, Fei Wu, Hanwang Zhang
Reasoning Physical Video Generation with Diffusion Timestep Tokens via Reinforcement Learning
Under Review (2025) · (Co-first authors)

[4] Bohan Wang*, Mingze Zhou*, Zhongqi Yue*, Wang Lin, Kaihang Pan, Liyu Jia, **Wentao Hu**, Wei Zhao, Hanwang Zhang
Selftok: Discrete Visual Tokens of Autoregression, by Diffusion, and for Reasoning
NeurIPS 2025

[5] Hao Fei*, Yuan Zhou*, Juncheng Li*, ..., **Wentao Hu**, Jiebo Luo, Tat-Seng Chua, Shuicheng Yan, Hanwang Zhang
On Path to Multimodal Generalist: Levels and Benchmarks
ICML 2025

RESEARCH EXPERIENCE

Huawei Singpore Graphics & Image Intelligence Lab
Research Intern
Supervisor: *Prof. Hanwang Zhang*
Aug. 2024 - May. 2025
Singapore

Selftok: Discrete Visual Tokens of Autoregression, by Diffusion, and for Reasoning
- **Background**: Existing visual tokenizers rely heavily on spatial priors or continuous representations, which limit compatibility with reinforcement learning and hinder seamless integration with LLMs.
- **Method**: Developed Selftok, a discrete visual tokenizer that leverages the reverse diffusion process to generate discrete visual tokens.
- **Results**: MLLMs built on Selftok achieve state-of-the-art visual comprehension and generation performance, exhibiting strong generalization and reinforcement learning compatibility.

- **My Involvement**: Developed the algorithmic framework for Selftok, including a unified NPU/GPU inference pipeline.

National University of Singapore   Apr. 2024 - Dec. 2024
Research Intern   Singapore

On Path to Multimodal Generalist: General-Level and General-Bench
- **Background**: Existing evaluation methods for MLLMs assume higher task scores imply stronger multimodal capabilities, which fails to capture the generality required for MLLMs.
- **Method**: Proposed General-Level, a framework with 5 levels that evaluates both performance and cross-modal synergy.
- **Results**: Introduced General-Bench, one of the largest multimodal benchmarks to date, covering 700+ tasks, 325k+ instances, and 100+ state-of-the-art MLLMs across diverse skills and modalities.
- **My Involvement**: Curated 40 image comprehension datasets and conducted systematic evaluations several leading MLLMs. Performed analysis using the proposed taxonomy, assigning each model to an appropriate generality level based on empirical evidence.

ACHIEVEMENTS

CUMCM, Provincial Third Prize   *Summer 2021*

SKILLS

Language	Mandarin(native), English(IELTS: 6.5, GRE: 333)
Programming Languages	Python, MATLAB, R, C++