Jinghan Zhang

+852-95314166 | jzhangjv@connect.ust.hk

Education

The Hong Kong University of Science and Technology — Hong Kong
Doctor of Philosophy in Computer Science and Engineering, adviced by Prof. Junxian He — *Jan. 2024 – now*

Southeast University — Jiangsu, China
Bachelor of Engineering, Artificial Intelligence, GPA:3.87/4.0 — *Sept. 2019 – Jul. 2023*

Research Interests

• VLM reasoning, Long context modeling, Model merging

Publications

* indicates equal contribution.

• Every Second Counts: Unlocking Long-Video Context in VLMs with Hour-long Data
J. Zhang, C. Li, T. Zhu, Y. Gao, Y. Deng, S. Chen, C. Ma, J. He, *submitted to ICCV 2025*
• Bring Reason to Vision: Will Model Merging Work for VLMs?
S. Chen*, **J. Zhang***, T. Zhu, W. Liu, S. Gao, M. Xiong, M. Li, J. He, *submitted to ICML 2025*
• Why Is Spatial Reasoning Hard for VLMs? An Attention Mechanism Perspective on Focus Areas
S. Chen, T. Zhu, R. Zhou, **J. Zhang**, S, Gao, JC. Niebles, M. Geva, J. He, J. Wu, M. Li, *submitted to ICML 2025*
• Compression Reflects Intelligence Linearly
Y. Huang*, **J. Zhang***, Z. Shan, J. He, *accepted by COLM 2024*
• Composing Parameter-Efficient Modules with Arithmetic Operations
J. Zhang, S. Chen, J. Liu, J. He, *accepted to NeurIPS 2023*
• C-Eval: A Multi-Level Multi-Discipline Chinese Evaluation Suite for Models
Y. Huang, Y. Bai, Z. Zhu, J. Zhang, **J. Zhang** et al., J. He, *accepted to NeurIPS 2023 D&B track*
• FELM: Benchmarking Factuality Evaluation of Large Language Models
S. Chen, Y. Zhao, **J. Zhang**, I-Chun Chern, S. Gao, P. Liu, J. He, *accepted to NeurIPS 2023 D&B track*

Research Projects

Every Second Counts: Unlocking Long-Video Context in VLMs with Hour-long Data

- We tackle the challenge that current VLMs struggle to capture transient moments and long-range dependencies in long video understanding.
- We introduce an hour-long video dataset consisting of synthetic video-description pairs generated solely from text.
- We extend the context window of short-context VLMs like LLaVA-OneVision using our dataset, resulting in a 6% improvement on Video-MME and other benchmarks.

Bring Reason to Vision: Will Model Merging Work for VLMs?

- We aim to bring the reasoning ability in LLMs into VLMs by merging.
- We further understand the inner workings of VLMs by adapting merging as a tool - we propose Masking-out the parameters to another model (e.g: before and after merging) to locate certain abilities.
- We find that CoT reasoning ability could be transfered via merging , and perception and reasoning ability could be decomposed in parameter space.

Compression Represents Intelligence Linearly

- Driven by the intuition of "compression leads to beyond human intelligence", we aim to quatify the concept of "compression" and "model intelligence" and explore their correlation.
- We find a nearly linear correlation between LLMs' ability to compress external text corpora and their intelligence measured by average benchmark scores.
- Our findings are consistent both across the three individual domains—knowledge, coding, and mathematical reasoning—and in their overall average, suggesting that compression efficiency is a reliable metric for evaluating LLMs.

Composing Parameter-Efficient Modules with Arithmetic Operations

- To reuse trained PEFT modules like LoRA efficiently for new tasks, we view them as plug-and-play components, and design arithmetic operations to implement PEFT merging.
- We perform various arithmetic operations to combine them for executing new tasks: negation for forgetting, addition for generalization and multitasking, and analogy for domain transfer.