

# YUNHAO GOU

✉ [ygou@connect.ust.hk](mailto:ygou@connect.ust.hk)

HKUST, Clear Water Bay, New Territories, Hong Kong, P.R.China

## EDUCATION

---

**Hong Kong University of Science and Technology**

Sep. 2022 – Jun. 2027 (Expected)

Ph.D. in Computer Science and Engineering

Advisor: [Prof. James Kwok](#) and [Prof. Yu Zhang](#) (co-supervision)

**National University of Singapore**

July. 2019 – Aug. 2019

Exchange student Overall grade: A

School of Computing (SoC)

**University of Electronic Science and Technology of China**

Sep. 2018 – Jun. 2022

B.S. in Software Engineering

GPA: 3.95/4.0 Average: 90.58

Advisor: [Prof. JingJing Li](#)

## RESEARCH INTERESTS

---

I'm interested in **Multimodal Large Language Models (MLLM)**, particularly in its **post-training strategies** with my recent work being [ECSO](#), [MoTE](#), [EMOVA](#) and [Corrupted but not broken](#).

I'm also passionate about **novel paradigms** for MLLMs. Relevant work are [MoCLE](#) and [RACRO](#).

## PUBLICATION

---

- Perceptual Decoupling for Scalable Multi-modal Reasoning via Reward-Optimized Captioning.  
**Yunhao Gou\***, Kai Chen\*, Zhili Liu\*, Lanqing Hong, Xin Jin, Zhenguo Li, James T. Kwok, Yu Zhang  
*Arxiv* [\[link\]](#)
- Mixture of insightful Experts (MoTE): The Synergy of Thought Chains and Expert Mixtures in Self-Alignment.  
Zhili Liu\*, **Yunhao Gou\***, Kai Chen\*, Lanqing Hong, Jiahui Gao, Fei Mi, Yu Zhang, Zhenguo Li, Xin Jiang, Qun Liu, James T. Kwok.  
*Annual Meeting of the Association for Computational Linguistics (ACL)*, 2025. [\[link\]](#)
- Corrupted but Not Broken: Rethinking the Impact of Corrupted Data in Visual Instruction Tuning.  
**Yunhao Gou\***, Hansi Yang\*, Zhili Liu, Kai Chen, Yihan Zeng, Lanqing Hong, Zhenguo Li, Qun Liu, James T Kwok, Yu Zhang..  
*Arxiv* [\[link\]](#)
- EMOVA: Empowering Language Models to See, Hear and Speak with Vivid Emotions.  
Kai Chen\*, **Yunhao Gou\***, Runhui Huang\*, Zhili Liu\*, Daxin Tan\* and other 26 authors.  
*IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2025 [\[link\]](#)
- ECSO: Eyes closed, safety on: Protecting multimodal llms via image-to-text transformation.  
**Yunhao Gou\***, Kai Chen\*, Zhili Liu\*, Lanqing Hong, Hang Xu, Zhenguo Li, Dit-Yan Yeung, James T. Kwok, Yu Zhang.  
*European Conference on Computer Vision (ECCV)*, 2024 [\[link\]](#)
- Mixture of Cluster-conditional LoRA Experts for Vision-language Instruction Tuning.  
**Yunhao Gou\***, Zhili Liu\*, Kai Chen\*, Lanqing Hong, Hang Xu, Aoxue Li, Dit-Yan Yeung, James Kwok, Yu Zhang..  
*Arxiv* [\[link\]](#)
- EPIC: Leveraging per image-token consistency for vision-language pre-training.  
**Yunhao Gou**, Tom Ko, Hansi Yang, James Kwok, Yu Zhang, Mingxuan Wang. .  
*IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023 [\[link\]](#)

- Exploring Hierarchical Graph Representation for Large-Scale Zero-Shot Image Classification.  
Kai Yi, Xiaoqian Shen, **Yunhao Gou**, Mohammed Elhoseiny.  
*European Conference on Computer Vision (ECCV), 2022* [\[link\]](#)
- Region Semantically Aligned Network for Zero-Shot Learning.  
Ziyang Wang\*, **Yunhao Gou\***, Jingjing Li, Yu Zhang, Yang Yang.  
*Conference on Information and Knowledge Management (CIKM 2021 full paper)* [\[link\]](#)

## HORNORS

---

<b>National scholarship</b>	(1%, by Ministry of Education of P.R. China)	2019
HKUST Postgraduate Scholarship		2022 - Now
Research Travel Grant HKUST		2023
Oversea Visiting Student Stipend of UESTC		2021
Excellent students scholarship	(Awarded to Top 10 % students)	2019 - 2021

## ACADEMIC SERVICES

---

### Conference reviewer

International Conference on Machine Learning (ICML)	2025
Neural Information Processing Systems (NeurIPS)	2025
European Conference on Computer Vision (ECCV)	2022
Annual Meeting of the Association for Computational Linguistics (ACL) (ARR Feb)	2025
Empirical Methods in Natural Language Processing (EMNLP) (ARR May)	2025

## TALKS

---

[TechBeat Online] Mixture of Cluster-conditional LoRA Experts for Vision-language Instruction Tuning.  
[\[Recording\]](#)  
 [Tsinghua Sustainable Innovation Workshop 2024] Exploring the Potential of Language Expert in Vision-Language Models

## INTERNSHIP

---

<b>King Abdullah University of Science and Technology</b>	<b>July. 2021 – Dec. 2021</b>
Visiting student, working with Prof. Mohammed Elhoseiny	Zero-shot learning
<b>Bytedance AI Lab</b>	<b>Jan. 2022 – Mar. 2023</b>
Research intern, working with Tom Ko	Vision-language pretraining
<b>Huawei Noah's Ark Lab (AI Theory group)</b>	<b>Oct. 2023 – Now</b>
Research intern, working with Lanqing Hong	Multimodal Large Language Models (MLLMs)

## TEACHING

---

**I have been the TA for the following courses in HKUST:**

COMP4211: Machine Learning (Fall 2023)  
 COMP4641: Social Information Network Analysis and Engineering (Spring 2023)  
 MSBD5008: Introduction to Social Computing (Spring 2023)

## Society Services

---

<b>Teaching Volunteer in Bali</b>	<b>Jan. 2019</b>
Taught local middle school students mathematics, English, and Chinese.	
<b>COVID-19 Community Volunteer</b>	<b>Feb. 2020 – Mar. 2020</b>
Assisted the community in distributing relief supplies	
<b>Participant in Mental Health First Aid Training of HKUST</b>	<b>June. 2023</b>
Qualified to identify individuals who may be experiencing depression and provide immediate assistance.	