Haodong Duan

✉ dhd.efz@gmail.com · ⌂ HomePage · 🏛 Google Scholar · ⏻ GitHub

Short BIO

Haodong Duan is a Post-Doctoral Researcher at Shanghai AI Laboratory. He received his Bachelor's degree from Peking University and his Ph.D. from The Chinese University of Hong Kong. His research experience spans **video understanding**, **large language models**, and **multimodal learning**. He has published 30+ papers in top-tier conferences in computer vision and artificial intelligence, including CVPR, ECCV, ICCV, ACL, and NeurIPS, with more than 6,000 citations. He led the development of MMBench, one of the most influential benchmarks in multi-modal evaluation. He has also spearheaded the development of several well-known open-source projects, including MMAction2, PYSKL, and VLMEvalKit. Additionally, he launched the Open VLM Leaderboard, one of the most authoritative rankings for multimodal performance. He is currently leading the multi-modal evaluation team at Shanghai AI Laboratory.

Professional Experience

Post-Doctoral Researcher, Shanghai AI Laboratory   **2023 – Present**

Work on large language model evaluation, multi-modal learning and multi-modal evaluation, supervised by Prof. Dahua Lin, mentored by Dr. Kai Chen.

Applied Scientist Intern, AWS AI   **2022**

Work on skeleton action recognition in the wild, mentored by Dr. Mingze Xu and Dr. Alessandro Bergamo.

Research Intern, Shanghai AI Laboratory   **2021 – 2022**

Develop and maintain the OpenSource codebases including MMAction2, MMCV, and MIM, mentored by Dr. Kai Chen.

Research Intern, SenseTime   **2017 – 2019**

Work on human pose estimation (skeleton keypoints and contour keypoints), mentored by Dr. Wentao Liu.

Education

The Chinese University of HongKong, HongKong   **2019 – 2023**

Work on *Video Understanding*, supervised by Dahua Lin
Ph.D. in Information Engineering

Peking University, Beijing   **2015 – 2019**

GPA 3.77/4.00, rank 1st in Data Science students
Undergraduate in Data Science, Yuanpei College

Projects

OpenSource Codebases
The main contributor and maintainer of MMAction2, PYSKL, OpenCompass, VLMEvalKit.

- MMAction2: An open-source codebase for video understanding and action recognition. (★ 4.7K, ⑂ 1.3K)

- PYSKL: An open-source codebase for skeleton-based action recognition. (★ 1.1K, ⑂ 205)

- OpenCompass: An open-source codebase for evaluating large language models. (★ 5.6K, ⑂ 616)
- VLMEvalKit: An open-source codebase for evaluating multi-modal models (★ 2.7K, ⑂ 440). We also host the OpenVLM Leaderboard (on huggingface space, ❤ 814, ⬈364K) based on this toolkit.

LLM / VLM Projects

- A main contributor of InternLM series: Mainly works on Subjective Evaluation, Long-Context Evalution and Multi-Turn Dialogue Evaluation.
- A main contributor of InternLM-XComposer series: Mainly works on the evalution of InternLM-XComposer.

PROFESSIONAL SERVICES

- **Conference Area Chair:** NeurIPS 2025.
- **Conference Reviewer:** ICCV (2021 ~), AAAI (2022 ~), CVPR (2022 ~), ECCV (2022 ~), NeurIPS (2022 ~), ICML (2024 ~), ICLR (2024 ~), ACL ARR (2024 ~), WACV 2023, EuroGraphics 2023, AISTATS 2025, COLM 2025, ACMMM 2025.
- **Journal Reviewer:** TPAMI, IJCV, TIP, PR, TMM, TOMM, TCSVT, Sensors, SPL, JVCIR.

AWARDS & GRANTS

- **China Postdoctoral Science Fund** (Special Program, No.2025T180412), 180,000 CNY, 2025
- **Outstanding Employee (#1 / 20)**, Shanghai AI Laboratory, 2024
- **Outstanding Project for OpenCompass (#2 / 6)**, Shanghai AI Laboratory, 2024
- **Shanghai Magnolia Award**, 2024
- **China Postdoctoral Science Fund** (No. 2024M751559), 80,000 CNY, 2024
- **Shanghai Postdoctoral Excellence Program** (No. 2023023), 300,000 CNY, 2023

PREPRINTS

[2025.P.9] GOBench: Benchmarking Geometric Optics Generation and Understanding of MLLMs

*Xiaorong Zhu, Ziheng Jia, Jiarui Wang, Xiangyu Zhao, **Haodong Duan**, Xiongkuo Min, Jia Wang, Zicheng Zhang, Guangtao Zhai*

[2025.P.8] Affordance Benchmark for MLLMs

*Junying Wang, Wenzhe Li, Yalun Wu, Yingji Liang, Yijin Guo, Chunyi Li, **Haodong Duan**, Zicheng Zhang, Guangtao Zhai*

[2025.P.7] MMSI-Bench: A Benchmark for Multi-Image Spatial Intelligence

*Sihan Yang, Runsen Xu, Yiman Xie, Sizhe Yang, Mo Li, Jingli Lin, Chenming Zhu, Xiaochen Chen, **Haodong Duan**, Xiangyu Yue, Dahua Lin, Tai Wang, Jiangmiao Pang*

[2025.P.6] Visual Agentic Reinforcement Fine-Tuning

*Ziyu Liu, Yuhang Zang, Yushan Zou, Zijian Liang, Xiaoyi Dong, Yuhang Cao, **Haodong Duan**, Dahua Lin, Jiaqi Wang*

[2025.P.5] GDI-Bench: A Benchmark for General Document Intelligence with Vision and Reasoning Decoupling

*Siqi Li, Yufan Shen, Xiangnan Chen, Jiayi Chen, Hengwei Ju, **Haodong Duan**, Song Mao, Hongbin Zhou, Bo Zhang, Pinlong Cai, Licheng Wen, Botian Shi, Yong Liu, Xinyu Cai, Yu Qiao*

[2025.P.4] Envisioning Beyond the Pixels: Benchmarking Reasoning-Informed Visual Editing

*Xiangyu Zhao, Peiyuan Zhang, Kexian Tang, Hao Li, Zicheng Zhang, Guangtao Zhai, Junchi Yan, Hua Yang, Xue Yang, **Haodong**

Duan[†]

[2025.P.3] LEGO-Puzzles: How Good Are MLLMs at Multi-Step Spatial Reasoning?
*Kexian Tang, Junyao Gao, Yanhong Zeng, **Haodong Duan**[‡], Yanan Sun, Zhening Xing, Wenran Liu, Kaifeng Lyu, Kai Chen*

[2025.P.2] Visualprm: An effective process reward model for multimodal reasoning
*Weiyun Wang, Zhangwei Gao, Lianjie Chen, Zhe Chen, Jinguo Zhu, Xiangyu Zhao, Yangzhou Liu, Yue Cao, Shenglong Ye, Xizhou Zhu, Lewei Lu, **Haodong Duan**, Yu Qiao, Jifeng Dai, Wenhai Wang*

[2025.P.1] BoostStep: Boosting mathematical capability of Large Language Models via improved single-step reasoning
*Beichen Zhang, Yuhong Liu, Xiaoyi Dong, Yuhang Zang, Pan Zhang, **Haodong Duan**, Yuhang Cao, Dahua Lin, Jiaqi Wang*

[2024.P.3] MME-Survey: A Comprehensive Survey on Evaluation of Multimodal LLMs
*Chaoyou Fu, Yi-Fan Zhang, Shukang Yin, Bo Li, Xinyu Fang, Sirui Zhao, **Haodong Duan**, Xing Sun, Ziwei Liu, Liang Wang, Caifeng Shan, Ran He*

[2024.P.2] CompassJudger-1: All-in-one Judge Model Helps Model Evaluation and Evolution
*Maosong Cao, Alexander Lam, **Haodong Duan**, Hongwei Liu, Songyang Zhang, Kai Chen*

[2024.P.1] MG-LLaVA: Towards Multi-Granularity Visual Instruction Tuning
*Xiangyu Zhao, Xiangtai Li, **Haodong Duan**, Haian Huang, Yining Li, Kai Chen, Hua Yang*

[2022.P.1] DG-STGCN: Dynamic Spatial-Temporal Modeling for Skeleton-based Action Recognition
***Haodong Duan**, Jiaqi Wang, Kai Chen, Dahua Lin*

[†] Corresponding Author, [‡] Project Lead

PUBLICATIONS

[2025.C.13] MM-IFEngine: Towards Multimodal Instruction Following (ICCV, 2025)
*Shengyuan Ding, Shenxi Wu, Xiangyu Zhao, Yuhang Zang, **Haodong Duan**, Xiaoyi Dong, Pan Zhang, Yuhang Cao, Dahua Lin, Jiaqi Wang*

[2025.C.12] Creation-MMBench: Assessing Context-Aware Creative Intelligence in MLLM (ICCV, 2025)
*Xinyu Fang, Zhijian Chen, Kai Lan, Shengyuan Ding, Yingji Liang, Xiangyu Zhao, Farong Wen, Zicheng Zhang, Guofeng Zhang, **Haodong Duan**[†], Kai Chen, Dahua Lin*

[2025.C.11] Information Density Principle for MLLM Benchmarks (ICCV, 2025)
*Chunyi Li, Xiaozhe Li, Zicheng Zhang, Yuan Tian, Ziheng Jia, Xiaohong Liu, Xiongkuo Min, Jia Wang, **Haodong Duan**[†], Kai Chen, Guangtao Zhai*

[2025.C.10] Visual-RFT: Visual Reinforcement Fine-Tuning (ICCV, 2025)
*Ziyu Liu, Zeyi Sun, Yuhang Zang, Xiaoyi Dong, Yuhang Cao, **Haodong Duan**, Dahua Lin, Jiaqi Wang*

[2025.C.9] Towards Storage-Efficient Visual Document Retrieval: An Empirical Study on Reducing Patch-Level Embeddings (ACL Findings, 2025)
*Yubo Ma, Jinsong Li, Yuhang Zang, Xiaobao Wu, Xiaoyi Dong, Pan Zhang, Yuhang Cao, **Haodong Duan**, Jiaqi Wang, Yixin Cao, Aixin Sun*

[2025.C.8] InternLM-XComposer2.5-Reward: A Simple Yet Effective Multi-Modal Reward Model (ACL Findings, 2025)
*Yuhang Zang, Xiaoyi Dong, Pan Zhang, Yuhang Cao, Ziyu Liu, Shengyuan Ding, Shenxi Wu, Yubo Ma, **Haodong Duan**, Wenwei Zhang, Kai Chen, Dahua Lin, Jiaqi Wang*

[2025.C.7] OmniAlign-V: Towards Enhanced Alignment of MLLMs with Human Preference (ACL, 2025)
Xiangyu Zhao, Shengyuan Ding, Zicheng Zhang, Haian Huang, Maosong Cao, Weiyun Wang, Jiaqi Wang, Xinyu Fang, Wenhai Wang,

*Guangtao Zhai, Hua Yang, **Haodong Duan**[†], Kai Chen*

[2025.C.6] Redundancy Principles for MLLMs Benchmarks (ACL, 2025)

*Zicheng Zhang, Xiangyu Zhao, Xinyu Fang, Chunyi Li, Xiaohong Liu, Xiongkuo Min, **Haodong Duan**[†], Kai Chen, Guangtao Zhai*

[2025.C.5] Condor: Enhance LLM Alignment with Knowledge-Driven Data Synthesis and Refinement (ACL, 2025)

*Maosong Cao, Taolin Zhang, Mo Li, Chuyu Zhang, Yunxin Liu, **Haodong Duan**[†], Songyang Zhang, Kai Chen*

[2025.C.4] VideoRoPE: What Makes for Good Video Rotary Position Embedding? (ICML Oral, 2025)

*Xilin Wei, Xiaoran Liu, Yuhang Zang, Xiaoyi Dong, Pan Zhang, Yuhang Cao, Jian Tong, **Haodong Duan**, Qipeng Guo, Jiaqi Wang, Xipeng Qiu, Dahua Lin*

[2025.C.3] Image Quality Assessment: From Human to Machine Preference (CVPR Spotlight, 2025)

*Chunyi Li, Yuan Tian, Xiaoyue Ling, Zicheng Zhang, **Haodong Duan**, Haoning Wu, Ziheng Jia, Xiaohong Liu, Xiongkuo Min, Guo Lu, Weisi Lin, Guangtao Zhai*

[2025.C.2] OVO-Bench: How Far is Your Video-LLMs from Real-World Online Video Understanding? (CVPR, 2025)

*Yifei Li, Junbo Niu, Ziyang Miao, Chunjiang Ge, Yuanhang Zhou, Qihao He, Xiaoyi Dong, **Haodong Duan**, Shuangrui Ding, Rui Qian, Pan Zhang, Yuhang Zang, Yuhang Cao, Conghui He, Jiaqi Wang*

[2025.C.1] MIA-DPO: Multi-Image Augmented Direct Preference Optimization For Large Vision-Language Models (ICLR, 2025)

*Ziyu Liu, Yuhang Zang, Xiaoyi Dong, Pan Zhang, Yuhang Cao, **Haodong Duan**, Conghui He, Yuanjun Xiong, Dahua Lin, Jiaqi Wang*

[2024.C.12] MMBench-Video: A Long-Form Multi-Shot Benchmark for Holistic Video Understanding (NeurIPS D&B, 2024)

*Xinyu Fang, Kangrui Mao, **Haodong Duan**[†], Xiangyu Zhao, Yining Li, Dahua Lin, Kai Chen*

[2024.C.11] Prism: A Framework for Decoupling and Assessing the Capabilities of VLMs (NeurIPS, 2024)

*Yuxuan Qiao, **Haodong Duan**[†], Xinyu Fang, Junming Yang, Lin Chen, Songyang Zhang, Jiaqi Wang, Dahua Lin, Kai Chen*

[2024.C.10] InternLM-XComposer2-4KHD: A Pioneering Large Vision-Language Model Handling Resolutions from 336 Pixels to 4K HD (NeurIPS, 2024)

*Xiaoyi Dong, Pan Zhang, Yuhang Zang, Yuhang Cao, Bin Wang, Linke Ouyang, Songyang Zhang, **Haodong Duan**, Wenwei Zhang, Yining Li, Hang Yan, Yang Gao, Zhe Chen, Xinyue Zhang, Wei Li, Jingwen Li, Wenhai Wang, Kai Chen, Conghui He, Xingcheng Zhang, Jifeng Dai, Yu Qiao, Dahua Lin, Jiaqi Wang*

[2024.C.9] GMAI-MMBench: A Comprehensive Multimodal Evaluation Benchmark Towards General Medical AI (NeurIPS D&B, 2024)

*Pengcheng Chen, Jin Ye, Guoan Wang, Yanjun Li, Zhongying Deng, Wei Li, Tianbin Li, **Haodong Duan**, Ziyan Huang, Yanzhou Su, Benyou Wang, Shaoting Zhang, Bin Fu, Jianfei Cai, Bohan Zhuang, Eric J Seibel, Junjun He, Yu Qiao*

[2024.C.8] Sharegpt4video: Improving video understanding and generation with better captions (NeurIPS D&B, 2024)

*Lin Chen, Xilin Wei, Jinsong Li, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Zehui Chen, **Haodong Duan**, Bin Lin, Zhenyu Tang, Li Yuan, Yu Qiao, Dahua Lin, Feng Zhao, Jiaqi Wang*

[2024.C.7] Are We on the Right Way for Evaluating Large Vision-Language Models? (NeurIPS, 2024)

*Lin Chen, Jinsong Li, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Zehui Chen, **Haodong Duan**, Jiaqi Wang, Yu Qiao, Dahua Lin, Feng Zhao*

[2024.C.6] ProSA: Assessing and Understanding the Prompt Sensitivity of LLMs (EMNLP Findings, 2024)

*Jingming Zhuo, Songyang Zhang, Xinyu Fang, **Haodong Duan**, Dahua Lin, Kai Chen*

[2024.C.5] VLMEvalKit: An Open-Source Toolkit for Evaluating Large Multi-Modality Models (ACMMM OS, 2024)

***Haodong Duan**, Junming Yang, Yuxuan Qiao, Xinyu Fang, Lin Chen, Yuan Liu, Xiaoyi Dong, Yuhang Zang, Pan Zhang, Jiaqi Wang, Dahua Lin, Kai Chen*

[2024.C.4] Mmbench: Is your multi-modal model an all-around player? (ECCV Oral, 2024)

*Yuan Liu, **Haodong Duan**[‡], Yuanhan Zhang, Bo Li, Songyang Zhang, Wangbo Zhao, Yike Yuan, Jiaqi Wang, Conghui He, Ziwei Liu, Kai Chen, Dahua Lin*

[2024.C.3] MathBench: Evaluating the Theory and Application Proficiency of LLMs with a Hierarchical Mathematics Benchmark (ACL Findings, 2024)

*Hongwei Liu, Zilong Zheng, Yuxuan Qiao, **Haodong Duan**, Zhiwei Fei, Fengzhe Zhou, Wenwei Zhang, Songyang Zhang, Dahua Lin, Kai Chen*

[2024.C.2] Ada-LEval: Evaluating long-context LLMs with length-adaptable benchmarks (NAACL, 2024)

*Chonghua Wang, **Haodong Duan**[†], Songyang Zhang, Dahua Lin, Kai Chen*

[2024.C.1] BotChat: Evaluating LLMs' Capabilities of Having Multi-Turn Dialogues (NAACL Findings, 2024)

***Haodong Duan**[†], Jueqi Wei, Chonghua Wang, Hongwei Liu, Yixiao Fang, Songyang Zhang, Dahua Lin, Kai Chen*

[2023.C.3] Journeydb: A benchmark for generative image understanding (NeurIPS D&B, 2023)

*Junting Pan, Keqiang Sun, Yuying Ge, Hao Li, **Haodong Duan**, Xiaoshi Wu, Renrui Zhang, Aojun Zhou, Zipeng Qin, Yi Wang, Jifeng Dai, Yu Qiao, Hongsheng Li*

[2023.C.2] SkeleTR: Towards Skeleton-based Action Recognition in the Wild (ICCV, 2023)

***Haodong Duan**, Mingze Xu, Bing Shuai, Davide Modolo, Zhuowen Tu, Joseph Tighe, Alessandro Bergamo*

[2023.T.1] Towards Effective Spatio-Temporal Feature Learning for Human Action Understanding (PhD Thesis)

Haodong Duan

[2023.C.1] Self-supervised Action Representation Learning from Partial Spatio-Temporal Skeleton Sequences (AAAI, 2023)

*Yujie Zhou, **Haodong Duan**, Anyi Rao, Bing Su, Jiaqi Wang*

[2022.C.5] Mitigating Representation Bias in Action Recognition: Algorithms and Benchmarks (ECCVW, 2022)

***Haodong Duan**, Yue Zhao, Kai Chen, Yuanjun Xiong, Dahua Lin*

[2022.C.4] PYSKL: Towards Good Practices for Skeleton Action Recognition (ACMMM OS, 2022)

***Haodong Duan**, Jiaqi Wang, Kai Chen, Dahua Lin*

[2022.C.3] OCSampler: Compressing Videos to One Clip with Single-step Sampling (CVPR, 2022)

*Jintao Lin, **Haodong Duan**, Kai Chen, Dahua Lin, Limin Wang*

[2022.C.2] TransRank: Self-supervised Video Representation Learning via Ranking-based Transformation Recognition (CVPR Oral, 2022)

***Haodong Duan**, NanXuan Zhao, Kai Chen, Dahua Lin*

[2022.C.1] Revisiting Skeleton-based Action Recognition (CVPR Oral, 2022)

***Haodong Duan**, Yue Zhao, Kai Chen, Dahua Lin, Bo Dai*

[2020.C.1] Omni-sourced Webly-supervised Learning for Video Recognition (ECCV, 2020)

***Haodong Duan**, Yue Zhao, Yuanjun Xiong, Wentao Liu, Dahua Lin*

[2019.C.1] TRB: A Novel Triplet Representation for Understanding 2D Human Body (ICCV Oral, 2019)

***Haodong Duan**, Kwanyee Lin, Sheng Jin, Wentao Liu, Chen Qian, Wanli Ouyang*

[†] Corresponding Author, [‡] Project Lead