Ronghao Dang

17317813949 | drh1710013267@163.com | drh1710013267 | Google Scholar | Rh-Dang

Research Interests: Embodied AI, Multimodal Foundation Models, 3D Vision, Reinforcement Learning

EXPERIENCE

Alibaba DAMO Academy May 2024 – Present
AI Researcher Hangzhou

- Investigated the limitations of MLLMs embodied cognitive capabilities, and proposed **ECBench** and **EOCBench**
- Open-sourced the embodied cognitive models **RynnEC** (400 GitHub stars; 1.5k+ Hugging Face downloads)
- Open-sourced the embodied foundation models **RynnBrain** (500 GitHub stars; 4k+ Hugging Face downloads)

SenseTime May 2023 – Sept. 2023
Research Intern Shanghai

- Built an LLM/VLM-driven data engine to scale referring expression comprehension (REC) instruction data by **20x+**
- Released the **InDET** datasets and **InstructDET** model that advances SOTA semantic understanding and vision language alignment (demo).

PROJECTS

RynnBrain *(First Contributor | Team Leader)*

- Integrated **four core capabilities** in a unified framework: comprehensive egocentric understanding, diverse spatiotemporal localization, physically grounded reasoning, and physics-aware planning.
- RynnBrain family (2B / 8B / 30B-A3B MoE) outperforms SOTA models such as **Google's Gemini-ER 1.5** and **NVIDIA's Cosmos-Reason2** on **28** benchmarks.
- Released four post-trained variants tailored for downstream embodied tasks (**RynnBrain-Nav**, **RynnBrain-Plan**, **RynnBrain-VLA**) and complex spatial reasoning tasks (**RynnBrain-CoP**).

RynnEC *(First Contributor | Team Leader)*

- Proposed a region-centric video paradigm for embodied-agent cognition, enabling fine-grained perception and more precise physical interactions.
- Outperformed Gemini 2.5 Pro by **23.5%** on object cognition, object segmentation, and spatial reasoning.

Visual Object Navigation *(First Contributor)*

- Introduced DOA, DAT, and MT, improving indoor object navigation success rate from **60%** to **80%**.

EDUCATION

Tongji University Sept. 2021 – May 2024
Master of Control Science and Engineering Shanghai

- **Reward**: National Scholarship (Ministry of Education of China), 2023
- **Research**: Published over ten top-tier conference papers in the Vision-Language Navigation (VLN) field

Tongji University Sept. 2017 – Jun 2021
Bachelor of Automation Shanghai

PUBLICATIONS

Selected Publications

- **Ronghao Dang**, Jiayan Guo, Bohan Hou, *et al. . RynnBrain: Open Embodied Foundation Models*. **Under Review Nature Machine Intelligence**, 2026. arxiv | Code
- **Ronghao Dang**, Yuqian Yuan, Yunxuan Mao, *et al.. RynnEC: Bringing mllms into embodied world*. 2025. arxiv | Code
- **Ronghao Dang**, Yuqian Yuan, Wenqi Zhang, *et al.. ECBench: Can multi-modal foundation models understand the egocentric world? a holistic embodied cognition benchmark*. **CVPR**, 2025. arxiv | Code
- Yuqian Yuan, **Ronghao Dang**, Long Li, *et al.. EOC-Bench: Can mllms identify, recall, and forecast objects in an egocentric world?*. **NeurIPS**, 2025. arxiv | Code
- **Ronghao Dang**, Jiangyan Feng, Haodong Zhang, *et al.. Instructdet: Diversifying referring object detection with generalized instructions*. **ICLR**, 2023. arxiv | Code
- **Ronghao Dang**, Lu Chen, Liuyi Wang, *et al.. Multiple thinking achieving meta-ability decoupling for object navigation*. **ICML**, 2023. arxiv
- **Ronghao Dang**, Liuyi Wang, Zongtao He, *et al.. Search for or navigate to? dual adaptive thinking for object navigation*. **ICCV**, 2023. arxiv | Code
- **Ronghao Dang**, Zhuofan Shi, Liuyi Wang, *et al.. Unbiased directed object attention graph for object navigation*. **ACMMM**, 2022. arxiv | Code
- Liuyi Wang, Zongtao He, **Ronghao Dang**, *et al.. Vision-and-language navigation via causal learning*. **CVPR**, 2024. arxiv | Code
- Liuyi Wang, Zongtao He, Jiagui Tang, **Ronghao Dang**, *et al.. A Dual Semantic-Aware Recurrent Global-Adaptive Network For Vision-and-Language Navigation*. **IJCAI**, 2023. arxiv | Code
- Minghao Zhu, Zhengpu Wang, Mengxian Hu, **Ronghao Dang**, *et al.. Mote: Reconciling generalization with specialization for visual-language to video knowledge transfer*. **Neurips**, 2024. arxiv | Code
- Liuyi Wang, Zongtao He, **Ronghao Dang**, *et al.. RES-StS: Referring expression speaker via self-training with scorer for goal-oriented vision-language navigation*. **TCSVT**, 2023.
- Xiao Lin, Minghao Zhu, **Ronghao Dang**, *et al.. Clipose: Category-level object pose estimation with pre-trained vision-language knowledge*. **TCSVT**, 2024.
- Zongtao He, Liuyi Wang, **Ronghao Dang**, *et al.. Learning depth representation from RGB-D videos by time-aware contrastive pre-training*. **TCSVT**, 2023.
- Minghao Zhu, Xiao Lin, **Ronghao Dang**, *et al.. Fine-grained spatiotemporal motion alignment for contrastive video representation learning*. **ACMMM**, 2023.
- Liuyi Wang, Jiagui Tang, Zongtao He, **Ronghao Dang**, *et al.. Enhanced language-guided robot navigation with panoramic semantic depth perception and cross-modal fusion*. **IROS**, 2024.
- Qinyang Zeng, **Ronghao Dang**, Xun Zhou, *et al.. Contrastive feedback vision-language for 3d skeleton-based action recognition*. **TMM**, 2025.
- Long Li, Weiwen Xu, Jiayan Guo, Ruochen Zhao, Xingxuan Li, Yuqian Yuan, Boqiang Zhang, Yuming Jiang, Yifei Xin, **Ronghao Dang**, *et al.. Chain of Ideas: Revolutionizing Research in Idea Development with LLM Agents*. **ACL**, 2025.
- **Ronghao Dang**, Chengju Liu, Ming Liu, *et al.. Channel attention and multi-scale graph neural networks for skeleton-based action recognition*. **AI Communications**, 2022.