Haomin Wang

kiyotakawang@sjtu.edu.cn | Homepage | Google Scholar

First-year Ph.D. student at the School of Artificial Intelligence, Shanghai Jiao Tong University. My core research primarily focuses on the application of MLLMs to vector graphics. More recently, I have also been exploring world models in the context of embodied intelligence.

Education

Shanghai Jiao Tong University Sept. 2025 – Present
Ph.D. Student, School of Artificial Intelligence Shanghai, China

Nanjing University Sept. 2021 – Jun. 2025
B.Eng. in Software Engineering Nanjing, China
GPA: 4.579/5.00 | **Rank**: 9/254 (Top 3.54%) | **English**: CET-4 636, CET-6 578

Internship Experience

Shanghai AI Laboratory Nov. 2024 – Jun. 2026
Research Intern, Vision-Language Multimodal Team, LLM Center Shanghai, China

OpenDriveLab Jul. 2024 – Nov. 2024
Research Intern Shanghai, China

Selected Projects

- **[ICLR 2026] InternSVG: Towards Unified SVG Tasks with Multimodal Large Language Models** *First Author*
  - Led the development of a unified SVG modeling framework, including SAgoge dataset, SArena benchmark, and InternSVG-8B, a unified MLLM for SVG understanding, editing, and generation.
  - Constructed **SAgoge**, the largest existing SVG dataset with 16M samples, covering three tasks (understanding, editing, and generation) and four data categories (icons, illustrations, chemical structural formulas, and vector animations). Proposed the **SArena** benchmark to provide a rigorous and systematic metric for evaluating general-purpose SVG modeling capabilities.
  - Developed **InternSVG-8B** based on InternVL3-8B. By designing SVG-specific special tokens with a subword-based embedding initialization strategy, combined with a two-stage progressive training scheme, the model achieves unified modeling of SVG semantic understanding, instruction-based editing, and cross-modal generation.

- **[ECCV 2026] Reliable Reasoning in SVG-LLMs via Multi-Task Multi-Reward Reinforcement Learning** *First Author*
  - Proposed **CTRL-S**, which introduces Chain-of-Thought reasoning into SVG generation and defect repair. By leveraging the grouping property of SVG, it achieves explicit alignment between reasoning steps and structured code, improving the readability, editability, and aesthetic quality of generated SVGs.
  - Built the **SVG-Sophia** dataset. In addition to Text-to-SVG and Image-to-SVG tasks, introduced a novel **SVG code refinement** task to train the model to generate more readable and editable SVG code while being capable of repairing defective SVGs.
  - Unlike prior RL methods that train on a single task, we developed a customized reward function based on **verl**, combining four reward signals (DINO similarity, text-image similarity, format compliance, and code efficiency) for multi-task multi-reward RL training. Using only 145K training samples in SFT and RL stages, the method outperforms prior works that rely on millions of samples, achieving **SOTA** results on the SArena benchmark.

- **[NeurIPS 2025] Point or Line? Using Line-based Representation for Panoptic Symbol Spotting in CAD Drawings** *Co-First Author*
  - Proposed **VecFormer**, a Point Transformer-based framework that replaces point-based primitive features with line segments, preserving geometric continuity in CAD floor plan drawings and better aligning with their line-dominated nature.
  - Achieved new SOTA on the FloorPlanCAD benchmark with **91.1 PQ**, significantly outperforming previous methods.
  - Responsible for line feature design, and all model training and evaluation tasks in this project.

- **Intern-S1-Pro: Scientific Multimodal Foundation Model at Trillion Scale** *Core Feature Contributor*
  - Led the integration of SVG generation as a core capability of the Intern-S1-Pro project, responsible for training data preparation and evaluation. Constructed training data for both CPT and SFT stages, and performed CPT and SFT training on a preview model built upon Qwen3-VL-30B-A3B using **XTuner**, validating the performance gains brought by the data.
  - Integrated the SArena benchmark evaluation code into **VLMEvalKit**.

- **InternVL3.5: Advancing Open-Source Multimodal Models in Versatility, Reasoning, and Efficiency** *Major Contributor*
  - Constructed SVG icon generation and understanding data for the SFT training stage. Evaluated the full InternVL3.5 model series against mainstream baselines on SVG icon generation and understanding tasks, with results showing that models at each parameter scale achieved the best performance among contemporary mainstream models of comparable size.
  - Constructed understanding and QA data for CAD floor plan drawings, enhancing the model's ability to recognize floor plans and perform object detection on the furniture within them.

Publications

- **Wang, H.**, Yin, J., Wei, Q., Zeng, W., Gu, L., Ye, S., ... & Zhang, H. (2026). InternSVG: Towards Unified SVG Tasks with Multimodal Large Language Models. *The Fourteenth International Conference on Learning Representations (ICLR)*.
- **Wang, H.**, Wei, Q., Ma, Q., Ding, S., Yin, J., Chen, K., & Zhang, H. (2026). Reliable Reasoning in SVG-LLMs via Multi-Task Multi-Reward Reinforcement Learning. *The Nineteenth European Conference on Computer Vision*.
- Wei, X., **Wang, H.**, Ye, S., Luo, R., Zhang, Y., Gu, L., ... & Zhang, H. Point or Line? Using Line-based Representation for Panoptic Symbol Spotting in CAD Drawings. *The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS)*.
- Deng, N., Gu, L., Ye, S., He, Y., Chen, Z., Li, S., ... & Wang, W. (2025). InternSpatial: A Comprehensive Dataset for Spatial Reasoning in Vision-Language Models. *The Fourteenth International Conference on Learning Representations (ICLR)*.
- Luo, R., Liu, Z., Cheng, T., Wang, J., Wang, T., Cheng, F., ... & Zhao, X. (2025). ArchCAD-400K: A Large-Scale CAD Drawings Dataset and New Baseline for Panoptic Symbol Spotting. *The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS)*.
- Wang, W., Gao, Z., Gu, L., Pu, H., Cui, L., Wei, X., ... & Luo, G. (2025). InternVL3.5: Advancing Open-Source Multimodal Models in Versatility, Reasoning, and Efficiency. *arXiv preprint arXiv:2508.18265*.
- Zhu, J., Wang, W., Chen, Z., Liu, Z., Ye, S., Gu, L., ... & Wang, W. (2025). InternVL3: Exploring Advanced Training and Test-Time Recipes for Open-Source Multimodal Models. *arXiv preprint arXiv:2504.10479*.
- Zou, Y., Zhu, D., Zhu, L., Zhu, T., Zhou, Y., Zhou, P., ... & Fan, Y. (2026). Intern-S1-Pro: Scientific Multimodal Foundation Model at Trillion Scale. *arXiv preprint arXiv:2603.25040*.
- Tian, C., Yang, D., Chen, G., Cui, E., Wang, Z., Duan, Y., ... & Zhang, H. (2026). InternVL-U: Democratizing Unified Multimodal Models for Understanding, Reasoning, Generation and Editing. *arXiv preprint arXiv:2603.09877*.

Skills

- **LLM Training & Fine-tuning**: Extensive experience in LLM training. Proficient with mainstream frameworks including **LLaMA-Factory** and **XTuner** for pre-training and SFT. Experienced in training large models on clusters of 100+ GPUs.
- **Reinforcement Learning**: Familiar with RL algorithms and proficient with the **verl** framework, with hands-on experience training models using the GRPO algorithm.
- **Foundation Models & Data Engineering**: Deeply involved in the development of foundation models such as **InternVL3**, **InternVL3.5**, and **Intern-S1-Pro**. Experienced in cleaning and constructing multimodal datasets at the tens-of-millions scale, with comprehensive capability in building model evaluation benchmarks.
- **Research & Algorithms**: Solid skills in research planning and algorithm implementation, with proven experience in writing high-quality academic papers. Strong deep learning foundation; proficient in Python, C++, and the PyTorch framework.
- **Vibe Coding**: Skilled at leveraging tools such as Claude Code to assist in daily development.

Honors & Awards

- **Outstanding Graduate**, Nanjing University 2025
- **Outstanding Student**, Nanjing University 2023
- **China Merchants Bank Scholarship**, Nanjing University 2022 – 2023
- **Ruli Scholarship**, Nanjing University 2021 – 2022