Jixuan Ying

Mobile: +86 133 9241 4376

Email: yingjx23@mails.tsinghua.edu.cn

Address: Room 601B, Main Building, Tsinghua University, Beijing

EDUCATION

Tsinghua University   Weiyang College   Beijing, China

B.S. in Mathematics and Physics   Sept. 2023 –June 2027 (Expected)

Academic Advisor: Prof. Gao Huang (CVPR 2017 Best Paper Award Winner)

EXPERIENCE

SenseTime Research / Multimodal Large Model Group   Beijing, China

Research Intern   Sept. 2025 –Present

Research Advisor: Lewei Lu (CVPR 2023 Best Paper Award Winner)

PUBLICATIONS

* denotes equal contribution. # denotes corresponding author.

Interleaved Multimodal Reasoning:

1. Z. Ning, W. Tong, X. Kong, S. Ma, Z. Shang, J. Ni, T. Hu, Y. X. Chng, **Jixuan Ying**, Z. Wu, J. Yang, W. Liu, H. Deng, L. Lu$^{#}$, "InterSketch: Interleaved Visual–Textual Chain-of-Thought with Tool-Generated Sketches and Reinforcement Learning", In *International Conference on Machine Learning (**ICML**), 2026, under review.*

   Abstract: Proposed InterSketch, an agentic vision–language framework for long-horizon interleaved visual–text chain-of-thought reasoning, leveraging tool-generated intermediate sketches to support multi-step reasoning. This work was conducted during my research internship at SenseTime Research / Multimodal Large Model Group.

   Contribution: Investigated unified understanding–generation models for interleaved image–text reasoning with long-horizon Chain-of-Thought. Built scalable data transformation pipelines for interleaved multimodal supervision, and conducted controlled evaluation experiments to analyze multi-step reasoning behaviors on visual reasoning benchmarks.

DiT / ViT Architecture Design:

2. Y. Pu*, **Jixuan Ying***, T. Ye, D. Han, Z. Wang, Q. Li, X. Shao, X. Wang, G. Huang$^{#}$, X. Li$^{#}$ "Linear Differential Vision Transformer: Learning Visual Contrasts via Pairwise Differentials", In *Neural Information Processing Systems (**NeurIPS**), 2025, accepted.*

   Abstract: Proposed a Linear Differential Vision Transformer that achieves linear complexity via visual contrast attention, enhancing the model's perception of local differences, with minimal added parameters and no extra compute, and strong performance in image classification and generation.

   Contribution: Co-designed the linear attention mechanism, optimized the model architecture, and independently conducted multiple experiments to verify effectiveness and generalization.

Representation Learning:

3. Y. Wang*, Z. Ni*, Y. Pu, C. Zhou, **Jixuan Ying**, S. Song, G. Huang$^{#}$, "InfoPro: Locally Supervised Deep Learning by Maximizing Information Propagation", In *International Journal of Computer Vision (**IJCV**), 2025, accepted.*

   Abstract: Proposed a locally supervised deep learning framework to maximize information propagation, improving information flow and feature expression, boosting generalization and training efficiency, and achieving superior results on multiple mainstream datasets.

   Contribution: Independently conducted multiple experiments and systematically validated the improvements in generalization and training efficiency.

Controllable Visual Generation:

4. Yue Ma*, Y. He*, H. Wang, A. Wang, L. Shen, C. Qi, **Jixuan Ying**, C. Cai, Z. Li, H.-Y. Shum, W. Liu[#], Q. Chen[#], "Follow-Your-Click: Open-domain Regional Image Animation via Motion Prompts", In *AAAI Conference on Artificial Intelligence (**AAAI**), 2025, accepted.*

   Abstract: Proposed a motion-prompted interactive regional image animation method that generates animations for arbitrary areas via mouse clicks, improving controllability and realism.

   Contribution: Conducted experiments and validated improvements in interactive experience and animation quality.

HONORS & AWARDS

The First-Class Congyou Weiyang Scholarship for Scientific and Technological Innovation   2025
Awarded by Tsinghua University

SELECTED PROJECTS

Tsinghua Student Research Training (SRT):

1. Image Generation and Multimodal AI   Advisor: Prof. Gao Huang   Sept. 2024 –Sept. 2025

   Description: Focused on advanced multimodal AI and image generation. Designed and implemented novel linear attention mechanisms based on DiT and ViT architectures to optimize neural network efficiency and performance for vision tasks. Conducted comprehensive experiments exploring applications in image generation and efficient editing on edge devices. **Published a paper in NeurIPS 2025 as Co-First Author.**

2. Pipeline Defect Detection based on Computer Vision   Advisor: Prof. Biqing Huang   Mar. 2024 –Sept. 2024

   Description: Participated in a computer vision-based project for pipeline defect detection. Primarily responsible for technical research and feasibility analysis. Contributed to improving model generalization and detection accuracy for minute defects, providing theoretical support and viable solutions for industrial inspection scenarios.