Kaiwen Zhang

kaiwen008@e.ntu.edu.sg · kevin-thu.github.io/homepage · Kevin-thu · Google Scholar

Education

Nanyang Technological University, Singapore — Jan. 2026 – Present

Ph.D. Student at MMLab@NTU, S-Lab, College of Computing and Data Science
Nanyang President's Graduate Scholarship (NTU's most prestigious PhD scholarship)
Advisor: Prof. Xingang Pan

Tsinghua University, Beijing, China — Sept. 2021 – Jun. 2025

B.Eng. in Computer Science and Technology
GPA: 3.9 / 4.0, Rank: Top 5 / 180 **(Graduated with Highest Honors)**

Publications & Preprints

Research interest: Generative AI, Multimodal AI, World Model, Visual Intelligence
Citations: 550+, GitHub stars: 1.6k+

ViDecide: Teaching Video Diffusion Models to Reason and Plan Step by Step — Under Review

***Kaiwen Zhang**, Yiming Gao, Zeqi Xiao, Xintao Wang, Pengfei Wan, Kun Gai, Xingang Pan*

- Project: `https://kevin-thu.github.io/ViDecide/`
- Paper: `https://arxiv.org/abs/2611.26389`

DeCoPrune: Efficient KV-Cache Pruning for Autoregressive Video Diffusion via Denoising Consistency — Under Review

Zeqi Xiao, Qingle Liu*, **Kaiwen Zhang**, Yifan Zhou, Zihan Ding, Xingang Pan (*Equal contribution)*

- Project: `https://decoprune.github.io/`
- Paper: `https://arxiv.org/abs/2609.39096`

StoryMem: Multi-shot Long Video Storytelling with Memory — Tech Report

***Kaiwen Zhang**, Liming Jiang, Angtian Wang, Jacob Zhiyuan Fang, Tiancheng Zhi, Qing Yan, Hao Kang, Xin Lu, Xingang Pan*

- Project: `https://kevin-thu.github.io/StoryMem/`
- Paper: `https://arxiv.org/abs/2512.19539` **(40+ citations)**
- Code: `https://github.com/Kevin-thu/StoryMem` **(780+ stars)**

Epona: Autoregressive Diffusion World Model for Autonomous Driving — ICCV 2025

Kaiwen Zhang, Zhenyu Tang*, Xiaotao Hu, Xingang Pan, Xiaoyang Guo, Yuan Liu, Jingwei Huang, Li Yuan, Qian Zhang, Xiaoxiao Long, Xun Cao, Wei Yin (*Equal contribution)*

- Project: `https://kevin-thu.github.io/Epona`
- Paper: `https://arxiv.org/abs/2506.24113` **(180+ citations)**
- Code: `https://github.com/Kevin-thu/Epona` **(390+ stars)**

DiffMorpher: Unleashing the Capability of Diffusion Models for Image Morphing — CVPR 2024

***Kaiwen Zhang**, Yifan Zhou, Xudong Xu, Xingang Pan§, Bo Dai (§Corresponding author)*

- Project: `https://kevin-thu.github.io/DiffMorpher_page`
- Paper: `https://arxiv.org/abs/2312.07409` **(130+ citations)**
- Code: `https://github.com/Kevin-thu/DiffMorpher` **(510+ stars)**

A Survey: Learning Embodied Intelligence from Physical Simulators and World Models — arXiv

Xiaoxiao Long, Qingrui Zhao*, Kaiwen Zhang*, Zihao Zhang*, Dingrui Wang*, Yumeng Liu*, Zhengjie Shu*, Yi Lu*, Shouzheng Wang*, Xinzhe Wei*, …, Qionghai Dai (*Equal contribution)*

- Paper: `https://arxiv.org/abs/2507.00917` **(90+ citations)**

Experience

Kling AI, Singapore (Remote) Mar. 2026 – Present

Research Intern Mentors: Dr. Xintao Wang, Yiming Gao

- Led research on *ViDecide*, a decision-centric framework that teaches video diffusion models to reason and plan step by step, generating one state transition at a time and deciding when to stop.
- Proposed *Corrective Fine-tuning (CFT)*, a new post-training method for autoregressive video diffusion. CFT maps the noise of each failed decision to the model's own successful continuation, achieving stable improvement without policy-gradient optimization.
- Introduced *RoboTwin-Plan*, a new benchmark of multi-step dual-arm robot planning. Raised the average strict success rate across abstract games, planning tasks, and robot planning from 23.3% to 83.8%. First-author paper under review.

ByteDance / TikTok, San Jose, US Jul. 2025 – Dec. 2025

Research Intern Mentors: Dr. Liming Jiang, Dr. Xin Lu

- Led research on *StoryMem*, a memory-conditioned video diffusion framework for coherent multi-shot, long-form video storytelling.
- Developed *Memory-to-Video (M2V)* with dynamic keyframe memory, latent conditioning, and negative RoPE shifts to enable consistent and flexible multi-shot generation.
- Introduced *ST-Bench*, featuring 30 multi-scene story scripts and 300 shot-level prompts to evaluate cross-shot consistency, prompt adherence, and visual quality.

Horizon Robotics, Beijing, China Sept. 2024 – Jun. 2025

Research Intern Mentors: Prof. Xiaoxiao Long, Dr. Wei Yin

- Led research on *Epona*, an autoregressive diffusion world model for autonomous driving trained from scratch, enabling long-horizon, trajectory-controllable video generation and end-to-end planning.
- Developed decoupled spatiotemporal modeling and chain-of-forward training for stable long-horizon prediction, with emergent traffic-rule understanding.
- Achieved state-of-the-art world modeling performance and competitive end-to-end trajectory planning (action prediction) results. First-listed co-first author accepted at ICCV 2025.

Shanghai AI Lab, Shanghai, China Jul. 2023 – Jan. 2024

Research Intern Advisor: Prof. Xingang Pan

- Proposed *DiffMorpher*, the first diffusion-based image morphing framework that enables smooth semantic transitions through interpolation in the LoRA parameter and latent space, without explicit correspondence.
- Developed attention interpolation and injection to improve visual fidelity and temporal consistency.
- Introduced *MorphBench*, the first benchmark for general-object image morphing. First-author paper accepted at CVPR 2024.

♡ Honors and Awards

Nanyang President's Graduate Scholarship Jan. 2026 – Present
Outstanding Graduate of Beijing Jun. 2025
Outstanding Graduate of Tsinghua University Jun. 2025
China Computer Federation (CCF) Outstanding Student Sept. 2024
SenseTime Scholarship (25 recipients nationwide) Jun. 2024
National Scholarship (Highest honor for undergraduates in China) Oct. 2023
Tsinghua University Comprehensive Excellence Scholarship Oct. 2022

Service

- **Reviewer:** CVPR, NeurIPS, TPAMI, SIGGRAPH, SIGGRAPH Asia
- **Teaching Assistant:** AI6132 Generative AI for Visual Synthesis, SC1303 Programming with Data Structure and Algorithm, SC1001 Computational Thinking and Programming