

Cheng Chen

cheng021@e.ntu.edu.sg | +65 8898 8363 | [Google Scholar](#)
<https://steller-chen.github.io/> | [LinkedIn](#)

Researcher in generative models of the physical world: models that simulate how scenes, objects and agents change under interaction, and that let robots and agents act on what they simulate. First-author work includes interactive video simulation (DiVA), 3D and CAD generation (CVPR 2024, 2025), world models that turn human video into robot experience (AnyWorld, CoRL 2026), and efficient VLA policies (SwiftVLA, CVPR 2026). Industry research at XPENG Robotics and Ant Group.

EDUCATION

Nanyang Technological University, Singapore | Jan. 2023 - present

Ph.D. in Computer Science. Supervisors: Prof. Guosheng Lin and Dr. Fayao Liu.

Thesis submitted; defense expected December 2026; degree expected January 2027.

Fudan University, Shanghai, China

M.Eng. in Information Engineering, 2019 - 2022; B.Eng. in Electronic Engineering, 2015 - 2019.

SELECTED RESEARCH CONTRIBUTIONS

AnyWorld | First author | CoRL 2026

Turned unpaired human egocentric video into robot-native training experience by factorizing action, camera and embodiment control in a video world model. Used the model to generate targeted counterfactual experience that repaired real-robot policy failures: it removed a systematic false-completion behavior and enabled instruction-conditioned target selection under identical observations. [Project](#) / [paper](#).

Physical Probing for Frontier-Model Robot Agents | First author | In preparation, 2026

Frontier agents such as GPT-6 can plan robot tasks but lack physical properties that are not observable from pixels, such as mass, friction and compliance. Proposed a learned, task-shared probing policy that decides which physical interactions to perform, and when to stop, to gather decision-sufficient evidence for a frozen frontier agent at minimal physical cost. Introduced the Physical Adaptation Cost (PAC) benchmark.

SwiftVLA | Co-first author | CVPR 2026

Lightweight vision-language-action model that injects spatiotemporal geometry representations, connecting visual geometry and temporal dynamics to efficient robot-action generation at minimal overhead. [Paper](#).

XPACE | Co-author | XPENG Robotics technical report, 2026

Company-scale unified model for joint future-video and executable-action prediction and action-conditioned simulation, trained with joint flow matching, chunk-wise autoregressive generation and self-gradient forcing. Built world-model training and the heterogeneous robot and human data pipeline, including full-parameter fine-tuning of 10B+ video models. [Project](#) / [paper](#).

DiVA | First author | Under review at IJCV

Interactive video simulation across multi-turn interactions. Developed Anchored Video Continuation to connect action, waiting and transition segments while preserving identity and spatial consistency, and characterized the trade-off between visual quality and interaction latency. [Paper](#).

RESEARCH EXPERIENCE

XPENG Robotics | Research Intern | Dec. 2025 - present

World-model training and data for XPACE; first-author AnyWorld, from experience generation to real-robot policy interventions.

Ant Group | Research Intern | Apr. 2025 - Oct. 2025

Interactive video generation and editing; first-author DiVA.

Nanyang Technological University | Ph.D. researcher | Jan. 2023 - present

Generative 3D and CAD modeling (Sculpt3D, CADCrafter), then embodied agents: physical probing for frontier-model robot agents, and slow-fast execution for LLM agents acting in real time.

Research engineering: Full-parameter fine-tuning of 10B+ video models; multi-node distributed training.

Available: February 2027.

SELECTED PUBLICATIONS AND PREPRINTS

Full list on [Google Scholar](#).

AnyWorld: Factorized Egocentric World Models for Cross-Embodiment Generalization | First author

Cheng Chen et al. *CoRL* 2026.

SwiftVLA: Unlocking Spatiotemporal Dynamics for Lightweight VLA Models at Minimal Overhead | Co-first author

Chaojun Ni, **Cheng Chen** (co-first author), et al. *CVPR* 2026.

XPACE: Joint World and Action Modeling from Heterogeneous Experience | Co-author

Jiacheng Wei, Jerry Bai, Xiaoyu Yue, Zidong Wang, Xiaoyang Guo, **Cheng Chen**, et al. *arXiv preprint*, 2026.

DiVA: Enabling Interactive Digital Life Simulation via Video Models | First author

Cheng Chen et al. *arXiv preprint*, 2026; *under review at IJCV*.

CADCrafter: Generating Computer-Aided Design Models from Unconstrained Images | First author

Cheng Chen et al. *CVPR* 2025.

Text-to-Image Rectified Flow as Plug-and-Play Priors | Co-author

Xiaofeng Yang, **Cheng Chen**, Xulei Yang, Fayao Liu, Guosheng Lin. *ICLR* 2025.

Sculpt3D: Multi-View Consistent Text-to-3D Generation with Sparse 3D Prior | First author

Cheng Chen et al. *CVPR* 2024.

Learn to Optimize Denoising Scores: A Unified and Improved Diffusion Prior for 3D Generation | Co-author

Xiaofeng Yang, Yiwen Chen, **Cheng Chen**, et al. *ECCV* 2024.

UTC: A Unified Transformer with Inter-Task Contrastive Learning for Visual Dialog | First author

Cheng Chen et al. *CVPR* 2022.

TALKS, SERVICE AND MENTORING

Invited talks: AnyWorld and XPACE, 具身智能之心, 2026; AnyWorld, AnySyn3D, 2026.

Top Reviewer, NeurIPS 2026. **Conference reviewer:** NeurIPS, ICLR, CVPR, ICML, ICCV, ECCV, ACM MM, CoRL.

Mentoring: mentoring junior researchers and students on video generation and embodied-agent research projects.

AWARDS AND HONOURS

SINGA Scholarship, Singapore International Graduate Award	Jan. 2023
Outstanding Graduate of Fudan University	Apr. 2022
National Scholarship (1/72)	Oct. 2021
First-Class Scholarship, Fudan University	Oct. 2019
Outstanding Student, Information School of Fudan University	May 2017

SKILLS

Research: Video world models, interactive video simulation, 3D/CAD generation; human-to-robot transfer, VLA policies, real-robot evaluation; frontier-model embodied agents.

Training: Large video-model fine-tuning, multi-node distributed training, heterogeneous robot and human data pipelines, autoregressive generated-context learning.

Tools: Python, PyTorch, TensorFlow, Linux, LaTeX. **Languages:** Mandarin Chinese, English.