

HAODONG LI

<https://haodong2000.github.io>

EDUCATION

University of California San Diego

09/2025 – Now

Department of Computer Science and Engineering

La Jolla, California

- **Doctor of Philosophy**
- Supervisor: Prof. [Manmohan Chandraker](#)
- Research topic: **Video Generation, 3D/Video World Models**
- Grades: 4.0/4.0, Credits: 37.0

Hong Kong University of Science and Technology

09/2023 – 06/2025

Information Hub, Guangzhou Campus

Clear Water Bay, Hong Kong & Guangzhou, China

- **Master of Philosophy**
- Supervisor: Prof. [Ying-Cong Chen](#)
- Research topic: **3D/Image Generation, 3D Perception**
- Grades: 3.7/4.0, Credits: 21.0

Zhejiang University

09/2019 – 06/2023

College of Control Science and Engineering

Hangzhou, China

- **Bachelor of Engineering**
- Grades: 3.7/4.0, Credits: 191.0

WORKING EXPERIENCE

Tencent Hunyuan

05/2025 – 09/2025

Center of 3D Generation

Shenzhen, China

- **Research Scientist Intern**
- Mentor: Dr. [Chunchao Guo](#)
- Research Topic: **3D Perception**

SELECTED PUBLICATIONS (FULL LIST IN [GOOGLE SCHOLAR](#), *EQUAL CONTRIBUTION)

Rolling Sink: Bridging Limited-Horizon Training and Open-Ended Testing in Autoregressive Video Diffusion

<https://rolling-sink.github.io>

arXiv 2026

Haodong Li, Shaoteng Liu, Zhe Lin, Manmohan Chandraker

TL;DR: Built on Self Forcing (trained on only 5s clips), Rolling Sink effectively scales the autoregressive video synthesis to ultra-long durations (e.g., 5-30min) at test time, with consistent subjects, stable colors, and smooth motions.

DA²: Depth Anything in Any Direction

<https://depth-any-in-any-dir.github.io>

ICLR 2026

Haodong Li, Wangguangdong Zheng, Jing He, Yuhao Liu, Xin Lin, Xin Yang, Ying-Cong Chen, Chunchao Guo

TL;DR: Powered by large-scale curated data and a sphere-aware ViT, DA² predicts dense distance from a single 360° panorama in an end-to-end manner, with remarkable geometric fidelity and strong zero-shot generalization.

Lotus-2: Advancing Geometric Dense Prediction with Powerful Image Generative Model

<https://lotus-2.github.io>

arXiv 2025

Jing He, **Haodong Li**, Mingzhi Sheng, Ying-Cong Chen

TL;DR: Lotus-2 is an advanced monocular geometric estimator built upon FLUX. By effectively analyzing the DiT-based rectified-flow formulation, Lotus-2 achieves SoTA performance while producing significantly finer details.

Lotus: Diffusion-based Visual Foundation Model for High-quality Dense Prediction

<https://lotus3d.github.io>

ICLR 2025

Jing He*, **Haodong Li***, Wei Yin, Yixun Liang, Leheng Li, Kaiqiang Zhou, Hongbo Zhang, Bingbing Liu, Ying-Cong Chen

TL;DR: Based on Stable Diffusion, Lotus delivers SoTA performance on monocular depth & normal estimation with a simple yet effective fine-tuning protocol that better fits the pre-trained visual prior for dense prediction.

DisEnvisioner: Disentangled and Enriched Visual Prompt for Customized Image Generation

<https://disenvisioner.github.io>

ICLR 2025

Jing He*, **Haodong Li***, Yongzhe Hu, Guibao Shen, Yingjie Cai, Weichao Qiu, Ying-Cong Chen

TL;DR: DisEnvisioner effectively identifies and enhances the subject-essential features while filtering out other irrelevant ones, enabling exceptional image customization in a tuning-free manner with only a single image.

DIScene: Object Decoupling and Interaction Modeling for Complex Scene Generation

<https://dl.acm.org/doi/10.1145/3680528.3687589>

SIGGRAPH Aisa 2024

Xiao-Lei Li*, **Haodong Li***, Hao-Xiang Chen, Tai-Jiang Mu, Shi-Min Hu

TL;DR: DIScene is capable of generating complex, high-fidelity 3D scene with decoupled objects and clear interactions, through a learnable scene graph and hybrid Mesh-Gaussian representation.

Bi-TTA: Bidirectional Test-Time Adapter for Remote Physiological Measurement

<https://bi-tta.github.io>

ECCV 2024

Haodong Li, Hao Lu, Ying-Cong Chen

TL;DR: Bi-TTA leverages spatial and temporal consistency with novel prospective and retrospective strategies, enabling pre-trained rPPG models to adapt effectively to target domains using only unannotated instance-level data.

LucidDreamer: Towards High-Fidelity Text-to-3D Generation via Interval Score Matching

<https://github.com/envision-research/luciddreamer>

CVPR 2024 Highlight

Yixun Liang*, Xin Yang*, Jiantao Lin, **Haodong Li**, Xiaogang Xu, Ying-Cong Chen

TL;DR: LucidDreamer is a text-to-3D generation framework that distills high-fidelity textures and shapes represented by 3D Gaussians from pre-trained Stable Diffusion with a novel Interval Score Matching objective.

ACADEMIC SERVICE

Reviewer: CVPR 2025 (1), ICLR 2026 (5), CVPR 2026 (2), ICML 2026 (6), SIGGRAPH 2026 (1), ECCV 2026, NeurIPS 2026.