

# Haoyi Duan

✉ haoyid@stanford.edu ·  ·  page ·  haoyi-duan

## Education Background



**Stanford University, California, USA** 09/2023 - Present  
M.S. Electrical Engineering (EE)  
**Zhejiang University, Zhejiang, China** 09/2019 - 06/2023  
B.Eng. Computer Science and Technology (CS), **Chu Kochen Honors College**  
· GPA: **3.97/4.00 (91.38/100)**  
· Major GPA: **3.99/4.00 (93.17/100)**  
**National University of Singapore, Singapore** 05/2022 - 07/2022  
School of Computing Summer Workshop   
· GPA: A<sup>+</sup>, Real-Time Graphics Rendering

## Research Interests




3D Vision; Computer Vision; Multimodal Learning.

My research interests primarily lies in **Computer Vision**, with a current focus on 3D/4D visual reconstruction and generation. I am particularly exploring the capabilities of generative models in both generation and perception scenarios. I also have a background in **Multimodal Learning**, working across various modalities, including 3D, video, audio, and language.



## Publications

**WonderWorld: Interactive 3D Scene Generation from a Single Image**   *CVPR'2025 submitted*

Hong-Xing (Koven) Yu\*, **Haoyi Duan\***, Charles Herrmann, William T. Freeman, Jiajun Wu

**WonderJourney: Going from Anywhere to Everywhere**    *CVPR'2024*

Hong-Xing (Koven) Yu, **Haoyi Duan**, Junhwa Hur, Michael Rubinstein, William T. Freeman, Forrester Cole, Deqing Sun, Noah Snavely, Jiajun Wu, Charles Herrmann



**Cross-modal Prompts: Adapting Large Pre-trained Models for Audio-Visual Downstream Tasks**   *NeurIPS'2023*

**Haoyi Duan\***, Yan Xia\*, Mingze Zhou, Li Tang, Jieming Zhu, Zhou Zhao

 **Outstanding Graduation Thesis, Zhejiang University**

**LaPE: Layer-adaptive Position Embedding for Vision Transformers with Independent Layer Normalization**  *ICCV'2023*

Runyi Yu\*, Zhennan Wang\*, Yinhuai Wang\*, Kehan Li, Chang Liu, **Haoyi Duan**, Xiangyang Ji, Jie Chen

**Timestamps as Prompts for Geography-Aware Location Recommendation**   *CIKM'2023*

Yan Luo, **Haoyi Duan**, Ye Liu, Chung Fu-Lai

**Beyond Two-Tower Matching: Learning Sparse Retrievable Interaction Models for Recommendation**  *SIGIR'2023*

Liangcai Su, Fan Yan, Jieming Zhu, Xi Xiao, **Haoyi Duan**, Zhou Zhao, Zhenhua Dong, Ruiming Tang

## Selected Research Experience

**WonderWorld: Interactive 3D Scene Generation from a Single Image** 01/2024 - 09/2024  
Advisor: Prof. Jiajun Wu, Stanford Vision & Learning Lab

· Presented a novel framework for interactive 3D scene generation that enables users to interactively specify scene contents and layout and see the created scenes in low latency.

· Introduced the Fast Layered Gaussian Surfels (FLAGS) as the scene representation and an algorithm

to generate it from a single view.

- Employed the guided depth diffusion that allows partial conditioning of depth estimation.

### **WonderJourney: Going from Anywhere to Everywhere**

09/2023 - 01/2024

Advisor: Prof. Jiajun Wu, Stanford Vision & Learning Lab

- Designed a modularized framework, which starts at any user-provided location and generate a journey through a long sequence of diverse yet coherently connected 3D scenes.
- Leveraged an LLM to generate textual descriptions of the scenes in this journey, a text-driven point cloud generation pipeline to make a compelling and coherent sequence of 3D scenes, and a large VLM to verify the generated scenes.

### **Adapting Large Pre-trained Models for Audio-Visual Downstream Tasks**

12/2022 - 08/2023

Advisor: Prof. Zhou Zhao, Zhejiang University DCD Lab

- Proposed a novel Dual-Guided Spatial-Channel-Temporal attention mechanism, which leverages audio and visual modalities as soft prompts to dynamically adjust the parameters of pre-trained models based on the current multi-modal input features.
- Achieved state-of-the-art results across multiple downstream tasks; exhibited promising performance in challenging few-shot and zero-shot scenarios.

## **Honors & Awards**

---

National Scholarship (top 0.2%)	2022
First-Class Scholarship of Zhejiang University (top 3%)	2020, 2021, 2022
Outstanding Graduation Thesis of Zhejiang University	2023
Zhejiang Provincial Outstanding Graduate	2023
Outstanding Graduate of Zhejiang University	2023

## **Languages & Skills**

---

- Languages: English (proficient); Chinese (native). TOEFL: 108(S23); GRE: (V155+Q169+3.5).
- Programming: C/C++, Python, CUDA, mini SQL, JavaScript, Verilog, x86, HTML.
- Tools/Software: Blender, Unity, Unreal Engine, Pytorch, OpenGL, MySQL, L<sup>A</sup>T<sub>E</sub>X, Vivado, Adobe.