Zhiding Yu

PERSONAL INFORMATION

Tel: (850) 567-2583 | Office: 2788 San Tomas Expy, Santa Clara, CA 95051
Email: scutchrisding@gmail.com | Homepage: https://chrisding.github.io

RESEARCH INTERESTS

I am interested in building general autonomy and intelligence across both virtual and physical domains. My recent focus lies in Vision Transformers, LLMs, multimodal LLMs, and vision-language-action (VLA) models, with applications spanning open-world understanding, reasoning, AV/robot perception-planning, and agentic systems. I have led or contributed to numerous flagship research efforts and products at NVIDIA, including **SegFormer** (Demo), **VoxFormer/FB-BEV** (CVPR23 3D Occ Pred Challenge winner, video), **Hydra-MDP** (CVPR24 E2E Driving Challenge winner, video), the **Eagle VLM** project, **Nemotron**, **Llama-Nemotron-VL**, and **GR00T N1/N1.5** (NVIDIA's foundation models for humanoid robots). I also participated in designing NVIDIA's next-generation end-to-end autonomous driving system. My works are characterized by state-of-the-art performance, scalable architectures, and data-centric strategies towards real-world generalization.

WORK EXPERIENCE

NVIDIA Research — Santa Clara, CA

Principal Research Scientist & Research Lead — 2024 - Present

- I conduct research in multimodal learning and intelligent data strategies. I lead the **Eagle** project which develops a family of frontier vision-language models with public training/data recipes and state-of-the-art performance matching or outperforming existing top-tier VLMs. Our work has laid the core VLM foundation and data strategy behind several flagship NVIDIA products/projects, including Llama-Nemotron-VL, Nemo Retriever Multimodal Embedding, GR00T N1, and GR00T N1.5.

Staff Research Scientist — 2023 - 2024

- Participated in a multi-org effort to design NVIDIA's next-generation AV system. Led a team to design and develop a Transformer-based 3D perception system for joint 3D object detection, tracking and online mapping at long distances. Developed Transformer-based neural planner and DriveVLM for E2E driving, with community-recognized works such as BEV-Planner, OmniDrive, and Hydra-MDP.

Senior Research Scientist — 2020 - 2023

- Led the earliest effort to develop Vision Transformers at NVIDIA which partially shaped the landscape of NVIDIA's internal AI product. Applications of my work include scene understanding, robust general purpose backbone, autonomous driving perception and scalable auto-labeling pipelines. Proposed multiple Transformer-based bird's-eye view (BEV) perception frameworks with SOTA results in 3D object detection, tracking, and 3D occupancy prediction. Some works from this period with community impact include SegFormer (Most Influential NeurIPS Papers, 3K Stars), VoxFormer (CVPR23 Highlight, 1.1K Stars), FB-BEV/FB-OCC (ICCV23, 735 Stars), and FocalFormer3D (ICCV23, Ranked 1st on nuScenes LiDAR 3D Detection and Tracking Leaderboard (Mar. 2023)).
- Successful tech transfers to numerous NVIDIA products, including AV and NVIDIA TAO Toolkit.

Research Scientist — 2018 - 2020

- Worked extensively in label-efficient learning and transfer learning. Proposed weakly supervised, semi-supervised and self-supervised learning frameworks, with SOTA performance in visual recognition applications. Proposed various unsupervised domain adaptation and synthetic-to-real generalization methods for improved model robustness and generalization "in the wild".

Mitsubishi Electric Research Laboratories — Cambridge, MA

Research Intern, Computer Vision Group — 07/11/16 - 11/18/16

Microsoft Research — Redmond, WA

Research Intern, Multimedia, Interaction, and Communication Group — 05/25/15 - 08/28/15

Adobe Research — San Jose, CA

Research Intern, Computer Vision Group — 06/03/13 - 08/30/13

EDUCATION

Carnegie Mellon University (CMU) — Pittsburgh, PA

Ph.D. in Electrical & Computer Engineering — 2012 - 2017

Hong Kong University of Science & Technology (HKUST) — Hong Kong

M.Phil. in Electronic & Computer Engineering — 2009 - 2012

South China University of Technology (SCUT) — Guangzhou, China

B.Eng. Information Engineering (Talented Student Program) — 2005 - 2008

HONORS & AWARDS

- Winner, CVPR24 Challenge on End-to-End Driving at Scale 2024
- 2nd Place, CVPR24 Challenge on Driving with Language 2024
- Winner, CVPR23 Challenge on 3D Occupancy Prediction 2023
- Winner, ECCV22 Robust Vision Challenge (RVC) on Semantic Segmentation 2022
- Winner, CVPR18 Autonomous Driving Challenge (WAD) on Domain Adaptation 2018
- 2nd Place, ICMI15 EmotiW Challenge on Static Facial Expression Recognition 2015
- Best Paper Award, BMVC20 2020
- Best Paper Award, WACV15 2015
- Best Student Paper Award, ISCSLP14 2014
- Carnegie Institute of Technology Dean's Tuition Fellowship, CMU 2012
- HKTIIT Post-Graduate Excellence Scholarship, HKUST 2012
- HKTIIT Post-Graduate Excellence Scholarship, HKUST 2010
- Research Postgraduate Studentship, HKUST 2009-2012

PUBLICATIONS

Please refer to Google Scholar for the list of my latest publications.