Shiyi Lan

PERSONAL INFORMATION

Tel: (240) 701-2979
Email: voidrank@gmail.com Google Scholar

RESEARCH INTERESTS

Foundation models and agents, focusing on efficient architectures and data curation for capability development through pretraining, target-domain mid-training, reward modeling, and reinforcement learning.

WORK EXPERIENCE

Waymo LLC Mountain View, CA 94043
Staff Research Engineer 7/27/2026 - present

- **Waymo Gemini.** Pretraining, target-domain mid-training, and RL post-training for end-to-end autonomous driving.
- **Reward modeling.** Led critic-based reward design under teacher-student information asymmetry. Distinguished outcome correctness from reasoning grounded in student-observable evidence to avoid reinforcing unsupported reasoning. Developed format and process rewards.
- **Accuracy-efficiency co-design.** Co-optimized pretraining architectures and kernels for task accuracy and end-to-end inference latency on target devices, developing and evaluating Gated DeltaNet and MuxNet variants, state experts, and multi-exit designs. Guided architectural decisions through empirical evaluation of model quality, training stability, and on-device efficiency.

NVIDIA Research Santa Clara, CA 95051
Senior Research Scientist 3/1/2023 - 7/21/2026

- **Multimodal Foundation Models.** Core member of **Nemotron-Diffusion-VL** (NVIDIA's first diffusion VLM), achieving state-of-the-art accuracy (MMMU, MMBench, MMMU-Pro-COT, MathVista). Developed training data recipes, model and evaluation implementations, and training infrastructure. Contributed to **Eagle 2** (post-training data strategies for frontier VLMs) and the **Cosmos** world foundation model platform for physical AI.
- **End-to-End Driving / Perception-Planning.** Led the development of the HydraMDP family (HydraMDP, Hydra-MDP++, Hydra-Next, DriveSuprim, HAD), building the full perception-planning loop from 3D scene understanding to multimodal trajectory planning via hydra-distillation and RL. **Led OmniDrive**, a holistic VLM-based driving framework integrating perception and planning with counterfactual reasoning. Contributed to **FB-OCC** (3D occupancy prediction) and **FocalFormer3D** (3D object detection) for robust scene understanding.
- **Challenges & Recognition. 1st place**, CVPR 2025 End-to-End Driving Challenge (Team Lead); **1st place**, CVPR 2024 End-to-End Driving Challenge (Team Lead); **1st place**, CVPR 2023 3D Occupancy Challenge. Featured in **NVIDIA GTC Keynote**; cited in **NVIDIA Earnings Call**.

NVIDIA Research Santa Clara, CA 95051
Research Scientist 05/20/2022 - 3/1/2023
(Manager: Anima Anandkumar)

- Led **NV-CLIP** pretraining, resolved training loss divergence, and transferred the model to the product team.
- Developed **MMAL** (Vision Transformers Are Good Mask Auto-Labelers, CVPR 2023), achieving 1st place accuracy in weakly-supervised instance segmentation on MS COCO.
- **1st place** in Robust Vision Challenge (RVC) 2022 Semantic Segmentation Track.

NVIDIA Research Santa Clara, CA
Research Intern, Machine Learning Group 01/27/2020 - 12/20/2020
(Manager: Anima Anandkumar)

- Drove independent research on weakly-supervised instance segmentation on MS COCO, achieving 1st place accuracy.

Bytedance AI Lab Beijing, China
Research Intern, AI Lab (Mentor: Yuning Jiang) 05/06/2018 - 08/18/2018

- Core team member of a warmup recommendation system that saves 5 million CNY per day in warmup costs.

EDUCATION

University of Maryland, College Park College Park, MD
Ph.D. Computer Science (Advisor: Prof. Larry S. Davis) 2018-2022

Fudan University Shanghai, China
B.S. Computer Science and Technology 2014-2018

SELECTED HONORS

- **1st place** in CVPR 2023 3D Occupancy Challenge
- **1st place** in Robust Vision Challenge (RVC) 2022 Semantic Segmentation Track
- **1st place** in MS COCO 2017 Object Detection, 2nd place in MS COCO Instance Segmentation

SELECTED PUBLICATIONS

1. W Yao, X Sun, Z Li, **S. Lan**, Z Wang, J Alvarez, Z Wu, HAD: Combining Hierarchical Diffusion with Metric-Decoupled RL for End-to-End Driving. Under submission.
2. W Yao, Z Li, **S Lan**, Z Wang, X Sun, JM Alvarez, Z Wu, DriveSuprim: Towards Precise Trajectory Selection for End-to-End Planning, *European Conf. on Comp Vision (arXiv preprint arXiv:2506.06659 1 2025)*.
3. K Li, Z Li, **S Lan**, Y Xie, Z Zhang, J Liu, Z Wu, Z Yu, JM Alvarez, Hydra-mdp++: Advancing end-to-end driving via expert-guided hydra-distillation, *arXiv preprint arXiv:2503.18280* 2025.
4. Z Li, S Wang, **S Lan**, Z Yu, Z Wu, JM Alvarez, Hydra-next: Robust closed-loop driving with open-loop training, *arXiv preprint arXiv:2503.12030* 2025.
5. Z Li, G Chen, S Liu, S Wang, V VS, Y Ji, **S Lan**, H Zhang, Y Zhao, Eagle 2: Building post-training data strategies from scratch for frontier vision-language models, *arXiv preprint arXiv:2501.14818* 2025.
6. N Agarwal, A Ali, M Bala, Y Balaji, E Barker, T Cai, P Chattopadhyay, S Lan, Cosmos world foundation model platform for physical ai, *arXiv preprint arXiv:2501.03575* 2025.
7. S Wang, Z Yu, X Jiang, **S Lan**, M Shi, N Chang, J Kautz, Y Li, JM Alvarez, Omnidrive: A holistic vision-language dataset for autonomous driving with counterfactual reasoning, *Proc. of the CVPR, 22442* 2025.
8. Z Li, K Li, S Wang, **S Lan**, Z Yu, Y Ji, Z Li, Z Zhu, J Kautz, Z Wu, YG Jiang, JM Alvarez, "Hydra-MDP: End-to-end Multimodal Planning with Multi-target Hydra-Distillation",
arXiv preprint arXiv:2406.06978
9. Z Li, Z Yu, D Austin, M Fang, **S Lan**, J Kautz, JM Alvarez. "FB-OCC: 3D Occupancy Prediction based on Forward-Backward View Transformation", *IEE Conf. on Comp Vision and Pattern Recognition (**CVPR**)*, 2024
10. J Xiao, Z Xu, **S Lan**, Z Yu, A Yuille, A Anandkumar, "1st Place Solution of The Robust Vision Challenge (RVC) 2022 Semantic Segmentation Track",
Arxiv preprint (https://arxiv.org/abs/2210.12852)
11. Y Chen, Z Yu, Y Chen, Shiyi Lan, Anima Anandkumar, Jiaya Jia, Jose M Alvarez, "Focalformer3d: focusing on hard instance for 3d object detection", *IEE Conf. on Comp Vision and Pattern Recognition (**CVPR**)*, 2024
12. Shiyi Lan, Xitong Yang, Zhiding Yu, Zuxuan Wu, Jose M. Alvarez, Anima Anandkumar, "Vision Transformers Are Good Mask Auto-Labelers", *IEEE Conf. on Comp Vision and Pattern Recognition (**CVPR**)*, 2023.
13. Lingchen Meng, Hengduo Li, Bor-Chun Chen, **Shiyi Lan**, Zuxuan Wu, Yu-Gang Jiang, Ser-Nam Lim, "AdaViT: Adaptive Vision Transformers for Efficient Image Recognition", *IEEE Conf. on Comp Vision and Pattern Recognition (**CVPR**)*, 2022.
14. **Shiyi Lan**, Zhou Ren, Yi Wu, Larry S Davis, Gang Hua, "SaccadeNet: A Fast and Accurate Object Detector" *IEEE Conf. on Comp Vision and Pattern Recognition (**CVPR**)*, 2020. This paper proposed a fast and accurate keypoint-based object detector, which achieves the state-of-the-art performance on the MS COCO dataset.
15. **Shiyi Lan**, Ruichi Yu, Gang Yu, Larry S Davis, "Modeling Local Geometric Structure of 3D Point Clouds using Geo-CNN", *IEEE Conf. on Comp Vision and Pattern Recognition (**CVPR**)*, 2019. This paper proposed a convolution-like operator for PointNet, which preserves local geometric relationships among points using a decomposition and aggregation module.
16. *Hexiang Hu, ***Shiyi Lan**, Yuning Jiang, Zhimin Cao, Fei Sha. "FastMask: Segment Multi-scale Object Candidates in One Shot" *IEEE Conf. on Comp Vision and Pattern Recognition (**CVPR**)*, 2017, Spotlight. It enables multi-scale object segmentation to be executed in one-shot.

For more information, please visit my google scholar