Zhe Liu

Postdoctoral Researcher, Department of Computer Science,
The Unversity of Hong Kong
Email: zheliu12@hku.hk URL: Google Scholar HomePage: MyPage

(i) Academic qualifications:

- 2025-Now, Postdoctoral researcher at the Department of Computer Science of the University of Hong Kong under the supervision of Prof. Hengshuang Zhao
- 2020-2025, PhD in the School of Electronic Information and Communications at Huazhong University of Science and Technology under the supervision of Prof. Xiang Bai

(ii) Previous academic positions held:

- Jan. 2022 – Dec. 2024 Intern at Baidu Vision Technology Department and research on multi-modal 3D object perception in autonomous driving
- Sep. 2020 - Dec. 2020 Intern at Huawei Noah's Ark and research on the design of end-to-end 3D object tracking algorithms

(iii) Present academic position held:

I received my Ph.D. from the School of Electronic Information and Communications, Huazhong University of Science and Technology, by Prof. Xiang-Bai (**Distinguished Young Scholars of China**). My previous research lies in 3D perception, prediction, planning based on multimodal fusion in autonomous driving. I have accumulated eight years of research experience in autonomous driving from 2018 to 2026. Besides, I have published more than 30 papers in top-tier journals and conferences, including TPAMI, TIP, ECCV, NeurIPS, CVPR, ICCV, AAAI, TITS, and ICRA. As of June 2026, my Google Scholar citations exceed **2,200**, with an h-index of 16.

Recently, my research has shifted toward **Embodied AI,** where I have accumulated one and a half years of research experience. Building on my expertise in 3D perception and autonomous driving, I aim to develop **general-purpose embodied intelligence systems** that integrate data, models, simulation, evaluation, and deployment. My current research focuses on:

1. **Embodied Data Engine**, including ego-centric human data collection, human-to-robot data transformation, and corner-case data feedback for scalable robotic learning;
2. **Vision-Language-Action Models, World-Action Models, and VLN Systems**, enabling embodied agents to understand multimodal instructions, perform spatial-temporal reasoning, and generate executable actions for manipulation, navigation, and long-horizon tasks;
3. **4D World Models and Embodied Simulation**, focusing on physical and attention-based world modeling for virtual evaluation, policy learning, and sim-to-real transfer;
4. **Efficient and Streaming Embodied Inference**, including model distillation, acceleration, and streaming policy inference for real-time robotic deployment;
5. **Intelligent Robotic OS and Cross-Embodiment Deployment**, developing unified input/output interfaces and foundation-model-driven robotic systems for manipulation, navigation, and cross-embodiment generalization.

(iv) Publication records: *(*: Equal contribution, +:Corresponding Author)*

Section A:

1. **Zhe Liu**, Tengteng Huang,Bingling Li,Xiwu Chen, Xi Wang, Xiang Bai. EPNet++: Cascade Bi-directional Fusion for Multi-Modal 3D Object Detection[J]. **IEEE TPAMI 2022.** (**Google Citation>200**)
2. **Zhe Liu,** Xin Zhao, Tengteng Huang, Ruolan Hu, Yu Zhou, Xiang Bai. TANet: Robust 3d object detection from point clouds with triple attention[C]. **AAAI 2020**. **(Oral, acceptance rate<5%, Google Citation>500)**
3. Tengteng Huang*, **Zhe Liu***, Xiwu Chen, Xiang Bai. EPNet: Enhancing point features with image semantics for 3d object detection[C]. **ECCV 2020**. (**Google Citation>650**)

4. **Zhe Liu,** Xiaoqing Ye, Zhikang Zou, Xinwei He, Xiao Tan, Errui Ding, Jingdong Wang, Xiang Ba, Multi-Modal 3D Object Detection by Box Matching[J]. **IEEE TITS 2024.(Q1)**
5. **Zhe Liu**, Jinghua Hou, Xinyu Wang, Xiaoqing Ye, Jingdong Wang, Hengshuang Zhao, Xiang Bai. LION: Linear Group RNN for 3D Object Detection in Point Clouds[C]. **NeurIPS 2024.**

Section B:（† : Project Leader, * : Co-first author, +: Corresponding author）

1. **Zhe Liu**, Runhui Huang, Rui Yang, Siming Yan, Zining Wang, Lu Hou, Di Lin, Xiang Bai, Hengshuang Zhao. DrivePI: Spatial-aware 4D MLLM for Unified Autonomous Driving Understanding, Perception, Prediction and Planning[C]. **CVPR 2026**. (VLA, 4D MLLM)
2. **Zhe Liu,** Jinghua Hou, et al. StreamPI: Streaming Multimodal Temporal Modeling for Vision-Language-Action Models. In submission 2026/05. (VLA)
3. Yuxiang Lu, **Zhe Liu**†, Xianzhe Fan, Zhenya Yang, Jinghua Hou, Junyi Li, Kaixin Ding, Hengshuang Zhao. FASTER: Rethinking Real-Time Flow VLAs[J]. arXiv:2603.19199, 2026. (VLA)
4. Ziyang Gong*, Zehang Luo*, Anke Tang*, **Zhe Liu***, Shi Fu, Zhi Hou, Ganlin Yang, Weiyun Wang, Xiaofeng Wang, Jianbo Liu, et al. ACE-Brain-0: Spatial Intelligence as a Shared Scaffold for Universal Embodiments[J]. arXiv preprint arXiv:2603.03198, 2026. (Embodied Brain)
5. Zhenya Yang, **Zhe Liu**†, Yuxiang Lu, Liping Hou, Chenxuan Miao, Siyi Peng, Bailan Feng, Xiang Bai, Hengshuang Zhao. GenieDrive: Towards Physics-Aware Driving World Model with 4D Occupancy Guided Video Generation[C]. **CVPR 2026**. (4D World Model)
6. Jingyu Li*, **Zhe Liu***, Wenxiao Wu, Li Zhang. MCNav: Memory-Aware Dynamic Cognitive Map for Zero-shot Goal-oriented Navigation[J]. arXiv:2605.19594, 2026. (VLN)
7. **Zhe Liu**, Jiachen Hou, Xuan Ye, Jingwen Wang, Heng Zhao, Xiang Bai. UniLion: Towards Unified Autonomous Driving Model with Linear Group RNNs[J]. arXiv:2511.01768, 2025. (Unified Model)
8. **Zhe Liu,** Jinghua Hou, Xiaoqing Ye, Tong Wang, Jingdong Wang, Xiang Bai, SEED: A Simple and Effective 3D DETR in Point Clouds[C]. **ECCV 2024**.
9. **Zhe Liu,** Xiaoqing Ye, Xiao Tan, Errui Ding, Xiang Bai. StereoDistill: Pick the Cream from LiDAR for Distilling Stereo-based 3D Object Detection[C]. **AAAI 2023**.
10. Hou, Jinghua*, **Zhe Liu***, Zhikang Zou, Xiaoqing Ye, Xiang Ba. Query-based Temporal Fusion with Explicit Motion for 3D Object Detection[C]. **NeurIPS 2023**.
11. Jingyu Li*, **Zhe Liu***, Jinghua Hou, Dingkang Liang. Dds3d: Dense pseudo-labels with dynamic threshold for semi-supervised 3d object detection[C]. **ICRA 2023.**
12. Runhui Huang, Jie Wu, Rui Yang, **Zhe Liu**, Hengshuang Zhao. AlphaGRPO: Unlocking Self-Reflective Multimodal Generation in UMMs via Decompositional Verifiable Reward[C]. **ICML 2026**.
13. Xin Zhao, **Zhe Liu+**, Ruolan Hu, Kaiqi Huang. 3D object detection using scale invariant and feature reweighting networks[C]. **AAAI 2019**.
14. Dingyuan Zhang, Dingkang Liang, Zhikang Zou, Jingyu Li, Xiaoqing Ye, **Zhe Liu**, Xiao Tan, Xiang Bai, A Simple Vision Transformer for Weakly Semi-supervised 3D Object Detection[C]. **ICCV 2023.**

(v) Professional qualifications, prizes and awards:

- Fall 2019: **Master National Scholarship Winner**, School of Artificial Intelligence and Automation, HUST.
- Spring 2020: **"NavInfo" Scholarship Winner**. **(Top 3 of 300+).**

(vi) Research output:

- Three-dimensional object detection method and system based on weighted channel features of point cloud. *Zhao Xin, Huang Kaiqi,* ***Liu Zhe.*** **(U.S. Patent)**
- A robust three-dimensional object detection method based on ternary attention mechanism. *Bai Xiang,* ***Liu Zhe****, Zhou Yu, Huang Tengteng.* **(Chineses Patent, Authorized)**
- A three-dimensional object detection method, device and system based on DETR structure. *Bai Xiang,* ***Liu Zhe****, Hou Jinghua.* **(Chineses Patent, Submitted)**
- A three-dimensional object detection method, device and system. *Bai Xiang,* ***Liu Zhe****, Hou Jinghua, Wang Xinyu, Zhao Hengshuang.* **(Chineses Patent, Submitted)**

(vii) Leading Projects (2025-2026):

1. **AI Glasses:** Exploring efficient spatial intelligence on edge devices to achieve fast and accurate 3D object perception, understanding, and memory retrieval (with Huawei)
2. **4D MLLM:** Unifying vision-action and vision-language-action frameworks into one 4D MLLM (one

model), surpassing existing 3D perception methods and VLA approaches with only a 0.5B model. (with Huawei)

3. **4D World Model:** Leveraging 4D occupancy for latent representation to enable high-resolution, long-term, physical-aware video generation. (with Huawei)
4. **Attention-based World Model + RL:** Achieving efficient world model on autonomous driving (with Changan)
5. **Streaming VLA:** A Multimodal Streaming Framework for Memory-Dependent and Spatially-Aware Vision-Language-Action Models (with Ace robotics)
6. **Ongoing:** Efficient world model, efficient VLA, Embodied OS (physical agent)

(viii) Academic services:

- ***Conference Reviewers:***

  The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)
  IEEE International Conference on Computer Vision (ICCV)
  European Conference on Computer Vision (ECCV)
  International Conference on Learning Representations (ICLR)
  Conference on Neural Information Processing Systems (NeurIPS）
  The Association for the Advancement of Artificial Intelligence (AAAI)
  International Joint Conference on Artificial Intelligence (IJCAI）
  IEEE/RSJ International Conference on Robotics and Automation (ICRA)
  International Conference on Acoustics, Speech and Signal Processing (ICASSP)
  ACM Multimedia Asia 2023 (ACMMM Asia)

- ***Selected Journal Reviewers:***

  IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)
  IEEE Transactions on Image Processing (IEEE TIP)
  IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT)
  IEEE Transactions on Intelligent Transportation Systems (IEEE TITS)
  IEEE Robotics and Automation Letters (IEEE RA-L)
  Science China-Information Sciences (SCIS)
  IEEE Transactions on Geoscience and Remote Sensing (IEEE TGRS)

- ***Invited Talks:***

  1."Jiqizhixin" invites to share our paper of TANet, 2020.
  2.“Midea Research Institute” invites to share the topic of “3D Object Detection in Point Cloud”, 2020.
  3.“3D CVer” invites to share our paper of LION, 2024.
  4.“Shuzihuanyu” invites to share our paper of LION and SEED, 2024.
  5.“The Heart of Autonomous Driving” invites to share our paper of “UniLION” and “DrivePI”
  6.“**Huawei Yinwang Intelligent Technology (formerly Huawei Car BU)**” invite to give a 2h talk about “LION”, “UniLION” and “DrivePI”, with **over 100 researchers attending**.

- ***Teaching：***

  “The Heart of Autonomous Driving” invites to teach the first course "Multimodal Fusion 3D Object Detection Tutorial" in China, 2023 **(>8h).**

- ***Volunteer:***

  Participated in the "China-Africa Text Recognition and Natural Language Processing Innovation Forum" as a main volunteer, 2023.