Xianzheng Ma

Embodied Agents · Robotic World Model · 3D Understanding

DPhil Student, VGG & AVL, University of Oxford

xianzheng@robots.ox.ac.uk · xianzhengma.github.io · Google Scholar · GitHub

Research Interests

3D computer vision and robotics. Previously focused on leveraging 3D Large Language Models to reason about and understand the 3D world. More recently, shifted towards enabling robots to understand and interact with the physical world through video and 3D modalities, with a current emphasis on utilizing world models to improve the generalization capability of general-purpose robotic systems.

Education

University of Oxford Oct 2023 – Present

DPhil, Visual Geometry Group (VGG) & Active Vision Lab (AVL) *Oxford, UK*

Supervised by Prof. Victor Prisacariu and Prof. Iro Laina.
Working closely with Prof. João F. Henriques and Prof. Ingmar Posner.

Wuhan University Sep 2014 – Jun 2021

Bachelor's and Master's Degrees *Wuhan, China*

Research Experience

Shanghai AI Laboratory 2021 – 2023

Full-time Researcher *Shanghai, China*

Supervised by Prof. Chao Dong.

Selected Publications (* equal contribution)

1. Yongchang Zhang*, **Xianzheng Ma***, Tianyi Liu, Guangquan Zhou, Yang Chen. *See It, Say It, Sorted: An Iterative Training-Free Framework for Visually-Grounded Multimodal Reasoning in LVLMs.* **CVPR**, 2026. [Paper]

2. Kohsuke Ide, Ryousuke Yamada, Yue Qiu, **Xianzheng Ma**, Yoshihiro Fukuhara, Hirokatsu Kataoka, Yutaka Satoh. *Beyond Single Object: Learning 3D Relations with Large Language Models.* **CVPR**, 2026 (Finding).

3. **Xianzheng Ma***, Tao Sun*, Shuai Chen, Yash Bhalgat, Jindong Gu, Angel X Chang, Iro Armeni, Iro Laina, Songyou Peng, Victor Adrian Prisacariu. *Do 3D Large Language Models Really Understand 3D Spatial Relationships?.* **ICLR**, 2026. [Project Page] [Code]

4. Guanqi Zhan*, **Xianzheng Ma***, Weidi Xie, Andrew Zisserman. *Inferring Dynamic Physical Properties from Video Foundation Models.* arXiv preprint, 2025. [Paper]

5. Yanbang Li, ZiYang Gong, Haoyang Li, Xiaoqi Huang, Haolan Kang, Guangpingbai, **Xianzheng Ma**. *Robotic Visual Instruction.* **CVPR**, 2025. [Paper] [Project Page]

6. **Xianzheng Ma***, Brandon Smart*, Yash Bhalgat*, et al.. *When LLMs Step into the 3D World: A Survey and Meta-Analysis of 3D Tasks via Multi-modal Large Language Models.* **IJCV**, under review. [Paper] [GitHub]

7. **Xianzheng Ma**, Zhixiang Wang, Yacheng Zhan, Yinqiang Zheng, Zheng Wang, Dengxin Dai, Chia-Wen Lin. *Both Style and Fog Matter: Cumulative Domain Adaptation for Semantic Foggy Scene Understanding.* **CVPR**, 2022 **Oral**. [Paper]

8. Min Shi, Zihao Huang, **Xianzheng Ma**, Xiaowei Hu, Zhiguo Cao. *Matching Is Not Enough: A Two-Stage Framework for Category-Agnostic Pose Estimation.* **CVPR**, 2023 **Highlight**. [Code]

9. Yulu Gan, **Xianzheng Ma**, Yan Bai, Yihang Lou, Renrui Zhang, Nian Shi, Lin Luo. *Decorate the Newcomers: Visual Domain Prompt for Continual Test Time Adaptation.* **AAAI**, 2023 **Outstanding Student Paper**. [Paper]

10. **Xianzheng Ma**, Hossein Rahmani, Zhipeng Fan, Bin Yang, Jun Chen, Jun Liu. *REMOTE: Reinforced Motion Transformation Network for Semi-supervised 2D Pose Estimation in Videos.* **AAAI**, 2022. [Paper]

Honors & Awards

- Outstanding Reviewer (top 3%), ICCV 2025
- Outstanding Student Paper Award, AAAI 2023
- Oral Presentation, CVPR 2022
- Highlight Paper, CVPR 2023

Skills

Programming: Python, PyTorch, C/C++, CUDA, Linux, Git. Multi-node distributed training on GPU clusters (SLURM, PBS).

Robotics Simulation: MuJoCo, ManiSkill, Genesis.

Robotics & VLA: Pre-training and reproduction of VLA models (e.g., xVLA) with multi-node multi-GPU training, achieving 40× inference acceleration; reproduced and benchmarked on CALVIN, RoboTwin, SIMPLER, etc.

Learning from Human: Built a Real-to-Sim-to-Real pipeline that extracts manipulation skills from human demonstration videos and transfers them to robotic policies via simulation; teleoperation data collection on Franka Arm via Meta Quest and handheld controllers.

Current Project: Video-based World Model for Robotics — utilizing video generation models to achieve closed-loop robotic control and tackle long-horizon manipulation tasks.

Academic Services

Conference Reviewer: CVPR (2022–2026), AAAI (2022–2026), ICLR (2023–2026), NeurIPS (2024–2025), ICCV (2025), ECCV (2024).

Journal Reviewer: IJCV, TPAMI, TMM.

Area Chair: IJCAI 2022 (China branch).