## 个人信息

**姓名：**韦小宝　　**政治面貌：**中共党员

**性别：**男　　**手机：**13928584687

**年龄：**25　　**邮箱：**weixiaobao0210@gmail.com

**本科学校：**北京航空航天大学机器人工程　**研究生学校：**中科院软件所

**个人主页：**https://ucwxb.github.io/　　**研究方向：**Agentic Robot、世界模型

## 个人陈述

本人现为中国科学院软件研究所直博生（预计 2028 年 6 月毕业），由中科院软件所陈辉老师与北京大学计算机学院仉尚航老师联合指导。本人研究方向为 Agents、世界模型、3D 视觉。截至目前，本人已在 CVPR、ICCV、ICLR、ECCV、AAAI 等顶级 AI 会议发表学术论文 30 余篇，其中以第一作者（含共同一作）发表 CCF-A 类论文 15 篇，并获得 ICCV Highlight、ICMR Oral。相关开源研究成果在 GitHub 累计获得 800+ Stars，学术引用突破 1000 余次 (H-Index=18)，担任 CVPR、NeurIPS、ICML 等十余个顶会顶刊审稿人，相关算法成果已实现业务验证与应用落地。此外，本人曾获国家奖学金、北京市优秀毕业生、优秀党员等 30 余项荣誉。

## 实习经历

**[1] AMD 超威半导体公司：端到端自动驾驶实习生 | 2023.07-2024.01**

算法研发：研究自驾场景下目标检测问题，提出了跨域目标检测与开放世界 3D 目标检测算法（ECCV、CVPR）。

业务落地：把算法核心技术贡献到公司业务管线中的一部分，解决真实场景中域偏移的问题。

**[2] 蔚来汽车：自动驾驶世界模型实习生 | 2024.01-2024.06**

**算法研发：**主导自动驾驶街景 4D 重建与生成式世界模型研究，提出基于 3DGS 的高速物体运动建模方法与场景压缩算法（EMD、S3Gaussian、SparseStreet），核心算法成果以一作/共一发表于 ICCV、ICRA 与 ICMR。

**业务落地：**面向业务数据，搭建高保真度街景仿真与 Corner Case 数据生成管道，给 NWM 训练交付 4000+ clips。

**[3] 小米汽车：自动驾驶 VLA 实习生（小米未来星实习） | 2024.06-2025.12**

**算法研发：**研究世界模型合成数据对感知任务的真实效果（ICLR）；同时作为核心成员之一，设计和训练轻量化自回归的规划模型，采用 DINOv3 搭配 AR Head（110M），结合业务数据探索自驾规划模型的 Scaling law。

**业务落地：**参与搭建可控视频生成与合成数据自动标注流水线，支撑业务合成数据工作；基于 500 万“五星司机”自驾 Clips 数据训练内部上车模型，参与完成车载芯片 FP8 模型量化，通过内部 SIL 评测集，实现发版上车。

**[4] 北京智源研究院：具身智能多模态实习生 | 2025.12-2026.02**

**算法研发：**负责增强 VLA 模型中的 ViT 视觉表征，增强模型对物理世界的空间感知能力（ACM MM）。

**业务落地：**将相关技术迁移至 VLN 上尝试并进行四足真机的部署，完善真机 infra。

**[5] 北京亦庄人形机器人研究院：北人-北大联合实验室负责人（主要负责数据 / Infra） | 2026.02-2026.10**

**算法研发：**研究 4D 先验引入世界模型，减少通用世界模型迁移至机器人后产生的碰撞（CoRL、AAAI）。探索 Unified Model 框架下面混合 Ego 与真机数据的 Scaling law。提出一个统一面向具身具身全面能力的 benchmark。

**业务落地：**主持构建更高质量的机器人与 Ego 数据采集、质检、清洗流程和维护代码 infra。维护全人形与半人形真机（天工、天轶）控制 infra。全流程设计和推动数据采集、数据清洗、模型训练（VA 混合预训练与后训练）、推理部署（DMD 加速、RTC）相关 infra 工作，相关 demo 在 WRC 上展示。

# 科研论文

### World Model：

[1] WAM4D: Fast 4D World Action Model via Spatial Register Tokens (一作)

[2] Manipdreamer3d: Synthesizing plausible robotic manipulation video with occupancy-aware 3d trajectory – AAAI 2026 (一作)

[3] Rethinking Driving World Model as Synthetic Data Generator for Perception Tasks – ICLR 2026 (Project Leader)

[4] ManipDreamer: Boosting Robotic Manipulation World Model with Action Tree and Visual Guidance – ICASSP 2026 (一作)

[5] PhyMix: Towards Physically Consistent Single-Image 3D Indoor Scene Generation with Implicit--Explicit Optimization – ECCV 2026

[6] Pelican-Unify 1.0: A Unified Embodied Intelligence Model for Understanding, Reasoning, Imagination and Action (技术报告，WorldArena 第一名)

[7] LAW: Unifying Language, Action, and World Modeling for Agentic Physical Intelligence – ICLR (Robotwin 开源第一)

[8] IOI: Decoupling Kinematics and Physics for Interactive World Models – AAAI

### VLA：

[9] EvoDriveVLA: Evolving Driving VLA Models via Collaborative Perception-Planning Distillation (一作)

[10] GraspFoM: Towards Reconstruction-Driven Robotic Grasping with 3D Foundation Priors – ACM MM 2026 (一作)

[11] VEGA: Visual Encoder Grounding Alignment for Spatially-Aware Vision-Language-Action Models – NeurIPS

[12] Fastdrivevla: Efficient end-to-end driving via plug-and-play reconstruction-based token pruning – AAAI 2026

[13] PIGEON: VLM-Driven Object Navigation via Points of Interest Selection

### 3D Vision：

[14] Dynamic Thermal Gaussians: Multimodal 4D Gaussian Splatting – ACM MM 2026 (一作)

[15] ParkGaussian: Surround-view 3D Gaussian Splatting for Autonomous Parking – CVPR 2026 (一作)

[16] GazeGaussian: High-Fidelity Gaze Redirection with 3D Gaussian Splatting – ICCV 2025 Highlight (一作)

[17] EMD: Explicit Motion Modeling for High-Quality Street Gaussian Splatting – ICCV 2025 (一作)

[18] Nto3d: Neural target object 3d reconstruction with segment anything – CVPR 2024 (一作)

[19] I-medsam: Implicit medical image segmentation with segment anything – ECCV 2024 (一作)

[20] Graphavatar: Compact head avatars with gnn-generated 3d gaussians - AAAI 2025 (一作)

[21] Embodiedocc++: Boosting embodied 3d occupancy prediction with plane regularization and uncertainty sampler – ACM MM 2025 Oral (一作)

[22] S3Gaussian: Self-Supervised Street Gaussians for Autonomous Driving – ICRA 2026 (一作)

[23] SparseStreet: Sparse Gaussian Splatting for Real-Time Street Scene Simulation – ICMR 2026 Oral

[24] Omniindoor3d: Comprehensive indoor 3d reconstruction – Pacific Graphics 2026 (一作)

[25] ThermalGaussian-X: RGB and Thermal Alignment via 3D Gaussian Splatting – ACM MM 2026

[26] Feed-Forward Gaussian Splatting from Sparse Aerial Views – ECCV 2026

[27] DNRSelect: Active Best View Selection for Deferred Neural Rendering – IEEE BigData Conf 2026 (Corresponding Author)

[28] RSATalker: Realistic Socially-Aware Talking Head Generation for Multi-Turn Conversation – IEEE VR 2026

[29] DiffusionTalker: Efficient and Compact Speech-Driven 3D Talking Head via Personalizer-Guided Distillation – ICME 2025

[30] MixedGaussianAvatar: Realistically and Geometrically Accurate Head Avatar via Mixed 2D-3D Gaussian Splatting – ACM MM 2025

[31] PLGS: Robust Panoptic Lifting with 3D Gaussian Splatting – IEEE TIP 2025

[32] Multi-scale full spike pattern for semantic segmentation - Neural Networks 2024

[33] Open-Vocabulary 3D Detection via Image-Level Class and Debiased Cross-modal Contrastive Learning – CVPR 2023

[34] MTTrans: Cross-domain object detection with mean teacher transformer – ECCV 2022

[35] A generalist foundation model and database for open-world medical image segmentation - Nature Biomedical Engineering 2025