Jianwen Cao

Zurich, Switzerland
jianwen.cao@uzh.ch | caojia@student.ethz.ch | +41 766395217 | +86 18094203447

Education

University of Zurich (UZH) — Zurich, Switzerland
M.Sc. in Informatics (Major: Artificial Intelligence, Minor: Data Science), GPA: 5.57/6 — 2024 – Present
Core Courses: Data Science (6.0/6.0), Machine Learning for Natural Language Processing (6.0/6.0), Vision Algorithms for Mobile Robotics (5.5/6.0), ETHZ 3D Vision (5.75/6.0), ETHZ Virtual Reality (6.0/6.0).
Nanjing University of Information Science and Technology (NUIST) — Nanjing, China
B.Eng. in Artificial Intelligence, GPA: 87.6/100 — 2020 – 2024

Publications

[Accepted] First Author: "Generative Event Pretraining with Foundation Model Alignment." IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026.
Proposed a multimodal pre-training method combining images and event cameras, achieving SOTA performance.
[Under Review] Second Author: "Memory Over Maps: 3D Object Localization Without Reconstruction." IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2026.
Responsible for 3D object localization (query input, 3D position output) and efficiency analysis. Ultimately deployed on the Boston Dynamics Spot robot dog.
[Published] First Author: "Fuse Tune: Hierarchical Decoder Towards Efficient Transfer Learning." Pattern Recognition and Computer Vision (PRCV), 2023.
Proposed an efficient transfer learning method that eliminates the need for backbone network backpropagation.

Projects & Awards

Robotics and Reinforcement Learning
- Complete vision-based racing drone RL training and real-world deployment experience (GitHub) (YouTube).
- Teleoperated the Inspire dexterous hand in Isaac Sim by synchronizing human hand movements from first-person view videos (GitHub).
- Familiar with forward/inverse kinematics, coordinate transformations, spline sampling, PID control, RL, IL, and Diffusion Policy.
- Implemented RL algorithms (DQN, PPO) and trained agents on Gymnasium (Atari games).
- Implemented IL algorithms (BC, Q-chunking, DAgger) and deployed them on Soarm101 for block grasping and placing with obstacle avoidance.
- Deep understanding of RL algorithm principles including DQN, Policy Gradient, TRPO, PPO, and SAC.

SLAM and 3D Gaussian Splatting (3DGS)
- Re-implemented Visual Odometry (VO) on the KITTI dataset based on the core concepts of PTAM/FAST-LIVO2.
- Successfully reproduced the original 3DGS paper, improved HI-SLAM2, and established a complete Novel View Synthesis pipeline.

Contemporary Undergraduate Mathematical Contest in Modeling (CUMCM) — 2022
- Responsible for designing robust mathematical models and optimization algorithms. Awarded Second Prize at the National Level.

Research & Internship Experience

Computer Vision and Geometry Group (ETHZ CVG) — Zurich, Switzerland
Student Researcher | Real-time Visual-Inertial-LiDAR Gaussian Splatting SLAM — Mar 2025 – Feb 2026
Reproduced and improved GS-LIVO to achieve better geometric accuracy and mapping efficiency, reaching SOTA performance.
Robotics and Perception Group (UZH RPG) — Zurich, Switzerland
Student Researcher | Event Camera-Enhanced Robot Motor Control Policies — Feb 2025 – Jan 2026
Conducted autoregressive pre-training on multimodal data, significantly improving the performance and robustness of downstream vision and control tasks.
IntSig Information Co., Ltd. (IntSig) — Shanghai, China
Computer Vision Algorithm Intern — Jul 2023 – Oct 2023
Developed object detection algorithms based on improved DETR series architectures.

Technical Skills

Robot Arm & Dog Deployment: Complete MuJoCo IL/RL training and ROS2 deployment experience on Soarm101, including data collection and state space optimization.
VLA Models: Familiar with RT-1, RT-2, Pi0, GR00T, ACT, and other Vision-Language-Action models.
Object Detection & Segmentation: RCNN series (Fast, Faster, Mask, Cascade), DETR, GLIP, GLIP2, SAM.
Pre-training & Contrastive Learning: Masked Modeling (BERT, MAE, BEiT, EVA), MoCo, CLIP, DINO.
LLMs & Multimodal Models: Familiar with GPT, MoE, RoPE, RAG, Engram; experience in pre-training Chinese corpora at the GPT-2 scale and RLHF post-training.
Generative Models: GAN, VAE, VQ-VAE, VQGAN, Latent Diffusion Models.
SLAM & 3D Vision: PTAM, FAST-LIVO2, 3DGS, HI-SLAM2, GS-LIVO.