Lei Fan

Bellevue, WA | leifan@u.northwestern.edu | 224-204-7107 | leifan95.github.io

Short Bio

Applied Scientist and Ph.D. researcher in computer vision and embodied perception, with experience in 3D scene understanding, visual localization, depth estimation, RGB-D perception, and real-time robotic perception systems. Published work spans multi-view localization, active embodied recognition, uncertainty-aware perception, depth completion, semantic mapping, and SLAM-related robotic vision.

Education

Northwestern University, Ph.D. in Computer Vision Lab, advised by Prof. Ying Wu — 2019 – 2024

Sun Yat-sen University, B.S. & M.S. in Computer Science — 2013 - 2017 & 2017 – 2019

Relevant Expertise

- **3D and embodied perception:** Visual localization, depth estimation, scene-structure understanding, and active viewpoint selection for robotic and embodied agents.
- **Robust visual recognition:** Open-vocabulary recognition, uncertainty estimation, and evidence-based modeling under ambiguous viewpoints, occlusion, and changing observations.
- **Multimodal visual reasoning** Egocentric video understanding, visual memory, and MLLM post-training for real-world visual reasoning tasks.

Publications

[1] *DACO: Dictionary-Aligned Concept Control for Safeguarding Multimodal LLMs*, Jinqi Luo, Jinyu Yang, Tal Neiman, **Lei Fan**, Bing Yin, Son Tran, Mubarak Shah, René Vidal. *CVPR 2026*.

[1] *GPVK-VL: Geometry-Preserving Virtual Keyframes for Visual Localization under Large Viewpoint Changes*, Yunxuan Li, **Lei Fan**, Xiaoying Xing, Jianxiong Zhou, Ying Wu. *CVPR 2025*.

[2] *Learning to Ask Denotative and Connotative Questions for Knowledge-based VQA*, Xiaoying Xing, Peixi Xiong, **Lei Fan**, Yunxuan Li and Ying Wu. *EMNLP Findings 2024*.

[3] *Active Open-Vocabulary Recognition: Let Intelligent Moving Mitigate CLIP Limitations*, **Lei Fan**, Jianxiong Zhou, Xiaoying Xing, Ying Wu. *CVPR 2024*.

[4] *Evidential Active Recognition: Intelligent and Prudent Open-World Embodied Perception*, **Lei Fan**, Mingfu Liang, Yunxuan Li, Gang Hua, Ying Wu. *CVPR 2024*.

[5] *Flexible Visual Recognition by Evidential Modeling of Confusion and Ignorance*, **Lei Fan**, Bo Liu, Haoxiang Li, Ying Wu, Gang Hua. *ICCV 2023*.

[6] *Avoiding Lingering in Learning Active Recognition by Adversarial Disturbance*, **Lei Fan**, Ying Wu. *WACV 2023*.

[7] *Unsupervised Depth Completion and Denoising for RGB-D Sensors*, **Lei Fan**, Yunxuan Li, Chen Jiang, Ying Wu. *ICRA 2022*.

[8] *FLAR: A Unified Prototype Framework for Few-sample Lifelong Active Recognition*, **Lei Fan**, Peixi Xiong, Wei Wei, Ying Wu. *ICCV 2021*.

[9] *Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning*, Xuehui Wang, Qing Wang, Yuzhi Zhao, Junchi Yan, **Lei Fan**, and Long Chen. *ACCV 2020*.

[10] *Monocular Outdoor Semantic Mapping with a Multi-task Network*, Yucai Bai, **Lei Fan**, Ziyu Pan, and Long Chen. *IROS 2019*.

[11] *Planecell: Representing Structural Space with Plane Elements*, **Lei Fan**, Long Chen, Kai Huang and Dongpu Cao. *IV 2018 - *Best Student Paper*.

[12] *RGB-T SLAM: A Flexible SLAM Framework by Combining Appearance and Thermal Information*, Long Chen, Libo Sun, Teng Yang, **Lei Fan**, Kai Huang, and Zhe Xuanyuan. *ICRA 2017*.

Experience

Applied Scientist, Amazon Alexa AI – Bellevue, WA May 2024 – Now

- Built long-horizon egocentric video QA and visual memory systems for multimodal perception.
- Post-trained multimodal assistant models through instruction fine-tuning on general-purpose and domain-specific vision-language data.
- Created synthetic and human-annotated multimodal datasets tailored for MLLM post-training and evaluation.
- Applied GRPO with composite reward functions to improve multimodal reasoning on challenging, verifiable visual tasks.

Applied Scientist Intern, Amazon Robotics – Seattle, WA June 2023 – Sep 2023

- Developed a monocular scene-structure reconstruction pipeline for robotic placement guidance, enabling geometric reasoning about stable support surfaces from a single image.
- Built a monocular multi-task perception model for depth estimation, surface-normal prediction, and scene segmentation.
- Co-drafted an invention disclosure/patent and submitted the work to an Amazon internal research conference.

Research Intern, Wormpex AI Research – Bellevue, WA June 2022 – Sep 2022

- Proposed a Dempster-Shafer-based method to jointly estimate ignorance and confusion through evidence aggregation.
- Authored an ICCV 2023 paper on flexible visual recognition via evidential modeling of confusion and ignorance.

Research Intern, Yosion Analytics – Chicago, IL June 2020 – Sep 2020

- Developed a camera-based pallet recognition and localization system for autonomous pallet-lifting guidance.
- Built a real-time vision system for empty pallet-space detection under varying lighting conditions.

Recent Research Projects

Geometry-Preserving Virtual Keyframes for Visual Localization

- Co-developed GPVK-VL, a structure-based visual localization pipeline for robust 6-DoF camera pose estimation under large viewpoint changes.
- Generated geometry-preserving virtual keyframes using 2D Gaussian Splatting with confidence-aware depth and normal priors, mesh extraction, and viewpoint sampling from a 3D occupancy map.
- Integrated virtual keyframes with NetVLAD retrieval, SuperGlue matching, and PnP-RANSAC pose estimation, improving robustness over HLoc, InLoc, and virtual-keyframe baselines on large-viewpoint-change localization.

Active Open-Vocabulary Recognition

- Studied how open-vocabulary vision-language models trained on fixed-view images fail in embodied settings due to viewpoint bias, occlusion, and incomplete observations.
- Proposed an active viewpoint-selection policy that guides an embodied agent toward informative perspectives for recognition.
- Designed a semantic-agnostic information-fusion module to improve generalization across open-vocabulary categories.

Visual token evidence-based pruning

- Investigated the computational burden of image and video inputs in MLLMs, where visual inputs produce far more tokens than text.
- Proposed an evidential learning method to dynamically prune visual tokens conditioned on the input question.
- Removed question-irrelevant or redundant visual tokens while preserving evidence needed for accurate answers.

Academic Services

Invited Conference & Journal Reviewer: CVPR, ICCV, ECCV, WACV, ICRA, IROS, IV, NeurIPS, ICPR, T-PAMI, IJCV, etc.