Khoa Vo

Postdoctoral Fellow · Computer Science

University of Arkansas, Fayetteville, AR 72701

khoavoho@uark.edu | vhvkhoa.github.io | khoa-vo-ho | Google Scholar

Research Interests

Embodied AI, vision-language-action models, video-language modeling, multimodal video understanding, object-centric reasoning, and robotic manipulation.

Professional Experience

2025-now **Postdoctoral Fellow**, Dept. of EECS, University of Arkansas; Advisor: Dr. Ngan Le
- Lead research on vision-language-action models for robotic manipulation.
- Contribute to funded projects in embodied AI, remote sensing, and vision-based livestock monitoring.

2020-2024 **Graduate Research Assistant**, Dept. of EECS, University of Arkansas
- Developed interpretable video-language models and video perception pipelines for action understanding, video captioning, visual question answering.

2019-2020 **Research Internship (6 months)**, National Informatics Institute, Tokyo, Japan
- Developed a framework for temporal action proposals in untrimmed videos.

2018-2020 **Research Fellow**, Ho Chi Minh University of Science
- Developed an anomaly detection system for traffic surveillance

Education

University of Arkansas *Fayetteville, AR, USA*

PhD in Computer Science *2020 - 2024*
- Advisor: Dr. Ngan Le
- Thesis: Towards Comprehensive and Interpretable Video Understanding

Ho Chi Minh University of Science *Ho Chi Minh City, Vietnam*

MSc in Computer Science *2018 - 2021*
- Advisor: Dr. Minh-Triet Tran
- Thesis: Agent-Environment Network for Temporal Action Proposal Generation
- Thesis Grade: 10.0/10.0

Ho Chi Minh University of Science *Ho Chi Minh City, Vietnam*

BSc in Honors Program of Information Technology *2014 - 2018*
- Advisor: Dr. Minh-Triet Tran
- Thesis: Image Captioning with Attention Mechanism and Concepts Augmentation
- Thesis Grade: 10.0/10.0

Awards and Fellowships

2019 **Vietnam National Master/PhD Scholarship**, Vingroup Innovation Foundation
Competitive rate: 150/400 (37.5%) graduate-level applicants receive scholarships

2018 **Best AI Project Award**, Eureka, Vietnam Nation-wide Awards for Student Scientific Research
Competitive rate: 12/903 (1.33%) teams have highest awards in 12 categories

Publications

Full list: Google Scholar

Clutter-Resistant Vision-Language-Action Models through Object-Centric and Geometry Grounding
Khoa Vo, Taisei Hanyu, Yuki Ikebe, Trong Thang Pham, Nhat Chung, Minh Nhat Vu, Duy Nguyen Ho Minh, Anh Nguyen, Anthony Gunderman, Chase Rainwater, and Ngan Le
IEEE T-RO, under review (2nd round)

CodeGraphVLP: Code-as-Planner Meets Semantic-Graph State for Non-Markovian Vision-Language-Action Models
Khoa Vo*, Sieu Tran*, Taisei Hanyu, Yuki Ikebe, Duy Nguyen, Bui Duy Quoc Nghi, Minh Vu, Anthony Gunderman, Chase Rainwater, Anh Nguyen, and Ngan Le
IROS 2026, under review ; * indicates equal contributions

1. **SemLT3D: Semantic-Guided Expert Distillation for Camera-only Long-Tailed 3D Object Detection**
Hao Vo, **Khoa Vo**, Thinh Phan, Ngo Xuan Cuong, Gianfranco Doretto, Hien Nguyen, Anh Nguyen, and Ngan Le
CVPR 2026

2. **SlotVLA: Towards Modeling of Object-Relation Representations in Robotic Manipulation**
Taisei Hanyu, Nhat Chung, Huy Le, Toan Nguyen, Yuki Ikebe, Anthony Gunderman, Duy Nguyen Ho Minh, **Khoa Vo**, Tung Kieu, Kashu Yamazaki, Chase Rainwater, Anh Nguyen, and Ngan Le
ICRA 2026

3. **Rethinking Progression of Memory State in Robotic Manipulation: An Object-Centric Perspective**
Nhat Chung, Taisei Hanyu, Toan Nguyen, Huy Le, Frederick Bumgarner, Duy Minh Ho Nguyen, **Khoa Vo**, Kashu Yamazaki, Chase Rainwater, Tung Kieu, Anh Nguyen, and Ngan Le
AAAI 2026

4. **CT-ScanGaze: A Dataset and Baselines for 3D Volumetric Scanpath Modeling**
Trong-Thang Pham, Akash Awasthi, Saba Khan, Esteban Duran Marti, Tien-Phat Nguyen, **Khoa Vo**, Minh Tran, Son Nguyen, Cuong Tran, Yuki Ikebe, Anh Totti Nguyen, Anh Nguyen, Zhigang Deng, Carol C. Wu, Hien Nguyen, and Ngan Le
ICCV 2025

5. **HENASY: Learning to Assemble Scene-Entities for Interpretable Egocentric Video-Language Model**
Khoa Vo, Thinh Phan, Kashu Yamazaki, Minh Tran, and Ngan Le
NeurIPS 2024

6. **Open-Fusion: Real-time Open-Vocabulary 3D Mapping and Queryable Scene Representation**
Kashu Yamazaki, Taisei Hanyu, **Khoa Vo**, Thang Pham, Minh Tran, Gianfranco Doretto, Anh Nguyen, and Ngan Le
ICRA 2024

7. **ZEETAD: Adapting Pretrained Vision-Language Model for Zero-Shot End-to-End Temporal Action Detection**
Thinh Phan, **Khoa Vo**, Duy Le, Gianfranco Doretto, Donald Adjeroh, and Ngan Le
WACV 2024

8. **Amodal Instance Segmentation with Diffusion Shape Prior Estimation**
Minh Tran, **Khoa Vo**, Tri Nguyen, and Ngan Le
ACCV 2024

9. **ShapeFormer: Shape Prior Visible-to-Amodal Transformer-based Amodal Instance Segmentation**
Minh Tran, Winston Bounsavy, **Khoa Vo**, Anh Nguyen, Tri Nguyen, and Ngan Le
IJCNN 2024

10. **CapsNet for Medical Image Segmentation**
Minh Tran, **Khoa Vo**, Kyle Quinn, Hien Nguyen, Khoa Luu, and Ngan Le
Deep Learning for Medical Image Analysis, 2nd ed., book chapter, 2024

11. **AOE-Net: Entities Interactions Modeling with Adaptive Attention Mechanism for Temporal Action Proposals Generation**
Khoa Vo, Sang Truong, Kashu Yamazaki, Bhiksha Raj, Minh-Triet Tran, and Ngan Le
IJCV 2023

12. **DNA: Deformable Neural Articulations Network for Template-Free Dynamic 3D Human Reconstruction From Monocular RGB-D Video**
Khoa Vo, Trong-Thang Pham, Kashu Yamazaki, Minh Tran, and Ngan Le
CVPR Workshops 2023

13. **VLTinT: Visual-Linguistic Transformer-in-Transformer for Coherent Video Paragraph Captioning**
Kashu Yamazaki*, **Khoa Vo***, Sang Truong, Bhiksha Raj, and Ngan Le
AAAI 2023 **Oral**; * indicates equal contributions

14. **CLIP-TSA: CLIP-Assisted Temporal Self-Attention for Weakly-Supervised Video Anomaly Detection**
Hyekang Joo, **Khoa Vo**, Kashu Yamazaki, and Ngan Le
ICIP 2023

15. **Neural Architecture Search for Medical Image Applications**
Khoa Vo, Kashu Yamazaki, Hieu Hoang, Minh-Triet Tran, and Ngan Le

Meta Learning with Medical Imaging and Health Informatics Applications, book chapter, 2023

16. **Spiking Neural Networks and Their Applications: A Review**
Kashu Yamazaki, **Khoa Vo**, Darshan Bulsara, and Ngan Le
Brain Sciences 2022 **Editor's Choice**

17. **AISFormer: Amodal Instance Segmentation with Transformer**
Minh Tran, **Khoa Vo**, Kashu Yamazaki, Arthur Fernandes, Michael Kidd, and Ngan Le
BMVC 2022

18. **VLCap: Vision-Language with Contrastive Learning for Coherent Video Paragraph Captioning**
Kashu Yamazaki, Sang Truong, **Khoa Vo**, Michael Kidd, Chase Rainwater, Khoa Luu, and Ngan Le
ICIP 2022

19. **ABN: Agent-Aware Boundary Networks for Temporal Action Proposal Generation**
Khoa Vo, Kashu Yamazaki, Sang Truong, Minh-Triet Tran, Akihiro Sugimoto, and Ngan Le
IEEE Access 2021

20. **Narrow Band Active Contour Attention Model for Medical Segmentation**
Ngan Le, Toan Bui, **Khoa Vo**, Kashu Yamazaki, and Khoa Luu
Diagnostics 2021

21. **AEI: Actors-Environment Interaction with Adaptive Attention for Temporal Action Proposals Generation**
Khoa Vo, Hyekang Joo, Kashu Yamazaki, Sang Truong, Kris Kitani, Minh-Triet Tran, and Ngan Le
BMVC 2021 **Oral**

22. **Agent-Environment Network for Temporal Action Proposal Generation**
Khoa Vo, Ngan Le, Kashu Yamazaki, Akihiro Sugimoto, and Minh-Triet Tran
ICASSP 2021

23. **A Smart System for Text-Lifelog Generation from Wearable Cameras in Smart Environment Using Concept-Augmented Image Captioning with Modified Beam Search Strategy**
Khoa Vo, Quoc-An Luong, Duy-Tam Nguyen, Mai-Khiem Tran, and Minh-Triet Tran
Applied Sciences 2019

24. **Vehicle Re-identification with Learned Representation and Spatial Verification and Abnormality Detection with Multi-Adaptive Vehicle Detectors for Traffic Video Analysis**
Khac-Tuan Nguyen, Trung-Hieu Hoang, Minh-Triet Tran, Trung-Nghia Le, Ngoc-Minh Bui, Trong-Le Do, **Khoa Vo**, Quoc-An Luong, Mai-Khiem Tran, Thanh-An Nguyen, Thanh-Dat Truong, Vinh-Tiep Nguyen, and Minh N. Do
CVPR Workshops 2019

25. **Personal Diary Generation from Wearable Cameras with Concept Augmented Image Captioning and Wide Trail Strategy**
Khoa Vo, Quoc-An Luong, Duy-Tam Nguyen, Mai-Khiem Tran, and Minh-Triet Tran
SoICT 2018 **Best paper award**

Patent Application

1. **Artificial Intelligence and Vision-Based Broiler Body Weight Measurement System and Process**
Michael T. Kidd, Thi Hoang Ngan Le, and **Khoa Ho Viet Vo**
US20250259325A1, published patent application, 2025

Teaching Experience

Fall 2022–2025	**Deep Learning, University of Arkansas**, Teaching Assistant
Spring 2023	**Algorithms, University of Arkansas**, Teaching Assistant
Summer 2021	**Applied ML Intensive, NACME-Google Summer Bootcamp, University of Arkansas**, Teaching Assistant

Service and Proposal Support

Funded proposal support: Technical writing and concept development for funded proposals totaling $575,000, including NSF CAREER ($500,000) (PI: Ngan Le) and Arkansas Research Alliance Impact Grant ($75,000) (PI: Chase Rainwater; Co-PIs: Ngan Le, Anthony Gunderman).

Conference reviewing: ICCV 2025, CVPR 2023–2026, AAAI 2025, ECCV 2024, ECCV 2026, IROS 2026, ICML 2021, WACV 2019–2024, ICASSP 2020–2023, ICIP 2023.

Journal reviewing: IJCV, TIP, T-RO, CVIU, ACM Computing Surveys, and Expert Systems with Applications.