

Khoa Vo

POSTDOCTORAL FELLOW · COMPUTER SCIENCE

University of Arkansas, Fayetteville, AR 72701

✉ khoavoho@uark.edu | 🏠 [vhvkhoa.github.io](https://github.com/vhvkhoa) | 🌐 [khoa-vo-ho](https://khoa-vo-ho.github.io) | 📄 Google Scholar

Summary

I am a postdoctoral fellow in Computer Science with 7+ years of progressive research experience in Computer Vision & Deep Learning, spanning vision-language models and video understanding with CNN-based and Transformer-based architectures. Currently, I lead a funded project focus on advancing vision-language-action model for robotic manipulation.

I serve as first or lead author on publications at high-impact venues, including NeurIPS, ICRA, AAI, BMVC, WACV, ICASSP, ICIP, and IJCV journal, and co-invented a patent on vision-based broiler-weight measurement system.

Professional Experience

- 2025-now **Postdoctoral Fellow**, Dept. of EECS, University of Arkansas
 - Manage funded projects in robotics, satellite imagery, and vision-based chicken weight estimation
 - Research project: vision-language action models for robotic arm manipulation
- 2020-2024 **Graduate Research Assistant**, Dept. of EECS, University of Arkansas
 - Conduct research in video language modeling, dynamic 3D object reconstruction
- 2019-2020 **Research Internship (6 months)**, National Informatics Institute, Tokyo, Japan
 - Develop a framework for temporal action proposals in untrimmed videos
- 2018-2020 **Research Fellow**, Ho Chi Minh University of Science
 - Develop an anomaly detection system for traffic surveillance

Selected Publications

For a complete list of publications, please visit my [Google Scholar](#) page.

JOURNALS

1. **Khoa Vo**, Sang Truong, Kashu Yamazaki, Bhiksha Raj, Minh-Triet Tran, and Ngan Le.
AOE-Net: Entities Interactions Modeling with Adaptive Attention Mechanism for Temporal Action Proposals Generation.
IJCV, Vol. 131, 2023. **Impact factor: 11.6**
2. **Khoa Vo**, Kashu Yamazaki, Sang Truong, Minh-Triet Tran, Akihiro Sugimoto, and Ngan Le.
ABN: Agent-Aware Boundary Networks for Temporal Action Proposal Generation.
IEEE Access, 2021.

CONFERENCES

1. **Khoa Vo**, Thinh Phan, Kashu Yamazaki, Minh Tran, Ngan Le.
HENASY: Learning to Assemble Scene-Entities for Interpretable Egocentric Video-Language Model.
NeurIPS 2024
2. Kashu Yamazaki, Taisei Hanyu, **Khoa Vo**, Thang Pham, Minh Tran, Gianfranco Doretto, Anh Nguyen, Ngan Le.
Open-Fusion: Real-time Open-Vocabulary 3D Mapping and Queryable Scene Representation.
ICRA 2024.
3. Thinh Phan, **Khoa Vo**, Duy Le, Gianfranco Doretto, Donald Adjeroh, Ngan Le.
ZEETAD: Adapting Pretrained Vision-Language Model for Zero-Shot End-to-End Temporal Action Detection.
WACV 2024.
4. Kashu Yamazaki*, **Khoa Vo***, Sang Truong, Bhiksha Raj, and Ngan Le. (*: *equal contributions*)
VLTinT: Visual-Linguistic Transformer-in-Transformer for Coherent Video Paragraph Captioning.
AAAI 2023. **Oral presentation**.
5. Hyekang Joo, **Khoa Vo**, Kashu Yamazaki, and Ngan Le.
CLIP-TSA: CLIP-Assisted Temporal Self-Attention for Weakly-Supervised Video Anomaly Detection.
ICIP 2023. **Oral presentation, top 18% submissions**.
6. **Khoa Vo**, Hyekang Joo, Kashu Yamazaki, Sang Truong, Kris Kitani, Minh-Triet Tran, and Ngan Le.
AEI: Actors-Environment Interaction with Adaptive Attention for Temporal Action Proposals Generation.
BMVC 2021. **Oral presentation, top 3.33% submissions**.

Patents

1. Michael Kidd, Ngan T.H. Le, **Khoa H.V. Vo**, “Artificial Intelligence And Vision-Based Broiler Body Weight Measurement System And Process”. *Worldwide patent WO 2023034834A1*.

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

Teaching Experience

| | |
|-------------|--|
| Spring 2023 | Algorithms, University of Arkansas, Teaching Assistant |
| Fall 22, 23 | Deep Learning, University of Arkansas, Teaching Assistant |
| Summer 2021 | Applied ML Intensive, NACME-Google Summer Bootcamp, University of Arkansas, Teaching Assistant |

Awards, Fellowships, & Grants

| | |
|------|---|
| 2019 | Vietnam National Master/PhD Scholarship , Vingroup Innovation Foundation <i>Competitive rate: 150/400 (37.5%) graduate-level applicants receive scholarships</i> |
| 2018 | Best AI Project Award , Eureka, Vietnam Nation-wide Awards for Student Scientific Research <i>Competitive rate: 12/903 (1.33%) teams have highest awards in 12 categories</i> |

Outreach & Professional Development

Contribution to the Development of Successful Proposals:

1. **NSF CAREER (\$500,000)**. "Trustworthy, Robust, and Efficient Multimodal Framework for Video Analytics". PI: Ngan Le.
2. **Arkansas Research Alliance Impact Grant (\$75,000)**. "Smart Robotics Arm for Packaging". PI: Chase Rainwater. Co-PI's: Ngan Le, Anthony Gunderman.

Conference reviews: ICCV 2025, CVPR 2023-2025, AAAI 2025, ECCV 2024, ICML 2021, WACV 2019-2024, ICASSP 2020-2023, ICIAP 2023.

Journals reviews: TIP, CVIU, and Expert Syst. Appl..