

Akash Kumar

 github.com/AKASH2907  akash2907.github.io  linkedin.com/in/akumar  akash.kumar@ucf.edu

EDUCATION

University of Central Florida

Aug. 2020 - Dec. 2025 (Expected)

Ph.D. Computer Science

GPA: 3.8/4.0

INTERNSHIP EXPERIENCE

Open-Vocabulary Video Object Detection | Applied Scientist Intern, Amazon

May'24 - Aug'24

- Global-Local Soft Prompting. Devised approach to associate visual and motion cues respective to different objects to help recognize objects for in-the-wild settings.

RESEARCH EXPERIENCE

Data-Efficient Learning for Dense Video tasks | Supervisor: **Dr. Yogesh S. Rawat**

Aug'20 - Present

- Data-Efficient Multimodal Foundation Models for Spatio-Temporal Video Understanding (Fall'23 - Present)
 - * Contextual self-paced learning for Weakly Supervised STVG (**ICLR'25**)
 - Developed *first* vision language models (VLMs) for dense video task without any labels.
 - * Spatial and Temporal Progressive Learning for Weakly STVG (**CVPR'25**)
 - Improved VLMs grounding capabilities via action composition and complex dense scene understanding.
 - * Scene Graphs generation with VLMs for sequential task planning (under review **COLM'25**)
 - Iterative refinement scene graphs with VLMs to solve multiple sequential planning tasks.
 - * IASEB: Interaction-Aware Spatiotemporal Evaluation Benchmark (under progress)
 - Evaluation benchmark to understand spatio-temporal relation understanding of foundation models.
 - * Training free STVG via Multimodal foundation models (under progress)
 - Adaptation of VLMs leveraging Large Language Models (LLMs) via attribute composite relationship.
- Limited-Label Learning for Dense video tasks (Fall'20 - Spring'23)
 - * Stable Mean Teacher for Semi-supervised Video Action Detection (**AAAI'25**)
 - Introduced class-agnostic spatio-temporal refinement module and temporal coherency constraint improving spatio-temporal localization.
 - * Semi-supervised Active Learning for Video Action Detection (**AAAI'24**)
 - Proposed a simple frame utility based informative sample selection and frequency based localization.
 - * End-to-End Semi-Supervised Learning for Video Action Detection (**CVPR'22**)
 - *First* end-to-end semi-supervised work tackling video action detection. Devised short-term and long-term smoothness constraints to exploit spatio-temporal coherency.
 - * Benchmarking Self-Supervised Video Representation Learning (**NeurIPSW'23**)
 - *First* exhaustive study on impact of pre-training in self-supervised learning for videos. Proposed a simple knowledge distillation approach outperforming previous works with 90% less videos.

FUNDING PROJECTS

Fine-grained Video Understanding Tasks | Advisors: **Dr. Yogesh S. Rawat, Dr. Mubarak Shah**

Aug'20 - Spring'24

- GAIT recognition in extremely challenging conditions. (BRIAR program, IARPA) (*Team Lead*)
 - * Secured 2^{nd} rank out of 7 teams including Michigan State, John Hopkins, Kitware, etc.
- Activity Detection in surveillance videos (DIVA program, IARPA)
 - * Achieved 1^{st} rank out of 10 teams including Stanford, Columbia, John Hopkins, Kitware, etc.

SKILLS

Languages & Frameworks: C/C++, Python, HTML/CSS, PyTorch, Keras, OpenCV

Tools: Git, L^AT_EX, Vim, Sublime, VS Code, Linux

ACHIEVEMENTS

AAAI, WACV 2025 Student Travel Grant, IEEE WACV Doctoral Consortium, 2025; 2^{nd} place, IARPA BRIAR, 2023; 2^{nd} place, NIST TRECVID, 2021; 1^{st} place, ActivityNet CVPR, 2021; 8^{th} HLF, 2021; Doctoral Fellowship, 2020.

PROFESSIONAL SERVICE

Conference Reviewer: NeurIPS'23,'24,'25, ICLR'23,'24,'25, CVPR'23,'24,'25, ECCV/ICCV'22,'23,'24, '25

Teaching Assistant: CAP5415 (Fall'24):Computer Vision, CDA3103C (Spring'25): Digital Logic Design