Akash Kumar

Google Scholar | GitHub | Medium | LinkedIn

Email : akash.kumar@ucf.edu
Mobile : +1-(321)-276-9387

Education

- **University of Central Florida (UCF)** — Orlando, FL
  PhD in Computer Science; GPA: 3.8/4.0 — *Aug. 2020 - Dec. 2025 (Expected)*
- **Delhi Technological University (DTU)** — Delhi, India
  Bachelor of Technology in Electronics and Communications; CGPA: 8.6/10.0 — *Aug. 2015 - June. 2019*

Internship Experience

- **Open-Vocabulary Video Object Detection** — Palo Alto, CA
  Applied Scientist Intern, Amazon | Supervisor: Dr. Shan Yang, Senior Applied Scientist — *May'24 - Aug'24*
  - Problem Statement: Open-Vocabulary Video Object Detection.
  - Solution: 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** — Orlando, FL
  Research Assistant | Supervisor: Dr. Yogesh Singh Rawat — *Aug.'20 - Present*
  - Research Area: Data-Efficient Multimodal Foundation Models for Spatio-Temporal Video Understanding (Multimodal) (Fall'23 - Present)
    * Contextual self-paced learning for Weakly Supervised STVG (**ICLR'25**).
      - Developed first vision language models (VLMs) for dense multimodal video detection task without any labels.
      - Devised context aware and self-paced progressive scene learning approach.
    * Spatial and Temporal Progressive Learning for Weakly STVG (under review **CVPR'25**).
      - Improved VLMs grounding capabilities via action composition and complex spatio-temporal scene understanding.
    * Training free STVG via Multimodal foundation models (under review).
      - Adaptation of VLMs leveraging Large Language Models (LLMs) via spatio-temporal composite relationship.
    * Benchmarking Vision Language Models on STVG (under review).
      - Studied impact of image and video dense VLMs on fine-grained video tasks to analyze attribute relation understanding capabilities.
  - Research Area: Limited Label Learning for Dense video tasks (Unimodal) (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 for better 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 spatio-temporal localization.
    * End-to-End Semi-Supervised Learning for Video Action Detection (**CVPR'22**).
      - 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.
    * Video Action Detection: Analyzing Limitations and Challenges (**CVPRW'22**).
      - Developed new spatio-temporal surveillance based dataset to incorporate real-world challenges.
    * GabriellaV2: Towards better generalization surveillance videos for Action Detection (**WACVW'22**).
      - Proposed real-time online action detection system for open-world surveillance videos.

Funding Projects

- **Fine-grained Video Understanding Tasks** Orlando, FL
  Research Assistant | Supervisor: Dr. Yogesh Singh Rawat *Aug.'20 - Spring'24*
  - Funding Project: GAIT recognition in extremely challenging conditions. (BRIAR datasets) (BRIAR program, IARPA). (Jan'22-April'24) (Project Lead - (Jan'22 - March'23)).
    - Achieved 2^{nd} rank out of 7 teams including Michigan State, John Hopkins, Kitware, etc.
  - Funding Project: Activity Detection in multi-camera environments (MEVA dataset) (DIVA program, IARPA). (Jan'21-Dec'21)
    - Achieved 1^{st} rank out of 10 teams including Stanford, Columbia, John Hopkins, Kitware, etc.

Conference Publications

- **Training free Spatio-Temporal Video Grounding via Multimodal Foundation models**
  In Review
- **Benchmarking Dense Vision Language Models on STVG**
  In Review
- **Spatial and Temporal Progressive Learning for Weakly Supervised STVG**
  In Review CVPR'25
- **Contextual self-paced learning for Weakly Supervised Spatio-Temporal Video Grounding**
  ***A. Kumar**, Z.Kira, Y.S. Rawat*
  International Conference on Learning Representations (ICLR), 2025. *(Link)*
- **Stable Mean Teacher for Semi-supervised Video Action Detection**
  ***A. Kumar**, S. Mitra, Y.S. Rawat*
  Association for the Advancement of Artificial Intelligence (AAAI), 2025. *(Link)*
- **Semi-supervised Active Learning for Video Action Detection**
  *A. Singh, A. Rana, **A. Kumar**, S.Vyas, Y.S. Rawat*
  Association for the Advancement of Artificial Intelligence (AAAI), 2024. *(Link)*
- **End-to-End Semi-Supervised Learning for Video Action Detection**
  ***A. Kumar**, Y.S. Rawat*
  Computer Vision and Pattern Recognition (CVPR), 2022. *(Link)*

Workshop Publications

- **Benchmarking Self-Supervised Learning for Video Representation Learning**
  ***A. Kumar**, Ashlesha Kumar, V. Vineet, Y.S. Rawat*
  4th Workshop on Self-Supervised Learning, NeurIPSW, 2023 *(Link)*
- **Video Action Detection: Analysing Limitations and Challenges**
  *R. Modi, A.J. Rana, **A. Kumar**, P. Tirupattur, S. Vyas, Y.S. Rawat, M. Shah*
  1st Workshop on Vision Datasets Understanding, CVPRW, 2022 *(Link)*
- **GabriellaV2: Towards better generalization in surveillance videos for Action Detection**
  *I.Dave, Z. Scheffer, **A. Kumar**, S. Shiraz, Y.S. Rawat, M. Shah*
  Human Activity Detection in Multi-Camera Long-Duration Video, IEEE WACVW, 2022 *(Link)*
- **Syn2Real: Forgery Classification via Unsupervised Domain Adaptation**
  ***A. Kumar**, A. Bhavsar*
  Deepfakes and Presentation Attacks in Biometrics, IEEE WACVW, 2020 *(Link)*
- **IceBreaker: Solving Cold Start Problem for Video Recommendation Engines**
  A. Kumar, A. Sharma*, A. Khaund*, Y. Kumar, P. Kumaraguru, R.R. Shah, R. Zimmerman*
  MR2AMC Workshop, 20th IEEE International Symposium on Multimedia(ISM) 2018 *(Link)*

Research Experience (Before Ph.D.)

- **Fine-grained classification and segmentation, MANAS Lab, IIT Mandi** — Himachal Pradesh, India
  Research Assistant | Supervisor: Dr. Arnav Bhavsar — *April'20 - July'20*
  - Improved classification accuracy via advanced data augmentation practices and part-wise attention localisation. *(Github)*

- **Image Forgery Detection & Localization, MANAS Lab, IIT Mandi** — Himachal Pradesh, India
  Research Assistant | Supervisor: Dr. Arnav Bhavsar — *June'19 - April'20*
  - Explored the online negative triplet mining for deepfakes classification in low resolution videos. *(Github)*
  - Detection and localization of Copy-Move Forgery in Images. Created a synthetic tampered dataset using semantic inpainting and copy-move forgery on COCO dataset. Employed domain adaptation to learn the representation from synthetic to real-world images. *(Github)*

- **Indian Landmark Recognition, DTU** — New Delhi, India
  Research Assistant | Supervisor: Dr. S. Indu — *Oct. 2018 - Jan. 2019*
  Devised an architecture to predict landmark labels directly from image pixels using Graph-based saliency approach to help better understanding and organizing photos of diverse Indian monuments style. Deployed saliency detection in conjunction with transfer learning, ML classifiers and ensembling methods. *(Github)*

- **Content-based Video Relevance Prediction, MIDAS Lab, IIIT Delhi** — New Delhi, India
  Research Assistant | Supervisor: Dr.Rajiv Ratn Shah(IIITD) & Dr.Roger Zimmerman(NUS) — *April 2018 - Aug. 2018*
  Developed a recommender system to solve the problems of "cold-start" videos and generate a personalized recommendation based on user's history. Built the system using Data Augmentation, Random Forest Regression & Deep Learning based Linear Discriminant Analysis. *(Github)*

- **Bird Species Classification, NuTech Labs (CVIP'18 Challenge Winner)** — New Delhi, India
  Conference Challenge — *July. 2018 - Sept. 2018*
  Implemented an end-to-end deep learning model for bird detection and inter-species classification in high resolution images. Worked on Transfer learning, Multistage training, Object detection via Mask R-CNN and Model ensemble on a very small dataset (150 images). *(Github)*

- **Rooftop Assessment for Solar Installation Using Satellite Imagery** — Himachal Pradesh, India
  Computer Vision Intern, The Solar Labs — *June. 2017 - July. 2017*
  Formulated an online rooftop assessment system for solar installations using Satellite Imagery. Implemented algorithms to identify individual rooftops of buildings and optimal area to place solar panels. *(Github)*

Selected Projects

- **Plant Pathology (Kaggle:Top 20%)**: Misdiagnosis of agricultural crops diseases leads to pathogen strains and increased input cost. Employed EfficientNet, Noisy Student weights, label smoothing, focal loss and test time augmentation to achieve 97% accuracy. (CVPR'20 Challenge)
- **Indian Driving Dataset Segmentation Challenge**: Devised solutions for unstructured driving scenarios on Indian roads. Traffic participants behaviours are highly diverse in India. Surveyed various segmentation models such as FCN, UNet, SegNet, Efficient Net, Pyramid Scene Parsing Net and DeepLabV3 to increase the performance for semantic labeling of pixels. *(Github)*
- **Understanding Clouds from Satellite Images (Kaggle:Top 30%)**: Addressed the problem of building climate models by analyzing cloud organization patterns from satellite images. Explored various Transfer learning and data augmentation techniques to boost the classification accuracy. *(Github)*
- **Amazon Product Review System**: Devised architectures to use Review titles and statements for sentiment analysis. Used NLTK and scikit-learn for vectorization and embeddings. Applied RNN, LSTM and BiLSTM models to improve the classification accuracy. *(Github)*
- **Bothoven, IIT Bombay Robotics Competition**: Assembled a line follower robot that process audio and strikes the rod based on musical sequence subject to various geometric and movement constraints. Worked on Audio Processing, Wireless Communication, Line Follower & Path Planning.

Programming Skills

- **Languages**: Python, C++, Basic HTML
- **Frameworks**: PyTorch, Keras, OpenCV
- **Tools & Platforms**: Git, LaTeX, Vim, Sublime, Visual Studio, Linux

Service

- **Conference Reviewer**: NeurIPS'23,'24,'25, ICLR'23,'24,'25, CVPR'23,'24,'25, ECCV/ICCV'22,'23,'24
- **Journal Reviewer**: MVAP'23, CVIU'23, TIP'24

Achievements

- **Doctoral Consortium**: Selected for IEEE/CVF Winter Conference on Applications of Computer Vision Doctoral Consortium, 2025.
- **8th HLF**: Selected for Heidelberg Laureate Forum 2021.
- **Student Travel Grant**: Received grant to attend 7th NCVPRIPG'19, Karnataka, India.
- **Winner**: Conference Challenge Winner organized in 3rd ICCVIP'18, Madhya Pradesh, India.
- **Top 5/15**: Conference Challenge organized in 26th ACM Multimedia Conference'18, Seoul, South Korea.
- **Top 50/2500**: e-Yantra Robotics Competition, IIT Bombay.
- **National Top 0.01%**: IIT-JEE Mains, 2015
- **Top 14/6k**: NTSE Stage-I, 2013
- **National Top 1%**: International Maths Olympiad, 2012.
- **1st Rank**: Military School Entrance Exam.

Extracurricular

- **Department Head**: Embedded & Machine Vision dept., INFERNO, Go-Kart Team DTU.
- **Team Captain**: Inter-Hostel Football Team. (Runner-Ups)
- **Organizing Member**: Organized Robowars event in DTU TechFest'17.
- **Team Member**: Organized various events as a core member of Stratazenith society.