AKSHITA GUPTA

akshitac8.github.io | Google Scholar | akshita.sem.iitr@gmail.com

EDUCATION

Technical University of Darmstadt — Darmstadt, Germany
PhD in Computer Science, ELLIS PhD Program — Starting in Mar'25
Supervisor: Prof. Marcus Rohrbach
ELLIS co-supervisor: Dr. Federico Tombari (Google Zurich)

University of Guelph & Vector Institute — Toronto, Canada
MASc in Computer Engineering with Specialization in AI — Sep'22 – Sep'24
Supervisor: Graham W. Taylor
Thesis: Advancing Temporal Action Localization: Efficient Large Model Adaptation and Open-Vocabulary Recognition in Videos

DIT University — Dehradun, India
BTech in Computer Science Engineering — Aug'14 – Dec'18

Research Internships

Research Intern, Apple — Cupertino, CA
Mentor: Tatiana likhomanenko, Navdeep Jaitly — Jun'24 - Feb'25

- Working on video to speech generation pipeline and submitted the work at CVPR 2025.

Research Intern, Microsoft Research — Remote
Mentor: Gaurav Mittal, Dr. Mei Chen — Jun'23 - Mar'24

- Worked on developing memory efficient end-to-end training algorithm for Temporal Action Localization.

Applied Machine Learning Intern, Vector Institute — Toronto, Canada
Supervisor: Dr David Emerson — Jan'23 - May'23

- Working on the prompting for Large language models and developing implementations towards the prompt engineering lab

PUBLICATIONS

1. Visatronic: A Multimodal Decoder-Only Model for Speech Synthesis [paper]
   Akshita Gupta, Tatiana Likhomanenko, Karren Dai Yang, Richard He Bai, Zakaria Aldeneh, Navdeep Jaitly
   Arxiv 2025

2. Long-Short-range Adapter for Scaling End-to-End Temporal Action Localization [paper]
   Akshita Gupta*, Gaurav Mittal*, Ahmed Magooda, Ye Yu, Graham W. Taylor, Mei Chen
   WACV 2025 - Oral!

3. Open-Vocabulary Temporal Action Localization using Multimodal Guidance [paper]
   Akshita Gupta, Aditya Arora, Sanath Narayan, Salman Khan, Fahad Shahbaz Khan, Graham W. Taylor
   BMVC 2024

4. Generative Multi-Label Zero-Shot Learning [paper, code, webpage]
   Akshita Gupta*, Sanath Narayan*, Salman Khan, Fahad Shahbaz Khan, Ling Shao, Joost van de Weijer
   TPAMI 2023

5. OW-DETR: Open-world Detection Transformer [paper, code]
   Akshita Gupta*, Sanath Narayan*, Joseph KJ, Salman Khan, Fahad Shahbaz Khan, Mubarak Shah
   CVPR 2022

6. Discriminative Region-based Multi-Label Zero-Shot Learning [paper, code, webpage]
   Sanath Narayan*, **Akshita Gupta***, Salman Khan, Fahad Shahbaz Khan, Ling Shao, Mubarak Shah
   ICCV 2021

7. Latent Embedding Feedback and Discriminative Features for Zero-Shot Classification [paper, code, webpage]
   Sanath Narayan*, **Akshita Gupta***, Fahad Shahbaz Khan, Cees G.M. Snoek, Ling Shao
   ECCV 2020

8. iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images [paper, webpage, code]
   Syed Waqas Zamir*, Aditya Arora*, **Akshita Gupta**, Salman Khan, Guolei Sun, Fahad Shahbaz Khan, Fan Zhu, Ling Shao, Gui-Song Xia, Xiang Bai
   CVPR-W Oral 2019

9. Acoustic features fusion using attentive multi-channel deep architecture [paper, ppt, code]
   Gaurav Bhatt, **Akshita Gupta**, Aditya Arora, Balasubramanian Raman
   InterSpeech-W 2018

(* denotes equal contribution)

Work EXPERIENCE

Scientist in Residence, NextAI — Toronto, Canada
Supervisor: Prof. Graham Taylor — May'24 - Aug'24

* Technical consultant for AI-based startups. Provided support on the implementation of state-of-the-art deep learning algorithms for various industry applications.

Data Scientist, Bayanat for Mapping & Surveying — Abu Dhabi, UAE
Supervisor: Dr Meng Wang — Jan'22 - Aug'22

* Working towards delivering Computer Vision models based on object detection, segmentation, satellite imagery, and autonomous driving.

Research Engineer, Inception Institute of Artificial Intelligence — Abu Dhabi, UAE
Supervisors: Dr Sanath Narayan, Dr Salman Khan, Dr Fahad Shahbaz Khan — Dec'18 – Jan'22

* Developing deep learning algorithms for low- (Few- and zero-) shot detection and classification, generative adversarial models and open-world object detection problems.
* Developed rock & seismic layer classification system.
  1. Refined classification algorithm for the task of rock and texture classification.
  2. Maintained & deployed GUI Interface for the algorithm presenting real-time results.
* Worked on High resolution imagery object detection and object counting system.
  1. Lead and maintained the main deployment codes working along with imagery team.
  2. Object detection: Improved object detector models like PANet and Mask-RCNN for their collected data.
  3. Object Counting: Implemented and combined different counting algorithms like LCFCN and LPNs.

Research & Development Intern, Mozilla — Remote
Supervisor: Mrs. Emma Irwin — May'18 – Aug'18

* Developed open-source analytics dashboard, metrics to evaluate diversity & inclusion across diff. communities.

Research Intern, IIT Roorkee — Roorkee, India
Supervisor: Dr R Balasubramanian — May'17 – Dec'18

* Worked on acoustic scene recognition & audio tagging system using channel & spatial attention modules.

PROFESSIONAL ACTIVITIES

* **Conference and Journal Reviewing**
  CVPR (2022-25) and ECCV (2022, 2024), ICCV 2021, TPAMI
* **Invited Talks and Panels**
  ComputerVision talks Dec'21 Call, Mozilla Open-source community Call
* Undergraduate Teaching Assistant, TCS821: Cloud Computing

ACHIEVEMENTS

* Travel Scholarship for ALL-Hands Mozilla, San Francisco. (Awarded to top 1% candidates)
* Outreachy Scholarship recipient (2018- 2019). (Awarded to top 2% candidates)
* Selected for Bertelsmann Data Science Scholarship. (Awarded to top 1500 students)
* Scored among top 150 globally at Cognizant Mastercode Hackathon

PROGRAMMING SKILLS

* **Languages:** Python, C++, SQL, HTML, Javascript
* **Libraries:** Pytorch, Tensorflow, Keras
* **Frameworks:** Flask, Bootstrap
* **Software:** GIT, Docker, Latex