ISHAN RAJENDRAKUMAR DAVE (PH.D., APPLIED SCIENTIST)

Seattle, WA ishandave95@gmail.com GitHub Webpage LinkedIn Google Scholar

Research Interests: Diffusion-based Image and Video Foundation Models, Large Video–Language Models (V-LLMs), Self-Supervised Video Representation Learning, Video Retrieval

Education

Ph.D. in Computer Science University of Central Florida, USA **Aug 2019 – Dec 2024**

- Advisor: Dr. Mubarak Shah
- Dissertation: Towards Label Efficiency and Privacy Preservation in Video Understanding

B.Tech in Electronics and Communication S.V. National Institute of Technology, India **2013 – 2017**

Work Experience

Adobe Inc., Seattle, WA, USA *Applied Scientist* **Jan 2025 – Present**

↪ Adobe Firefly team

- Extensive experience training state-of-the-art image and video diffusion models across open research and in-house.
- **Initiated and led** the *Generative Image Upscaler* project end-to-end: from proposing the idea and developing the research prototype to training the in-house production model and driving performance improvements beyond competitive systems. Generative Upscaler shipped in general-release of *Photoshop* as the default upsampling solution[1] and a core component of Adobe's flagship *Firefly Custom Models* ecosystem[2].
- **Initiated and led** the effort of *Generative Video Refiner*: improving cutting-edge video generation models like Veo3, Sora2, Ray3, Pika2.2, and achieving state-of-the-art performance over existing restoration methods [1].

Apple Inc., Cupertino, CA, USA *PhD AI/ML Intern* **May 2024 – Aug 2024**

↪ Dr. Hao Pan, Video Engineering Group

- Enhanced Stable Diffusion models for image editing by leveraging vision-language multimodal foundation models.
- Trained diffusion models on a large-scale, high-resolution dataset of 10M samples.
- Reproduced and outperformed state-of-the-art image editing methods using a novel approach.

Adobe Inc., San Jose, CA, USA *Research Scientist Intern* **May 2023 – Nov 2023**

↪ Dr. Simon Jenni and Dr. Fabian Caba

- Improved the fine-grained capability of the existing video retrieval methods.
- Experience working with large-scale video galleries (millions of videos).
- Filed Patent, Paper accepted in ECCV (Oral, Top 3%) [5].

Adobe Inc., San Jose, USA *Research Scientist Intern* **May 2022 – Nov 2022**

↪ Dr. Simon Jenni

- Developed a novel self-supervised video representation framework by reformulating temporal self-supervision as frame-level recognition tasks and introducing an effective augmentation strategy to mitigate shortcuts.
- Achieved state-of-the-art performance on 10 video understanding benchmarks of linear classification (Kinetics400, HVU, SSv2, Charades), video retrieval (UCF101, HMDB51), and temporal correspondence (CASIA-B).
- Published paper at AAAI [8].

Selected Publications

1. **Ishan Dave***, Tejas Panambur*, Chongjian Ge, Ersin Yumer, Xue Bai. Diffusion-Prior-Guided Approach for Structure and Motion Restoration in Generative and Real Videos, 2025 (*= equal contribution) (ArXiv: 2512.12060).
2. Animesh Gupta, Jay Parmar, **Ishan Dave**, and Mubarak Shah. From Play to Replay: Composed Video Retrieval for Temporally Fine-Grained Videos. Neural Information Processing Systems (**NeurIPS**), 2025.
3. David Shatwell, **Ishan Dave**, Sirnam Swetha, and Mubarak Shah. GT-Loc: Unifying When and Where in Images Through a Joint Embedding Space. Proceedings of the IEEE/CVF International Conference on Computer Vision (**ICCV**), 2025. **Oral (top 0.6% of papers)**
4. Joseph Fioresi, **Ishan Dave**, and Mubarak Shah. ALBAR: Adversarial Learning approach to mitigate Biases in Action Recognition, International Conference on Learning Representations (**ICLR**), 2025.
5. **Ishan Dave**, Fabian Caba, Mubarak Shah, and Simon Jenni. Sync from the Sea: Retrieving Alignable Videos from Large-Scale Datasets, European Conference on Computer Vision (**ECCV**), 2024. **Oral (top 3% of papers)**
6. **Ishan Dave**, Mamshad Nayeem Rizve, and Mubarak Shah. FinePseudo: Improving Pseudo-Labelling through Temporal-Alignablity for Fine-Grained Action Recognition. European Conference on Computer Vision (**ECCV**), 2024.

[1] https://helpx.adobe.com/photoshop/desktop/repair-retouch/clean-restore-images/enhance-image-quality-with-generative-upscale.html

[2] https://business.adobe.com/products/firefly-business/custom-models.html

7. **Ishan Dave**, Tristan de Blegiers, Chen Chen and Mubarak Shah. CodaMal: Contrastive Domain Adaptation for Malaria Detection in Low-Cost Microscopes, 31st IEEE International Conference on Image Processing (**ICIP**), 2024. **Oral**.
8. **Ishan Dave**, Simon Jenni, and Mubarak Shah. No More Shortcuts: Realizing the Potential of Temporal Self-Supervision, AAAI Conference on Artificial Intelligence (**AAAI**), Main Technical Track, 2024.
9. **Ishan Dave**, Mamshad Nayeem Rizve, Chen Chen, and Mubarak Shah. TimeBalance: Temporally-Invariant and Temporally-Distinctive Video Representations for Semi-Supervised Action Recognition, Conference IEEE Computer Vision and Pattern Recognition (**CVPR**), 2023.
10. **Ishan Dave**, Chen Chen, and Mubarak Shah. SPAct: Self-supervised Privacy Preservation for Action Recognition, Conference IEEE Computer Vision and Pattern Recognition (**CVPR**), 2022.
11. **Ishan Dave**, Rohit Gupta, Mamshad Nayeem Rizve, and Mubarak Shah. TCLR: Temporal Contrastive Learning for Video Representation, Computer Vision and Image Understanding (**CVIU**), 2022. **[240+ citations!]**
12. **Ishan Dave***, Tristan de Blegiers*, Adeel Yousaf, and Mubarak Shah. EventTransAct: A video transformer-based framework for Event-camera based action recognition, IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS**), 2023. (*= equal contribution)
13. Joseph Fioresi, **Ishan Dave**, and Mubarak Shah. TeD-SPAD: Temporal Distinctiveness for Self-supervised Privacy-preservation for video Anomaly Detection, IEEE/CVF International Conference on Computer Vision (**ICCV**), 2023.
14. Tushar Sangam, **Ishan Dave**, Waqas Sultani, and Mubarak Shah. TransVisDrone: Spatio-Temporal Transformer for Vision-based Drone-to-Drone Detection in Aerial Videos. IEEE International Conference on Robotics and Automation (**ICRA**), 2023.
15. **Ishan Dave**, Zacchaeus Scheffer, Akash Kumar, Sarah Shiraz, Yogesh Singh Rawat, Mubarak Shah. GabriellaV2: Towards better generalization in surveillance videos for Action Detection, 4th International Workshop on Human Activity Detection in multi-camera, Continuous, long-duration Video (HADCV'22), at the IEEE Winter Conf. on Applications of Computer Vision (**WACV**), 2022.
16. Mamshad Nayeem Rizve, Ugur Demir, Praveen Tirupattur, Aayush Jung Rana, Kevin Duarte, **Ishan Dave**, Yogesh Singh Rawat, and Mubarak Shah. Gabriella: An online system for real-time activity detection in untrimmed surveillance videos, 25th International Conference on Pattern Recognition (**ICPR**), 2020 **[Best Scientific Paper Award]**

This is a publicly available version of my CV, for articles under-review contact me

Patents

1. Action Recognition System Preserves Privacy in Video Sharing. Inventors: **Ishan Dave**, Mubarak Shah, Chen Chen. The University of Central Florida. US Patent: US12142053B2. (Status: **Granted**)
2. Identifying and aligning video clips from large-scale video datasets System. Inventors: Simon Jenni, **Ishan Dave**, Fabian Caba, US Patent: US20250342699A1. (Status: Filed)

Major Research Projects

Image and Video Foundation Models **Aug 2023 – present**

- *Video Foundational models*[2]: Introduced a large-scale benchmark and training paradigm enabling video foundational models to perform temporally fine-grained composed retrieval across subtle, fast-paced action differences.
- *Image Foundational models*[3]: Developed a retrieval-driven image foundational model that predicts when and where a photo was captured by learning a shared embedding space for time, location, and visual cues.
- *Fine-grained Video Understanding* [6]: Improved the visual encoding of Large Video-Language model for the label-efficient Fine-grained Action recognition in a semi-supervised setting.
- *Privacy Preservation in Large Video Models*[??]: Proposed a latent-space anonymization module for video foundation models that removes sensitive attributes while preserving downstream task utility, reducing privacy leakage.

Label-Efficient Video Representation Learning **May 2020 – Dec 2024**

- *TCLR Framework*[11] (May 2020- June 2021): Proposed novel temporal contrastive losses to explicitly increase the temporal distinctiveness at two temporal aggregation steps in video tasks: (1) clip-level (2) feature level.
- *TimeBalance Framework*[9] (Aug 2022 - Jan 2023): Studied two complementary self-supervised video representations: (1) Temporally-Invariant (2) Temporally-Distinctive. Proposed a dual teacher-based framework for semi-supervised action recognition using a novel temporal-similarity based reweighting strategy.
- *Mitigating Shortcuts in temporal self-supervision* [8] (May 2022- Jan 2023): Internship work at Adobe.

Privacy Preserving Video Understanding **June 2021 – present**

- *Privacy Preserving Action Recognition*[10]: Implemented a privacy-preserving action recognition framework that removes private attributes (gender, skin color, etc.) without labels, maintaining competitive performance and achieving best generalization across novel action and attributes.

- *Privacy Preserving Video Anomaly Detection*[13]: Developed a privacy-aware video anomaly detection framework utilizing temporally-distinctive video representations, achieving state-of-the-art tradeoff between privacy protection and utility performance on three popular weakly supervised VAD datasets.
- *Action Fairness*[4]: Studying the bias of private attributes (gender, skin color, etc.) in SOTA action recognition models.

Video Understanding for Robotics **May 2022 – present**

- *TransVisDrone Framework*[14]: Proposed an end-to-end drone detection framework to tackle various challenging real-world scenarios by learning spatio-temporal dependencies of drone motion.
- *EventTransAct Framework*[12]: Proposed a video transformer-based framework for event-camera based action recognition, which leverages event-contrastive loss and augmentations to adapt the network to event data.
- *Egocentric Multimodal Action Recognition* : Proposed a framework for recognizing actions from egocentric RGB and Depth modalities in an industry-like environment.

Awards and Honors

Selected for Doctoral Consortium (**ECCV**)	**2024**
Outstanding Reviewer Ranked in the top 2% for review quality among 10,000 reviewers (**CVPR**)	**2024**
Nomination The Order of Pegasus Award, University of Central Florida	**2024**
1st place Multi-modal Action Recognition challenge (**ICIAP**)	**2023**
2nd place, ActivityNet ActEV Challenge (**CVPR**)	**2022**
2nd place TRECVID ActEV: Activities in Extended Video	**2021**
1st place & Jury Prize, VI-Priors Action Recognition Challenge (**ICCV**)	**2021**
1st place, PMiss@0.02tfa, ActivityNet ActEV SDL (**CVPR**)	**2021**
1st place, VI-Priors Action Recognition Challenge (**ECCV**)	**2020**
1st place, PMiss and nAUDC, ActivityNet ActEV SDL (**CVPR**)	**2020**
2nd place, TRECVID ActEV: Activities in Extended Video	**2020**
ORCGS Doctoral Fellowship,	**2019-2020**
Top 0.5%, Joint Engineering Entrance-Mains exam, India	**2013**

Skills

Programming Languages	Python,
Deep learning frameworks	PyTorch, Keras
Tools/Frameworks	AWS, OpenCV, SciKit, MATLAB

Coursework

- Advance Computer Vision (*CAP 6412*)
- Advance Machine Learning (*CAP 6614*)
- Computer Vision Systems (*CAP 6411*)

Professional Services

- Reviewer of CVPR, ECCV, ICCV, WACV, ICRA, IROS, TPAMI, TIMM, CVIU, Pattern Recognition, TCSVT, IEEE Access, Multimedia Tools and Application, etc.
- Program Committee Member of AAAI 2025.
- Technical Committee, BMVC Privacy and Fairness Workshop 2024.
- Mentored students of NSF Research Experience for Undergrad (REU) 2020, 2021 & 2022

Character Referees

Available upon request