Prateksha Udhayanan

pudhayan@umd.edu | +1 (240)-501-2679 | GitHub | LinkedIn | Website

Research Interests

My research interests lie in **computer vision**, with a focus on **generative models for images and videos**. My recent projects span **pixel-space generation**, from multi-resolution diffusion for image generation to adapting pretrained latent-space transformer backbones for direct pixel-space video generation. Previously, I have also worked on controllable image generation and motion quality evaluation in generated videos.

Education

- **Ph.D. in Computer Science** — College Park, MD, USA
  University of Maryland, College Park; CGPA: 4.00/ 4.00 — *2024 – Present*
  - **Advisor:** Prof. Abhinav Shrivastava
  - **Dean's Fellowship:** 2024, 2025
  - **Awarded the Certificate of Outstanding Achievement:** 2025–2026
  - **Teaching Assistant:** CMSC412 (Fall 2024): Operating Systems
    CMSC335 (Spring 2025): Web Application Development with JavaScript

- **B.Tech + M.Tech in Computer Science (Integrated Master of Technology)** — Bangalore, India
  International Institute of Information Technology, Bangalore; CGPA: 3.76/4.00 — *2017 – 2022*
  - **Dean's Merit List:** 2018, 2020, 2021, 2022
  - **Teaching Assistant:** ESS 201 (Fall 2020): Programming II (C++, Java)
    ESS 112 (Spring 2021): Programming in Python
    AI 511 (Fall 2021): Machine Learning

Publications and Preprints

- [CVPR 2026] Soumik Mukhopadhyay*, *Prateksha Udhayanan** and Abhinav Shrivastava. **Scale Space Diffusion.** [LINK]
- [WACV 2025] Sahil Goyal, Abhinav Mahajan, Swasti Mishra, *Prateksha Udhayanan*, Tripti Shukla, KJ Joseph and Balaji Vasan Srinivasan. **Design-o-meter: Towards Evaluating and Refining Graphic Designs**. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 5676-5686), 2025. [LINK]
- [AAAI 2024] Koustava Goswami, Srikrishna Karanam, *Prateksha Udhayanan*, KJ Joseph and Balaji Vasan Srinivasan. **CoPL: Contextual Prompt Learning for Vision-Language Understanding**. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 16, pp. 18090-18098), 2024. [LINK]
- [WACV 2024] KJ Joseph, *Prateksha Udhayanan*, Tripti Shukla, Aishwarya Agarwal, Srikrishna Karanam, Koustava Goswami, and Balaji Vasan Srinivasan. **Iterative Multi-Granular Image Editing using Diffusion Models**. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 8107-8116), 2024. [LINK]
- [WACV 2023] *Prateksha Udhayanan*, Suryateja BV, Parth Laturia, Dev Chauhan, Darshan Khandelwal, Stefano Petrangeli, and Balaji Vasan Srinivasan. **Recipe2Video: Synthesizing Personalized Videos from Recipe Texts**. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2268-2277), 2023. [LINK]
- [SN Computer Science] Chinchu Thomas, *Prateksha Udhayanan*, Ayush Yadav, Seethamraju Purvaj, and Dinesh Babu Jayagopi. **Multimodal Unsupervised Domain Adaptation for Predicting Speaker Characteristics from Video** - SN Computer Science, 5(5), article 531. 2024 [LINK]
- [AI-ML Systems 2021] Arjun Verma, *Prateksha Udhayanan*, Rahul Murali Shankar, Nikhil KN, and Sujit Kumar Chakrabarti. **Source-Code Similarity Measurement: Syntax Tree Fingerprinting for Automated Evaluation**, The First International Conference on AI-ML-Systems (pp. 1-7). 2021 [LINK]
- [Physica A: Statistical Mechanics and its Applications] *Prateksha Udhayanan*, Swasti S Mishra, and Shrisha Rao. **Firm dynamics and employee performance management in duopoly markets** Physica A: Statistical Mechanics and its Applications. (583, p.126298). 2021 [LINK]
- *Prateksha Udhayanan*, Srikrishna Karanam, and Balaji Vasan Srinivasan. **Learning with Multi-Modal Gradient Attention for Explainable Composed Image Retrieval** - [arXiv:2308.16649]

Patents

- Suryateja BV, *Prateksha Udhayanan*, Dev Chauhan, Parth Laturia, Darshan Khandelwal, Stefano Petrangeli, and Balaji Vasan Srinivasan. **Auto-Generating Video to Illustrate a Procedural Document**. US Patent Number: 12027184
- *Prateksha Udhayanan*, Kuldeep Kulkarni, Aishwarya Agarwal, Balaji Vasan Srinivasan. **Method for Evaluating Motion Videos Generated by Neural Network Models** (Filed with USPTO, US Patent App. 19/206,528)
- *Prateksha Udhayanan*, Srikrishna Karanam, and Balaji Vasan Srinivasan. **Text-Conditioned Visual Attention for Multimodal Machine Learning Models** (Filed with USPTO, US Patent App. 18/351,211)
- Koustava Goswami, Srikrishna Karanam, KJ Joseph, *Prateksha Udhayanan*, and Balaji Vasan Srinivasan. **Generating Text Prompts for Digital Images Utilizing Vision-Language Models and Contextual Prompt Learning** (Filed with USPTO, US Patent App. 18/342,954)

Experience

- **Institute of Foundation Models – MBZUAI, Sunnyvale, USA** Manager: Dr. Matthew Gwilliam
  Research Intern *May 2026 – Present*
  - Developing and training pixel-space video generation architectures that adapt pretrained latent-space transformer backbones across model scales ranging from 2B to 14B parameters.
- **Adobe Research, Bangalore, India** Manager: Dr. Balaji Vasan Srinivasan
  Research Associate *July 2022 – July 2024*
  - Worked on computer vision and multi-modal content understanding projects as part of the Collaborative Creativity team, contributed to papers, filed patents, and developed research technologies that were integrated into products.
  - Projects spanned image retrieval, editing and generation, automated graphic design creation and editing, layout and color optimization, and cinemagraph generation and evaluation.
- **Adobe Research, Bangalore, India** Mentor: Dr. Balaji Vasan Srinivasan
  *Research Intern [**Published in WACV 2023**]* *May 2021 – August 2021*
  - Built a novel end-to-end deep-learning based architecture to convert instructional documents into multimodal videos.
  - Proposed re-ranking methods to retrieve and select the optimal combination of multimodal visuals aligned with constraints.
  - Used Viterbi algorithm-based optimization for multi-modal frame selection, resulting in a coherent video composition.
  - Established evaluation metrics based on cognitive models of procedural text understanding.
- **Siemens Technology and Services Pvt. Ltd. Bangalore** Mentors: Varghese Alex, Vinay Sudhakaran
  Research Intern *May 2020 – July 2020*
  - Implemented an LSTM-based hierarchical model in TensorFlow for Anomaly Detection in procedural videos.
  - The model is trained to learn the sequential structure of procedural activities from text domain and transfer it to visual domain which enables visual appearance learning and future prediction.
- **Multimodal Perception Lab (MPL) - IIITB** Mentor: Dr. Dinesh Babu J
  Research Intern at MPL *June 2019 – September 2019*
  - Indian Sign Language Synthesis - implemented 3D CNN models to estimate 3D hand joint coordinates from depth images.
  - Developed a pipeline to map the predicted coordinates to a virtual avatar for gesture generation.
  - Reduced the gesture generation time for each sign from 20 minutes to less than 2 minutes by automating the process.

Selected Projects

- **Scale Space Diffusion**
  University of Maryland | Published in CVPR 2026
  - Studied the relationship between information degradation in diffusion processes and image scale-spaces.
  - Developed the mathematical formulation for a family of generalized linear diffusion processes.
  - Designed a novel architecture supporting both resolution-changing and resolution-preserving reverse diffusion steps across multiple scales.
- **Evaluating Video Generation Models for Motion Quality**
  Adobe Research
  - Proposed a benchmark to evaluate motion modeling in image-to-video generation models, focusing on linear, circular, and oscillatory motion types. Designed metrics for motion smoothness, direction, speed, and overall video quality.
  - Helped in product integration of automatic animation for fluid elements such as water, smoke, and fire, from a single image.
- **Composed Image Retrieval**
  Adobe Research
  - Developed a novel multi-modal gradient attention computation mechanism for composed image retrieval.
  - Proposed an end-to-end learning scheme for vision-language siamese transformers with our attention-based loss function.
  - Improved the performance by $\approx 3\%$ on the *Recall*@1 metric and by $\approx 20\%$ on the $Recall_{subset}$@1 metric.
- **Design Co-pilot**
  Adobe Research | Integrated into product | Published in WACV 2025
  - Implemented novel methodologies to score and refine designs based on color, layout and content.
  - Designed and built an end-to-end pipeline to edit and transform a design based on a user prompt.
  - Implemented a transformer-based masked field prediction approach for layout and color optimization in design.

Achievements

- **Winner of CodeHers Coding Challenge 2021:** One of the **top 15** winners out of 50,000+ women participants.
- **Rails Girls Summer of Code Scholar 2018:** One of the 7 full-time teams selected out of 200+ international teams.

Service

- **Reviewer:** CVPR 2026; AAAI 2026, 2025; ECCV 2026 Workshop on Women in Computer Vision (WiCV)
- **Founding Member, Lean In Chapter IIIT-B**