Isht Dwivedi

718-679-3159 | isht.dwivedi@gmail.com | Google Scholar | LinkedIn | github.com/isht7

EDUCATION

Columbia University — New York, USA
M.S. in Computer Science, Specialization: Machine Learning, GPA: 3.94/4.00 — *Aug 2017 – Dec 2018*

Indian Institute of Technology Guwahati — Guwahati, India
B. Tech., Major: Electronics and Electrical Engg., Minor: Computer Science, GPA: 8.36/10 — *Jul 2012 – Jun 2016*

WORK EXPERIENCE

Honda Research Institute USA — Feb 2019 – Present
Senior Research Engineer — *San Jose, USA*

- Created an extensive inhouse HD map dataset for training models to estimate HD map using only camera data, and to predict lane connectivity information at intersections. Some challenges included obtaining highly accurate lidar-camera (more accurate than possible using checkerboard) and lidar-INS calibrations.
- Led the Road Condition Monitoring Project where the goal is to automate geo-tagging of damaged/degraded road infrastructure. Key achievements included (1) creating a system that can create and maintain a map of traffic signs using a fleet of cars to identify damaged and missing traffic signs, (2) creating segmentation models to segment out lane markings and cable barriers by quality, (3) research on anomaly classification and reasoning related to this work was published at CVPR 2025.
- Our work on zero-shot video classification (for action videos) achieved state of the art results for unseen classes by using "leaf augmentations" and "shadow negative" concepts. This work was published at WACV 2025.
- Road scene understanding with emphasis on unstructured events like construction zones. Bird's Eye View (BEV) semantic segmentation map is directly predicted from perspective view video stream captured using a front-facing camera mounted on a car. Used COLMAP to create colored 3D point clouds from video snippets which were annotated to synthesize ground truth. Achieved state of the art performance on BEV segmentation. (Paper at BMVC 2021). Automated captions are generated to help the ego driver navigate around construction zones.
- Created a new benchmark dataset and method to identify risky agents on the road. A caption describing the risk is also produced - this can be used to alert the driver of oncoming risks. The work is published at WACV 2023.
- Future trajectory prediction for all agents in a scene in a single shot in constant time. Single shot nature of this work makes it faster than other works in crowded situations (published at IROS 2020). Single shot nature achieved by predicting composite fields from neural network. Non-Local interaction block used to capture interactions between agents. Semantic segmentation features used to make semantically aware predictions.
- Proposed a 2-stage road scene understanding algorithm (road place classification, construction zone detection) to be run on car to obtain predictions in real-time. Improved road place classification performance by redesigning neural network architecture and adding semantic segmentation as auxiliary task. F-score improved from 28% to 40%. The work has been published at ICRA 2019).

Active Interpretation of Disparate Alternatives (AIDA) Project — Jan – May 2018
Research credit with Prof. Shih-Fu Chang, Columbia University — *New York City, USA*

- Worked on the AIDA project with Prof. Shih-Fu Chang in the DVMM lab.
- My work involved discovering relevant visual concepts from a given scenario (body of text). Using the discovered visual concepts, I experimented with different methods to create a weakly supervised object detection method for these visual concepts using the Open Images dataset.

Indian Institute of Science, Video Analytics Lab — Aug 2016 – May 2017
Research Intern, Prof. Venkatesh Babu — *Bangalore, India*

- Hierarchical CNN used for semantic segmentation of hand drawn sketches with pose estimation as auxiliary task.
- Fully automatic, works without class information on over 11 classes including cat, car, bird, bicycle and airplane.
- Sketches not required for training: Sketchification process used to convert natural photos to sketch like images.
- Published at **ACMMM 2017**.

Ecole Polytechnique, Laboratoire Leprince-Ringuet — May 2015 – Jul 2015
Summer Intern, Prof. Vincent Boudry — *Paris, France*

- High energy particles decay in a particle detector generating new particles in a graph like pattern (shower). Designed and implemented an approach to reconstruct vertices of this graph.
- Written in C++, using PCA and Arbor Particle Flow Algorithm, experiments done on Monte Carlo simulations.

PUBLICATIONS

Jiacong Xu, Shao-Yuan Lo, Bardia Safaei, Vishal M Patel, **Isht Dwivedi**. "*Towards Zero-Shot Anomaly Detection and Reasoning with Multimodal Large Language Models*". Conference on Computer Vision and Pattern Recognition (**CVPR**) 2025. Link to paper.

Reza Ghoddoosian, Nakul Agarwal, **Isht Dwivedi**, Behzad Dariush. "*ACE: Action Concept Enhancement of Video-Language Models in Procedural Videos*". Winter Application on Computer Vision (**WACV**) 2025. Link to paper.

Reza Ghoddoosian, **Isht Dwivedi**, Nakul Agarwal, Behzad Dariush. "*Weakly-Supervised Action Segmentation and Unseen Error Detection in Anomalous Instructional Videos*". International Conference on Computer Vision (**ICCV**) 2023. Link to paper.

Srikanth Malla, Chiho Choi, **Isht Dwivedi**, Joon Hee Choi, Jiachen Li. "*DRAMA: Joint Risk Localization and Captioning in Driving*". Winter Conference on Applications of Computer Vision (**WACV**) 2023. Link to paper.

Srikanth Malla, **Isht Dwivedi**, Behzad Dariush, Chiho Choi. "*NEMO: Future Object Localization Using Noisy Ego Priors*". International Conference on Intelligent Transportation Systems (**ITSC**) 2022. Link to paper.

Reza Ghoddoosian, **Isht Dwivedi**, Nakul Agarwal, Behzad Dariush "*Weakly-Supervised Online Action Segmentation in Multi-View Instructional Videos*" Conference on Computer Vision and Pattern Recognition **(CVPR) 2022**.

Isht Dwivedi, Srikanth Malla, Yi-Ting Chen, Behzad Dariush. "*Bird Eye View Segmentation Using Lifted 2D Semantic Features*", British Machine Vision Conference **(BMVC) 2021**. Link to paper.

Isht Dwivedi, Srikanth Malla, Behzad Dariush, Chiho Choi. "*SSP: Single Shot Future Trajectory Prediction*", International Conference on Intelligent Robots and Systems **(IROS) 2020**. Link to paper.

Isht Dwivedi*, Athmanarayanan Narayanan*, Behzad Dariush. "*Dynamic Traffic Scene Classification with Space-Time Coherence*", International Conference on Robotics and Automation **(ICRA) 2019**. Link to paper.

Kaveri A Thakoor, Qian Zheng, Linyong Nan, Xinhui Li, Emmanouil Manos Tsamis, Rashmi Rajshekhar, **Isht Dwivedi**, Iddo Drori, Paul Sajda, Donald C Hood. "*Assessing the Ability of Convolutional Neural Networks to Detect Glaucoma from OCT Probability Maps*", Journal on Investigative Ophthalmology & Visual Science (**IOVS** 2019). Link to paper.

Iddo Drori, **Isht Dwivedi**, Pranav Shrestha, Jeffrey Wan, Yueqi Wang, Yunchu He, Anthony Mazza, Hugh Krogh-Freeman, Dimitri Leggas, Kendal Sandridge, Linyong Nan, Kaveri Thakoor, Chinmay Joshi, Sonam Goenka, Chen Keasar, Itsik Pe'er. "*High-Quality Protein Q8 Secondary Structure Prediction by Diverse Neural Network Architectures*", **NeurIPS 2018 Workshop** on Machine Learning for Molecules and Materials. Link to paper.

Ravi Kiran, **Isht Dwivedi**, Abhijat Biswas, Sahil Manocha, V. Babu. "*SketchParse: Towards Rich Descriptions for Poorly Drawn Sketches Using Multi-Task Deep Networks*", ACM Multimedia **(ACM MM) 2017**. Link to paper.

Isht Dwivedi, Swapnil Gupta, V. Venugopal, S. Sundaram. "*Online Writer Identification using Sparse Coding and Histogram based Descriptors*", **Oral** presentation at International Conference on Frontiers in Handwriting Recognition **(ICFHR, oral) 2016**. Link to paper.

PATENTS

US Patent 16374205, "Scene classification", Athmanarayanan Lakshmi Narayanan, **Isht Dwivedi**, Behzad Dariush

US Patent 11034357, "Scene classification prediction", Athmanarayanan Lakshmi Narayanan, **Isht Dwivedi**, Behzad Dariush

US Patent 16917864, "Composite field based single shot prediction", **Isht Dwivedi**, Chiho Choi, Srikanth Malla, Behzad Dariush

US Patent App. 17/710,807, "Systems and methods for birds eye view segmentation", **Isht Dwivedi**, Yi-Ting Chen, Behzad Dariush

US Patent App. 17/590,379, "System and method for providing weakly-supervised online action segmentation", Reza Ghoddoosian, **Isht Dwivedi**, Nakul Agarwal, Chiho Choi, Behzad Dariush

Research Projects

Online writer identification   Sept – May 2016
Bachelor Thesis Project   *Guwahati, India*
- Improvement of 10% obtained over state-of-the-art results for the IAM On-Line Handwriting Database when using only online features at the text-line level.
- Developed novel histogram-based features to represent handwriting characteristics of a writer.
- Codebook developed using sparse coding. Modified Tf-Idf approach used to create document descriptor. One vs. all SVM model used as classifier. Paper (oral) published at **ICFHR 2016**.

Snake Robot with passive wheels and active joints [Video of robot]   Jan – Apr 2015
Design Course Project   *Guwahati, India*
- Designed and developed a wirelessly controlled snake robot with passive wheels and active joint modules. Motion of robot specifically designed to imitate that of a snake.
- Zigbee protocol used for wireless communication and Arduino Mega 2560 used for control of servo motors.

Teaching Experience

Teaching Assistant, CS 4995: Deep Learning, Columbia University, Spring 2018
Instructor: Prof. Sameer Maskey

Teaching Assistant, CS 4995: Deep Learning, Columbia University, Fall 2018
Instructor: Prof. Iddo Drori

Relevant Coursework

Graduate level courses: Deep Learning, Machine Learning, Natural Language Processing, Unsupervised Methods of Learning, Bayesian Models for Machine Learning, Visual Interfaces to Computers

Undergraduate courses: Computer Vision, Image Processing, Pattern Recognition and Machine Learning, Probability and Random Processes, Digital Signal Processing, Linear Algebra and Real Analysis, Differential Equations and Multi-variable Calculus, Speech Technology