Saumya Sinha | Curriculum Vitae

1300 30th Street, Apartment B4 -16, Boulder, CO 80303

+7204211144 • saumya.sinha@colorado.edu • saumyasinha.github.io

I am a 4th year PhD candidate at the University of Colorado Boulder. My research focuses on building Machine Learning (ML) based solutions to address challenges related to climate change impacts and sustainability. Through my research, I have made contributions to real world applications such as detecting extreme events with the help of satellite imagery, providing probabilistic forecasts of renewable energy at longer lead-times, and currently, I am developing a ML pipeline to predict the future trends of sea level change.

Education

- **University of Colorado, Boulder** — **Boulder, CO**
  PhD in Computer Science, with Prof. Claire Monteleoni — *2019 -*
- **University of Colorado, Boulder** — **Boulder, CO**
  Masters in Computer Science, GPA - 3.98/4 — *2017-2019*
- **Indian Institute of Technology, Kharagpur** — **Kharagpur, India**
  Integrated BS+MS in Mathematics and Computing, GPA - 8.4/10 — *2009–2014*

Publications

- **Sea-Level Projections via Spatiotemporal Deep Learning from Altimetry and CESM Large Ensembles.**
  Saumya Sinha, Claire Monteleoni, John Fasullo, R Steven Nerem. *Abstract accepted at AGU2022: AI for Ocean and Climate Change Oral.*
- **Week-ahead Solar Irradiance Forecasting with Deep Sequence Learning.**
  Saumya Sinha, Bri-Mathias Hodge, Claire Monteleoni. *CI 2022: Proceedings of the 11th International Conference on Climate Informatics (to appear in Environmental Data Science journal).*
- **Subseasonal Solar Power Forecasting via Deep Sequence Learning. (link)**
  Saumya Sinha, Bri-Mathias Hodge, Claire Monteleoni. *Tackling Climate Change with Machine Learning Neurips Workshop, 2021*
- **Variational Autoencoder Anomaly-Detection of Avalanche Deposits in Satellite SAR Imagery. (link)**
  Saumya Sinha, Sophie Giffard-Roisin, Fatima Karbou, Michael Deschatres, Anna Karas, Nicolas Eckert, Cécile Coléou, Claire Monteleoni. *CI2020: Proceedings of the 10th International Conference on Climate Informatics (ACM 2020).*
- **Detecting Avalanche Deposits using Variational Autoencoder on Sentinel-1 Satellite Imagery. (Spotlight Talk link)**
  Saumya Sinha, Sophie Giffard-Roisin, Fatima Karbou, Michael Deschatres, Anna Karas, Nicolas Eckert, Cécile Coléou, Claire Monteleoni. *Tackling Climate Change with Machine Learning Neurips Workshop, 2019*
- **Can Avalanche Deposits be Effectively Detected by Deep Learning on Sentinel-1 Satellite SAR Images? (link)**
  Saumya Sinha*, Sophie Giffard-Roisin*, Fatima Karbou, Michael Deschatres, Anna Karas, Nicolas Eckert, Cécile Coléou, Claire Monteleoni. *CI 2019: Proceedings of the 9th International Workshop on Climate Informatics.*

Employment

- **AI for Good Research Lab, Microsoft** — **Redmond, US**
  Research Intern — *May 2022–Aug 2022*

  Self supervised learning for geospatial imagery: Developed a self supervised learning algorithm that uses a contrastive loss particularly designed to learn meaningful representations from satellite images and transfers well to semantic segmentation tasks. These representations were evaluated for the downstream task of land cover mapping.
- **American Express** — **Bangalore, India**
  Data Analyst, Big Data Capabilities and Algorithms division — *March 2016–May 2017*

  Text Mining/Modelling to build merchant database: Worked to improve data quality of merchant (businesses using Amex services) attributes in a consumer-friendly database by integrating data from miscellaneous information sources.
- **Accenture** — **Bangalore, India**
  Data Analyst — *June 2014 - Feb 2016*

  Predictive Maintenance: Worked for an Oil and Natural Gas resources client, creating ML pipelines for failure detection of equipments in the plant.

Professional service and Awards

- **Awards at CU Boulder**:
  - Conference support fellowship 2022.
  - Departmental Summer Research fellowship for 2020-2021.
  - Annual Grad Students' Research Expo award 2021.
  - Outstanding TA award for the year 2018-2019.
- **Reviewer**:
  - IEEE Transactions on Sustainable Energy, Journal of Renewable and Sustainable Energy.
  - ICLR, ICML, Climate Informatics.
  - Program Committee member of ICLR & Neurips 2020 workshop on "Tackling Climate Change with Machine Learning."
- **Teaching**:
  - Guest Lecturer for ASEN5307 Engineering Data Analysis, CSCI 7000 Advanced Topics in Machine Learning, CSCI 4802/5802 Data Science Team.
  - Teaching Assistant for Machine Learning, Data Structures.

Course Projects

- **Learning to synthesize adaptive camouflage textures to hide an object by only making use of it's local background (in changing environments when one static camouflage is not enough.)**
  - Worked primarily on two major texture synthesizing models - data driven parametric approach (from the literature of style transfer) and a GAN-based method.
  - To measure the success, used a MaskRCNN detector to find out the extent to which we can adversarially attack them.
  - Used a simulation engine (Unreal) with high-fidelity 3D vehicle models to apply the camouflage pattern on the car's body and test it under a diverse set of environmental conditions.
- **Augmented Reality based IOS app using text detection and manipulation for people with reading disabilities.**
  - Worked on Augmented Reality development using ARkit.
  - Represented virtual 3D text obtained from an deep-learning based OCR model and anchored it over the real text.
  - Added hit test (single/double tap)functionality on the virtual 3D text to navigate to a new screen or display objects relevant to the text. Also enabled anchoring the objects to a plane and interacting with it through hand gestures.
- **Building a GAN based creative assistant for Logo Generation**
  - Explored GAN generative models to create a creative assistance to help generate new and unique preliminary logo ideas.
  - Generated a set of semantically meaningful clusters that could improve in conditional GAN training.
  - Achieved best results (and mitigated mode collapse) with improved distance metric in the WGAN combined with ACGAN, conditioned on the above mentioned clusters.
- **Develop a scalable tool for Trend Classification in Twitter**
  - Built a Machine Learning pipeline with a Real-time Classifier which takes twitter data from a Kafka queue and writes into Cassandra. It was able to classify 6k tweets per minute with a pretty high accuracy(93%).
  - Tweets were classified into topics - Sports, Politics, Entertainment,Technology and Mood. Every trend was assigned a topic based on the majority classification of it's tweets.
- **Depth Map Estimation from Monocular Images**
  - Developed and compared various CNN based architectures from fully connected models to purely convolutional and also used transfer learning with pretrained models to estimate depth map from monocular images.
  - Used the depth map to identify and differentiate between the foreground and background of an image.
  - Visualised features formed by kernels.
  - Developed models on data with indoor scenes and objects. Achieved a very low RMSE on the object data, around 0.4.
- **Motion Planning with Rapidly Exploring Random Trees**
  - Explored and compared various variants of rapidly-exploring randomized-tree algorithm (RRT) to sub-sample the free space and find a path between two points in a 2D grid world with multiple obstacles.
  - Used GUI APIs (built by my team) to visualize and demonstrate the strengths and weaknesses of RRT, A-star and their variants like RRT*, Bi-directional RRT.

Technical skills

Proficient in: Python, C++, Deep Learning Frameworks - Pytorch, Keras, Sklearn
Big data tools – Hive, PySpark, Statistical Skills - R, SQL