

Anshul Nasery

Pre-Doctoral Researcher, Google Research

🌐 Homepage @ anshulnasery@gmail.com 🐙 Github 🎓 Google Scholar

Research Interests

I am interested in fundamental deep learning research for training deployment ready models with strong guarantees.

- › **Out-Of-Distribution Robustness:** Designing training algorithms and model architectures to handle distribution shift.
- › **Understanding deep learning:** Gaining theoretical and empirical insights into neural networks' representations.
- › **Efficient Machine Learning:** Models and algorithms for inference on compute and memory constrained devices.

Education

Aug 2021	Indian Institute of Technology, Bombay	GPA: 9.58/10
Jul 2017	Bachelor of Technology (With Honors) in Computer Science and Engineering, Minor in Statistics <i>Thesis: Continuously Indexed Domain Generalization Along Time</i> <i>Advisor: Prof. Sunita Sarawagi</i>	

Research Experience

Jul 2021	Google Research	Bangalore, India
Present	Pre-Doctoral Researcher Advisor: Dr. Prateek Jain, Dr. Praneeth Netrapalli Worked on research problems around inference efficient and generalizable neural networks.	
Apr 2020	Adobe Research	Bangalore, India
Jul 2020	Research Intern Advisor: Dr. Balaji Vasani Srinivasan Worked on a research problem around Multi-Modal Question Answering [NAACL'21, US Patent].	
May 2019	Aarhus University	Aarhus, Denmark
Jan 2020	Research Intern Advisor: Prof. Davide Mottin Worked on research problems around generative modelling and efficient matrix multiplication.	

Publications

- [C.5] **Training for the Future: A Simple Gradient Interpolation loss to Generalize Along Time** [🔗]
Anshul Nasery*, Soumyadeep Thakur*, Vihari Piratla, Abir De, Sunita Sarawagi
34th Conference on Advances in Neural Information Processing Systems [NeurIPS '21]
- [C.4] **What if Neural Networks had SVDs?** [🔗]
Alexander Mathiasen, Frederik Hvilshøj, Jakob Rødsgaard Jørgensen, Anshul Nasery, Davide Mottin
Spotlight at 33rd Conference on Advances in Neural Information Processing Systems [NeurIPS'20]
- [C.3] **CogCNN: Mimicking Human Cognition to resolve Texture Shape Bias** [🔗]
Satyam Mohla*, Anshul Nasery*, Biplob Banerjee
2022 IEEE International Conference on Acoustics, Speech and Signal Processing [ICASSP'22]
- [C.2] **MIMOQA: Multimodal Input Multimodal Output Question Answering** [🔗]
Hrituraj Singh, Anshul Nasery*, Denil Mehta*, Jatin Lamba, Aishwarya Agarwal, Balaji Vasani
2021 Conference of the North American Chapter of the Association for Computational Linguistics [NAACL'21]
- [C.1] **Rule Augmented Unsupervised Constituency Parsing** [🔗]
Anshul Nasery*, Atul Sahay*, Ayush Maheshwari, Ganesh Ramakrishnan, Rishabh Iyer
Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021 [Findings of ACL'21]

Pre-prints

- [W.2] **Learning an Invertible Mapping can Mitigate Simplicity Bias** [🔗]
Sravanti Addepalli*, Anshul Nasery*, R Venkatesh Babu, Praneeth Netrapalli, Prateek Jain
DistShift Workshop, NeurIPS 2022, Under Submission at ICLR'23 [NeurIPS-W'22]
- [W.1] **DAFT: Distilling Adversarially Finetuned Teachers for better OOD generalization** [🔗]
Anshul Nasery, Sravanti Addepalli, Praneeth Netrapalli, Prateek Jain
Principles of Distribution Shifts Workshop, ICML 2022, Preparing for Submission to TMLR [ICML-W'22]

Selected Research Projects

Out-of-Domain Robustness of Neural Nets

Sept'21 - Present

Advisors: *Dr. Prateek Jain, Dr. Praneeth Netrapalli*

- › Developed a novel feature reconstruction regularizer to alleviate simplicity bias and improve OOD generalization.
- › Obtained upto **1% gain** in accuracy over state-of-the-art methods on the DomainBed benchmark. [NeurIPS-W'22]
- › Combined adversarial fine-tuning and knowledge distillation to boost the OOD robustness of small models. [ICML-W'22]
- › Using the proposed technique, a **ResNet-50 can outperform a ResNet-101 by 2.5%** on the DomainBed benchmark.

Training For the Future

Jul'20 - Jul'21

Advisor: *Prof. Sunita Sarawagi*

- › Investigated gradient based techniques for better **domain generalization on temporally varying** data. [NeurIPS'21]
- › Achieved over **20% relative improvements** over sota on 5 real world datasets including M5 and HousePrice.
- › Proved upper bounds on generalization error of the proposed method for regression tasks with temporal drift.

What If Neural Networks had SVDs?

Nov'19 - May'20

Advisor: *Prof. Davide Mottin*

- › Developed a parallelizable algorithm for matrix multiplication via Householder decompositions of orthogonal matrices.
- › Achieved a **29x speedup** over prior work by implementing the algorithm in CUDA for running on GPUs.
- › Work presented as a **Spotlight paper (top 3% of all submissions)** at NeurIPS 2020 [C.4].

Inference Efficient ML Models

Jul'21 - Present

Advisors: *Dr. Prateek Jain, Dr. Praneeth Netrapalli, Dr. Gaurav Aggarwal*

- › **NAS**. Achieved **0.8% gain in ImageNet accuracy** for no extra FLOPs on MobileNetV3 using a novel FLOPs regularizer.
- › **Conditional Computation**. Obtained **1% gain in ImageNet accuracy** for MobileNetV2 by introducing decision trees to route examples. Introduced a skip-and-branch architecture for **25% savings in amortized FLOPs** with MobileNetV3.
- › **Compressing LLMs**. Adapting algorithms from the compressed sensing literature to prune weight matrices of large language models by over 50%, resulting in **latency reduction of 30%**.

Academic Achievements

- › Awarded Institute Academic Prize for exceptional academic performance (top 10% of class) in IIT Bombay in 2017-2018.
- › Ranked **137** in 110000 candidates in JEE Advanced 2017 and 265 in 1.5 million candidates JEE Mains 2017.
- › Placed among the **top 35 students** in Indian National Astronomy Olympiad 2017 and qualified for Indian National Olympiad of Informatics, Indian National Physics Olympiad & Indian National Chemistry Olympiad 2017.

Other Projects

Generative Modelling using Invertible Neural Networks

Summer 2019

Advisor: *Prof Davide Mottin*

- › Formulated a novel algorithm to **compute Wasserstein Distance** between distributions for generative modelling. Implemented constant memory backprop and variational dequantization to train very deep networks efficiently.

Better parsing with background knowledge

Fall 2020

Advisor: *Prof Ganesh Ramakrishnan*

- › Improved F-1 score by **1% on constituency parsing** for WSJ dataset by regularizing model with linguistic rules [C.1].

CognitiveCNN: Mimicking Human Cognitive Models to resolve Texture-Shape Bias

Fall 2019

Advisor: *Prof Biplab Banerjee*

- › Quantified the **shape-texture bias** of neural networks using techniques from cognition and image processing. Achieved a **5 % gain** in accuracy under miscue on Office-31 dataset using a novel attention matching based regularizer [C.3].

Key Courses Undertaken

Machine Learning Theoretical ML, Advanced ML, Natural Language Processing, Intelligent Learning Agents
Math And Stats Linear Algebra, Statistical Inference, Probability and Measure Theory, Regression Analysis

Miscellaneous

- › Reviewer for ICML'22, NeurIPS'22
- › Teaching assistant for undergraduate course on Artificial Intelligence and Machine Learning at IIT Bombay
- › Competed and won various national quizzes, and received recognition from IIT Bombay for these.
- › As hobby projects, built a bot to play word-games over messaging apps including Discord.