Suraj Srinivas

Contact Information

📍 6.220, 150 Western Ave
Harvard University, Boston, MA

✉ suuraj.srinivas@gmail.com
ssrinivas@seas.harvard.edu

🌐 suraj-srinivas.github.io

Research Interests

Robustness, Interpretability & Computational Efficiency of Deep models; Generative modelling (Vision-Language models, Diffusion models); Continual and Lifelong learning

Work Experience

01/2022
- Present
Postdoctoral Research Fellow,
Harvard University, USA,
Advisor: Prof. Hima Lakkaraju
Duties: Academic research · Technical guidance & mentoring · Teaching.

Education

2017
- 2021
Doctor of Philosophy,
École Polytechnique Fédérale de Lausanne &
Idiap Research Institute, Switzerland,
Advisor: Prof. François Fleuret.

2014
- 2017
Master of Science (Engineering),
Indian Institute of Science, Bangalore, India,
Advisor: Prof. R. Venkatesh Babu.

Internships

08/2020
- 01/2021
Research Intern, *Qualcomm AI Research, Netherlands*,
Research on algorithms to sparsify neural networks.

06/2016
- 08/2016
Research Intern, *DataGrokr, India / Verisk Analytics, USA*,
Speeding up inference on deep neural networks using tensor factorization.

01/2014
- 06/2014
Engineering Intern, *Tonbo Imaging, Bangalore*,
Implemented image processing algorithms on FPGA for a thermal imaging camera.

06/2013
- 08/2013
Research Intern, *Indian Institute of Science, Bangalore*,
Research on computational photography to perform camera jitter compensation.

Selected Publications

2022   **Suraj Srinivas***, Kyle Matoba*, Hima Lakkaraju, François Fleuret. (*co-first-author)
"Efficient Training of Low-Curvature Neural Networks"
Neural Information Processing Systems (NeurIPS)
Code: github.com/kylematoba/lcnn (Jointly authored)

2022   Tessa Han, **Suraj Srinivas**[†], Hima Lakkaraju. ([†]advising role)
"Which Explanation Should I Choose? A Function Approximation Perspective to Characterizing Post hoc Explanations"
Neural Information Processing Systems (NeurIPS)
ICML Interpretable ML for Healthcare Workshop - **Best Paper Award**

2022   Marwa El Halabi, **Suraj Srinivas**, Simon Lacoste-Julien.
"Data-Efficient Structured Pruning via Submodular Optimization"
Neural Information Processing Systems (NeurIPS)

2022   **Suraj Srinivas**, Andrey Kuzmin, Markus Nagel, Mart van Baalen,
Andrii Skliar, Tijmen Blankevoort.
"Cyclical Pruning for Sparse Neural Networks"
Computer Vision and Pattern Recognition Workshops (CVPRW) - **Oral**

2021   **Suraj Srinivas**, François Fleuret.
"Rethinking the Role of Gradient-based Attribution Methods in Model Interpretability"
International Conference on Learning Represetations (ICLR) - **Oral**
Code: github.com/idiap/rethinking-saliency

2019   **Suraj Srinivas**, François Fleuret.
"Full-Gradient Representation for Neural Network Visualization"
Neural Information Processing Systems (NeurIPS)
Code: github.com/idiap/fullgrad-saliency

2018   **Suraj Srinivas**, François Fleuret.
"Knowledge Transfer with Jacobian Matching"
International Conference on Machine Learning (ICML)
NeurIPS Learning with Limited Data (LLD) Workshop - **Best Paper Award**

2017   **Suraj Srinivas**, Akshayvarun Subramanya, R. Venkatesh Babu.
"Training Sparse Neural Networks"
Computer Vision and Pattern Recognition Workshops (CVPRW) - **Oral**

2016   **Suraj Srinivas**, R. Venkatesh Babu.
"Learning the Architecture of Deep Neural Networks"
British Computer Vision Conference (BMVC)

2015   **Suraj Srinivas**, R. Venkatesh Babu.
"Data-free Parameter Pruning for Deep Neural Networks"
British Computer Vision Conference (BMVC)

Book Chapters

2017 **Suraj Srinivas**, Ravi Kiran Sarvadevabhatla, Konda Reddy Mopuri, Nikita Prabhu, Srinivas SS Kruthiventi, R. Venkatesh Babu.
"A taxonomy of deep convolutional neural nets for computer vision",
Book chapter: *Deep Learning for Medical Image Analysis, Elsevier*
Journal version: *Frontiers in Robotics and AI*

Talks

03/2023 *Pitfalls and Opportunities with Feature Importance Methods*
MERL seminar series, Boston

07/2022 *Pitfalls and Opportunities with Feature Attribution Methods*
Simons Institute, UC Berkeley

06/2022 *Pitfalls and Opportunities with Feature Attribution Methods*
Vanderbilt University, USA

03/2022 *Cyclical Pruning for Neural Network Sparsity*
Google Sparsity Reading Group

08/2021 *Pitfalls of Saliency Map Interpretation in Deep Neural Networks*
HES-SO, Sierre, Switzerland

05/2021 *Pitfalls of Saliency Map Interpretation in Deep Neural Networks*
Harvard University, USA

04/2021 *Rethinking the Role of Gradient-based Attribution Methods for Model Interpretability*
ICLR (virtual)

01/2020 *Neural Network Interpretability using Full-Gradient Representation*
Indian Institute of Science, Bangalore

01/2020 *Full-Gradient Representation for Neural Network Visualization*
ML for Astrophysicists Club

11/2019 *Full-Gradient Representation for Neural Network Visualization*
Swiss Machine Learning Day, Lausanne

05/2019 *Complete Saliency Maps using Full-Jacobians*
Valais / Wallis AI workshop, Martigny

07/2018 *Knowledge Transfer with Jacobian Matching*
ICML, Stockholm

07/2016 *Making Deep Neural Networks Smaller and Faster*
Deep Learning Conf, Bangalore

Reviewing

Conferences AAAI, CVPR, ECCV, NeurIPS (2020) ; WACV, ICML, ICCV, NeurIPS (2021); ICLR, ICML, NeurIPS (2022); ICLR, AISTATS (2023)

Journals IEEE SP-Letters, Elsevier Neural Networks, IEEE T-PAMI, Nature Communications

Teaching

2023	Co-instructor for *Interpretability and Explainability in ML* (CS-282BR) Harvard University, USA
2018, '19, '21	Teaching Assistant for *Deep Learning* (EE-559) EPFL, Switzerland
2021	Guest Lecturer on Interpretability for *Deep Learning for Computer Vision* (DS-265) Indian Institute of Science, Bangalore

Awards and Honors

2022	Best paper award at ICML *Interpretable ML for Healthcare* Workshop
2022	Highlighted Reviewer at *International Conference on Learning Representations (ICLR)*
2021	EPFL PhD Thesis Distinction Award for top 8% thesis in the dept. of EE
2017	Best paper award at NeurIPS *Learning with Limited Data* Workshop
2014	Ranked **399** (out of $\sim$ 200k candidates) nation-wide in the Graduate Aptitude Test in Engineering for entrance to graduate school in electronics and communications engineering
2012	Won first place at the E-Yantra nation-wide robotics contest held at IIT-Bombay, and featured in The Times of India, New Indian Express and DH Education
2010	Ranked **191** (out of $\sim$ 100k candidates) state-wide in the Karnataka Common Entrance Test for entrance to undergraduate engineering programmes.