

Suraj Srinivas

suraj-srinivas.github.io | ssrinivas@seas.harvard.edu | suuraj.srinivas@gmail.com

Summary

I am a machine learning researcher with research interests in the **robustness**, **interpretability** and **computational efficiency** of deep neural networks.

Work Experience

Jan 2022 - **Postdoctoral Research Fellow**,
Current Harvard University, USA,
Advisor: Prof. Hima Lakkaraju
Research Focus: Foundations of Post-hoc Interpretability.

Education

2017 - 2021 **Doctor of Philosophy**,
École Polytechnique Fédérale de Lausanne &
Idiap Research Institute, Switzerland,
Advisor: Prof. François Fleuret
Thesis: Gradient-based Methods for Deep Model Interpretability.
(EPFL Thesis Distinction Award for Top 8% thesis in EDEE)

2014 - 2017 **Master of Science (Engineering)**,
Indian Institute of Science, Bangalore, India,
Advisor: Prof. R. Venkatesh Babu
Thesis: Learning Compact Architectures for Deep Neural Networks.

2010 - 2014 **Bachelor of Engineering**,
Electronics and Communication Engineering,
PES Institute of Technology (Now PES University), Bangalore, India.

Internships

Aug-Dec 2020 **Research Intern**, *Qualcomm AI Research, Netherlands*,
Research on algorithms for improving neural network sparsity.

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

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

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

Selected Publications

[Google Scholar Profile](#) | Citations: 1400+ | h-index: 9

- 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, “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)
(Mentoring Role)
- 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)
- 2021 **Suraj Srinivas**, François Fleuret. “Rethinking the Role of Gradient-based Attribution Methods in Model Interpretability”
International Conference on Learning Representations (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 (169 stars)
- 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)

Talks

- Jul 2022 Title: “**Pitfalls and Opportunities for Feature Attribution Methods**”
Venue: Simons Institute, UC Berkeley
- Jun 2022 Title: “**Pitfalls and Opportunities for Feature Attribution Methods**”
Venue: Vanderbilt University, USA
- Mar 2022 Title: “**Cyclical Pruning for Neural Network Sparsity**”
Venue: Google Sparsity Reading Group (Virtual)

- Aug 2021 Title: "**Pitfalls of Saliency Map Interpretation in Deep Neural Networks**"
Venue: HES-SO, Sierre, Switzerland
- May 2021 Title: "**Pitfalls of Saliency Map Interpretation in Deep Neural Networks**"
Venue: Harvard University, USA
- Apr 2021 Title: "**Rethinking the Role of Gradient-based Attribution Methods for Model Interpretability**"
Venue: ICLR (Virtual)
- Jan 2020 Title: "**Neural Network Interpretability using Full-Gradient Representation**"
Venue: Indian Institute of Science, Bangalore
- Jan 2020 Title: "**Full-Gradient Representation for Neural Network Visualization**"
Venue: [ML for Astrophysicists Club](#) (virtual)
- Nov 2019 Title: "**Full-Gradient Representation for Neural Network Visualization**"
Event: Swiss Machine Learning Day, Lausanne
- May 2019 Title: "**Complete Saliency Maps using Full-Jacobians**"
Event: Valais / Wallis AI workshop, Martigny
- Jul 2018 Title: "**Knowledge Transfer with Jacobian Matching**"
Event: ICML, Stockholm
- Jul 2016 Title: "**Making Deep Neural Networks Smaller and Faster**"
Event: 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

- Spring 2023 Teaching Fellow for "Interpretability and Explainability in ML" at Harvard University
- 2018/'19/'21 Teaching Assistant for Deep Learning (EE-559) at EFPL, Lausanne
- Apr 2021 Guest Lecture on Interpretability for Deep Learning for Computer Vision Course (DS-265) at IISc, Bangalore

Awards and Honors

- 2022 Best paper award at ICML Interpretable ML for Healthcare (IMLH) Workshop
- 2022 Highlighted Reviewer at *International Conference on Learning Representations (ICLR)*
- 2021 EPFL PhD Thesis Distinction Award for top 8% thesis in EDEE
- 2017 Best paper award at NeurIPS LLD Workshop
- 2014 Ranked **399** (out of ~ 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 was featured in The Times of India, New Indian Express and DH Education
- 2010 Ranked **191** (out of ~ 100k candidates) state-wide in the Karnataka Common Entrance Test for entrance to undergraduate engineering programmes.