Suraj Srinivas

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

Summary

I am a machine learning researcher, with active research interests in **interpretability** and **robustness** of machine learning models, and **compression** of deep neural networks.

Work Experience

Jan 2022 - Current **Postdoctoral Research Fellow**,
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.

Internships

Aug 2020 - Jan 2021 **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

2021 **Suraj Srinivas** and François Fleuret. "Rethinking the Role of Gradient-based Attribution Methods in Model Interpretability", International Conference on Learning Representations (ICLR) **[Oral]** (PyTorch Implementation)

2019 **Suraj Srinivas** and François Fleuret. "Full-Gradient Representation for Neural Network Visualization.", Neural Information Processing Systems (NeurIPS)
(PyTorch Implementation)

2018 **Suraj Srinivas** and François Fleuret. "Knowledge Transfer with Jacobian Matching.", International Conference on Machine Learning (ICML)

2017 **Suraj Srinivas** and François Fleuret. "Local Affine Approximations for Improving Knowledge Transfer.", NeurIPS Workshop on Learning with Limited Data **[Best Paper Award]**

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

2016 **Suraj Srinivas** and R. Venkatesh Babu. "Learning Neural Network Architectures using Backpropagation." British Machine Vision Conference (BMVC)

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

Talks

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 (Virtual)

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 (2022)

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

Teaching

2018-2021 Teaching Assistant for Deep Learning Course (EE-559) at EFPL, Lausanne

Apr 2021 Guest Lecture on Interpretability for Deep Learning for Computer Vision Course (DS-265) at IISc, Bangalore

Miscellaneous

2022 Highlighted Reviewer at ICLR 2022

2021 EPFL PhD Thesis Distinction Award for top 8% thesis in EDEE

2014 Ranked **399** (out of $\sim$ 200k candidates) nation-wide in the Graduate Aptitude Test in Engineering for entrance to IITs / IISc for graduate studies in electronics and communications engineering

2012 Won first place in 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 $\sim$ 100k candidates) state-wide in the Karnataka Common Entrance Test for entrance to state engineering colleges for undergraduate studies.