Ramanan Sekar, Senior Machine Learning Engineer at Qualcomm

Webpage: ramanans1.github.io, GitHub: github.com/ramanans1 Interests: Deep Learning, Robotics, Artificial Intelligence

Education

University of Pennsylvania

Master of Science in Engineering in Robotics; GPA: 4.00 Aug 2018-May 2018 Relevant Courses: Computer Vision and Computational Photography, Introduction to Robotics, Machine Perception, Machine Learning, Learning in Robotics, Integrated Intelligence with Robotics, Operating Systems, Master's Thesis

Anna University

Bachelor of Electrical Engineering; GPA: 3.9 (8.89/10.0 - top 1% of Graduating Class) Aug 2014 - Ma Relevant Courses: Advanced Control Systems, System Identification and Adaptive Control, Object Oriented Programming, Applied Soft Computing, Computer Programming

SKILLS SUMMARY

- Languages: Python, C++, C, MATLAB
- Machine Learning Tools: PyTorch, TensorFlow, Keras, Scikit-Learn, ONNX

• Tools: Linux, Git, Docker, OpenCV, NumPy, ROS, Matplotlib, Google Cloud Platform, MuJoCo, LATEX

EXPERIENCE

Qualcomm

- Senior Machine Learning Engineer, Qualcomm R&D
 - $\circ\,$ Harware-Software Co-design for Autonomous Driving:
 - Working in the AI Systems team in the Autonomous Driving R&D unit at Qualcomm Research on developing deep learning models for the driving stack and efficiently deploying them on Qualcomm's SnapDragon Ride platform, collaborating with cross-functional teams internally (compiler, hardware, AI research, customer engineering, AD research) and externally (Automotive customers, Google)
 - Leading effort on Transformer based model optimization focused towards perception and motion prediction models. Innovated on architecture and efficient attention design to result in BEV perception models that are 60 times faster than baseline models, and motion prediction models that are 5 times faster than baseline. Filed Invention Disclosures and Patent applications for this work
 - Leading effort on Large-Language Model, Vision-Language Model optimization, focusing on end-to-end autonomous driving with VLAs (SigLip, LLaMa, OpenVLA,.). Filed Invention disclosures towards optimizing these set of models
 - Leading the Neural Architecture Search R&D work in the BU. Work involves developing i: effective search spaces that trade-off efficiency vs complexity, ii: efficient search algorithms such as Reinforcement Learning, evolutionary search, etc., iii: fast evaluation procedures to develop deep models that have high task performances as well as efficient deployment on target hardware. This work has been used to optimize different components of the perception stack, including CNN based models, Transformer based models, Multi-task networks. Results from this work has been included in product offerings at Qualcomm's ADAS Software Stack
 - Worked on effective quantization and compression of internal models as well as external customer models for improving real-time performance. This involves using internally developed tools like AIMET to understand compressibility, or quantizability of models using post-training quantization or quantization-aware training methods. This work was directly utilized and commercialized in Auto OEM products
 - Developed a cutting-edge Reinforcement Learning based hardware-aware Automatic Mixed Precision assignment algorithm to optimize models for inference.
 - Worked on developing tools to understand performance bottlenecks such as analysis of quantization errors of models through a layer-wise breakdown, analysis of profiling tools by enabling tracing bottlenecks all the way from the hardware execution to the source code, analysis of accuracy degradations when executing models on target devices using different error metrics
 - Key contributor in the development of a model zoo inside the Automotive BU, collaborating with internal and external stakeholders in curating a model list and taking them through model onboarding process to have optimized and fully functional set of models to benchmark results on.
 - Work has so far led to several invention disclosures and innovations internal to Qualcomm, with one invention currently being filed for patent by Qualcomm.
 - Mentored interns in the team on research projects that lead to several useful and impactful results. Mentored students externally in Universities through the Qualcomm Innovation Fellowship. Mentored full-time new-grad team members
 - Part of the reviewing committee for the Qualcomm Innovation Fellowship applications, providing expert opinions on proposals.

Email : ramanansekar.22@gmail.com

LinkedIn: in/ramanansekar

Philadelphia, PA Aug 2018-May 2020

Chennai, India Aug 2014 - May 2018

Santa Clara, CA June 2020 - Present

Qualcomm

Deep Learning Intern, Qualcomm R&D

• Denoising with Recurrent Neural Networks:

- Designed large, custom recurrent neural networks for optimally denoising signals at the Shannon decoding limit, by decoding signals on graphs
- Implemented and refactored vectorized Tensorflow code and a modular scalable codebase for high frequency design iterations. Produced a reproducible codebase by attacking non-determinism problems in GPU processing
- Scaled performance to real-time, full scale models and achieved significant performance boosts with the developed method

GRASP Lab, University of Pennsylvania

Research Assistant - Prof. Kostas Daniilidis' Group

- Planning to Explore with Self-Supervised World Models (ICML 2020):
 - Master's thesis focusing on learning world-models in a self-supervised manner directly from pixels for robot control through model-based reinforcement learning, with underlying world-model built on top of Dreamer.
 - Evaluated on Deepmind Control Suite, self-supervised reinforcement learning agents are taught to quickly solve new tasks at test time without access to them at train time, learning directly from pixels through effective state-space coverage from extensive exploration
 - Effective world models are learned through an exploration policy optimizing for intrinsic rewards based on Ensemble Disagreement of forward dynamics predictors
 - Learnt world model demonstrates impressive zero-shot and few-shot capabilities, through training expert task-specific policies on experience

• Perception Driven Curiosity with Bayesian Surprise (RSS 2019):

- Contributed to developing an intrinsic rewards based on-policy RL agent based on Bayesian formulation of surprise
- Developed a C-VAE based perception module to control RL agents from vision, that is also trained to predict forward dynamics
- A PPO based agent was trained with a combination of extrinsic and intrinsic rewards on Atari suite, with significant improvements in sparse-reward environments. Integrated simulated Baxter tasks for testing future iterations.

• Unsupervised few-shot classification with Meta-Learning:

• Developed a novel meta-learning based few-shot unsupervised classification module with MAML, using a novel CNN + MLP architecture, in Tensorflow. Used Bayesian Optimization to tune hyperparameters. Achieved similar performance with 85% less labeled data

University of Pennsylvania

Teaching Assistant - Machine Learning

Spring 2019, Fall 2019, Spring 2020 • TA for CIS 520: Machine Learning, CIS 519: Applied Machine Learning: Teaching Assistant for graduate level machine learning courses focusing on both theory and applications. In charge of designing homeworks, conducting office hours, grading and project feedbacks.

Indian Institute of Technology

Robotics Research Intern

- Kalman Filters for efficient Localization: Developed a novel KF algorithm using correlations between measurements, proved superiority to original EKF, and validated on a Mobile Robot using SONARs. Achieved Faster and accurate convergence
- Sensor Fusion Algorithm: Developed a new efficient sensor fusion algorithm using Maximum likelihood estimators and measurement noise correlations. Achieved an error reduction of 70% compared to existing methods
- SLAM: Implemented 2D Feature based EKF-SLAM for AmigoBot, using SONAR ring and wheel encoders, and Particle Filter based 2D localization for mobile robots. Worked with ARIA C++ libraries to deploy on mobile robots like P3DX and Amigobot

Raman Research Institute

- Research Fellow
 - Decoy-state Quantum Key Distribution: Studied about the loss of security in practice of the original BB84 QKD protocol due to Photon Number Splitting Attacks in multi-photon signal states. Studied about Decoy State QKD as a solution to improve security, increase distance and rate of key distribution. Described mathematical background and state-of-the-art advances in Decoy State QKD in a comprehensive final report

PUBLICATIONS

- [1]: R. Sekar, O. Rybkin, K. Daniilidis, P. Abbeel, D. Hafner, and D. Pathak. (2020) Planning to explore via self-supervised world models, International Conference on Machine Learning (ICML), 2020
- [2]: Beradette Bucher, Anton Arapin, Ramanan Sekar, Feifei Duan, Marc Badger, Kostas Daniilidis, Oleh Rybkin (2019) Perception-driven Curiosity with Bayesian Surprise, Workshop on Combining Learning and Reasoning Towards Human-Level Robot Intelligence, RSS 2019
- [3]: Sekar R, Sai Shankar N., Shiva Shankar B., Manivannan P.V. (2018) Modified Extended Kalman Filter Using Correlations Between Measurement Parameters, Advances in Intelligent Systems and Computing, Springer
- [4]: Sekar R, Sai Shankar N., Shiva Shankar B., Manivannan P.V. (2017) Use of measurement noise correlations for an improved SONAR model, IEEE

Philadelphia, PA

Nov. 2018 - Nov. 2020

Madras, India

Dec 2016 - April 2018

Philadelphia, PA

Bengaluru, India

May 2017 - July 2017

PATENTS

- [1]: Under Application: Sundar S, Ravi Konduru, *Ramanan Sekar*, Soyeb N. (2024) Improving Transformer Attention on Matrix Multiplication Accelerators
- [2]: Under Application: Ramanan Sekar, Soyeb N. (2025) Learnt Kernel Attention for BEV Perception
- [3]: Under Application: Ramanan Sekar, Soyeb N. (2025) Efficient Vision Transformers via Controlled Invariance

AWARDS

- [1]: Qualcomm CEO Award: (2023) Received a recognition and award directly from the CEO of the company for my contributions
- [2]: IAS Summer Research Fellow Award: (2017) Research Fellowship award from the Indian Academy of Sciences for research on Quantum Communications
- [3]: Merit Scholarship Awards: (2016-2018) Undergraduate scholarship awards awarded by SSN College of Engineering, Anna University, for outstanding academic performance

RESEARCH GRANTS

- [1]: Honda Research Institute: Curious Minded Machines (2019) Part of the research group that got awarded the Honda Research Institute Curious Minded Machines grant for research on Reinforcement Learning, Exploration and Curiosity
- [2]: Research Scholarship, SSN College of Engineering (2017) Awarded a research scholarship to conduct research on robotics and multi-robot collaboration for search and rescue missions

REVIEW WORK

- [1]: Qualcomm Innovation Fellowship (2023, 2024) Part of the committee at Qualcomm Innovation Fellowship that reviews fellowship proposals and provides expert opinion and rating on whether to award funding to the proposal
- [2]: Teaching Assistant, Machine Learning, University of Pennsylvania (2019-2020) Teaching Assistant for Graduate level Machine Learning courses CIS 520 and CIS519 at the University of Pennsylvania. Duties include creating and grading homeworks, creating and grading exams, mentoring and evaluating student projects, and conducting office hours

TALKS

- [1]: Machine Learning Summit, Qualcomm (2022, 2023) Presentation of cutting edge Research and Development work carried out in the AI Systems team in Autonomous Driving organization in the Auto BU at the annual Machine Learning Summit organized at Qualcomm.
- [2]: ICML Oral Presentation (2020) Oral presentation at ICML 2020 on the Planning to Explore via Self-Supervised World Models paper
- [3]: Intel Research (2020) Oral presentation at an invited talk from Intel on the Planning to Explore via Self-Supervised World Models paper
- [4]: International Conference on Computational Intelligence (2017) Oral presentation at an ICCI 2017 conference to present the work on Modified Extended Kalman Filter
- [5]: IEEE International Conference on Technological Advancements in Power and Energy (2017) Oral presentation at an TAPEnergy 2017 conference to present the work on an effective SONAR model that uses measurement noise correlations

Other Projects

- **Operating System**: Contributed to the building of a functioning, simple operating system with a working process scheduler and file system, in a team of 4. Specifically worked on the creation of the process scheduler
- Real-time Object Detection: Contributed to the deployment of a real-time object detection model on an embedded NVIDIA Jetson device, running on a Turtlebot. Worked on optimizing the trained model for real-time perf. on-device.
- Qualcomm's Hackathon: Unsupervised Depth Estimation: Integrated a real-time unsupervised monocular depth estimation system, coupled with an off-the-shelf object detection module, for vision based navigation guidance for visually impaired people, in Tensorflow
- Multi-Robot Vision based Collaboration: Designed an Aerial and Ground mobile robot collaboration with SIFT based aerial imagery stitching and object detection for ground robot motion planning with Probabilistic Roadmaps
- **3D Reconstructions with Structure from Motion**: Developed an SfM pipeline from two views with the 8 point RANSAC algorithm
- **Deep Fakes**: Developed a pipeline in Python for swapping faces between videos. Used DLib Library to detect faces and generate features. Used a KLT Tracker to track features in both source and target video, and used Delaunay Triangulation and Affine Transforms to swap faces with Seamless Blending
- Hindsight Experience Replay for Manipulators:: Controlled a simulated 2DOF manipulator with Deep Reinforcement Learning by learning its Inverse Kinematics with sparse rewards using Hindsight Experience Replay (HER) on top of DDPG in Tensorflow
- Manipulator Motion Planning:: Implemented Motion Planning on the 5DOF manipulator Lynx physical robot with PRMs and novel Potential Field Planners, after implementing Forward, Inverse and Velocity kinematics for the Lynx.
- Sparse Optical Flow: Developed from scratch the KLT feature tracking for Optical Flow of sparse features in slow to medium paced videos, in Python.