Shaan Luthra

luthrashaan@gmail.com · 2068989413 · github.com/ShaanLuthra1

Experience

PrecisionHawk Raleigh, NC
Data Scientist Jun 2022 - Present

- Refactored core Object Detection Deep Learning Model (YOLO architecture, PyTorch) and reintegrated all dependencies after +1 year leading to an mAP increase of 5%. Integrated Bayesian hyperparameter optimization to updated model.
- Created Aerial-based oriented bounding box modeling for Object Detection Model, leading to an IOU increase of 8%.
- Implemented Distributed Data caching using custom Samplers to allow for parallel GPU computing on AWS EC2 instances. Projected to save over $15,000 annually in computational expenses.
- Led data annotation effort for multiple projects and expedited workflow with active learning based tools.

Hewlett Packard Enterprise Seattle, WA
Data Science Intern May 2021 - Aug 2021

- Developed a full model pipeline that took unstructured log messages and leveraged Hierarchical Clustering based template extraction to transform 4 million unique log messages into 84 templates.
- Increased efficiency of support engineers identifying harmful log messages by over 30% in weekly ticket completions.
- Implemented an Unsupervised anomaly detection algorithm to identify harmful log messages based on auto-embeddings of extracted templates.

Automatic Child Speech Recognition Research, Purdue University West Lafayette, IN
Machine Learning Researcher Aug 2020 - May 2022

- Key leader and contributor in all phases of project life-cycle including data preprocessing, augmentation, feature selection, modeling, validation, and inference.
- Identified and implemented methods like noise reduction, filter transformations, spectrograms, Mel-Frequency Cepstral Coefficient features, and more in order to build out signal processing infrastructure.
- Led Deep Learning team of 4 researchers and built a Deep Convolutional-LSTM Model with Tensorflow achieving a phonetic accuracy of 87%.

Safkan Health Seattle, WA
Machine Learning Engineer Intern Oct 2020 - Feb 2021

- Developed Deep-CNN model to detect the presence of tympanic membrane infections with 90% accuracy using Tensorflow for the Safkan OtoSet.
- Responsible for the full model pipeline including data acquisition, preprocessing, augmentation, inference, and validation.

Education

Purdue University West Lafayette, IN
B.S. Data Science Aug 2018 - May 2022
Department of Computer Science

Skills

Languages:	Python, C++, Java, SQL, R, C#, Matlab
Libraries:	PyTorch, Tensorflow, Scikit-Learn, Numpy, Scipy, Pandas
Tools/Services:	AWS, GCP, WandB, Spark, Nvidia Cuda/Nccl

Awards

Presidential Volunteer Service Award

2015, 2016

Put on year-round community service events, performed charitable work, planned and executed several summer camps totaling 220+ volunteer hours.