## Experiences

#### **Sharad Chitlangia**

Email | Website

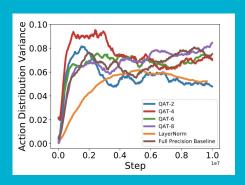
# Research Intern @ Harvard Edge Computing Lab

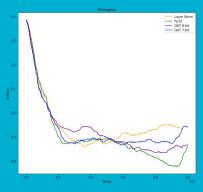
 Deep Reinforcement Learning for Embedded Devices like UAVs, etc.

Example use case produced during the internship

2. Studied effects of techniques like Quantization on RL during the internship algorithms e.g. DQN, PPO2, DDPG, A2C.

- 3. Showed Quantization acts as a regularizer. Adding quantization loss to the loss metric enables <u>higher exploration</u> by creating <u>higher entropy policies</u>.
- 4. Showed policies can be quantized upto 6-8 bits of precision without loss of accuracy.
- 5. Showed with quantization, one can achieve as much as 18x speedup in the loop rate and 4x memory reduction
- 6. Ongoing remote work. For more results check, Research Pre-print: <a href="https://arxiv.org/abs/1910.01055">https://arxiv.org/abs/1910.01055</a>.



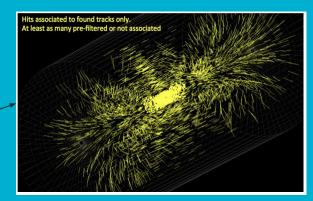


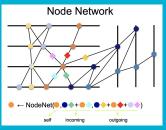
Higher action distribution variance and higher entropy policies. QAT denotes Quantization aware trained policies.

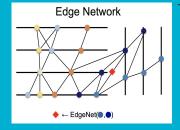
## **Google Summer of Code** with CERN-HSF

Problem statement

- Particle Track Reconstruction using Machine Learning
- 2. Ported, studied and developed top reconstruction algorithms to ACTS-Framework.
- Models used included Graph Neural Networks, NNs operating on top of DAGs, Logistic Regression.
- Explored approaches to better techniques using optimization of Gradient Boosting Classifiers, Logistic Regression, etc
- 5. Developed end-to-end pipeline of deploying and testing pytorch models in ACTS.







#### Information Flow

- · Graph is sparsely connected from layer to layer
- InputNet + EdgeNet + NodeNet only correlates hits information on triplet of layers
  - The information from the outer hits and inner hits are not combined
- Several possible ways to operate the connection
- Correlates hits information through multiple iterations of (EdgeNet+NodeNet)



GNN approach

### **Others**





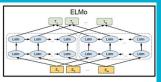
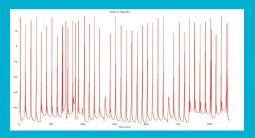


Figure 1: Differences in pre-training model architectures. BERT uses a bidirectional Transformer. OpenAI GPT uses a left-to-right Transformer. ELMo uses the concatenation of independently trained left-to-right and right-to-left LSTM to generate features for downstream tasks. Among three, only BERT representations are jointly conditioned on both left and right context in all layers.

ML Intern @ Unfound.ai (summer 2018) Work with contextual embeddings, Information Retrieval, API microservices, Stance Detection in NLP



Pneumonia Detection with TCS Research: Implemented SOTA approaches and fine tuned for RSNA dataset (Fall 2018) Remote unofficial setting with Prof. Ashwin Sriniyasan



Formulating simple RL approaches for Neuromodulated STDP (Dopamine based learning) on Visual thalamus plus basal ganglion response learning using Spiking Neural Networks (Ongoing) As DOP with Prof. Basabdatta Sen Bhattacharya

#### Miscellaneous

- Leading group of exceptional undergraduate researchers - SAIDL
- Intel AI Student Ambassador Publishing blogs on latest research in AI.
- 3. Teaching <u>course</u> on Deep Learning. Course structure similar to CS231n
- Teaching course on RL, this semester in part to implement the Rainbow RL algorithm under TIP.
- 5. Project Implementation of various SOTA papers on Github.
- Checkout: https://sharad24.github.io/