Harshavardhan Adepu

(+1)608-686-2694 | harshavardhan163@gmail.com | linkedin.com/in/adepu-harshavardhan/

EDUCATION

University of Wisconsin-Madison, Madison, WI
    Ph.D. in Electrical and Computer Engineering   Jan'24 - May'26
    MS. in Electrical and Computer Engineering, GPA: 3.85/4   Sep'21 - May'26
Indian Institute of Technology Madras, Chennai, India
    B.Tech.(Honours) in Electrical Engineering, GPA: 8.75/10   Jul'14 - May'18

RESEARCH INTERESTS

I am interested in efficient Deep learning, particularly making Large Language models and Vision models more efficient through quantization, pruning, and PEFT. I am also interested and have experience in modern Deep Learning paradigms such as Multimodal models, LLMs, and, INRs.

PUBLICATIONS

FrameQuant: Flexible Low-Bit Quantization for Vision Transformers   ICML 24
Harshavardhan Adepu, Zhanpeng Zeng, Li Zhang, Vikas Singh
A PTQ method for quantizing LLMs and Vision Transformers to effectively two bits with only a small drop in performance. Our method is based on classical frame analysis and provides flexibility to quantize a model to fractional bit widths on average, uniformly across all layers.

Operator-theoretic Implicit Neural Representation   ICML 24
Sourav Pal, Harshavardhan Adepu, Clinton Wang, Polina Golland, Vikas Singh
An operator theoretical reformulation of the INR model, able to use operators such as CNNs as INRs(which are mainly dominated by MLPs). This formulation allows interpolation between INRs trained on different images while also being able to represent 3D volumes, videos, and 3D MRI scans.

Offset Correction in High-Speed Serial Link Receivers	US Patent 17/363,855
Lane Adaptation in High-Speed Serial Links	US Patent 17/205,456
Digital Upconverter for Radio Frequency Sampling Transmitter	US Patent 17/493,943

PROFESSIONAL EXPERIENCE

Google DeepMind   Seattle, WA
Student Researcher, Perception: Humans & Interactions team   Apr'24 - Oct'24
- Training ultra-low precision Mixture-of-Experts model for approximating general floating point models.
- Designing efficient quantization schemes for multi-modal generative models with theoretical guarantees.

Texas Instruments India   Bangalore, India
Systems Engineer, High Speed Signal Conditioning team   Mar'19 - Aug'21
- Designed an Adaptive Equalizer in RX of Retimer SoCs for mitigating the ISI in High-Speed Lanes supporting a data rate of 56GBPS.
- Modeled the entire TX-RX link, including a High-Speed Lane, cross talk, and device impairments, and tuned the adaptation algorithm for fast convergence and a steady-state BER of <1e-15.
- Developed Firmware for a swift device bring-up and channel tracking in steady-state operating conditions.

Texas Instruments India   Bangalore, India
Systems Engineer, Wireless Infrastructure Team   July'18 - Feb'19
- Developed fixed point models for various modules in the DDC and DUC chains of 5G base station Transceiver SoCs and analyzed their SNR and SFDR performance.
- Identified key contributors of spurs in the DDC chain and devised solutions to maintain SFDR below -100dBc.

RESEARCH PROJECTS

Data Augmentation for Image Classification using Latent Diffusion Models
Advisor: Prof. Sharon Li, UW-Madison   *Sep'22- Dec'22*
- Generated synthetic images using Stable Diffusion for CIFAR-10 dataset and trained a ResNet18 model, resulting in an accuracy of $\sim 18\%$ on the CIFAR-10 test set.
- Fine-tuned the Stable Diffusion with Textual Inversion with a few images from CIFAR-10 classes and retrained the ResNet18 model to improve its accuracy to $\sim 80\%$.
- Fine-tuned the U-Net of Stable Diffusion using the technique in Dream-Booth to improve the accuracy to $\sim 82\%$.
- Observed that the FID scores of the generated images using different methods correlated with the accuracies of the downstream classifier

Performance of Multi-Image Comparison Queries
Advisor: Prof. Ramya Vinayak, UW-Madison   *Jan'22 - Jun'22*
- Proved that the Cross Cluster Edge Density for queries with 3 to 8 images per query is better than the classical pair comparison queries under the Conditional Block Model.
- Implemented a Python program to derive the expressions for Cross cluster edge density given the number of images per query and generate the mathematical proof.
- Observed that the performance with human annotators peaks at 4-5 images per query, which could be limited by the working memory (Miller 1956).

Efficiency Model for Buck Converter
Advisor: Prof. Qadeer Khan, IIT Madras   *Dec'17 - May'18*
- Developed an Empirical model for the total power loss in a Buck converter with maximum error $<3\%$.
- Derived an analytical expression for the Optimum sizes of switches and the frequency of operation of the Buck Converter using this model and validated the results with actual data.

COURSE PROJECTS

Robust Image Classification
Course: Machine Learning, with Prof. Fred Sala, UW-Madison   *Sep'21-Dec'21*
- Studied the effects of noisy images and adversarial attacks on Image classification and defenses against them.
- Processed the images with a combination of DCT and DWT followed by filtering to tackle different types of noises applied on the images and improved the classification accuracy by 15% on the CIFAR-10 dataset.
- Implemented the Bayesian uncertainty-based two-layer defense by Alshemali,2019 to defend a classifier against Adversarial attacks such as FGSM, BIM, and CW L2.

Systematic Encoding and Decoding of BCH Codes
Course: Error Control Coding, with Dr. Pradeep Sarvepalli, IIT-Madras   *Feb'17-May'17*
- Developed a Systematic encoder for binary BCH code with a code rate of 0.614 and correction capacity of 7.
- Designed a decoder for this code with error and erasure correcting capabilities within the correction capacity.
- Implemented the simplified Berlekamp-Massey algorithm for binary codes to determine the error location polynomial.

TECHNICAL SKILLS

Deep Learning, Transformers, Computer Vision, LLMs, Diffusion Models, INR, Quantization, PEFT, Machine Learning, Python (Pytorch, Numpy), Shell scripting(Bash, Fish), Computer Science, Matlab, Wavelets, Communication, Signal Processing, Digital System Design, Firmware, C, C++.

ACHIEVEMENTS

- Recipient of GCP Credit Award in support of my research on Efficient Transformer Models
- Graduated with "Honours" in Electrical Engineering from IIT Madras, 2018 (Top 7%)
- Secured an All India Rank of 2293(Top 1.5%) in engineering Joint Entrance Exam Advanced 2014.

RELEVANT COURSEWORK

- Advanced Deep Learning
- Matrix Methods in Machine Learning
- Theoretical Foundations of Machine Learning
- Machine Learning
- Learning with Less Supervision
- Probability and Random Processes
- Mathematical Foundations of Machine Learning
- Introduction to Data Structures and Algorithms