Sumanth R Hegde

s1hegde@ucsd.edu (858) 257-8544 San Diego, CA SumanthRH LinkedIn

EDUCATION

- **M.S. in Computer Science, University of California, San Diego** *Exp. Mar 2024*
  Specialization: Artificial Intelligence GPA: 4.0
  Teaching Assistant: Principles of Database Systems, Mathematics for Algorithms & Systems Analysis
  Graduate Student Researcher: Working in Prof. Julian McAuley's lab on evaluating the impact of negative/ positive dialog context on chatbots/LLMs
- **B.Tech(Hons.), Electrical Engineering, Indian Institute of Technology(IIT) Madras** *Jul 2021*
  University Rank: 1/778 (Gold medalist) CGPA: 9.76/10
  Relevant Coursework: Statistical Natural Language Processing, Machine Learning, Deep Generative Models, Deep Learning, Graduate Networked Systems, Scalable Data Systems, Operating Systems, Database Systems

INTERNSHIPS

C3.AI | PyTorch, Accelerate, DeepSpeed, Python *Jun 2023 - Present*
Data Science Intern.

- Building a finetuning codebase from scratch for C3.AI's generative AI applications leveraging HuggingFace's Accelerate.
- Implemented support for different decoder-only and encoder-decoder models, multiple training and evaluation datasets, control over generation metrics, and better monitoring.
- Added parameter-efficient fine-tuning, quantization (int4, int8 support) aware training and flash attention support.
- Trained LLama-2-13B and Flan-UL2 models on Orca-style datasets (3M+ samples) leveraging DeepSpeed ZeRO to improve performance on instruction-following tasks.

HyperVerge Inc. | Tensorflow, Python, C++ *May 2019 - Jul 2019*
Computer Vision Intern

- Implemented a learning-based face detection algorithm for Know-Your-Customer services, reduced false positives 10 times and false negatives by 2.5 times on HyperVerge's benchmark.
- Trained a Multi-task Cascaded Convolutional Neural Network using $>$ 200,000 images, employing data augmentation and hard positive mining to beat the previous model with $>$99.5% accuracy.

PROJECTS

Performance-Efficient Fine Tuning (PEFT) for Language Models *Apr 2023 - Present*

- Implemented and benchmarked $(IA)^3$, a new parameter-efficient finetuning method, and integrated our code with HuggingFace's PEFT library. [Pull Request] [Report]
- Implemented $(IA)^3$ to support different encoder-decoder and decoder-only models, multiple adapters, 8-bit quantization.
- Benchmarked against other PEFT methods on tasks like causal language modelling and sequence classification on GLUE datasets.

SurfStore: A DropBox-like Service | gRPC, Go *Jan 2023 - Mar 2023*

- Built a scalable, distributed DropBox-like cloud storage service for file syncing, with fault-tolerance support.
- Designed a MetaStore service for storing file metadata and a BlockStore service for storing file blocks using gRPC.
- Implemented the RAFT consensus protocol and employed consistent hashing for horizontal scaling.

Jester: A Text-to-Meme Generation Engine | HuggingFace, Streamlit *Oct 2022 - Dec 2022*

- Built a novel two-stage system to generate relevant meme templates and meme captions given any user text.[Demo]
- Implemented a flexible softmax-free transformer model to serve candidate meme templates for given user text, achieving a top-5 accuracy of 71% on a dataset of 300,000 captions. [Poster][Report][Code]
- Utilized GPT-3 and designed custom prompts for 100 templates to generate relevant meme captions from user text.

Megapixel Image Restoration for Under-Display Cameras | Pytorch, OpenCV *May 2020 - Jul 2020*
Student Researcher under Prof. Kaushik Mitra, IIT Madras

- Created a novel deep learning based model for image restoration, resulting in a publication at ECCV Workshops 2020 and placed 2^{nd} /150 teams at the Under Display Camera Challenge. [Presentation]
- Rectified severe blur and low light conditions in the images using a two-stage pipeline, obtaining $>$12% improvement in image quality with 88% (7.8M) lesser parameters than existing work. [Paper] [Website][Code]

TECHNICAL SKILLS

Languages : Python, C++, Go, SQL, C, R
Frameworks/Tools : Pytorch, Pandas, Tensorflow, Transformers, Accelerate, DeepSpeed, Keras, Docker, Git

PUBLICATION

Sundar, Varun†, Sumanth Hegde†, Divya Kothandaraman, and Kaushik Mitra. "Deep Atrous Guided Filter for Image Restoration in Under Display Cameras". In: *Computer Vision - ECCV 2020 Workshops*. Springer International Publishing, 2020, 379-397.

† Equal Contribution