

Karan Shah

🏠 Stony Brook, NY 11790 | 📞 (+1) 631-438-6950 | ✉ karan.shah@stonybrook.edu | 🌐 the-karanshah.github.io | 📷 the-karanshah | 📺 the-karanshah

Education

Stony Brook University

MS IN COMPUTER SCIENCE | GPA : 3.68/4

Coursework: Data Science, Probability and Statistics, Big Data Analytics, Theory of Database Systems, Machine Learning, Computer Vision

Stony Brook, NY

Aug. 2019 - Exp. Dec. 2020

Gujarat Technological University

B.E. IN COMPUTER ENGINEERING | GPA : 8.03/10

Coursework : Artificial Intelligence, Compiler Design, Data Structures, Algorithms, Software Engineering, Distributed Operating System

Ahmedabad, India

Aug. 2014 - May. 2018

Work Experience

Software Developer Intern

QUANTIPHI

- Building software tools in Python and React.js with Machine Learning Engineers to accelerate production-grade Conversational AI use cases in Docker Container over GCP cluster
- Using Flask based routing methods with multiprocessing to trigger underlying capabilities of **NVIDIA's** end-to-end ML framework and Socket+GRPC for live audio transcription

Sept. 2020 - Current

Marlborough, MA

Software Engineer

SIMFORM

- Developed an end to end Video Streaming system; allows user to upload and scan the content on the cloud, stream on cross-platform desktop application over globally separated clients; using AWS services (Lambda, S3, Cloudfront, MQTT), Electron.js, and React.js
- Built a software system to semi-automate Company's marketing work with Chrome Extension and REST API Web backend using Python and Node.js, efficient in reducing ~ **40 %** of Marketing team's manual work
- Designed and automated ETL jobs to process a huge realtime data using AWS services (Glue, Lambda, S3) and Spark

June 2018 - April 2019

Ahmedabad, India

Research Intern, Machine Learning

ISRO - INDIAN SPACE RESEARCH ORGANIZATION

- Designed generalized **CNN** architecture to handle multimodal and/or multispectral imagery along with adaptability to varying application requirements; trained on Potsdam benchmark dataset producing State of the art results to **Semantic Segmentation**
- Built software tools in **Python** for preprocessing of geographic remote sensing data, visualization, real-time performance evaluation, post-processing and output generation as a semi-automation approach
- Manipulated raster GIS data as a part of the pre-processing task using Python and QGIS application

July 2017 - May 2018

Ahmedabad, India

Skills

Programming Python, Matlab, Javascript, JAVA, C/C++

Cloud/Big Data AWS, Google Cloud Platform, MapReduce, Spark

Tools/Technologies Docker, Kubernetes, Git, REST, GRPC, SQL, NoSQL

Frameworks/Libraries PyTorch, Tensorflow, Keras, OpenCV, Node.js, Django, Flask, React.js

Publication

Band-wise Independent Pansharpening Using Neural Networks with Shared Weights

PANKAJ BODANI, KARAN SHAH, SHASHIKANT SHARMA (ISRO)

SUBMITTED TO **GEOCARTO INTERNATIONAL JOURNAL**

Projects

Child Mortality Analysis, Under Supervision of Prof. Andrew Schwartz (Python, Dask, Spark)

April 2020 - May 2020

- Analysed secondary factors contributing to child mortality through Big Data Techniques
- Implemented **Linear Regression and Similarity Search** running on women lifestyle real data distributed on Hadoop clusters

Covid Data Analysis, Instructed by Prof. Anshul Gandhi (Python, Pandas, Numpy)

March 2020 - April 2020

- Analysed Covid's impact using a Covid dataset and an associated (non-covid) dataset
- Performed AR, EWMA, Wald's test, Z-test, K-S test, permutation test, correlation, and Bayesian inference along with suitable visualizations

Retail Sales Data Analysis, Under Supervision of Prof. Steven Skiena (Python, Pandas)

Oct. 2019 - Dec. 2019

- Provided insights to a local chain retail stores to help the company with their business by working on their market basket dataset
- Analysed data to find reasons for drop in sales with additional help of external datasets
- Implemented **Word2Vec and LSTM model with k-means clustering** to improve Shelf Space Management and Inventory management

Video Action Classification - Recognition (Python, PyTorch, Matlab)

Nov. 2019 - Dec. 2019

- Trained CNN for human action recognition task on the UCF101 data
- Achieved **86%** (1st in class) on image test data, and **84%** (2nd in class) on video frames test data
- Trained LSTM in RNN to actions classification on data collected by Kinect v2
- Used Transfer Learning to compute features for 60000 video frames with limited compute resources.

Ultrasound Nerve Segmentation (Python, Tensorflow, Keras)

Nov. 2017 - Jan. 2018

- Designed and trained CNN based on U-net to segment a collection of Nerves in Ultrasound Images taken from a Kaggle competition
- Used pre-trained weights of VGG model in the Encoder of the network (transfer learning), gained **68 %** Test Accuracy (Top 100)