SAHIL MODI

sahilkmodi@gmail.com | 847-890-3506 | U.S. Citizen | sahilmodi.dev | sahil-modi | sahilmodi

EDUCATION

M.S. Computer Science **May 2022**
University of Illinois at Urbana-Champaign, Thesis Topic: Computer Vision & Deep Learning *GPA: 4.00/4.00*

B.S. Computer Science, Minor in Statistics **May 2021**
University of Illinois at Urbana-Champaign *GPA: 3.96/4.00*
Coursework Computational Photography, Computer Vision, Machine & Deep Learning

TECHNICAL SKILLS

Languages Python, C++, C, Java, Javascript, Typescript, SQL, Bash
Frameworks PyTorch, OpenCV, TensorFlow, Linux, Git

PROFESSIONAL EXPERIENCE

NVIDIA | *Inference Software Engineer, TensorRT* *Santa Clara, CA* | **Sep 2022 – Present**
- Prototyped LLM weight and offloading to enable **large-language-model** inference on a single GPU
- Reduced system memory usage by over 50% for **generative transformer models** (GPT3, BART)
- Extended the TensorRT Fill layer to include a wrap mode

NVIDIA | *Software Engineering Intern, TensorRT* *Santa Clara, CA* | **May 2021 – Aug 2021**
- Introduced a software-based kernel timing heuristic (DLSim) for **neural network optimization**
- Bridged **C++** TensorRT and **Python** DLSim by implementing a server-client interface handling 10s of queries/s with convolutional layer parameter translation
- Compared 5 networks (**ResNet-50**, **MobileNet**, **Inception-V4**, etc), across 3 batch sizes and FP16, INT8 precision, meeting the <10% throughout reduction goal

Amazon | *Software Development Engineer Intern* *Seattle, WA* | **May 2020 – Aug 2020**
- Reduced aggregate **Javascript** asset build time by 18.5% and decreased memory usage by 11%
- Analyzed code syntax trees for unfavorable behavior, decreasing final asset size by 5%
- Designed a variant generation algorithm an **order of magnitude faster** for server built variants and client responsive variants

Distributed Autonomous Systems Laboratory **Jan 2020 – May 2020**
Undergraduate Research Assistant | Advisors: Dr. Girish Chowdhary, Dr. Saurabh Gupta *Urbana, IL*
- Investigated vision-based robot heading estimation with a **self-supervised network** on **PyTorch** achieving 2 degrees error
- Devised a **supervised network** for autonomously calculate pose and drive a robot with distance to intervention of 30 meters
- Augmented video data with **homographic transformations** to simulate robot variance and increase dataset coverage

EarthSense | *Computer Vision Research Intern | Advisor: Dr. Girsh Chowdhary* *Champaign, IL* | **Sep 2019 – Dec 2019**
- Ascertained intrinsic camera matrices of Terrasentia robot cameras
- Achieved 92% accuracy for corn ear height estimation from video by fusing a **neural network** with **single view metrology**

Northrop Grumman | *Software Engineering Intern* *Rolling Meadows, IL* | **May 2019 – Aug 2019**
- Developed a C# application to configure and test missile warning algorithms and pulled in project schedule by 2 months

EarthSense | *Computer Vision Intern* *Champaign, IL* | **Sep 2018 – May 2019**
- Trained a **convolutional neural network** with **TensorFlow** on a biased dataset to classify lodging of wheat with 80% accuracy
- Deployed a **TensorFlow ML** model to detect and count plant stems with 96% accuracy

PROJECT HIGHLIGHTS

HackIllinois Stock Analysis
- A python package discovering sentiment about a company from its tweets using NLTK and correlating it with stock price
- **Linear, ridge regression**, and a **convolutional neural network** are used for prediction and compared against each other

CU-Recycle
- Devised an Android application to report an item's recyclability status in the Urbana-Champaign area, **winning 2nd** at PygHacks
- Trained a convolutional neural network for **object recognition** with **Keras** to overcome lighting and object variance

PUBLICATIONS

Learned Visual Navigation for Under-Canopy Agricultural Robots | *Robotics: Science and Systems* **2021**
Arun Sivakumar, **Sahil Modi**, Mateus Gasparino, Che Ellis, Andres Velasquez, Girish Chowdhary, Saurabh Gupta