Mustafa Shaikh

mustafanshaikh@gmail.com | linkedIn/mustafa-shaikh-716ba8140 | shaikh58.github.io/website

Machine learning practitioner focusing on object tracking with experience in computer vision, deep learning and classical ML. I have 5 years of industry experience in data science working closely with business stakeholders. I can dive deep into the details, and I have also lead projects end to end, from identifying a problem and designing a solution, to implementation and delivery.

EDUCATION

M.S. Electrical and Computer Engineering — San Diego, CA | 2022 - 2024
University of California, San Diego
Specialization: Robotics and Intelligent Systems

B.A.Sc. Engineering Science — Toronto, ON | 2013 - 2017
University of Toronto

WORK EXPERIENCE

RESEARCH ENGINEER | Salk Institute for Biological Studies — San Diego, CA | 2024 -

Key Skills: Multi-object tracking, deep learning, transformers, ResNet, rotary embeddings, segmentation, Kalman filters, optical flow, Hungarian matching, camera distortion, video encoding, ffmpeg, parallel video loading

Summary: Implement a deep learning based multiple object tracking system applied to animal and microscopy videos, to study the impact of neurodegenerative diseases on animal and cell movement

- Implemented features to boost learning capacity and usability of the model, including: **rotary embeddings**, support for **sparsely annotated** training data, mixed dataset training, customizable **feature extractor factory**
- Acquired and processed public cell tracking video datasets 50x larger than in-house data
- Created a **pretrained microscopy model** that **improves key metrics** ID switches and tracking accuracy by **>40%** vs. in-house baseline, and **>8%** vs. state of the art
- Optimized tracker code to reduce **GPU memory** usage by **50%** and reduce **runtime** by **30%** during inference
- **Created** and **implemented global tracking accuracy** metric to address gaps in existing multi-object tracking metrics; this helped address a key failure mode of the model

GRADUATE RESEARCH ASSISTANT | Existential Robotics Lab, UC San Diego San Diego, CA | 2023 - 2024

Key Skills: Model predictive control, control barrier functions (CBF), casADi, CVX, JAX, automatic differentiation, QP, Extended Kalman Filter, RRT*, collision avoidance, ROS, Jackal robot, LiDAR, depth camera, HectorSLAM

Summary: Determine an optimal, collision-free trajectory for a robot to keep a target within its limited field of view

Publication: *"Control Strategies for Pursuit-Evasion Under Occlusion Using Visibility and Safety Barrier Functions"* accepted at the *IEEE International Conference on Robotics and Automation (ICRA) 2025*. Arxiv: https://arxiv.org/abs/2411.01321

- **Implemented** a **model predictive controller** with control barrier (CBF) constraints to compute the optimal trajectory of a robot in order to keep a target within the field of view
- Implemented features to improve the **stability, and reliability** of the controller: **smoothing** to reduce changes in direction; **adaptive tuning** of CBF to improve collision avoidance; **slack** to improve feasibility of solution
- Demonstrated **>95% tracking** in **real world experiments** with Jackal wheeled robot in a cluttered environment
- Achieved 50Hz control frequency on the robot, including SLAM and target detection
- Implemented **Extended Kalman Filter** to estimate target's position and velocity using camera detections

DATA SCIENTIST, SR. DATA SCIENTIST | Walmart Canada

Toronto, ON | 2019 - 2022

Key Skills: Natural Language Processing - Spacy, Named Entity Recognition, BERT, human-in-the-loop systems, AutoML, ARIMA, Prophet, xgboost, PySpark, SQL, MLOps, Python, Pandas, Numpy, Keras, Google Cloud Platform, Airflow

Project lead - Automated Attribute Assignment (2021 - 2022). **Goal:** Extract product features from item descriptions to populate missing data for products on walmart.ca to improve search quality

- Developed **named entity recognition pipeline** (Spacy, BERT) to learn **context-aware** features from product descriptions; led to **>$1MM CAD revenue increase** annually by populating features for over 500,000 items
- Recognized need for **high quality custom annotated data**; pitched and integrated a **human-in-the-loop** annotation tool (Prodigy); setup **active labelling** loop with **least confident predictions** annotated
- **Coordinated Jr. Data Scientist**, and guided the implementation of an **asynchronous orchestration layer**
- Worked closely with business stakeholders to guide problem framing, roadmap, execution and production support

Other Projects (2019 - 2021)

- Developed and deployed **hierarchical model factory** (xgboost) to categorize 3rd party vendor items on walmart.ca; increased **categorization rate from 90% to 97%** which increased product views for previously 'unfindable' items
- Lead developer for data-driven, rules based online **grocery substitutions recommendation** engine; **300bps improvement in customer satisfaction**
- Created and maintained **fulfillment centre forecast** (ARIMA, Prophet, AutoML) to optimize labour; >90% accuracy up from 75% previously, **>$1MM annual labour savings** by achieving high accuracy during holiday period
- Designed and implemented **randomized control test** to test optimal email voucher strategy for grocery customers; **saved $100,000 in marketing spend** for a single 3-month campaign

SR. DATA ANALYST | Shoppers Drug Mart

Toronto, ON | 2017 - 2019

Key Skills: Data engineering, systems engineering, performance metrics, worfklow improvements

- Developed **ETL pipeline** and a large SQL application that **transforms event triggers** into key pharmacy **workflow metrics** such as customer wait times, rework rates and labour heatmaps, to enable store performance reporting
- Used metrics to create business cases and guided rollout of **$1MM CAD in workflow improvements**

PROJECTS

DEEP LEARNING AND SOFTWARE

UC San Diego, 2023 - 2024

Key Skills: PointNet, instrinsics, odometry, depth images, SIFT, point cloud registration, PyTorch, C++, RAII, smart pointers, design patterns, templates, variants, STL

- **PointNet-based neural network** in PyTorch to solve the **point cloud registration** problem, and achieved performance close to **Iterative Closest Point** (ICP) method on basic trajectories in Eden dataset
- (C++) **String library** with underlying buffer manager. **Achieved 25% lower memory usage than C++ std::string** for common string operations: append, replace, insert, erase, search
- (C++) Graph-based **in-memory JSON datastore** with query capability
- (C++) **Compressing archive tool** with add, extract, retrieve capability

SKILLS

Python, C++, SQL, PyTorch, JAX, Numpy, Pandas, Keras, CVX, CasADi OpenCV, GCP, PySpark, Hadoop, Airflow, ROS, ROS2