Roger Wang

Los Angeles, CA | rogerwng9@gmail.com | (669) 207-7815 | LinkedIn: rogerwang00 | rogerwng.github.io

EXPERIENCE

LiveMechX Lab, University of Osaka *Osaka, Japan*
Urban Search and Rescue Robot, Data Science Intern *June 2023 - August 2023*
- Built a pipeline around robotic sensor data for end-to-end data collection, preprocessing, and model training
- Led project to optimize robotic localization accuracy to 76% with regression models + gradient boosting trees
- Assisted in training human detection models using low-resolution infrared cameras w/ linear SVM to 98% acc
- Presented applied machine learning strategies for real-time localization in a public research conference

PROJECTS

Exploring Crime with San Francisco's Police Report Database *Los Angeles, CA*
Personal Project - Python, Pandas, Matplotlib, Sk-learn, XGBoost, PyTorch *March 2025 - April 2025*
- Cleaned and visualized dataset of over 900k police incident reports in SF since 2018 using Pandas + Matplotlib
- Achieved 1.701 log-loss in 17-category classification using ensemble of RandomForests, XGBoost, and NN
- Derived insights by forecasting daily crime rates in 2025 using XGBoost-Regressor in time series analysis

Applying Generative Data Augmentation in Data Scarce Environments *Los Angeles, CA*
Academic Project - Python, Huggingface, PyTorch *November 2024 - December 2024*
- Fine-tuned Stable Diffusion using Dreambooth to generate realistic, subject specific images of Felis cat species
- Doubled size of the training dataset from ~500 to ~1200 using generated images for data augmentation
- Boosted image classification accuracy of ResNet18 by 12% and 3% on untrained and pretrained models

MiniPlaces Classification Challenge *Los Angeles, CA*
Academic Project - Python, PyTorch, XGBoost, GCP *October 2024 - November 2024*
- Placed in 10th percentile using CNN ensemble of ResNet, RegNet, and EfficientNet with top-1 accuracy of 64%
- Experimented with CNN-XGBoost hybrid models by passing high-level CNN features as XGB inputs (58% acc)
- Boosted evaluation accuracy of all models by >2% by implementing test-time augmentations (FiveCrop)
- Achieved 76% acc by replicating winning techniques (ensemble confidence, label smoothing, mixed precision)

Insurance Sentiment Classification and Used Car Price Regression *Los Angeles, CA*
Kaggle Competitions - Python, Optuna *July 2024 - September 2024*
- Developed models in Python with popular gradient boosting trees libraries (XGBoost, LightGBM, CatBoost)
- Created an automated ML pipeline including hyperparameter search in Optuna for fast and efficient prototyping
- Gained insight into competitive strategies including OOB scoring, hill climbing, and auto ML with AutoGluon

Convolutional Neural Networks on Image Classification *Los Angeles, CA*
Personal Project - Python, PyTorch, AWS EC2 *October 2023 - January 2024*
- Leveraged cloud compute to accelerate and optimize large model (>16GB) training with AWS EC2 instance
- Implemented CNN's in PyTorch from research papers: LeNet, AlexNet, VGG, NiN, Inception, and ResNet

EDUCATION

Georgia Institute of Technology - *Masters in Computer Science* *In Progress*
- Machine Learning Specialization

University of California, Los Angeles - *Aerospace Engineering, B.S., Data Science minor* *December 2024*
- Relevant Coursework: Deep Learning for Computer Vision, Artificial Intelligence and Search Algorithms, Data Structures and Algorithms, Data Science, C++, Computer Organization, Probability, Statistics, Linear Algebra
- Activities: *UCLA Rocket Project, UCLA DataRes Research, ASUCLA Supervisor*

SKILLS

Languages: C++, C, Python, MATLAB, SQL, HTML/CSS, JS
Libraries/Frameworks: Pandas, Sk-Learn, PyTorch, Huggingface, OpenCV, Matplotlib, XGBoost, AWS + GCP