Yihan Duan

duanyihan@cs.toronto.edu (519) 781-1471
github.com/airbagy
www.linkedin.com/in/yihan-dennis-duan

- Highly experienced in machine learning model training, evaluation, deployment and data pipelining.
- Extensive knowledge in graph clustering algorithms, object detection models and representation learning.
- Solid experience in distributed algorithm design, open-source software development and Android development.

EDUCATION

Candidate for MSc in Applied Computing, Data Science Concentration
University of Toronto, Department of Computer Science — Sep 2021 – Dec 2022 (expected)
Bachelor of Computer Science, Honours Data Science, Co-operative Program (Dean's Honours)
University of Waterloo, Department of Math — Sep 2016 – May 2021

EXPERIENCE

Google (Ads), *Software Developer Intern* — May 2021 – Aug 2021
- Designed, implemented and conducted multiple experiments on heterogeneous graph clustering algorithms.
- Increased recall by about 15% compared to current algorithm in production, improved clustering stability.
- Researched and implemented new distributed graph pre-processing technique with probabilistic sampling.
- Established workflow for configuring, launching and analyzing multiple parallel experiments using Google's internal tools.

Google (i18n), *Software Developer Intern* — May 2020 – Aug 2020
- Developed an open-source project for Unicode https://github.com/unicode-org/ml-confusables-generator.
- Designed and implemented end-to-end computer vision system for Chinese character spoof detection by automatically identifying large sets of visually similar characters.
- Utilized representation learning (triplet loss) and transfer learning techniques for model training.
- Prototyped custom metrics measuring similarity between any pair of characters based solely on their visual features.

DarwinAI Corporation, *Deep Learning Developer Intern* — Jan 2019 – May 2019, Sep 2019 – Dec 2019
- Implemented and trained networks including FasterRCNN-PVANet, FSSD-Resnet50, SqueezeDet, Yolo v3 etc.
- Built data pipeline for ImageNet, VOC and COCO datasets, integrated AutoAugment and RandAugment for generalization.
- Transferred pre-trained PVANet model from PyTorch to TensorFlow, verified output layer by layer.
- Improved inference speed using model trimming, quantization and TensorRT with Nvidia GPUs.
- Engaged with automotive and aerospace customers' research team to meet their performance requirements.

Sony Interactive Entertainment (PlayStation), *Data Analyst Intern* — May 2018 – Aug 2018
- Created an interactive web app for data visualization, comparative analysis and anomaly detection/monitoring.
- Designed and implemented scripts for efficient auto-scheduled data update and data aggregation.

PerkinElmer Innovation Lab, Full-stack Developer Intern — Jan 2018 – Apr 2018
- Patent inventor for ICP-OES spectral interference correction technology.
- Created and fine-tuned convolutional neural network models for interference correction, matching commercial algorithm MSF in terms of accuracy and performance.
- Utilized data pre-processing techniques including Savitzky-Golay and PCA for noise reduction.

PROJECTS

Skin Cancer Detection Android App
- Built compact classification network based on MobileNetV2 that achieved 73% accuracy despite data biases.
- Developed Android App for skin cancer detection with image capturing, luminance correction and analysis.

Motion Sensing Model for PS4 Dev Kit Control Panel
- Constructed rule-based motion sensing program for controlling PS4 dev kit.
- Utilized OpenPose deep learning model for human pose and hand gesture estimation on Jeston TX2.

TECHNICAL SKILLS

- **Programming Language**: Python, C++, SQL, Java, R, HTML, CSS, JavaScript
- **Tools**: TensorFlow, PyTorch, TF Object Detection API, Docker, Git, Bash, Google Colab, Hadoop, Spark, Linux