Fangyuan (Carrie) Xu

fx49@cornell.edu | 609-582-8635 | GitHub: carriex | carriex.github.io/me

Experience	**Twitter (San Francisco), Machine Learning Engineer**	July 2020 - Present

- Working at Ads Machine Learning Team which is responsible for building end to end machine learning system for Twitter's Ads (ETL pipelines, model training and serving)
- Designed and developed engagement features for engagement prediction model; designed experiment protocol and custom metrics for offline evaluation; developed Beam pipeline for feature generation, designed and launched online experiment
- Performed analysis which helped the team understand the performance of audience extension model and pinpointed headroom for improvements

CUHK T Stone Robotics Institute (Hong Kong), Research Assistant
June – Aug 2019

- Deep learning on video understanding, supervised by Professor Yun-hui Liu.
- Implemented self supervised learning algorithm with motion statistics [repo], built multi-processes data preprocessing pipeline on video datasets with OpenCV and FFmpeg.

Morgan Stanley (Hong Kong), Technology Associate
Aug 2017– June 2019

- Asia E-trading system's production management team; support lead for realtime inventory management system, worked with multiple stakeholders to ensure pipeline healthiness
- Tooling: built and deployed a set of tools (REST API and python CLI) for project provisioning

Education

Cornell University (New York), Master of Engineering (Computer Science)
Sep 2019 - May 2020

- GPA: 4.06/4.3 (Overall), 4.1/4.3 (CS).
- Relevant courseworks: Natural Language Processing, Intro to Computer Vision, Deep Learning, Applied Machine Learning, Optimisation Methods, Networked and Distributed Systems.

University of Hong Kong (Hong Kong), Bachelor of Engineering
Sep 2013 - June 2017

- Major in Computer Science and Minor in French.
- GPA: 3.73/4.3 (Overall), 3.9/4.3 (CS). Graduated with First Class Honour.
- Award: Kai Chong Tong Scholarship, Walter Brown Memorial Prizes in Mathematics

Skills

- Python, Java, Scala, C++/C, SQL.
- PyTorch, TensorFlow, scikit-learn, Pandas, NumPy.
- Hugging Face, NLTK, OpenAI Gym, OpenCV, FFmpeg.
- KubeFlow, Apache Beam, Docker, Scalding, AWS, GCP.

Projects

- ***Grounded Language Understanding:*** Developed a sequence to sequence model which maps situated sequential instruction to action on SCONE dataset. [repo] [blog]
- ***Named Entity Recognition on Twitter's dataset:*** Built a NER model on Twitter text using BERT. Achieved F1 score with 0.71 (Placed top 2/60 on class leaderboard).
- ***Image Retrieval on Google Landmark Dataset:*** Implemented and evaluated the DELG algorithm on image retrieval task on Google LandMarks Dataset v2. [repo]
- ***Large Scale Image Engine:*** Developed a Ridge Regression model for text to image retrieval using pre-trained image embeddings and various language model (BoW, word2vec, BERT).
- ***Distributed Chatroom:*** Implemented a raft-based distributed chatroom using C++.
- ***Landmark Recognition App:*** Built a landmark recognition mobile app using React and Google Vision API (in collaboration with Priceline).