Hao-Ning Wu

http://howardwu.space/

Email : haoning.wu@nyu.edu
Mobile : (917) 302-5403 | New York, NY

Education

New York University (Courant Institute) — New York, NY
Master of Science in Computer Science; GPA: 3.9/4.0 — Sept. 2019 – May 2021

National Tsing Hua University — Hsinchu, Taiwan
Master of Science in Computer Science; GPA: 4.0/4.0 — Sept. 2016 – Sept. 2018

National Taiwan University — Taipei, Taiwan
Bachelor of Science in Electrical Engineering — Sept. 2010 – June 2014

- **Coursework:** Deep Learning, Machine Learning, Deep Learning Systems, Computational Cognitive Modeling, Data Mining, Big Data, Computer Vision, Graphics Processing Units, Computer Graphics, Parallel Programming

Publication

H-N Wu and C-T Huang, "Data Locality Optimization of Depthwise Separable Convoluions for CNN Inference Accelerators," *2019 IEEE/ACM Design, Automation and Test in Europe (DATE)*

Work Experience

PicCollage — Taipei, Taiwan
Machine Learning Intern — July. 2020 – Sept. 2020

- Researched and implemented an internal image retrieval system for 20K in-app stickers on my own
- Performed semi-supervised classification on unlabeled stickers
- Enabled arbitrary search terms including color names, object names, adjectives, and etc

National Tsing Hua University — Hsinchu, Taiwan
Research Assistant — Jan. 2019 – June 2019

- Designed a tool in **Pytorch**, allowing users to apply function-preserving transformations to pre-trained CNNs
- Simplified the process of CNN pruning and neural architecture search with user-defined **JSON** configuration files
- Reduced 25% parameters in VGG16 with negligible accuracy drop on ImageNet by L1-norm filter pruning

Projects

Incorporating Prior Knowledge to RL Agents for Atari Games — Mar. 2020 - May 2020

- Designed learning curriculums by modifying the rules and components of **PyGame Learning Environment**
- Implemented and tuned the A3C-ICM model using **Pytorch** and **multiprocessing** package
- Speeded up 5x RL agents' training on unseen maps in a sparse-reward setup

Imbalanced Classification for Fake Review Detection — Apr. 2020 - May 2020

- Developed NLP pre-processing pipeline using **Scala** and **Spark** on 300K Yelp's reviews
- Solved imbalanced dataset problem by class weighting and various data re-sampling methods
- Built a gradient boosted trees model achieving 50% AP and 90% AUC with **SKlearn**, **XGBoost** and **Pandas**
- Improved the classification result by 5% AP and 2% AUC with 5 innovative new features

Data Locality Optimization of Convolutions — Mar. 2018 - Sept. 2018

- Invented a new loop transformation sequence to optimize data reuse in tiled convolutions
- Generalized existing algorithms by fusing consecutive layers to eliminate unnecessary data transfer
- Reduced 67% DRAM energy and 65% DRAM access latency for MobileNet V2 as reported by DRAMSim2
- Presented and answered questions to 100+ professionals at DATE conference

Text to Photo-Realistic Image Synthesis — Dec. 2017 – Jan. 2018

- Implemented StackGAN using **Tensorflow** and **Numpy** to generate 256×256 realistic flower images from text
- Improved data pre-processing pipeline using word embeddings extracted from a sequence-to-sequence model

Blocked Floyd-Warshall Algorithm on GPUs — Dec. 2016 – Jan. 2017

- Parallelized the algorithm using the combination of **C++**, **CUDA** and **MPI**
- Optimized the performance with selected tile sizes, pinned memory and non-blocking APIs, e.g. CUDA Stream
- Achieved 20x speedup compared to the CPU version as reported by **nvvp**

Skills

Languages: Python, Scala, C++, Java, Verilog, HTML, CSS, Bash Script
Tools: TensorFlow, PyTorch, Hadoop, Spark, XGBoost, MPI, CUDA, OpenMP, OpenGL, Scikit-learn, Matplotlib, Git