Hao Xu

+1 8622145874 hx8@njit.edu github.com/haoxuhao Newark, NJ

An ex software engineer with proven problem-solving skills, now focusing on machine learning algorithm innovations.

Education

New Jersey Institute of Technology **2023 - Now**
PhD in Computer Science; Interests: General Visual Representation Learning, Multi-Modality; GPA: 4.0/4.0 *NJ, US*

Southeast University **2018 - 2021**
MSc in Control Science and Engineering; Outstanding graduate thesis award *Nanjing, China*

Southeast University **2014 - 2018**
BEng in Automation; Overall ranking top 10% in class *Nanjing, China*

Learning at NJIT

- Data poisoning attack in web-scale training data: Investigated data poisoning in text2image latent diffusion models, and explored the relationship between the noisy nature of web-based text-image data and poisoning efficiency.
- Memorization ability of DNN: Investigated how the residual connection affects CNN's ability in memorizing noisy-labeled data on the CIFAR-10 dataset, demonstrating that residual connection reduces the memorization effect.
- Open Courses: Caltech's Learning From Data, Stanford NLP, Fastai-Practical Deep Learning.

Work Experience

PINDUODUO (a.k.a. Temu) **Jul 2021 - Aug 2023**
R&D Software Engineer - Recommender System - GPU optimization for DNN model *Shanghai, China*

ANT GROUP - Alibaba GROUP **Jun 2020 - Aug 2020**
Software Engineer Intern - Operations Research - Software framework for multi-objective optimization *Hangzhou, China*

Swinburne University of Technology **Sep 2019 - Dec 2019**
Research Intern - Evolutionary multitasking algorithm and parallel computing *Melbourne, Australia*

Fujitsu Nanjing **Mar 2018 - Jul 2018**
Software Engineer Intern - Distributed File System *Nanjing, China*

Selected Projects

GPU Optimization for Deep Recommendation Models **July 2021 - Apr 2023**
- Improved model training performance by about 18% over Nvidia's open source framework through redesigning the CUDA kernel based on the in-house data distributions.
- Reduced the online serving GPU cost by 20% through mixed precision matrix operations and thorough kernel selection strategy, surpassing Nvidia's TensorRT.
- Improved the GPU efficiency by 10% under high serving pressure through dynamic mini-batch packing.

Evolutionary Multi-Task Optimization with Adaptive Knowledge Transfer **Sep 2019 - Mar 2021**
- Utilized two multi-armed bandits to reduce the occurrence of negative knowledge transfer and enhance the positive one in evolutionary multitasking. The method is published in a top journal of this field.
- Outperformed 3 SOTA methods on 3 benchmark problems and 1 real-world problem regarding the convergence and accuracy.
- Designed 2 ablation studies to validate the behaviour of the proposed method and analyzed the parameter sensitivity.

Vehicle Detection in Aerial Images **Aug 2019 - Sep 2019**
- Fine-tuned the RetinaNet model, backboned by Resnext101, to detect vehicles in aerial images, achieving rank 2nd out of 74 teams from 7 Chinese top universities.
- Designed an overlapped sliding-window technique to deal with large resolution variation, which generates image patches for both training and inference, greatly improving the validation accuracy by 10%.
- Utilized the object size distribution and number of vehicles to determine the hard examples and conducted hard example resampling to improve accuracy by about 3%.

Crowd Counting with Segmentation Map Guidance **Oct 2018 - Jan 2019**
- Augmented the crowd density estimation model with an additional segmentation task to reduce false responses under complex backgrounds. This work is published as a conference paper.
- Compared the method with the SOTA method on 2 public datasets and showed a 7% improvement and comparable results on mean absolute error, respectively. The visualization results also demonstrate its superiority in complex backgrounds over baseline methods without the segmentation task.

Publications

H. Xu, A. K. Qin, and S. Xia, "Evolutionary Multitask Optimization with Adaptive Knowledge Transfer", IEEE Transactions on Evolutionary Computation, 2022.
H. Xu, C. Zheng, Y. Nie, and S. Xia, "Crowd Counting with Segmentation Map Guidance", Chinese Control Conference (CCC), 2019.

Technical Skills

Programming: Python, C++, CUDA, LATEX, SQL, Git
ML Frameworks: PyTorch, TensorFlow, Fastai
ML Techniques: YOLO, ResNet, RetinaNet, Pruning, Quantization, Latent diffusion model.

Awards

2022 SEU Distinguished Graduate Thesis Award (Top 1% University wide)
2019 First Prize in 2019 Object Detection and Recognition Competition, CSAA (2/74 Nationally)
2017 Second Prize in National Student Electrical Design Competition (Top 4.7% Nationally)
2016 National Encouragement Scholarship (Top 3% Nationally)
2015 Texas Instruments Scholarship (Top 5% University wide)