

Daheng Yin

daheng_yin@sfu.ca | Simon Fraser University

Education

Simon Fraser U

Computing Science PhD
2023.09~

Southeast University

Master of Engineering
Major: Computer Science
2020.09~2023.06
GPA 81.69/100

Cambridge University

Visiting Student
2017.08-2017.09

Jiangnan University

Bachelor of Engineering
Major: IoT Engineering
2016.09~2020.09
GPA 3.59/4

English

IELTS 6.5
CET6 576

Skills

Pytorch
CUDA • TensorRT
WebRTC • LibVPX
Docker • Kubernetes
Git • \LaTeX

Programming

(Lines of code)

Python	23,963
Golang	20,812
C/C++	8,478
Java	7,309
JavaScript	4,013
PHP	4,294
Matlab	1,395
C#	392

Links

Github@yindaheng98
Blog:yindaheng98.top

Research & Development

Collaborative live video super-resolution with edge computing 2021.06~now

- A system research that leverage distributed computing resources to maximize the latency-bounded quality of live video super-resolution in edge computing environments.
- Academic achievements: 1) Proposed a parallel-optimized DNN architecture to improve multi-device acceleration in edge computing environments. 2) Designed a distributed inference schedule mechanism based on adaptive batch size to optimize content quality and latency of distributed inference.
- Engineering achievements: 1) Low-latency video stream routing and dynamic topology control across multiple devices based on WebRTC. 2) Dimensional compression and int8 quantization of intermediate features to reduce transmission latency of distributed inference. 3) SR-integrated decoder based on LibVPX to accelerate video super-resolution using compressed video information.
- Related paper D. Yin et al., "WAEVSR: Enabling collaborative live video super-resolution in wide-area MEC environment," rejected by WWW 2023 and now editing for IWQoS 2023

Cooperated Research & Development

Adaptively computational routing based on environmental awareness in Compute First Network (CFN) 2020.10~2020.12

- Optimize the strategy of 1) DNN layer segmentation for distributed deployment, 2) computing device selection, 3) data transmission path selection.
- My contribution: Development of DNN inference control testbed (DNet), schedule and synchronize inference process among multiple computing devices.
- Related paper: X. Guo et al., "Exploiting the computational path diversity with in-network computing for MEC," 2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), 2022, pp. 1-9.

Projects

Contest TensorRT Hackathon 2022 Winner Prize 2022.03~2022.5

NVIDIA | Alibaba Cloud TIANCHI

- Quantized a speech recognition DNN WeNet and a super-resolution DNN ELAN to FLOAT16 and INT8 using TensorRT.
- Fixed the precision issue of FLOAT16 BatchNorm by implementing a TensorRT plugin.
- Optimized FLOAT16 quantization of ELAN by omitting layers with significant impact on precision, reducing error by 75%.
- Achieving 2x speedup by quantizing ELAN to INT8 with QAT.
- Github: github.com/liu-mengyang/trt-wenet and github.com/liu-mengyang/trt-elan

Contest TensorRT Hackathon 2021 Ranking 4/48 2021.03~2021.5

NVIDIA | Alibaba Cloud TIANCHI

- Quantized a multi-object tracking DNN FairMOT to FLOAT16 and INT8 using TensorRT.
- Implement FLOAT16 DCNv2 kernel as a TensorRT plugin, resulting in a $2.36\times$ speedup.
- Import and align model parameters from Pytorch into TensorRT through API.
- Github: github.com/liu-mengyang/trt-fairmot

Some Web Projects

- | | |
|-----------------|--|
| 2022.06~now | github.com/yindaheng98/dblp-crawler
Asynchronous dblp data collector and analyser. |
| 2019.03~2019.05 | github.com/yindaheng98/ExpertField
Collect and show field data collection. From Android and STM32 devices
9 cooperators, nearly 10,000 lines of code, a complete IoT system. |
| 2018.11~2018.12 | github.com/yindaheng98/iQRGenuine
Cloud-based anti-fake system. |
| 2018.12~2019.03 | github.com/yindaheng98/WebSiteAnalysisKit
A user behavior analysis platform based on Hadoop. |