

Yin Daheng

yindaheng98@seu.edu.cn | +86 188 5189 9135 | Southeast University, Nanjing, Jiangsu

Education

Southeast University

Master of Science
Major: Computer Science
2020.09~2023.06
GPA 3.59/4

Cambridge University

Overseas study tour
2017.08-2017.09

Jiangnan University

Bachelor of Engineering
Major: IoT Engineering
2016.09~2020.09
GPA 81.69/100

English

IELTS 6.5
CET6 576

Skills

Pytorch
CUDA • TensorRT
WebRTC • LibVPX
Docker • Kubernetes
Git • Github • GitLab
C++ • Golang • Java

Links

Github@yindaheng98
Blog:yindaheng98.top

Research & Development

Accelerate live video super-resolution with edge computing 2021.06~currently

- Derived a parallel-friendly DNN architecture from a multi-scale feature extraction structure for better multi-device acceleration in the edge environment.
- Dimensionally compressed and int8 quantized intermediate features of the DNN and encoded features into a video stream for transmission among multiple devices.
- Enhance the video decoder to combine low-framerate high-definition stream and high-framerate low-definition stream into high-framerate high-definition streams for smooth video playback when frame-by-frame super-resolution inference is not supported.
- Control the inference process adaptively based on the variable batch size of DNN input and enhanced video decoder to achieve the best video quality under a specific latency bound in dynamic edge environments.
- Implement low-latency video stream routing and dynamic topology control across multiple devices based on WebRTC.
- Related paper D. Yin et al., "WAEVSR: Enabling collaborative live video super-resolution in wide-area MEC environment," is submitted to WWW 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), which can 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

- Quantize a speech recognition DNN WeNet and a super-resolution DNN ELAN to FLOAT16 and INT8 using TensorRT.
- Implement FLOAT16 BatchNorm as a TensorRT plugin to replace native TensorRT BatchNorm kernel with precision issues.
- Search and omit the quantization on those layers that have a significant impact on precision in FLOAT16 quantized ELAN structure, which decreases the error by 75% and has a similar speedup.
- Quantized ELAN to INT8 with QAT, which achieved 2x speedup.
- 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

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

Contest & Scholarship During Undergraduate

2018.09	National College Mathematical Contest in Modeling	2nd Prize(National)
2017.11	9th National College Mathematical Contest	2nd Prize(Provincial)
2017.11	China National Scholarship (2016-2017)	
2017.05	14th Jiangsu College Mathematical Contest	1st Prize