

YOUNGMOK JUNG, Ph.D

✉ young.mok.jung.cs@gmail.com bit.ly/youngmokjung github.com/quito418

Work Experience

Inocras Inc. (formerly Genome Insight Inc.)

Aug 2023 - Present

AI team lead - Developing and applying machine learning methods to problems in science and health (1y 6m) San Diego

- Leading development of ML software and pipeline for product, MRDVision
- Development of a generative model-based approach for cancer subtyping using methylation sequencing
- Leading the design and implementation of deep learning based somatic mutation caller

Awards and Honors

Samsung Electronics 29th Humantech Paper Award (Silver Prize, Communication & Networks)

Feb. 2023

Samsung PhD Sponsorship

Feb. 2022

1st Place in Kiwoom US Stock Trading Competition (ROI 201%, out of 10K participants)

Sep. 2020

Education

KAIST (Korea Advanced Institute of Science and Technology)

Sep. 2018 - Feb. 2024

Ph.D. (Electrical Engineering)

Advisor: Dongsu Han and Young Seok Ju

KAIST (Korea Advanced Institute of Science and Technology)

Feb. 2014 - Aug 2018

Bachelor of Science in Electrical Engineering

Projects

1. Enhancing Genome Analysis Pipeline with AI and ML

Generalizing Deep Variant Callers via Domain Adaptation and Semi-supervised learning | Pytorch Aug 2023

- Deploying deep learning-based variant callers (DVCs) to a sequencing method with varying error profiles necessitates generalization which is challenging due to their reliance on extensive labeled data.
- Developed a generalization framework that enables DVCs (e.g., Google DeepVariant, Clair3) to accommodate diverse sequencing methods, leveraging semi-supervised learning and domain adaptation techniques.
- Improved SNP and INDEL F1-score by up to **6.40 %p** and **9.36 %p** or achieves the same variant calling accuracy using merely half of the labeled data compared to the supervised training approach in the sequencing method of interest.

BWA-MEME: Machine-learning Enhanced Read Alignment Software | C++, Rust, SIMD vector May 2022

- BWA MEM is an **industry-standard alignment software** developed by the **Broad Institute of MIT and Harvard** for next-generation sequencing data.
- Developed and open-sourced a short-read alignment software, BWA-MEME, that achieves up to **3.45x speedup** in seeding throughput over its' predecessor, BWA-MEM2 from Intel, while ensuring identical output. Our approach reduced the number of instructions by 4.60x, memory accesses by 8.77x, and LLC misses by 2.21x,
- BWA-MEME is now operational in the production environments of numerous institutions, projected to lower alignment costs by 35%. This efficiency translates into **millions of \$ in cost reductions** for projects on a million-genome scale.

2. Deep Learning for High-performance Network System

LiveNAS: Deep Learning-based Live Video Streaming (SIGCOMM'20) | C++, Python, Pytorch July 2020

- Developed live video streaming system based on Google WebRTC which involves; **1)** Online training and inference system for super-resolution DNN model during live video streaming. **2)** Bandwidth allocation algorithm to maximize user Quality of Experience (QoE).
- LiveNAS system delivers live video with the same quality as Google WebRTC **using only 45.9% bandwidth** on average or **enhances average QoE by 69%** compared to WebRTC using the same bandwidth.

TLT:Timeout-less Transport Protocol (EuroSys'21) | C++, Switch configuring March 2019

- Implemented data-center network protocols (TLT, PFC) in NS-3 network simulator. TLT protocol was also implemented in Linux kernel and tested in real-world testbed.
- TLT augments diverse datacenter transports, from widely-used (TCP, DCTCP, DCQCN) to state-of-the-art (IRN and HPCC), by achieving up to **81% lower tail latency**.

NAS:Deep Learning-based Internet Video Delivery (OSDI'18) | Python, Tensorflow, Google Cloud October 2018

- Developed an adaptive bit-rate streaming algorithm based on Reinforcement Learning (a3c) for NAS on top of MPEG-DASH—a video-on-demand system such as Youtube or Netflix.
- NAS enhanced the **average QoE by 43.08%** using the same bandwidth budget or **saving 17.13% of bandwidth** while providing the same user QoE compared to MPEG-DASH.

Technical Skills

Languages: Python, C++, Rust, SQL, shell scripting

Technologies/Frameworks: Pytorch, Tensorflow, SIMD vector processing, Linux, Git, Cloudstack, Docker, Spark, AWS

Publications

- TopFull: An Adaptive Top-Down Overload Control for SLO-Oriented Microservices** *ACM SIGCOMM 2024*
Jinwoo Park, Jaehyeong Park, Youngmok Jung, Hwijoon Lim, Hyunho Yeo, and Dongsu Han
- Generalizing deep variant callers via domain adaptation and semi-supervised learning** *Preprint 2023*
Youngmok Jung, Jinwoo Park, Hwijoon Lim, Jeong Seok Lee, Young Seok Ju, and Dongsu Han
- Co-optimizing for Flow Completion Time in Radio Access Network** *ACM CoNEXT 2022*
Jaehong Kim, Yunheon Lee, Hwijoon Lim, Youngmok Jung, Song Min Kim, and Dongsu Han
- Engorgio: Neural Video Enhancement at Scale** *ACM SIGCOMM 2022*
Hyunho Yeo, Hwijoon Lim, Jaehong Kim, Youngmok Jung, Juncheol Ye, and Dongsu Han
- BWA-MEME: BWA-MEM emulated with a machine learning approach** *Oxford Bioinformatics 2022*
Youngmok Jung and Dongsu Han
- Towards Timeout-less Transport in Commodity Datacenter Networks** *ACM EuroSys 2021*
Hwijoon Lim, Wei Bai, Yibo Zhu, Youngmok Jung, and Dongsu Han
- Enabling Neural-enhanced Video Streaming on Commodity Mobile Devices** *ACM MobiCom 2020*
Hyunho Yeo, Chan Ju Chong, Youngmok Jung, Juncheol Ye and Dongsu Han
- Neural-Enhanced Live Streaming: Improving Live Video Ingest via Online Learning** *ACM SIGCOMM 2020*
*Jaehong Kim, *Youngmok Jung, Hyunho Yeo, Juncheol Ye, and Dongsu Han
- Neural Adaptive Content-aware Internet Video Delivery** *USENIX OSDI 2018*
Hyunho Yeo, Youngmok Jung, Jaehong Kim, Jinwoo Shin, and Dongsu Han

Open-source

Main Contributor

Generalizing deep variant callers via domain adaptation and semi-supervised learning
<https://github.com/kaist-ina/RUN-DVC>

BWA-MEME: BWA-MEM emulated with a machine learning approach
<https://github.com/kaist-ina/BWA-MEME> *111 Github stars & 34K Conda and Docker*

Collaborative Projects

NeuroScaler: Neural Video Enhancement at Scale
<https://github.com/kaist-ina/neuroscaler-public>

Towards Timeout-less Transport in Commodity Datacenter Networks
<https://github.com/kaist-ina/ns3-tlt-rdma-public>

Enabling Neural-enhanced Video Streaming on Commodity Mobile Devices
<https://github.com/kaist-ina/nemo>

Neural Adaptive Content-aware Internet Video Delivery
https://github.com/kaist-ina/NAS_public