

Dr. Hyunho Yeo

 Website |  LinkedIn |  Google Scholar |  hyunho.yeo@kaist.ac.kr

WORK EXPERIENCE

Moloco Full-stack Machine Learning Engineer (ML-Infra Team)	Redwood City, CA, United States Jun 2023-
<ul style="list-style-type: none">• Infrastructure: Orchestrating cloud resources for our training and inference platforms using Terraform and Helm• Platform: Developing and maintaining custom training and inference platforms at a large scale, starting from scratch.• Modeling: Designing and implementing cutting-edge recommender models tailored for performance advertising	
Korea Advanced Institute of Science and Technology (KAIST) Graduate Research Assistant (INA Lab)	Daejeon, South Korea Feb 2017–Apr 2023
<ul style="list-style-type: none">• ML Networking: Developed a neural-enhanced adaptive streaming framework integrating super-resolution capabilities.• ML Acceleration: Increased the speed of super-resolution by utilizing temporal redundancy within video data.• ML Cost Optimization: Created an efficient and scalable inference engine designed for optimal performance at scale.	

EDUCATION

Korea Advanced Institute of Science and Technology (KAIST) Ph.D. in Electrical Engineering Thesis: Enabling Neural-enhanced Video Streaming Advisor: Dongsu Han	Feb 2017 - Apr 2023
Korea Advanced Institute of Science and Technology (KAIST) B.S. in Electrical Engineering (Magna Cum Laude)	Feb 2012 - Feb 2017

AWARDS

Google Conference Scholarship For the qualified graduate students who would attend selected conferences	2022
KAIST Breakthrough of the Year For the top 15 most significant research achievements	2021
KAIST Global Leader Scholarship For the graduate students with outstanding research achievements	2020
Microsoft Fellowship Asia Nomination Award For the top 25% graduate students among 101 highly competitive applicants from Asia universities	2019
KAIST EE Best Research Achievement For the graduate student with the best research achievement	2018

PUBLICATIONS

- (C1) AccelIR: Task-aware Image Compression for Accelerating Neural Restoration
Juncheol Ye, **Hyunho Yeo**, Jinwoo Park, and Dongsu Han
IEEE CVPR 2023 (Acceptance Rate: 25.7%; 2360/9155)
- (C2) NeuroScaler: Neural Video Enhancement at Scale
Hyunho Yeo, Hwijoon Lim, Jaehong Kim, Youngmok Jung, Juncheol Ye, and Dongsu Han
ACM SIGCOMM 2022 (Acceptance Rate: 19.5%; 55/281)
- (C3) NEMO: Enabling Neural-enhanced Video Streaming on Commodity Mobile Devices
Hyunho Yeo, Chan Ju Chong, Youngmok Jung, Juncheol Ye, and Dongsu Han
ACM MobiCom 2020 (Acceptance Rate: 16.1%; 62/384)
- (C4) Neural-Enhanced Live Streaming: Improving Live Video Ingest via Online Learning
Jaehong Kim*, Youngmok Jung*, **Hyunho Yeo**, Juncheol Ye, and Dongsu Han
ACM SIGCOMM 2020 (Acceptance Rate: 21.2%; 53/250)
- (C5) Neural Adaptive Content-aware Internet Video Delivery
Hyunho Yeo, Youngmok Jung, Jaehong Kim, Jinwoo Shin, and Dongsu Han
USENIX OSDI 2018 (Acceptance Rate: 18.2%; 47/257)
- (W1) Neural Cloud Storage: Innovative Cloud Storage Solution for Cold Video
Jinyeong Lim, Juncheol Ye, Jaehong Kim, Hwijoon Lim, **Hyunho Yeo**, and Dongsu Han
ACM HotStorage 2023
- (W2) SAND: A Storage Abstraction for Video-based Deep Learning
Uitaek Hong, Hwijoon Lim, **Hyunho Yeo**, Jinwoo Park, and Dongsu Han
ACM HotStorage 2023

(W3) How will Deep Learning Change Internet Video Delivery?

Hyunho Yeo, Sunghyun Do, and Dongsu Han

ACM HotNets 2017 (Acceptance Rate 28/124: 22.5%)

ISSUED PATENTS

(P1) Apparatus and method for accelerating super-resolution in real-time video streaming, 2022, US 11,399,201

(P2) Live video ingest system and method, 2022, US 2022/0368965

(P3) Server apparatus and method for content delivery based on content-aware neural network, 2020, US 10,560,731

PROJECTS

Neural-enhanced Live Video Ingest at Scale

Aug 2020 - Aug 2022

Graduate Research Assistant, KAIST

- Designed an inference engine that delivers efficient and scalable live neural enhancement.
- Implemented the end-to-end system on top of NVIDIA TensorRT, libvpx, and gRPC (~ **10.1K LoC**).
- Reduced computing cost by **3.0-22.3×** and improved processing throughput by **2.5-10×**.

Neural-enhanced Mobile Video Streaming

Nov 2018 - Jul 2020

Graduate Research Assistant, KAIST

- Designed an algorithm that accelerates neural enhancement using temporal redundancy across video frames.
- Implemented an end-to-end system upon Exoplayer, libvpx, and Qualcomm SNPE (~ **9.4K LoC**).
- Improved processing throughput by **11.5×** and reduced energy consumption by **88.6%**.
- Improved processing throughput by **11.5×** and reduced energy consumption by **88.6%**.

Neural-enhanced Adaptive Video Streaming

Mar 2017 - Oct 2018

Graduate Research Assistant, KAIST

- Designed adaptive streaming that applies neural enhancement to video utilizing client computation.
- Implemented an end-to-end system on top of MPEG DASH (dash.js) and TensorFlow (~ **13.6K LoC**).
- Improved user quality experience by **43.08%** or saved **17.13%** of network bandwidth.

INVITED TALKS

NeuroScaler: Neural Video Enhancement at Scale

- ACM SIGCOMM Conference, Amsterdam, the Netherlands

Aug 2022

NEMO: enabling neural-enhanced video streaming on commodity mobile device

- ACM MobiCom Conference, Virtual

Sep 2020

Neural Adaptive Content-aware Internet Video Delivery

- NVIDIA AI Conference, Seoul, South Korea
- USENIX OSDI Conference, Carlsbad, CA, USA

Oct 2020

Oct 2018

How will Deep Learning Change Internet Video Delivery?

- ACM HotNets Workshop, Palo Alto, CA, USA

Nov 2017

SERVICE

External reviewer: IEEE/ACM ToN, IEEE TPAMI, IEEE MM

2018-

SKILLS

Programming languages: C/C++, Python, JAVA

ML frameworks: Tensorflow, Pytorch, TensorRT, SNPE

Languages: Korean (native), English (fluent)

REFERENCES

Available upon request.