Jin Heo

jinheo44@gmail.com • https://www.linkedin.com/in/jinheo4 • https://jheo4.github.io

Summary

- Senior Researcher at Dolby and PhD in Computer Science (Georgia Tech) with a systems background across the full stack, from PhD research on real-time inference serving and adaptive rendering for XR and Cloud Gaming, to production engineering of cloud infrastructure, MLOps pipelines, and high-throughput data platforms at Dolby.
- Core areas: ML Inference Serving & Scheduling, Distributed & Edge Systems, Cloud Infrastructure (AWS, Kubernetes, Terraform), MLOps Orchestration, High-Throughput Data Pipelines (Kafka, Databricks)

Experience

Dolby Laboratories — Atlanta, GA
Senior Researcher — Aug 2025 - Present

- Led technical evaluation and deployment of ML workflow orchestrators (Metaflow, Argo, Flyte) on AWS; provisioned secure infrastructure via Terraform and Helm, enforcing network isolation and endpoint protection policies.
- Researching a novel inference serving system for real-time Streaming Perception on resource-constrained edge servers; developing an effectiveness-driven scheduling mechanism that dynamically estimates temporal utility of inference results to optimize multi-user serving allocation.
- Engineered resilient live-streaming telemetry pipelines (CMCD/CMSD) for Dolby Optiview's NFL streaming into Databricks, sustaining ~1M requests within 15 minutes with zero data loss via optimized ingestion windows and automated failure recovery workflows.

Georgia Institute of Technology — Atlanta, GA
Graduate Research Assistant — Aug 2019 - Aug 2025

- Developed Stimpack, an adaptive rendering optimization system that balances server-side rendering costs against user-perceived quality for scalable multi-user cloud gaming, achieving up to 24% higher service quality and 2x user capacity with the same resources; published at NSDI 2026 [1].
- Designed a low-latency, soft real-time inference serving system for XR perception models, serving more users with the same resource footprint on the resource-constrained server [2].
- Built FleXR, a distributed stream processing system enabling flexible distribution of XR workloads across deployment scenarios, achieving up to 50% lower end-to-end latency and 3.9x pipeline throughput compared to alternatives [4, 7].

Dolby Laboratories — San Francisco, CA
Research Intern — May 2024 - Aug 2024

- Developed a parser for the Universal Scene Description (USD) standard of 3D assets to facilitate real-time content handling.
- Designed a partial query and data retrieval methodology to enable efficient, low-latency streaming of interactable 3D graphics following the USD standard.

AT&T Labs — Austin, TX
Research Intern — May 2023 - Aug 2023

- Developed a fast-prototyping framework for city-scale geospatial digital twins in Unity 3D to simulate and analyze the wireless network performance of AT&T base stations [3].

Ericsson Research — Santa Clara, CA
Research Intern (remote, part-time) — Mar 2021 - Dec 2022

- Developed FLiCR, a lightweight LiDAR point cloud compression method for edge-assisted real-time perception, achieving 12x compression ratio with up to 80% lower end-to-end latency compared to state-of-the-art alternatives [5].
- Developed an interpolation method to recover geometric information lost due to compression, resulting in an international patent application (WO2024073084A1) [6].

University of California, Irvine — Irvine, CA
Research Assistant (Undergraduate Research Opportunities Program) — Jul 2017 - Jan 2018

- Designed and implemented an FPGA acceleration framework for computer vision algorithms using OpenVX graph pipelines [8].

CSIRO (Commonwealth Scientific and Industrial Research Organization) — Brisbane, Australia
Research Assistant (Undergraduate Research Opportunities Program) — Sep 2016 - Feb 2017

- Developed and deployed an optimized network flooding algorithm for efficient broadcasting in Wireless Sensor Networks (WSNs), validating performance on battery-constrained sensor devices in a field environment.

Education

Georgia Institute of Technology — Atlanta, GA
Doctor of Philosophy in Computer Science; GPA: 4.0 — 2019 - 2025

- Thesis: "Adaptively Serving XR Workloads from Resource-constrained Edge"

Ajou University — Suwon, South Korea
Bachelor of Science in Computer Engineering; Summa Cum Laude — 2016 - 2018

Publications

1. **Heo, J.**, Wang, V., Bhardwaj, K., & Gavrilovska, A. (2026). *Stimpack: An Adaptive Rendering Optimization System for Scalable Cloud Gaming*. In 23rd USENIX Symposium on Networked Systems Design and Implementation (NSDI 26) (pp. 397-414).
2. **Heo, J.** and Gavrilovska, A., 2024, December. Poster: *Adapting XR Perception Serving for Edge Server Scalability*. In 2024 IEEE/ACM Symposium on Edge Computing (SEC) (pp. 518-520). IEEE.
3. **Heo, J.**, Novlan, T., Akoum, S. and Gavrilovska, A., 2024, December. *GT-Craft: A Framework for Fast Prototyping Geospatial-Based Digital Twins in Unity 3D*. In 2024 IEEE/ACM Symposium on Edge Computing (SEC) (pp. 395-401). IEEE.
4. **Heo, J.**, Bhardwaj, K. and Gavrilovska, A., 2023, June. *FleXR: A system enabling flexibly distributed extended reality*. In Proceedings of the 14th Conference on ACM Multimedia Systems (pp. 1-13).
5. **Heo, J.**, Phillips, C. and Gavrilovska, A., 2022, December. *FLiCR: A fast and lightweight lidar point cloud compression based on lossy RI*. In 2022 IEEE/ACM 7th Symposium on Edge Computing (SEC) (pp. 54-67). IEEE.
6. **Heo, J.**, Phillips, G., Brodin, P.E. and Gavrilovska, A., 2022, December. Poster: *Making Edge-assisted LiDAR Perceptions Robust to Lossy Point Cloud Compression*. In 2022 IEEE/ACM 7th Symposium on Edge Computing (SEC) (pp. 293-295). IEEE.
7. **Heo, J.**, Bhardwaj, K. and Gavrilovska, A., 2021, December. Poster: *Enabling flexible edge-assisted XR*. In 2021 IEEE/ACM Symposium on Edge Computing (SEC) (pp. 465-467). IEEE.
8. Taheri, S., **Heo, J.**, Behnam, P., Chen, J., Veidenbaum, A. and Nicolau, A., 2018, April. *Acceleration framework for FPGA implementation of OpenVX graph pipelines*. In 2018 IEEE 26th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM) (pp. 227-227). IEEE.

Skills

- Programming Languages & Systems: C, C++, CMake, Makefiles, Python, Linux, Git
- Distributed Systems & Infrastructure: AWS, GCP, Kubernetes, Terraform, Helm, Flyte, Metaflow, Docker, gRPC, Databricks, Apache Spark, Apache Kafka, ZMQ
- Multimedia, ML & Scientific Computing: Universal Scene Description (USD), Unity, Unreal Engine, FFmpeg, GStreamer, OpenCV, OpenGL, PyTorch, Scikit-learn, Point Cloud Library (PCL)

Activities

- Academic Service: Program Committee (PC), ACM SIGOPS Annual Technical Conference (ATC) 2026, External Reviewer, IEEE OJ-COMS 2024.
- Teaching: Head Teaching Assistant, Advanced Operating Systems (CS6210/4210) at Georgia Tech (Spring 2022, Spring 2024).
- Open-Source Contributions: RaftLib (resolved pipeline scheduler bottlenecks for streaming efficiency); uvgRTP (implemented Linux build system support via pkg-config).