# JUNG HWAN (JOHN) HEO

jhjohnheo@gmail.com | https://johnheo.github.io | US Permanent Resident - Pasadena, CA

## EDUCATION

**University of Southern California, Viterbi School of Engineering (GPA 3.82/4.00)** Los Angeles, CA
*B.S., Computer Engineering & Computer Science* *Aug 2019 - May 2023*

## EXPERIENCE

**Korean Augmentation to the US Army (KATUSA)** *Osan Air Base,* S. Korea
*Human Resource Specialist* *Oct 2023 - Present*
Supervisors: Colonel Kevin P. Stonerook and Command Sergeant Major Antonio Golden

- As the dedicated Command Group KATUSA to the 35th Air Defense Artillery Brigade, aid US-Korea combined forces for a combat-ready army via translation missions and administrative maintenance

**Efficient AI Team, NAVER Cloud** *Gyeonggi,* S. Korea
*Research Intern* *May 2023 - Sep 2023*
Advisors: Dr. Dongsoo Lee and Dr. Sejung Kwon

- Conduct research in weight quantization for Large Language Models
- Develop AdaDim, a technique to adaptively switch the quantization dimension per layer, increasing MMLU by 4.7% and HumanEval coding by 10% with 1.53x lossless 3-bit inference acceleration

**SPORT Lab, Dept. of Electrical and Computer Engineering, USC** Los Angeles, CA
*Provost's Undergraduate Research Fellow* *Aug 2021 - May 2023*
Advisor: Dr. Massoud Pedram

- Conduct research in novel computer vision architectures for efficient deep learning
- **Efficient Transformers**: Apply token reduction techniques with spatial locality bias to accelerate off-the-shelf Vision Transformers even *without* training
- **Software-Hardware Co-design:** Lead the development of indexing-free convolutional neural network dataflow and compiler optimizations for designing an energy-efficient ML accelerator

**HAN Lab, Dept. of Electrical Engineering and Computer Science, MIT** Cambridge, MA
*Research Intern* *May 2022 - Nov 2022*
Advisor: Dr. Song Han

- Conduct research in classical ML and model compression methods such as weight pruning
- **ML with Signal Processing:** Led the development of a statistical sampling-based algorithm with Bayesian search to realize ultra-efficient heart stroke detection on commercial microcontrollers
- **Pruning Algorithms:** Study convex optimization methods for alternating direction method of multipliers (ADMM) and validate its efficacy on channel pruning for efficient deep learning

**BMM Lab, Dept. of Electrical Engineering and Computer Science, KAIST** Daejeon, S. Korea
*Visiting Research Fellow* *May 2020 - Aug 2020*
Advisor: Dr. Hyunjoo (Jenny) Lee

- Conduct research in ML-based methods for sensitivity optimization of neurochemical sensors
- Designed an objective function with electrochemical features and implemented evolutionary algorithms to optimize FSCV waveforms, improving adsorption efficiency by 11%

## PROJECTS

**Semantic Deduplication for Data-Efficient Learning** *Spring 2023*

- Developed a semantic deduplication method using k-means clustering to identify and remove redundant data samples
- Analyzed the efficiency tradeoff between data storage requirements and training efficiency

**Vision Transformer Acceleration with CUDA** *Nov 2022 - Dec 2022*

- Write CUDA kernels to accelerate Vision Transformer inference by 223x compared to CPU baseline
- Optimize parallel matrix multiplications using shared memory and linear attention

**Neurologue.ai** *Aug 2020 - May 2021*

- Led a team of 11 members at USC MEDesign to develop a multi-modal dementia detection model combining Speech Recognition and Natural Language Processing
- Source NIH-approved 200+ patient speech dataset, DementiaBank, and train Decision Trees, XGBoost, DNN, achieving up to 78% accuracy

## SERVICE and LEADERSHIP

**Reviewer: ICLR 2025 and WACV 2025**

**Ming Hsieh Institute Undergraduate Scholar, *USC ECE Dept.*** *Sep 2022 - May 2023*

- Selected as one of four scholars in the ECE department to serve as a liaison between faculty and undergraduate body to cultivate a research hub through a bi-weekly seminar series

**Curriculum Lead, *USC Center for Artificial Intelligence in Society (CAIS++)*** *June 2022 - May 2023*

- Develop and teach a 10-week AI curriculum that covers SVM, Neural Networks, CNN, Transfer Learning, etc.

**Co-founder and Curriculum Lead, Shift SC** *June 2021 - May 2023*

- Develop a 10-week AI Ethics curriculum covering the social implications of AI tools in education, media censorship, healthcare, and more
- Teach the self-developed curriculum to 150+ students as part of the Ethics in Engineering class

## PUBLICATIONS (* Denotes Equal Contribution)

[1] **Jung Hwan Heo***, Jeonghoon Kim*, Beomseok Kwon, Byongwook Kim, Sejung Kwon, and Dongsoo Lee
*Rethinking Channel Dimensions to Isolate Outliers for Low-bit Weight Quantization of Large Language Models*
International Conference on Learning Representations (ICLR), 2024.

[2] **Jung Hwan Heo**, Seyedarmin Azizi, Arash Fayyazi, and Massoud Pedram
*Training-Free Acceleration of ViTs with Delayed Spatial Merging*
International Conference on Machine Learning (ICML) Workshop on Efficient Systems for Foundation Models (ES-FoMo) II, 2024.

[3] **Jung Hwan Heo***, Arash Fayyazi*, Amirhossein Esmaili, and Massoud Pedram
*Sparse Periodic Systolic Dataflow for Lowering Latency and Power Dissipation of Convolutional Neural Network Accelerators*
International Symposium on Low Power Electronics and Design (ISLPED), 2022. Oral Presentation.

[4] **Jung Hwan Heo**
*Ethical Review in the Age of Artificial Intelligence*
AI Robotics Ethics Society (AIRES) Conference and Artificial Intelligence Ethics Journal Vol. II, 2021.

[5] **Jung Hwan Heo**, Andrew Kyung
*A Study on Functionalized Cancer Scanning Contrast Agents in Positron Emission Tomography (PET)*
Bulletin of the American Physical Society (APS), 2018.

## PREPRINTS and CONTESTS (* Denotes Equal Contribution)

[1] **Jung Hwan Heo***, Seyedarmin Azizi*, Arash Fayyazi, and Massoud Pedram
*CrAFT: Compression-Aware Fine-Tuning for Efficient Visual Task Adaptation*
ArXiv, 2023.

[2] Hanrui Wang*, **Jung Hwan Heo***, Wei-Chen Wang, Jessica Zheng, Ji Lin, Han Cai, and Song Han
*Efficient Heart Stroke Detection on Low-cost Microcontrollers*
TinyML Design Contest at the International Conference on Computer-Aided Design (ICCAD), 2022.

## AWARDS

- **Recipient:** General Paik Sun Yup Leadership Board Award — Sep 2024
  - Top 0.1% of soldiers per year for leadership, fitness, and military knowledge
- **Recipient:** Army Commendation Medal — Sep 2024
  - The highest US military decoration given to soldiers in the KATUSA program
- **Recipient:** USC Viterbi Award for Outstanding Research — May 2023
  - One graduating engineering student per year for research impact
- **Recipient:** USC Discovery Scholar — May 2023
  - Ten graduating university students per year for research impact
- **Recipient:** The Order of Arête, USC (Highest service honors) — May 2023
- **Recipient:** USC Provost Undergraduate Research Fellowship — Jan 2023
- **Honorable Mention:** CRA Outstanding Researcher Award — Dec 2022
- **Recipient:** USC Provost Undergraduate Research Fellowship — Nov 2022
- **First Place:** ICCAD TinyML Contest, Flash Occupation Track — Oct 2022
- **Recipient:** USC Ming Hsieh Institute Scholar — Sep 2022
- **Recipient:** USC Provost Undergraduate Research Fellowship — Aug 2022
- **Recipient:** USC Provost Undergraduate Research Fellowship — May 2022
- **Recipient:** USC Provost Undergraduate Research Fellowship — Aug 2021
- **Admitted:** Accelerated Mathematics **(**Graduate Linear Algebra & Probability) — Aug 2021
- **Recipient:** USC CURVE Research Fellowship — Aug 2021
- **First Place:** IEEE@USC IoT Hackathon — May 2021
- **Recipient:** Interdisciplinary prize, 23rd Annual USC Undergraduate Symposium — May 2021
- **Qualified:** USA Computing Olympiad, Gold Division — March 2020
- **Recipient:** USC Trustee Scholarship — Mar 2019
  - Full tuition merit-based scholarship (top 0.15% of the class of 2023)

## SKILLS and COURSEWORK

- **Programming Languages**: Python, C, C++, CUDA, OpenMP, MATLAB
- **Frameworks**: Pytorch, Hugging Face, Weights and Biases
- **Coursework:**
  - *Mathematics:* Graduate Linear Algebra; Graduate Probability Theory; Vector Calculus
  - *Machine Learning:* Machine Learning Theory; Machine Learning for Electrical Engineers
  - *Computer Science:* Data Structures; Theory of Algorithms; Discrete Mathematics
  - *Electrical Engineering*: Parallel and Distributed Computation; Computer Architecture