

Hyunsu Ye

✉ dpgustn1@unist.ac.kr ☎ (+82)010-7665-4342

Research Interest

AI Accelerator, HW/SW Co-simulation

Education

BS	UNIST , B.S. Computer Science and Engineering	Feb. 2020 – Feb. 2026
	• GPA: 4.11/4.3, Major 4.16/4.3	
MS	KAIST , M.S. School of Computing	Mar. 2026 – Present

- To Be Updated..

Research Experience

UVLL, UNIST Vision & Learning Lab , UNIST	Jun. 2023 – Dec. 2023
• Undergraduate Research Intern	
• Advisor : Prof. Seungryul Baek	
• Focus : <u>ICCV 2023 Hand Challenge</u>	
• Contribution :	
– ICCV 2023 Hand Challenge, 3rd place as team UVHAND (Role : Experiments and Visualization)	
– Best Undergraduate Paper Award(Top 10) in IEIE	
3D Vision & Robotics Lab , UNIST	Feb. 2024 – Dec. 2024
• Undergraduate Research Intern	
• Advisor : Prof. Kyungdon Joo	
• Focus : Event Camera Vision, SfM pipeline study, 3D Computer Vision Study, AICP project	
CASYS , KAIST	Jan. 2025 – August. 2025
• Undergraduate Research Intern	
• Advisor : Prof. Youngjin Kwon	
• Focus: LLM Serving System with Test Time Compute Search, Speculative Decoding on VLM	

Last

UIRP, UNIST (Apr. 2023 – Nov. 2023)

- UNIST Undergraduate Interdisciplinary Research Project
- **Topic** : Real-time Egocentric 3D Hand Pose Estimation via Instance Activation Map
- Role : Idea building & Implementation, Experiments
- Associated with ICCV2023 hand challenge

AICP, UNIST (Apr. 2024 – Nov. 2024)

- UNIST AI Challengers Program
- **Topic** : Vanishing Point Guided Monocular Depth Estimation for Event Camera
- Role : Leader of 3D_Luv Team

LLM Serving System Research, KAIST (Jan. 2025 - May. 2025)

- **Topic** : LLM Serving System for Test Time Compute Search

- **Role:** Designed pipeline to efficiently determine beam width and size during serving-time beam search.
- **Summary:**
 - Developed an "**Adaptive Beam-Search Pipeline**" that dynamically adjusts beam parameters based on PRM scores, allocating more compute to harder queries and scaling back for easier ones.
 - Extended traditional PRM-thresholding by integrating system overhead metrics as additional decision signals.
 - Implemented a **Contextual-Bandits Algorithm (Reinforcement Learning)** to balance accuracy and latency, by selecting discrete beam configurations per context for a high-performance serving system.
 - By integrating Experimental results, we submitted it to an **International Systems Conference.**

Accelerating VLM with Speculative Decoding, KAIST (May. 2025 - Aug. 2025)

- **Topic :** Lossless Acceleration of VLM Inference with DB Drafting
- **Summary:**
 - Proposed **VSPD** pipeline (**Vision SPeculative D**ecoding): A plug-and-play DB-drafting framework for VLMs that uses high-attention image embedding windows as keys to retrieve future tokens without any additional model training(Lossless).
 - **Intuition** : token generations which highly related with image (e.g. "dog") will have high attention values on corresponding image patch's embeddings.
 - Implemented attention-based DB pipeline integrated with LLaVA-Onevision 7B and achieve average accept length of **1.8 tokens per forward pass.**

Skills

- **Programming Language :** Python, C++, C, Scala, Verilog
- **Frameworks :** Pytorch, CUDA
- **Language :** Korean(mother tongue), English (fluent)

Works in Progress

- **LLM Serving Simulation Framework**

Awards and Achievements

- **Best Undergraduate Paper Award**, (Host: IEIE, The Institute of Electronics and Information Engineers) Nov. 2023
- **3rd place in ICCV2023 Hand Challenge** Sep. 2023
- **Superior semester grades, UNIST**
 - 1-1 Semester (3.92/4.3) Aug. 2020
 - 1-2 Semester (4.12/4.3) Feb. 2023
 - 2-2 Semester (4.15/4.3) Mar. 2024
 - 3-1 Semester (4.08/4.3) Aug. 2024
 - 3-2 Semester (4.24/4.3) Feb. 2025

Scholarship

- **Academic Performance Scholarship, UNIST** Whole semesters
- **WoonHae Scholarship Foundation(WHF), 2024, 11th Scholarship Student** Jan. 2024 - Feb. 2025
- **Seoul Scholarship Foundation, Seoul Hope University Career Scholarship Student** Sep. 2024 - Dec. 2024
- **Korea Scholarship Foundation, National Excellence Scholarship(Science and Engineering) Student** Mar. 2024 - Dec. 2024
- **WoonHae Scholarship Foundation(WHF), 2025, 12th Scholarship Student** Jan. 2025 - Dec. 2025

Other Experiences

Development of BAPU

- The Largest Use Platform on UNIST Campus
- Used Tool: Spring Boot framework, Flutter
- Role : Backend Development
- Launched in **APP Store & Play Store** and service a **Table of Meals in UNIST**

BAPU application URL [↗](#)

Sea Logistics Start-up Audition, 6th

Jun. 2023 - Oct. 2023

- Shipper service that improves JIT (Just In Time) by predicting Expected Time Arrival (ETA) at sea
- Role : Leader of ControlPort Team

Teaching Assistant (TA)

- Teaching Assistant(TA) of Probability and Random Process
- Teaching Assistant(TA) of Artificial Intelligence Programming I

Mar. 2023 - Dec. 2024

2023, 2024 Spring
semesters

ICT Piwooda Project : Development Contest

Sep. 2023 - Nov. 2023

- Development of public services in the Seungdong city
- Advance to the finals (18 teams among 122 teams, **14.75%**)

Student Council

Sep. 2022 - Dec. 2024

- A Festival Preparatory Committee
- "Yuja" App Manager, Issue Tracker