Foundations of Edge AI

Final Project

Lanyu (Lori) Xu

Email: lxu@oakland.edu
Homepage: https://lori930.github.io/
Office: EC 524

Final Project

- Find your project idea
- Form your team
- Work on your project
- Present your work
- Submit your accomplishments

Find Your Project Idea

- Choose a topic relevant to the course material, specifically, enable Edge AI
  - We'll talk about it.
- Ensure your idea is feasible within the time frame and resources available

Form Your Team

- Group up with your classmates. 2-3 members per team
  - I can help to form teams after seeing your proposal.
- Group Selection Spreadsheet

Work On Your Project

- Research and Implement your solution(s)
- Document your progress and challenges
- Prepare for the presentation

Present Your Work (15 mins / group)

Dec 5th, 2024 @ 3PM

- Showcase your project, focusing on your problem, solution, and results
- Highlight your key takeaways and any future improvements
- Demo video is highly encouraged
- Open source is highly encouraged

Submit Your Accomplishments

Dec 14, 2024 @ 11:59 PM

- All submissions must be in PDF format
  - Presentation slides
  - Written report, 5 page at least
    - You must format your submission using the ACM template: https://www.acm.org/publications/proceedings-template
- Grading rubric
  - Motivation, technical soundness, novelty, evaluation, oral presentation, written report, slides
  - 10 pts each

Need Inspirations?

- **Topics we already covered**
  - Pruning
  - Quantization
- **Topics we are going to cover**
  - Neural Architecture Search (Oct 22-24)
  - Knowledge Distillation (Oct 29)
  - TinyML on Microcontrollers (Oct 31 - Nov 5)
  - Transformer and LLM (Nov 7 - Nov 12)
  - Vision Transformer (Nov 14)
  - GAN, Video, and Point Cloud (Nov 19)
  - Diffusion Model (Nov 21)

Technique deployment

Paper reproduction

Practical application

Need More Inspirations?

- **Model optimization and compression**
  - Classic models / New fancy models
  - Focusing on improving efficiency and scalability of NN models through techniques like pruning, quantization, and optimization
- **Real-time applications and privacy**
  - Have some fun / practical scenarios where Edge AI will renovate the current situation?
  - Develop solutions for real-time processing tasks, including privacy-enhancing techniques and responsive systems for live applications
  - Create tools that improve user interaction, accessibility, or daily assistance
- **Novel architecture and techniques**
  - Push the boundaries of model architectures, including new approaches to efficiently training and using AI in various context

Aim High, Stay Grounded

https://www.instagram.com/p/C5lT8SdLQsP/?igsh=bzBoZzV4bDVuamc2