ECE 499-001: Machine Learning for Embedded Systems

Fall 2021

Instructor: Weiwen Jiang <wjiang8@gmu> (412)427-0695

Class Meeting: Lectures – Monday 4:30 pm – 7:10 pm

Location: Horizon Hall (HORIZON) 2010

Course Materials: Course materials will be posted before or after the class. No formal textbook is required. The book from http://www.deeplearningbook.org/ will be referred to on the course.

Prerequisites: CS 222 and ECE 231 and ECE 350 with the minimum grade of C

The course topics are self-contained so that a background in machine learning is not required. Students should be familiar with programming and embedded systems to complete the course projects.

Course Description: Machine learning (ML) has gradually become the core component of wide applications in different computing scenarios, ranging from edge computing to cloud computing. This course focuses on resource-constrained edge computing, in particular the embedded systems, and introduces techniques for developing energy/time efficient ML algorithms and models for the embedded systems. Topics that are covered include (i) commonly used ML algorithms, (ii) ML model compression techniques, (iii) hardware-aware machine learning, (iv) hardware and neural architecture co-design. The course also provides a comprehensive team-based development experience through projects. Offered by Electrical & Comp. Engineering. May not be repeated for credit.

Course Schedule: there are three sections of the course, including:

- Section I: Introduction of Machine Learning and Deep Neural Networks
- Section II: Automated Neural Network Design
- Section III: Optimization of both ML/DNN and Hardware Design

SECTION I: Introduction of Machine Learning and Deep Neural Networks

Date	Topic
Week 1	Course Information & Introduction to Machine Learning
Week 2	Train Neural Networks
Week 3	Deep Convolutional Neural Networks (CNN)
Week 4	Natural Langue Processing

Week 5	Reinforcement Learning

SECTION II: Automated Neural Network Design

Date	Topic
Week 6	ML Accelerator Design (1)
Week 7	ML Accelerator Design (2)
Week 8	Model Compression
Week 9	Neural Architecture Search (1)
Week 10	Neural Architecture Search (2)

SECTION III: Optimization of both ML/DNN and Hardware Design

Date	Topic
Week 11	Hardware-Aware Neural Architecture Search
Week 12	HW/SW Co-Design with Neural Architecture Search (1)
Week 13	HW/SW Co-Design with Neural Architecture Search (2)
Week 14	Course Project Demonstration

* The schedule might change during the semester depending on the progress of the class.

Goals and Outcomes:

- Understand the basic training and inference techniques of a neural network. (Section I)
- Get familiar with commonly used neural networks, such as CNN and RNN (Section I)
- Be able to implement neural networks using machine learning tools (Section I)
- Understand hardware accelerator design for neural networks, such as FPGA (Section II)
- Get familiar with the neural network architecture search (Section II)
- Be able to apply compression techniques, i.e., pruning and quantization (Section II)
- Understand the co-design of neural networks and hardware accelerators (Section III)

Homework Labs: There are 5 take-home labs assigned to the students to practice basic skills of machine learning implementation using TensorFlow or Pytorch. A total of 7 days after the due dates are permitted for all assignments. However, after the due date, each assignment will be deducted 10 points for each day late.

Project: Students will form teams to implement one open-topic project on mobile devices. Each team will be assigned one project, and the demonstration of the completed projects will be presented at the end of the course.

Grading:

- Homework assignments (5 take-home labs, 15%) 50%
- Paper Critiques 10%
- Project progress review (team based) 10%
- Project demonstration (team based) 30%

Course and University Policies:

- Email Communications: Students must use their MasonLive email account to receive important University information, including messages related to this class. See http://masonlive.gmu.edu for more information. Homework assignments and other course material will be emailed to your MasonLive email account. Also, when you send me an email, please write ECE499-001 on the subject line.
- Avoid Reposting Course Material: It is not allowed to reposting course material. The course materials (lecture notes, homework, projects, exams, solutions, and anything else posted on the course website) are copyrighted. You may not upload them to any other website or share them with any on-line or off-line test bank.
- Honesty and Integrity: Mason expects students to pursue their academic work with honesty and integrity. Students should feel free to work in groups to discuss lecture material and homework assignments; however, under no circumstance should a student represent another's work as his or her own. Copying solutions for assigned homework problems, from any source, constitutes a violation of the university honor code. Any form of cheating may cause penalties, from getting an F in this course to academic actions in accordance with university policy.
- Office of disability services: Mason provides accommodations through the Office of Disability Services (ODS) http://ods.gmu.edu. If you are a student with a disability and you need academic accommodations, please see me and contact ODS at 993-2474.
- University Policies: The University Catalog, http://catalog.gmu.edu, is the central resource for university policies affecting student, faculty, and staff conduct in university academic affairs. Other policies are available at http://universitypolicy.gmu.edu/. All members of the university community are responsible for knowing and following established policies.