Course Syllabus

ECE/CS528: Embedded Systems and Machine Learning

Instructor: Dr. Sudeep Pasricha (sudeep@colostate.edu), ENGR B119

Lectures: Tu/Thu 12:30pm – 1:45pm, TILT 221 (online recordings available via Echo360)

Office Hours: 9am – 10:30am, Fridays

Course TA: Febin Sunny (Febin.Sunny@colostate.edu)

TA Hours: Mon/Thu 2pm – 4pm (in-person, ENGR C1) or Zoom (contact TA to schedule time)

Course Description: Machine learning is becoming pervasive in embedded computing platforms, such as smart mobile systems, wearable IoT devices, and autonomous vehicles. This course will present recent advances towards the goal of enabling efficient implementation of deep machine learning models on embedded systems. Specifically, it will provide an overview of 1) the theoretical foundations and motivations behind various deep learning models, 2) software modeling and optimization techniques for these models, 3) hardware platforms and architectures to support efficient execution of machine learning models, and 4) hardware-software co-design approaches for machine learning. The course will cover emerging machine learning models, custom hardware accelerators, as well as paradigms including processing-in-memory, memristors, and photonics for machine learning. The course is very topical and relevant for graduate and senior undergraduate students in computer engineering, computer science, data science, and electrical engineering, as well as practitioners in industry and students in other engineering departments who are interested in data engineering, machine learning, and embedded systems. The course does not assume prior expertise in machine learning; **however knowledge of Python is essential**. Some background in computer architecture (ECE452 or similar) would also be helpful. Students will get hands-on experience in the design and optimization of deep machine learning models, as well as the analysis of these models on emerging hardware platforms. This is a unique course that is well suited for beginners and experts alike, who want to comprehend the state-of-the-art in deep machine learning software and hardware design, and understand future trends and opportunities in this exciting field.

Course Learning Objectives (CLO): Students successfully completing this course will be able to

1. Understand and apply the key design considerations for efficient processing of machine learning models on embedded systems;
2. Analyze tradeoffs between various embedded software model architectures and optimizations;
3. Optimize micro-architectural knobs such as precision, data reuse, and parallelism to architect embedded machine learning accelerators given target area-power-performance metrics;
4. Evaluate the utility of various memory and network centric design techniques for efficient machine learning processing across different embedded system platforms;
5. Demonstrate skills in hardware/software co-design of machine learning solutions for embedded platforms
6. Understand basics of unsupervised learning and its applications to embedded systems
7. Analyze challenges with security and adversarial examples for machine learning
8. Understand future trends and opportunities from machine learning algorithms, hardware accelerators, and emerging technologies in embedded systems

Prerequisites: ECE251 or CS270 or equivalent computer organization course; programming experience in Python

Textbook: None. The course will cover materials from various books and conference/journal articles.

Syllabus: Here is a tentative outline and syllabus for this course.

Specific Course Topics/Units/Weekly Schedule:

Week	Lecture Content	CLO Number
1	Intro to Machine Learning (ML) and Embedded Computing	1
2	Intro to Machine Learning (ML) and Embedded Computing	1
3	Foundations of Deep Neural Networks (DNNs/CNNs)	1, 2
4	Foundations of Deep Neural Networks (DNNs/CNNs)	1, 2
5	Software Optimizations for Embedded ML	2
6	Software Optimizations for Embedded ML	2
7	Hardware Acceleration for Embedded ML	3, 4
8	Hardware Acceleration for Embedded ML	3, 4
8	HW/SW Codesign for Embedded Systems	2, 3, 4, 5
9	Sequence and Time Series Embedded Processing	6
10	Unsupervised Learning Foundations	6
11	Anomaly Detection and Security in Embedded Systems	6, 7
12	Autoencoders and GANs for Generative Modeling	6, 7, 8
13	Advanced Embedded ML Models	8
14	Fall Recess	
15	Emerging Directions and Applications in Embedded ML	8
16	Finals Week	1, 2, 3, 4, 5, 6, 7, 8

Grading: The final grade will be on a curve. Grading is based on the following components:

- Homework/Lab Assignments (5): 30%
- Reading Assignments (8): 20%
- Quizzes (2): 10%
- Class Participation: 10%
- Final Project: 30%
    - Final presentation: 10%
    - Project report: 20%

Grading Scale:

>95%	90-94%	85-89%	80-84%	75-79%	70-74%	65-69%	55-64%	40-55%	<40%
A+	A	A-	B+	B	B-	C+	C	D	F

Lecture Attendance: If you are signed up for sections 001 or 002, you should attend the lectures in-person on your designated day (I expect that you attend almost all of the lectures in-person). Online 801 section students and 001/002 section students (on the day they cannot attend the in-person lecture) can access livestream videos of the lectures during the lecture hours on Tu/Thu, as well as recorded videos of the lectures via the Echo360 page on Canvas. Unfortunately, questions from students watching the livestreams cannot be answered in real-time. It is recommended that you note your questions and ask them during TA office hours. Try to watch the lecture videos within 24 hours of posting, as sometimes there is time critical information (e.g., related to quizzes, assignment deadlines, class project) discussed in the lectures.

Participation: Participation includes two components: 1) in-class engagement, and 2) Canvas discussion forum posts. I expect all students to ask and respond to questions in class, if you are attending lectures in-person. I expect all students to also post/respond thoughtfully and meaningfully ~10-15 times throughout the semester, to discussion forum posts that I create or those created by students (posts related to help with assignments do not count). Online (801) students must demonstrate participation only via the discussion forum.

Assignments: Homework assignments will involve working with Python and Tensorflow/Keras, as well as tools for embedded platform exploration. Reading assignments will involve reading technical research papers and summarizing their key contributions and a critique in around 500 words.

Submission Policy: Homework and reading assignments will be assigned throughout the semester. You are allowed late submission up to 3 days on one homework and one reading assignment. You can also skip one reading assignment of your choice, without impacting your grade. Otherwise all homework and reading assignments should be submitted before the deadline via Canvas, and ***late submissions will not be graded!***

Re-grading Policy: Re-grading requests should be made within a week from the date of the graded item (homework, exam, or project) becoming available.

Academic Integrity: All submitted work should be your own. Copying of language, structure, images, ideas, or thoughts of another, and representing them as one's own without proper acknowledgement (from github code repos, other web sites, books, papers, other students, etc) and failure to cite sources properly is not acceptable. Sources must always be appropriately referenced, whether the source is printed, electronic, or spoken. Minor first infraction in HWs and presentations will lead to a zero score + one letter level (e.g. A to B) reduction in course grade. Project or Major or repeated infractions in HWs and presentations will result in "F" grade for the course + report to Dean's Office. For more information see CSU's Academic Integrity Policy: https://tilt.colostate.edu/AcademicHI (Links to an external site.) and Student Conduct Code: https://resolutioncenter.colostate.edu/student-conduct-code/ (Links to an external site.)

COVID Reporting: All students are expected and required to report any COVID-19 symptoms to the university immediately, as well as exposures or positive tests from a non-CSU testing location. If you suspect you have symptoms, or if you know you have been exposed to a positive person or have tested positive for COVID, you are required to fill out the COVID Reporter (https://covid.colostate.edu/reporter/ (Links to an external site.)). If you know or believe you have been exposed, including living with someone known to be COVID positive, or are symptomatic, it is important for the health of yourself and others that you complete the online COVID Reporter. Do not ask your instructor to report for you. If you do not have internet access to fill out the online COVID-19 Reporter, please call (970) 491-4600. You may also report concerns in your academic or living spaces regarding COVID exposures through the COVID Reporter. You will not be penalized in any way for reporting. When you complete the COVID Reporter for any reason, the CSU Public Health office is notified. Once notified, that office will contact you and, depending upon each situation, will conduct contact tracing, initiate any necessary public health requirements and notify you if you need to take any steps. For the latest information about the University's COVID resources and information, please visit the **CSU COVID-19 site**: https://covid.colostate.edu/