

Efficient AI Model Design for Machine Learning and Inference 2024Fall

Chia-Chi Tsai (蔡家齊)

cctsai@gs.ncku.edu.tw

AI System Lab

Department of Electrical Engineering
National Cheng Kung University

Course Information



- Lecture:

- Chia-Chi Tsai 蔡家齊 cctsai@gs.ncku.edu.tw
 - 電機系館5樓 92510
- Location: 啟端館一樓階梯教室(96112)
- Wed. 13:10~16:00

- TAs:

- TA Group
 - 陳柏翰、王士逢、賴姿伶、陳喬雅、施尚甫、林言羲、權嘉
 - Email course.aislab@gmail.com
 - Please include [EAI2024] to the beginning of the email subject

Course Information



- Lecture slides will be available on Moodle
- Reference book
 - I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, 1st Ed., MIT Press, Dec. 2016
 - Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang, Joel Emer, Efficient Processing of Deep Neural Network, Morgan and Claypool Publisher, 2020
- One semester course, which include knowledge of building an Efficient AI Model
 - To understand the math of deep learning techniques
 - To familiarize with deep learning tools, such as Caffe, Tensor Flow, Torch, etc.
 - To understand the latest developments and applications of deep learning techniques
 - To develop practical working systems
- Prerequisites
 - Linear Algebra
 - Probability and Statistics
 - C/C++/Python

Course Grade



- Final Exams **20%**
- Paper Readings and Review **16%**
 - 4 paper reviews and 4% for each
- Lab Assignment **34%**
 - Lab exercise related to AI fundamentals and efficient AI model design
- Final Project **30%**
 - Practical working systems with efficient AI model

Course Outline



- Introduction
- Machine Learning Basics
- Deep Neural Networks
- Case Studies - Smart Sport with AI
- Efficient AI Model
- Pruning
- Knowledge Distillation
- Designing Efficient AI Model

Course Timetable



Week	Date	Lecture	Assignments	Paper Review
1	9/11	Introduction		
2	9/18	Machine Learning Basics	Lab1: Understand NN and Training Process	Paper Review1 - AI Models
3	9/25	Machine Learning Basics		
4	10/2	Deep Neural Networks	Lab2: AI Framework and Practices	
5	10/9	Deep Neural Networks		Paper Review2 – Advanced AI Models
6	10/16	Case Studies - Smart Sport with AI	Lab3: 2D post estimation	
7	10/23	Efficient AI Model	Final Project Proposals	
8	10/30	Efficient AI Model		Paper Review3 – Model Pruning
9	11/6	Pruning	Lab4: Model Pruning	
10	11/13	Final Project Proposals Presentation		
11	11/20	Final Project Proposals Presentation	Lab5: Knowledge Distillation	Paper Review4 – Knowledge Distillation
12	11/27	Pruning		
13	12/4	Knowledge Distillation		
14	12/11	Designing Efficient AI Model		
15	12/18	Final Exam		
16	12/25	Final Project Presentation		
17	1/1	Holiday		
18	1/8	Final Project Presentation		