

Data Science

HW 2: Model Compression

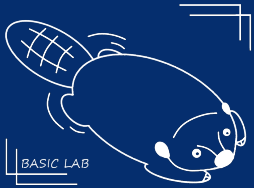
TA: 曾偉倫

Email: wlt seng.ee06@nycu.edu.tw

2023 Spring Data Science

Outline

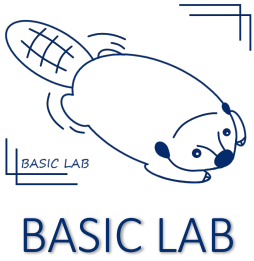
- Introduction – Model Compression
- Problem Description
- Kaggle Competition
- Grading Policy
- Report & Demo
- E3 Submission



BASIC LAB

NYCU

Introduction

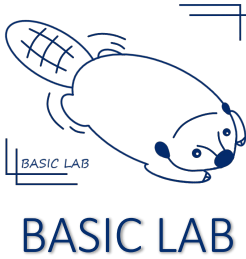


- Model Compression
 - Knowledge Distillation
 - Pruning
 - Model Architecture Design

Problem Description

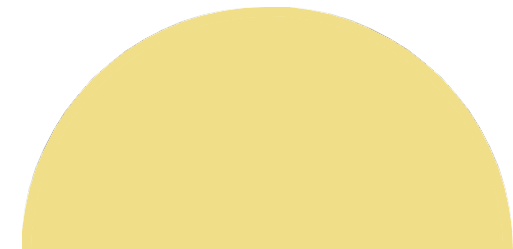
- Dataset: Fashion MNIST
- Input: Well-trained ResNet-50
- Output: compressed model
- Constrain:
 - number of parameter $\leq 100,000$
 - accuracy \geq **baseline benchmark** (update to Kaggle soon)
 - **DO NOT USE ANY TEST DATA, EXTERNAL DATA**

Grading Policy

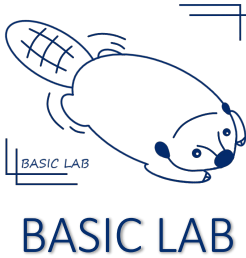


Model Compression (total: 100%)

- Kaggle Competition (75%)
- Report (20%)
- Demo (5%)



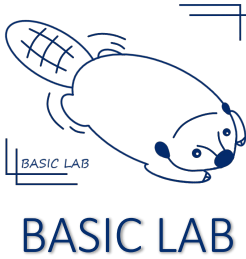
Grading Policy



Model Compression (total: 100%)

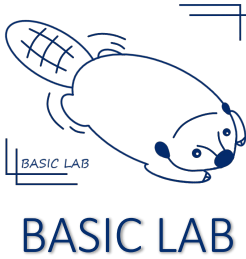
- Kaggle Competition (45%+30%)
 - Constrain: num_parameter \leq 10,0000
 - 45%: accuracy \geq baseline benchmark (update to Kaggle soon)
 - 30%: private leaderboard ranking

Kaggle Competition



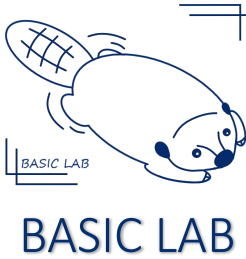
- Invitation Link:
<https://www.kaggle.com/t/ee36089663ee48cba68845dd1b791fba>
- A maximum of 5 submissions per day is allowed on Kaggle.
- Timeline:
 - 3/07 12:00 Competition Start
 - 3/20 23:59 Competition Finished

Grading Policy



- Report (20%)
 - torchsummary output (5%)
 - Brief Explanation of Compression Methods (15%)
 - Name, student_ID
 - Methods you used
 - Reference
 - ≤ 200 words
- Demo (5%)
 - TA will execute your code and reproduce the results.

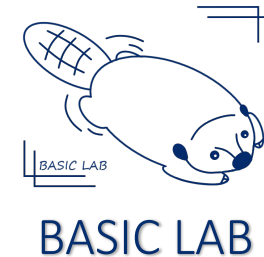
Special Rules



1. **Plagiarism** is prohibited.
2. **Sharing of code or submission files** is prohibited.
3. A maximum of **5 submissions per day** is allowed on Kaggle.
Please do not use any methods to bypass this limit.
4. Using testing data or external data is prohibited. TA will check the dataloader.
5. Using pre-trained models created by others as the final result is prohibited.
Please train your own model.
6. Using other models for compression is prohibited. Please use the trained model provided in the assignment release.

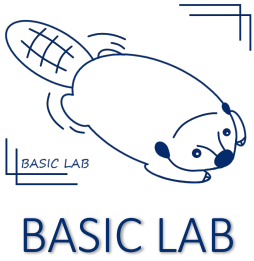
Violation of any of the above rules will result in a score of 0 for this assignment.

Demo Platform



- OS: Ubuntu Server 20.04
- CPU: AMD Ryzen Threadripper (will set num_worker=8)
- GPU: RTX 3080 (8GB) *1
- Python 3.8.10
- CUDA: 11.07
- Framework: PyTorch 1.13.1

E3 Submission



- Two File:
 1. <pdf file> hw2_report_[student_ID].pdf
 - Example: “hw2_report_311000123.pdf”
 2. <zip file> hw2_[student_ID].zip
 - Example: “hw2_311000123.zip”
 - Please make sure your submission contains the following items:
 - 1) All the code you used for training and testing
 - 2) The final weights used for testing
 - 3) A README file explaining how to execute your code (e.g., in txt or md format)