
GenAI-ML HW6

Deep Learning for Image Classification

TA: 劉建豐、馮柏翰、李梁玉軒

ntu-gen-ai-ml-2025-fall-ta@googlegroups.com

Deadline: 2025/11/21 23:59:59 (UTC+8)

Outline

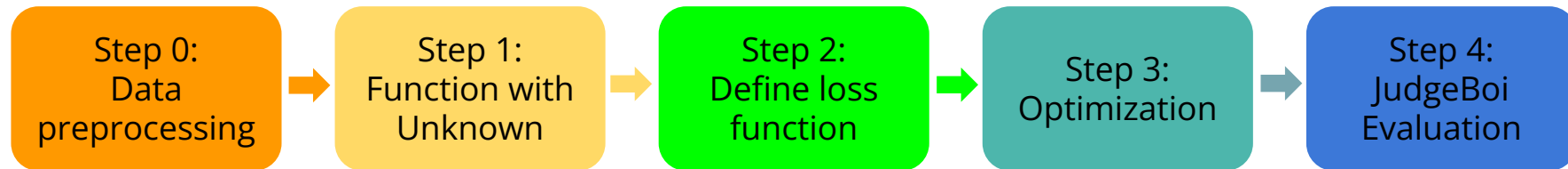
- Task Description
- Dataset
- Evaluation Metric
- Baseline
- Submission & Grading
- Hints

Links

- [JudgeBoi](#)
- [Colab sample code](#)
- [Kaggle sample code](#)
- [ML2025 Colab/Kaggle tutorial](#)
- [Intro. to CNN \(2021\)](#)

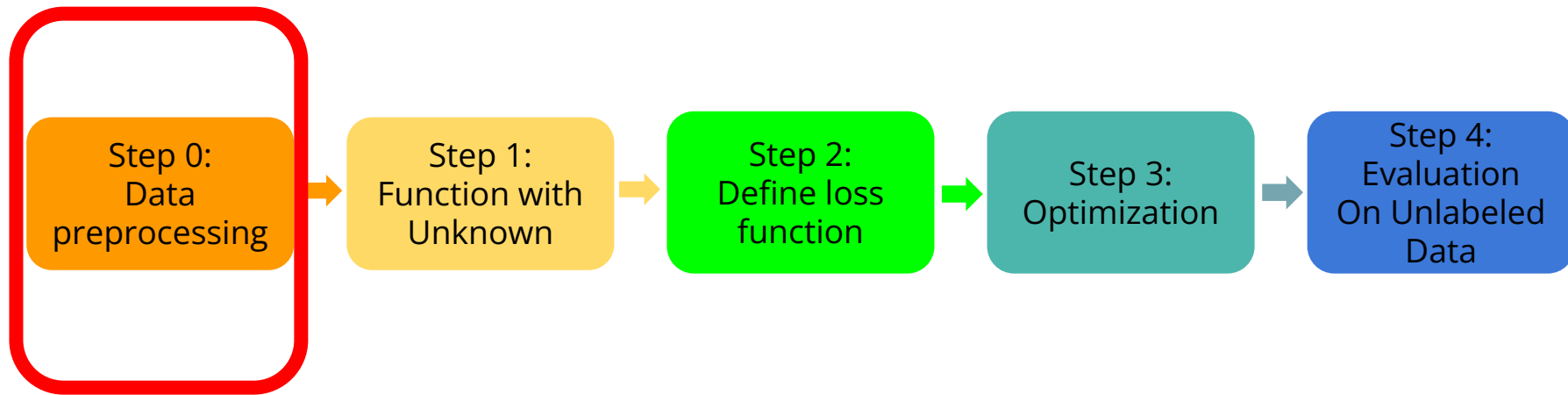
Homework Overview

- In this assignment, we will learn how to use a convolutional neural network (CNN) to solve an image classification task.
- We can decompose the whole assignment into 5 steps :



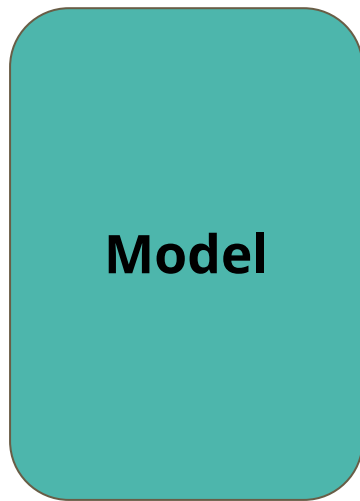
Homework Overview

- We expect you to know how to implement a convolutional neural network (CNN) with PyTorch in this assignment
- Be more familiar with convolutional neural networks (CNNs)
- Don't worry about building a CNN from scratch, we have provided a simple CNN template for you to refine it
- You can get at least **4 points** just by running the code, and **6 points** by simply adjusting some numbers.



Task Description

- Construct an image classifier with pytorch



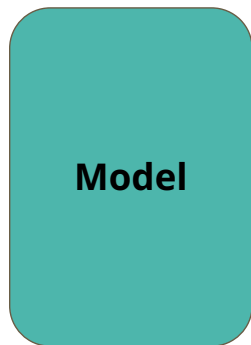
cat



dog

Task Description

- In this homework, you will train a image classifier to classify 10 main characters in 《BanG Dream! It's MyGO!!!!!》and 《BanG Dream! Ave Mujica》



Anon



Soyo



Dataset

Images are collected from《BanG Dream! It's MyGO!!!!!》and 《BanG Dream! Ave Mujica》

There are 2378 images in the dataset:

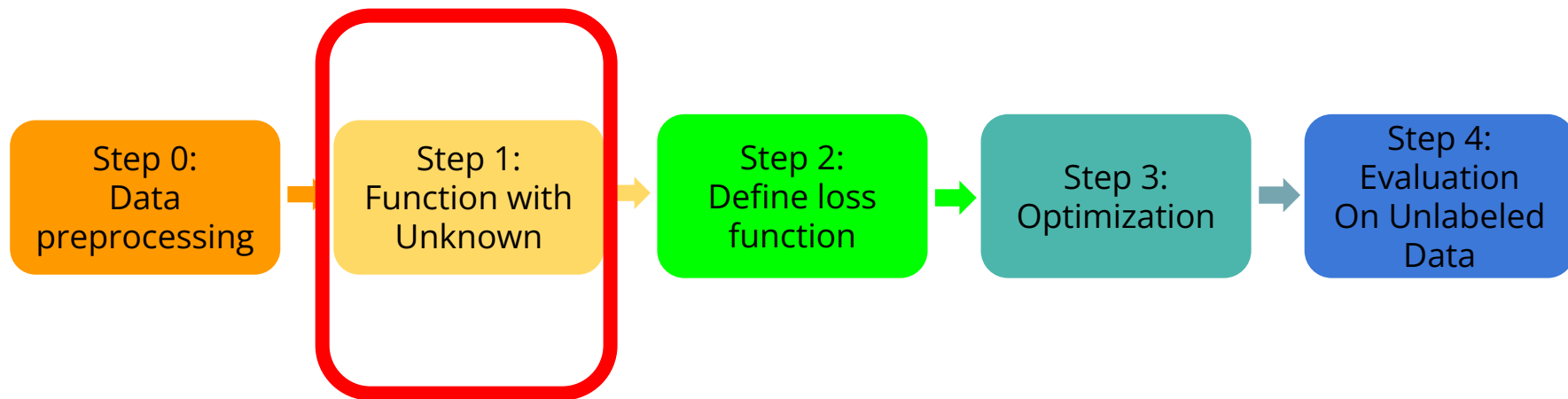
- Training set: 1578 labeled images
- Validation set: 300 labeled images
- Testing set: 500 unlabeled images



Data Augmentation

- Is training data enough for training your model? Try to “create” more data with existing data.





Properties of Images

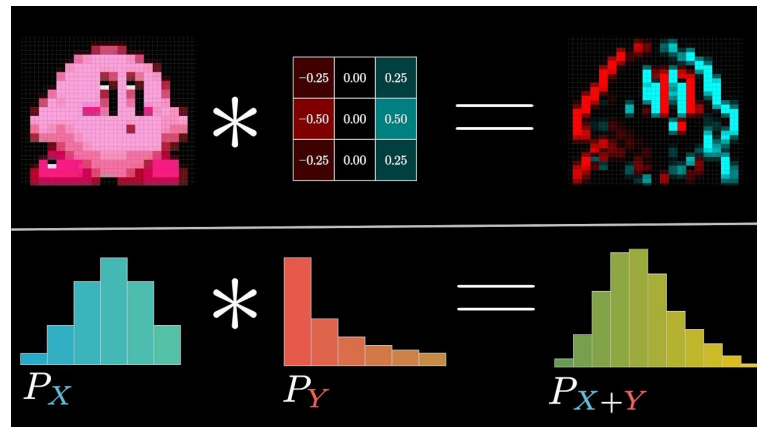
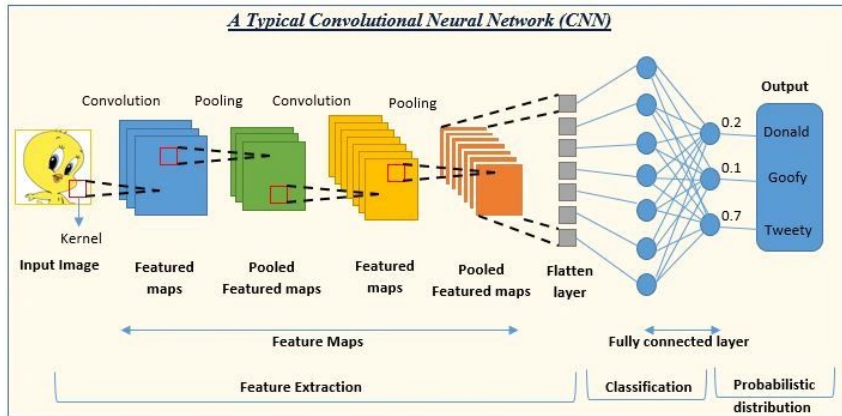
- Properties of images
 - Spatial Locality
 - Translation Invariant
- Can we leverage such properties to design other architecture than MLP?

Translation Invariance



Convolutional Neural Network (CNN)

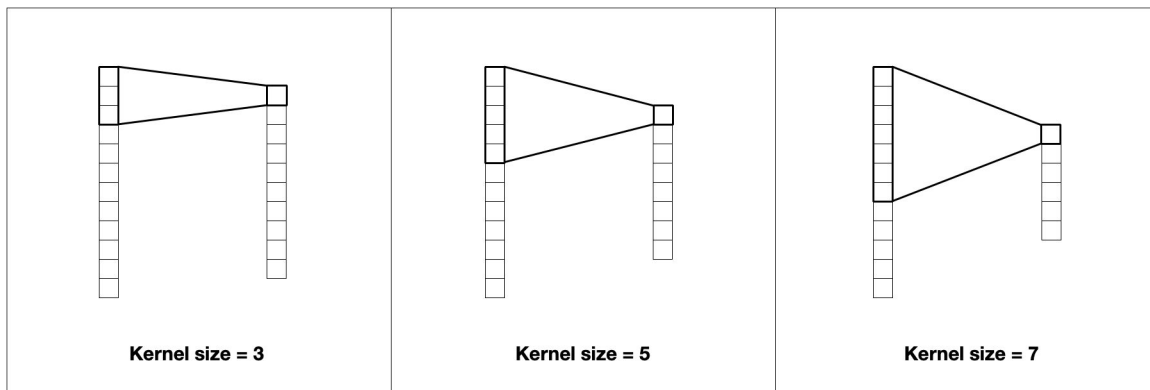
- Leverage the properties of image to capture the features in image more efficiently
 - Convolution
 - Pooling



Ref: <https://locusit.se/techpost/specialisation/deep-learning/convolutional-neural-networks/>
But what is a convolution?

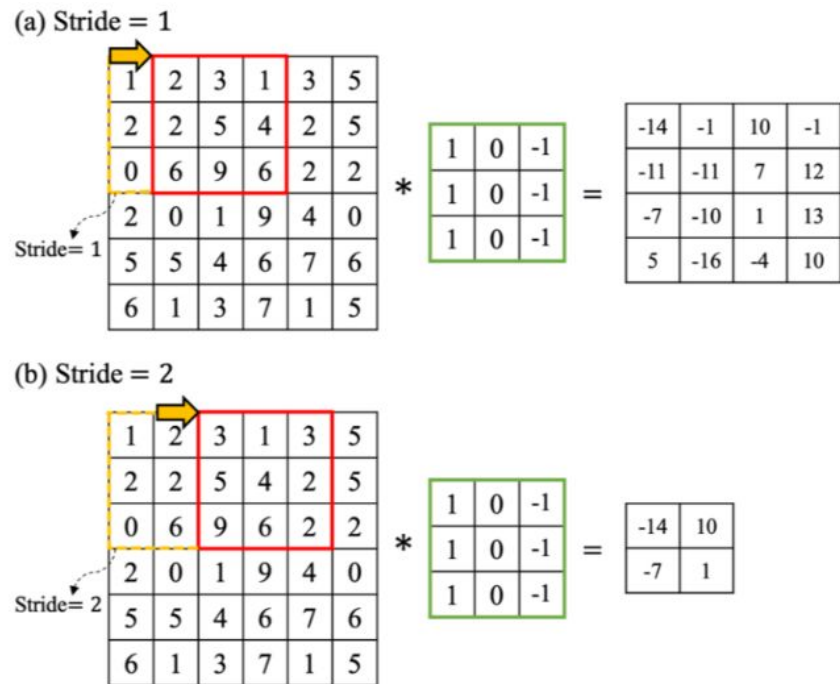
Different Setting of CNN

- Filter size of CNN
 - Larger filter
 - Broader area
 - Intermediate scale feature (eg. shape, contour)
 - Smaller filter
 - Narrower area
 - Local feature (eg. edge, texture)



Different Setting of CNN

- Stride for filter in CNN
 - Large stride
 - Less computation and memory
 - May skip some features
 - Small stride
 - More computation and memory
 - Capture more features



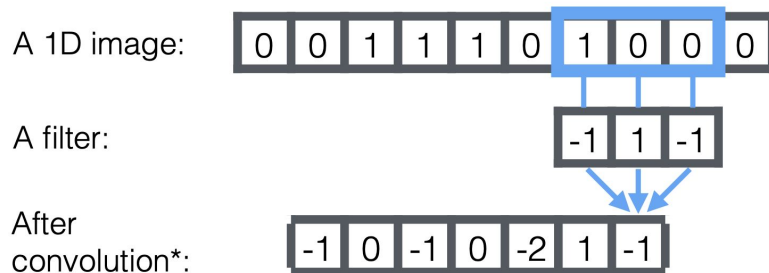
Ref:

<https://www.brilliantcode.net/1584/convolutional-neural-networks-1-convolution-layer-stride-padding-kernel/>

1D-CNN

- A simple example of 1D-CNN

Convolutional Layer: 1D example



After convolution*:



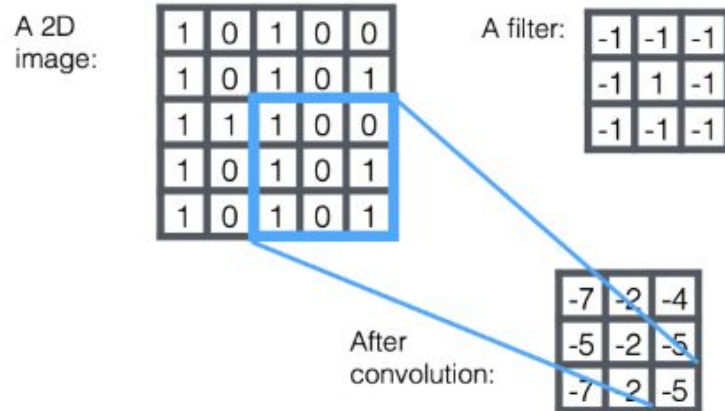
After ReLU:



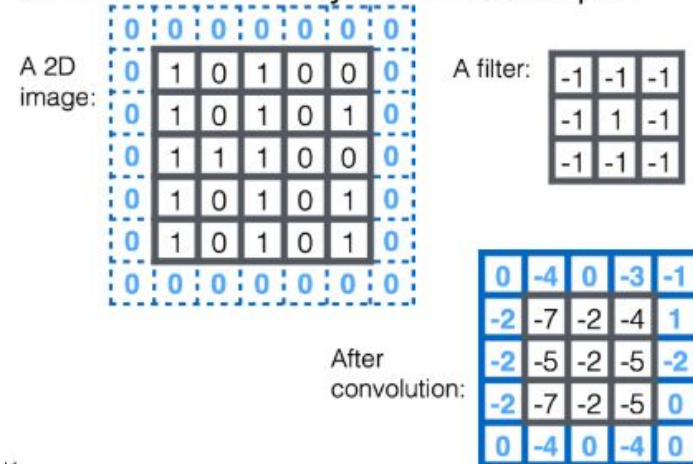
2D-CNN

- Images “shrinks” after convolution, so padding is crucial for it to make size consistent across layers

Convolutional Layer: 2D example



Convolutional Layer: 2D example

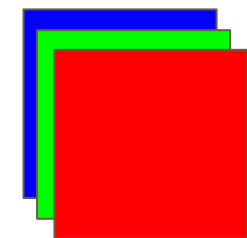


Pytorch Implementation

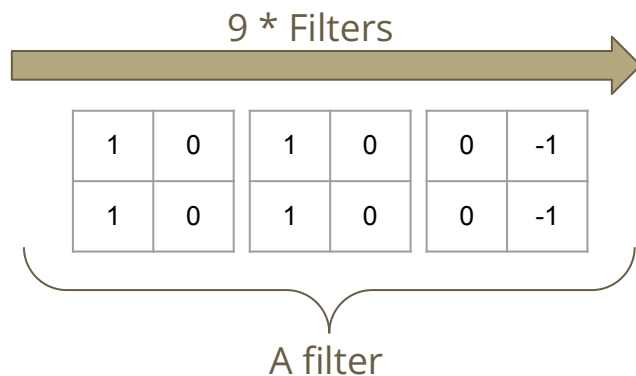
- Convolutional layer is already implemented in PyTorch!!
- An example of convolutional layer in PyTorch

- `nn.Conv2d(3, 9, 2, 1, 1)`

Input channel Output channel Kernel size stride padding



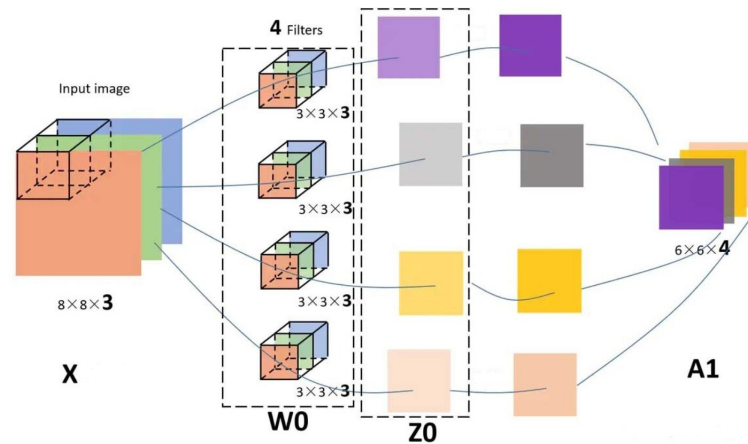
Input
(3 channels)



Output
(9 channels)

Number of Parameters for Filters

- What is the trainable number of parameters in `nn.Conv2d(3, 9, 2, 1, 1)`?
 - Input channel = 3
 - Kernel size = $2 * 2$
 - Filter size = (Input channel) * (kernel size)
 - Output channel = Filter number = 9
 - Total = (Input channel) * (kernel size) * (Filter number) = 108 parameters !!



Ref:
<https://discuss.pytorch.org/t/masking-the-intermediate-5d-conv2d-output/144026>

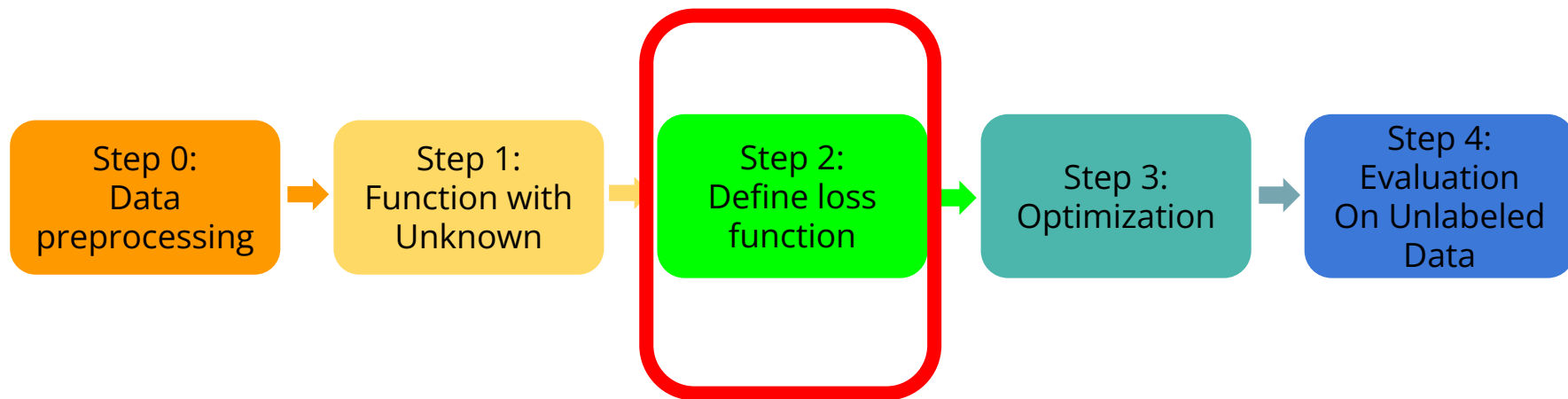
Pooling

- Dimension reduction
- Different pooling is used
 - Max pooling
 - Average pooling
 - Min pooling

2	2	7	3
9	4	6	1
8	5	2	4
3	1	2	6

Max Pool
→
Filter - (2 x 2)
Stride - (2, 2)

9	7
8	6



Loss function - Cross Entropy

- What is cross entropy loss?
- Why is cross entropy loss, not MSE?
 - Not continuous

$$\sum_{i=1}^N \sum_{c=1}^C -y_{i,c} \log_2(p_{i,c})$$

Groundtruth of i-th data for c-th class

Predicted prob. of i-th data for c-th class

Example - Cross Entropy

	Prediction			Groundtruth		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
Data 1	0.4	0.3	0.3	1	0	0
Data 2	0.3	0.4	0.3	0	1	0
Data 3	0.5	0.2	0.3	1	0	0
Data 4	0.8	0.1	0.1	0	0	1

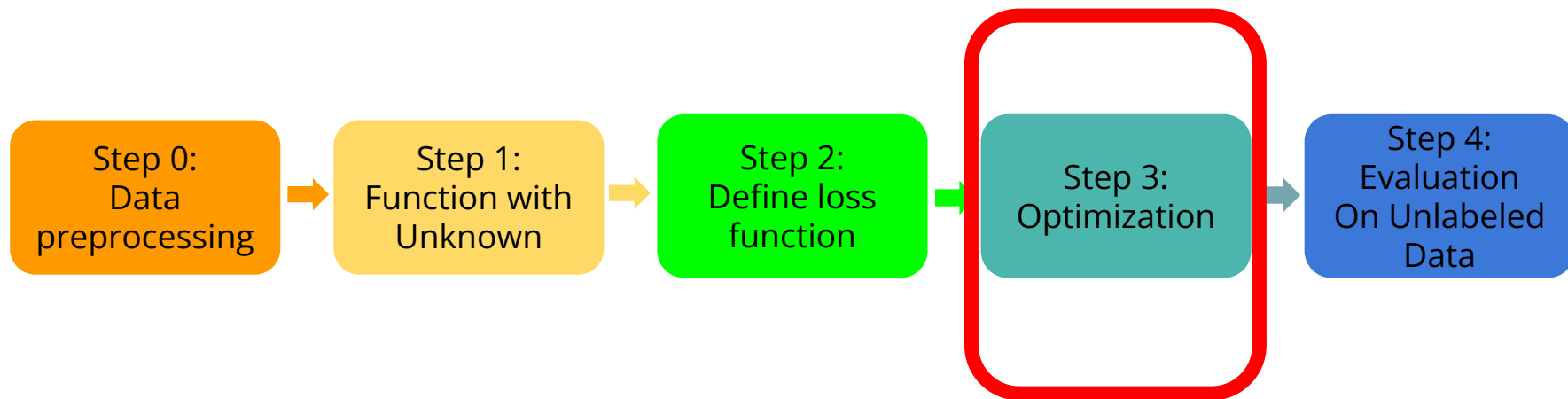
CE for data 1: $-(1 \times \log(0.4) + 0 \times \log(0.3) + 0 \times \log(0.3)) = 1.322$

CE for data 2: $-(0 \times \log(0.3) + 1 \times \log(0.4) + 0 \times \log(0.3)) = 1.322$

CE for data 3: $-(1 \times \log(0.5) + 0 \times \log(0.2) + 0 \times \log(0.3)) = 1$

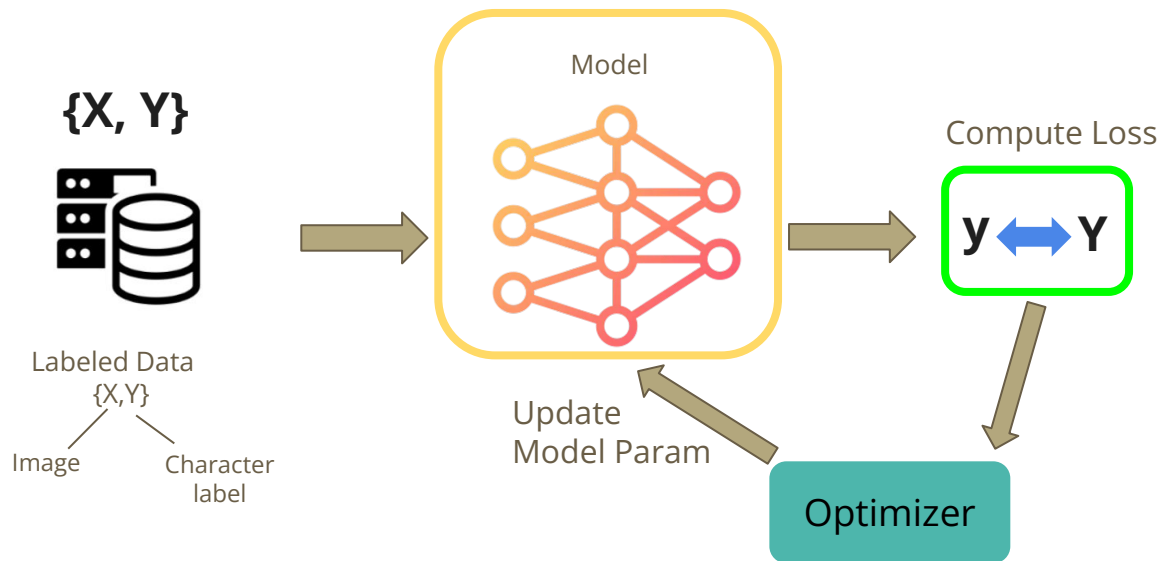
CE for data 4: $-(0 \times \log(0.8) + 0 \times \log(0.1) + 1 \times \log(0.1)) = 3.322$

Total cross entropy loss = $1.322 + 1.322 + 1 + 3.322 = 6.966$



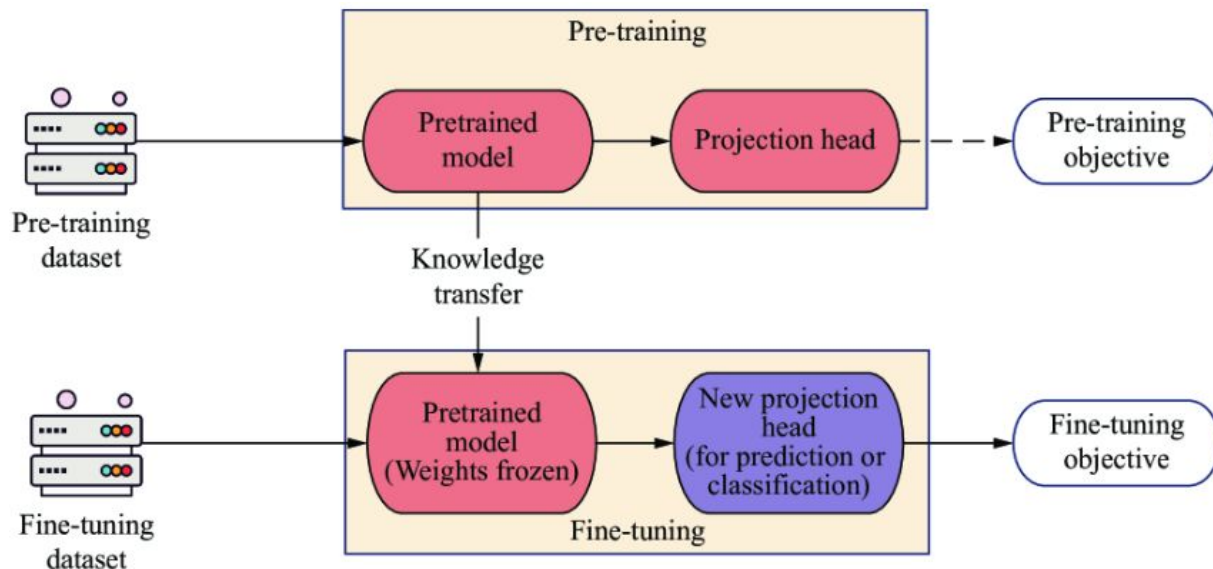
Training Loop

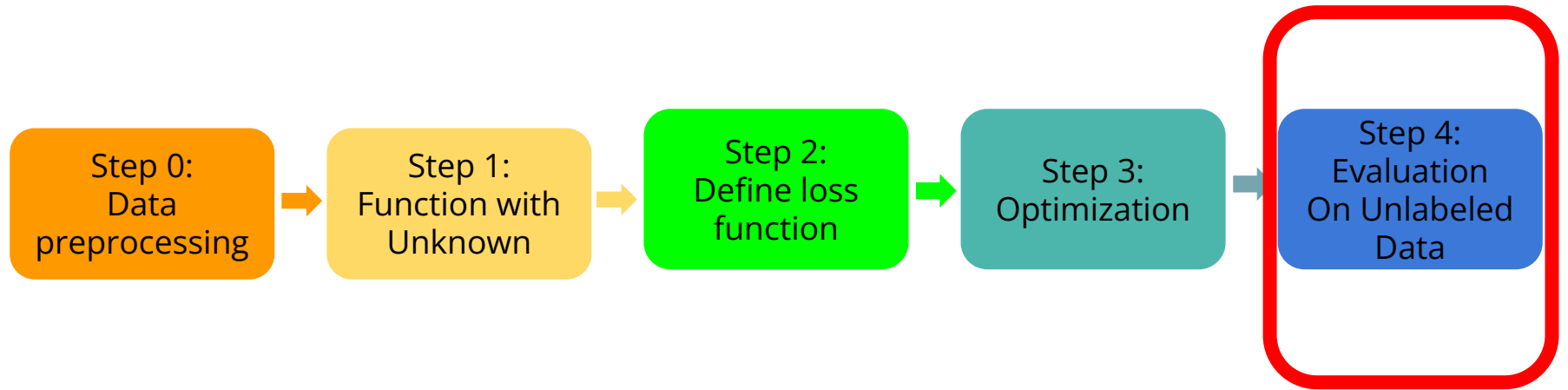
- Leverage the label and the output of current model to optimize the model parameters



Pre-trained model

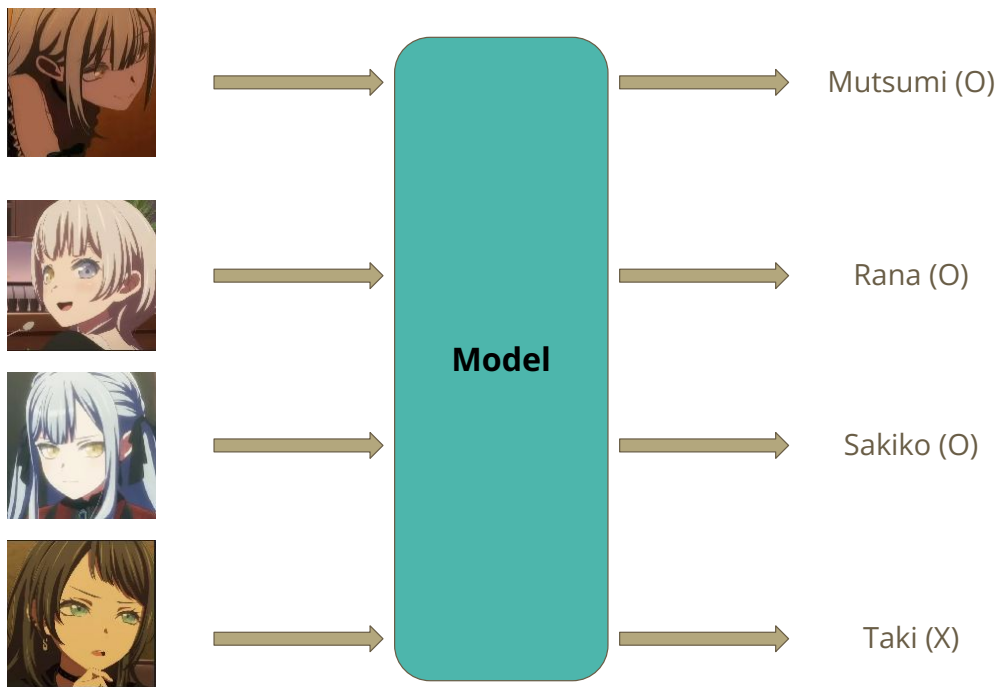
- Why pre-trained models?





Evaluation Metric

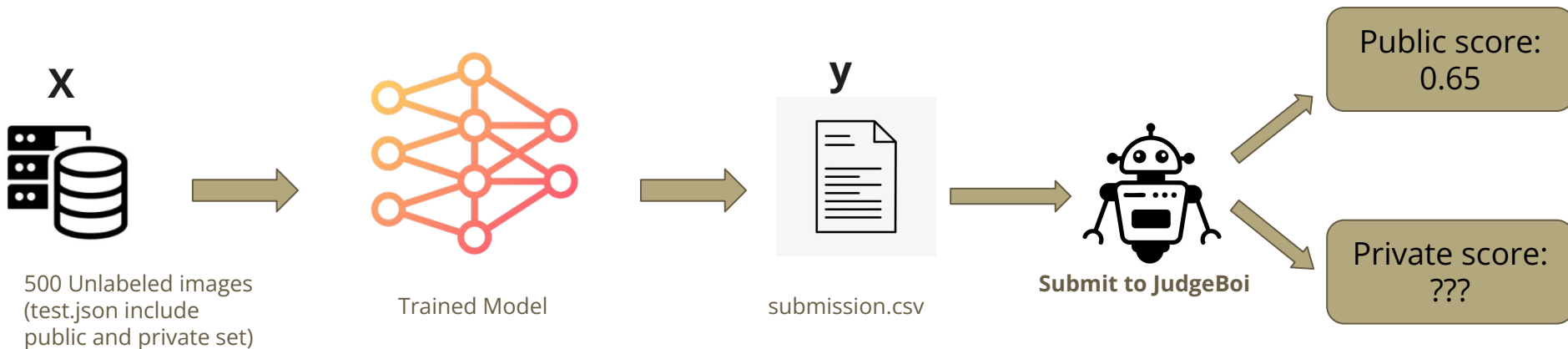
- Accuracy on the testing set



Accuracy: $\frac{3}{4} = 0.75$

Evaluation pipeline

- Divide the whole testing set into two parts:
 - Public testing set (scores can be seen before the deadline)
 - Private testing set (scores cannot be seen before the deadline)



**Release after
Deadline**

Grading & Baseline

- Any successful submission on Judgeboi: **2 pts**
- Any public baseline worth **1 pt**
 - Simple public baseline: **0.45**
 - Medium public baseline: **0.70**
 - Strong public baseline: **0.83**
 - Boss public baseline: **0.94**
- Any private baseline worth **1 pt**
 - There are also 4 baselines for private baseline
 - Announce the result after the deadline

Submission(1/2)

- Submit your prediction file to JudgeBoi
- You have 5 submission quota per day, reset at 23:59
- Deadline: 2025/11/21 23:59:59 (UTC+8)
- No late submission is allowed, please finish your homework as soon as possible

Submission(2/2)

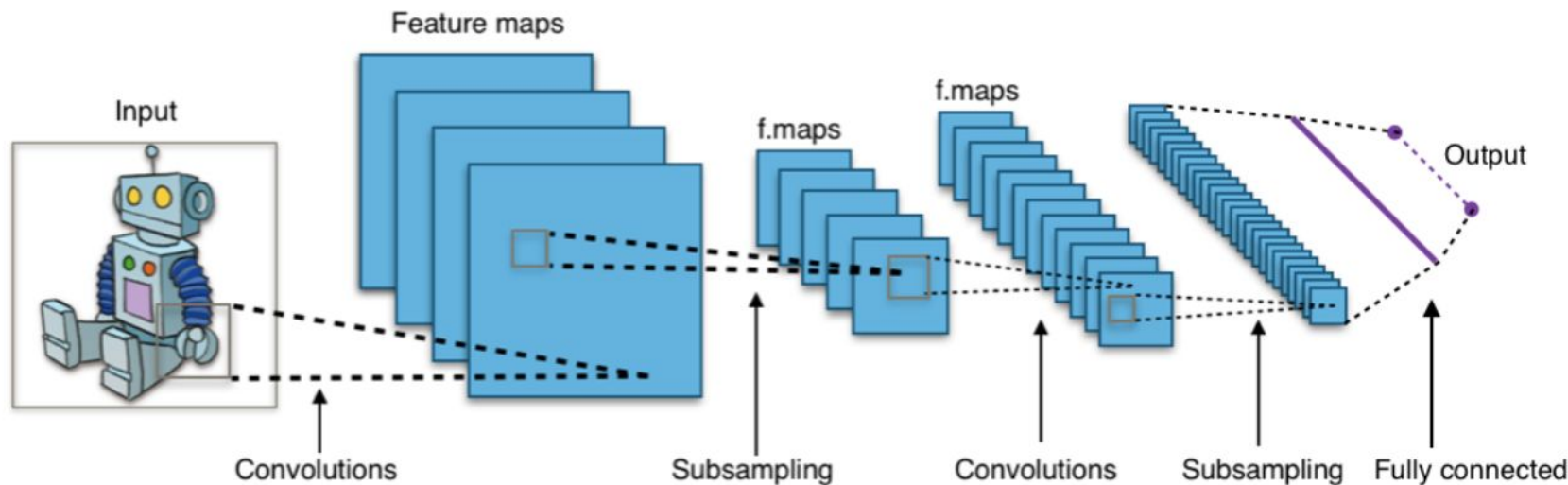
- Reminder : You can select up to 2 submissions on JudgeBot (If not selected, top 2 public score will be chosen by default)
- Beware of overfitting, high public score doesn't assure high private score
 - Refer to your validation score during training

Hints For Passing the Baseline

JudgeBoi Submission (+2 pt)	Submit predicted results to JudgeBoi	Estimated Training Time	Estimated Inference Time
Public Simple Baseline (+1 pt)	Just run the colab with MLP model and submit the predicted "submission.csv" to JudgeBoi	5-10 minutes	1 minutes
Private Simple Baseline(+1 pt)			
Public Medium Baseline (+1 pt)	Change model architecture to simple CNN in the sample code and simply refine it	10-15 minutes	1 minutes
Private Medium Baseline(+1 pt)			
Public Strong Baseline (+1 pt)	Refine your CNN model architecture again and data augmentation	10-15 minutes	1 minutes
Private Strong Baseline(+1 pt)			
Public Boss Baseline (+1 pt)	Use pre-defined models and fine-tuning models with pre-trained weights	10-20 minutes	1 minutes
Private Boss Baseline(+1 pt)			

Medium Baseline – Convolution Neural Network (CNN)

- Is linear layer enough for the classification? You may try convolution neural network (CNN).



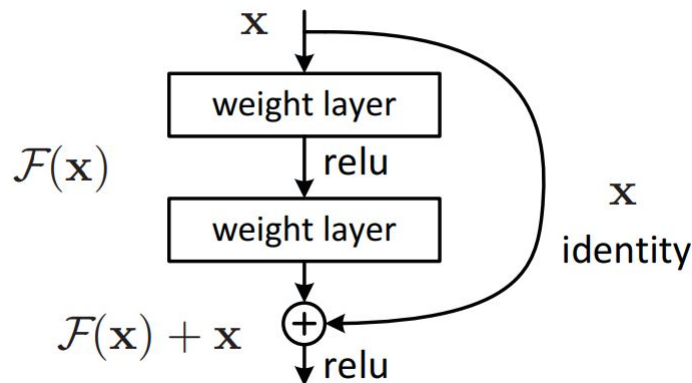
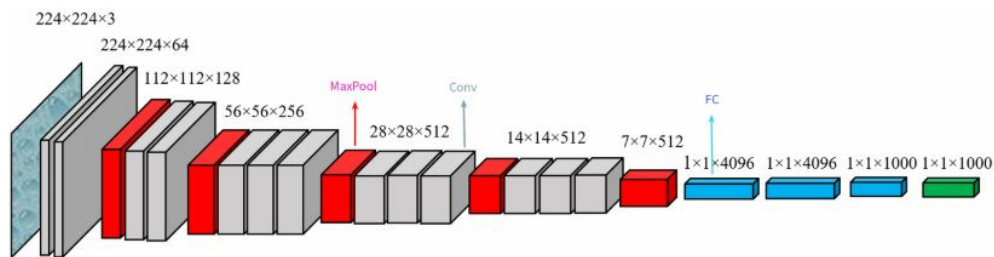
Strong Baseline – Data Augmentation

- Is training data enough for training your model? Try to “create” more data with existing data.



Boss Baseline – More Advanced Pre-defined Model

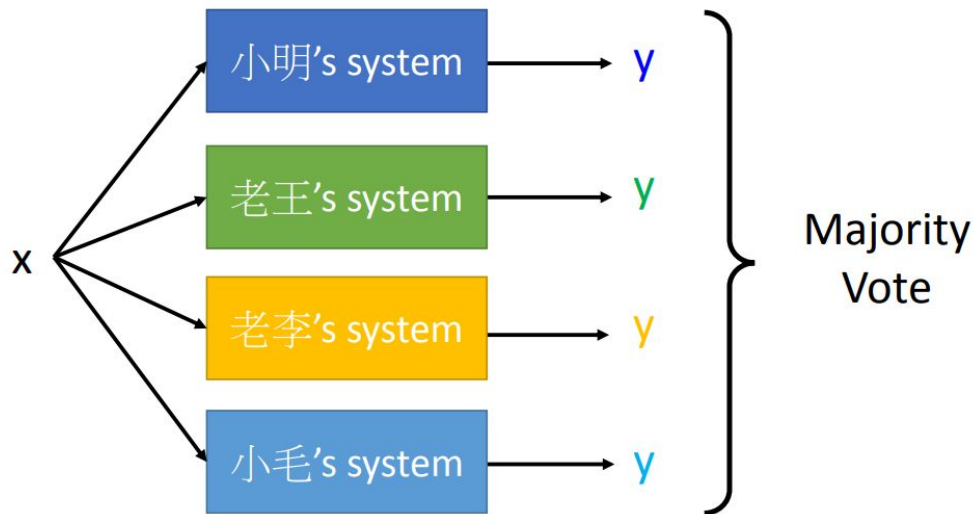
- Standing on the shoulders of giants.
 - VGG
 - ResNet
 - etc.



Ref: <https://zhuanlan.zhihu.com/p/107884876>
<https://arxiv.org/pdf/1512.03385>

Boss Baseline – Ensemble

- Two heads are better than one.
 - Majority vote
 - Logit sum



Other Hints – Techniques

You may try other techniques

- Dropout
- Test-time Augmentation
- Or others...?

Other Hints – Hyperparameter

Number of epochs	10~100 (default: 20)
Learning rate	1e-5 ~ 1e-2 (default: 2e-4)
Batch size	8, 16, 32, 64, 128 (default: 32)
Optimizer	Adam, AdamW, SGD (default: Adam)
Weight decay	1e-3 ~ 1e-5 (default: 1e-5)
Model size	5 MB ~ 150 MB

Grading Release Date

- The grading of the homework will be released by 2025/**12/12** 23:59:59 (UTC+8)

Regulations

- You should NOT plagiarize
- You should NOT modify your prediction files manually
- Do NOT share codes or prediction files with any living creatures
- Do NOT use any approaches to submit your prediction files more than 5 times a day
- Your final grade will x 0.9 and get a zero score for this homework if you violate above rules first time (within this semester)
- You will get a **F** as your final grade if you violate above rules multiple times
- Prof. Lee & TAs preserve the rights to change the rules & grade



If You Have Any Questions

- NTU Cool **HW6** 作業討論區
 - 如果同學的問題不涉及作業答案或隱私, 請**一律使用** NTU Cool 討論區
 - 助教們會優先回答NTU Cool討論區上的問題
- Email: ntu-gen-ai-ml-2025-fall-ta@googlegroups.com
 - Title should start with [GenAI-ML 2025 Fall **HW6**]
 - Email with the wrong title will be moved to trash automatically
- TA Hours
 - Time:
 - 11/3, 11/10, 11/17 Monday 20:00~22:00
 - 11/7, 11/14, 11/21 Friday 17:30~19:30
 - Location: Google Meet (<https://meet.google.com/qox-xubu-jak>)