



Institute of Electronics
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Final Project – Image Classification (Kaggle)



Student: Jyun-Hua Chen

Advisor: Juinn-Dar Huang, Ph.D.

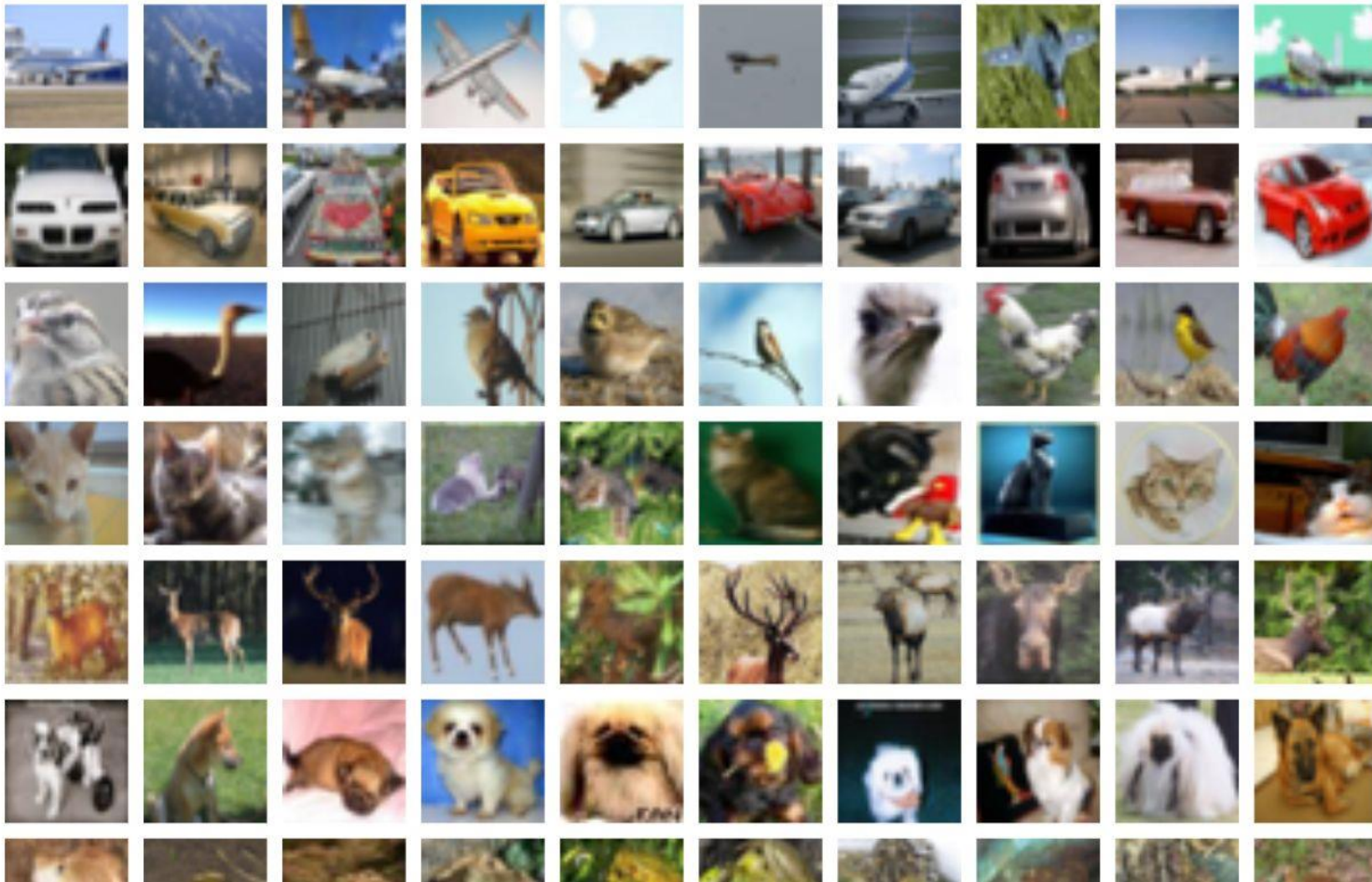
Aug 22, 2021

Outline

- Final project: Image Classification
 - Task Description
 - Grading and Requirement
 - Submission
 - Kaggle
 - Report
 - Hints

Task Description : Image Classification

- Dataset: Canadian Institute for Advanced Research



Task: Image Classification

- The images are collected from CIFAR-100 dataset
- Dataset provided by TA:
 - Train: 50000 images
 - Test: 10000 images
- **DO NOT** use the label of test data, it is regarded as **CHEATING**

Model Limitations and Grading

- Model limitations:
 - You have to use **Transformer**-based model (e.g., ViT 、 DeiT 、 LeViT, ...)
 - You can increase or decrease #layers in the model
 - Your model must include attention layers
 - You can refine the model through
 - › Adding layers before/after attention layers
 - › Using residual links, different activation functions, ...
 - › Applying model reduction (e.g., pruning)
- Scoring formula:
 - Unlike the midterm project, this time the grading process considers both accuracy and model size

Grading and Requirements

- Deadline:
 - Kaggle: 9/5 (Sun.) 23:59 in UTC+8
- Grading:
 - Scoreboard ranking (60%) :
 - › $\text{Score} = 100 * \text{ReLU}(\text{Acc} - 70) / \text{Model_Size}$
 - » Acc is Top-1 accuracy in % on Kaggle leaderboard
 - » Model_Size is in M (10^6)
 - Report (40%)
- Rules:
 - Discussion is allowed, but you should do your project by yourself
 - **DO NOT** label the testing data or use the labels of testing data!
 - Your code should be able to generate the results you submitted to the leaderboard (within 1%)

Submission

- Submit your code via email andychen10209@gmail.com
 - <your chinese name>_final_project.zip (Ex: 陳俊樺_final_project.zip)
 - Include:
 - > **Your code (.py or .ipynb)**
 - > **Model weight file (.pt or .pth)**
 - > **Report.pdf**
- **DO NOT**
 - Submit the dataset or model
 - Submit your results in wrong formats
 - > like **.rar/.7z** or **report.doc**

Kaggle

- Register a Kaggle account by yourself
- Join our in-class competition
 - Link: <https://www.kaggle.com/c/2021-ai-training-final-project/data>
 - Maximum daily submission: **20**
- Rules:
 - Your team name should be: < your Chinese name >
> **E.g.**, 陳俊樺
> **Otherwise, your submission will NOT be graded.**
 - You can NOT create multiple Kaggle accounts to submit more results
 - You can NOT upload results identical to other people's

Report Problems (40%)

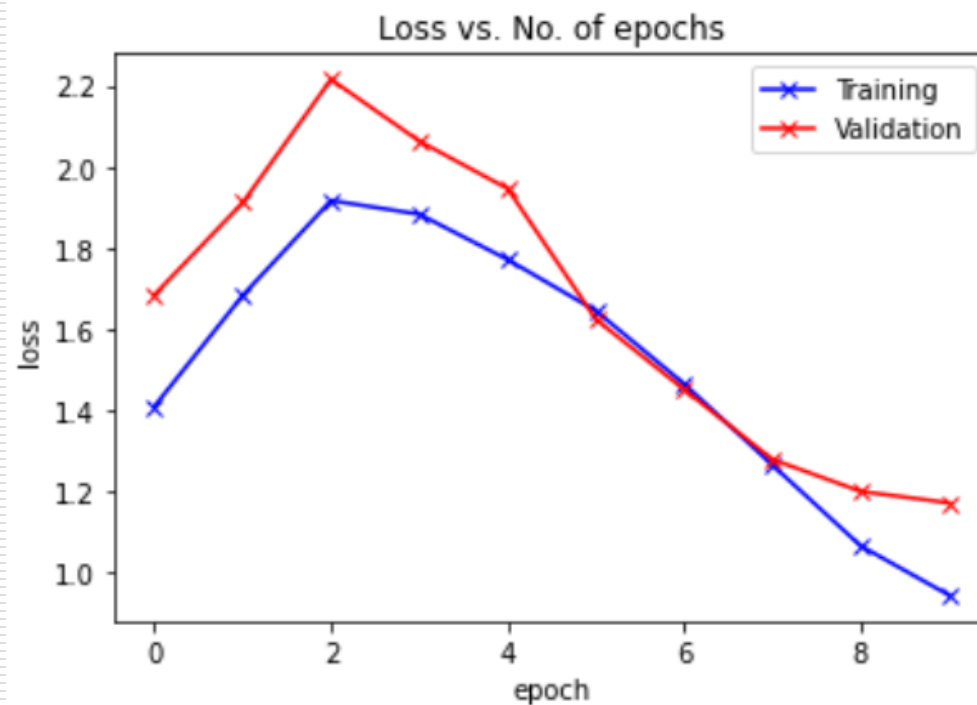
- Model Size (10%)
 - Print your model and show number of total size / number of parameters(5%)
[Hint]:

```
1 from torchsummary import summary
2 result = summary(model, (3, 32, 32))
3 print(result)
```
 - Print learning curve and briefly state your observations. (loss or accuracy for your CNN) (5%)
[Hint]: use plt ()
- How to improve the accuracy (list your method) (30%):
 - Loss function ?
 - Your network?
 - Activity function?
 - Etc...


Report Example

```
=====
Total params: 3,191,397
Trainable params: 3,191,397
Non-trainable params: 0
=====
```

```
Input size (MB): 0.01
Forward/backward pass size (MB): 69.50
Params size (MB): 12.17
Estimated Total Size (MB): 81.69
=====
```



Hints

- You can refine the items listed below to improve model accuracy
 - **Model architecture** (Prof. Adar's personal favorite)
 - Data normalization
 - Data augmentation
 - Residual connections
 - Batch normalization
 - Learning rate scheduling
 - Optimizer selection
 - Hyperparameter adjustment
 - Etc...
- 
- Equally Important!**

References

- An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale
 - ViT: <https://arxiv.org/abs/2010.11929>
- Training data-efficient image transformers & distillation through attention
 - DeiT: <https://arxiv.org/abs/2012.12877>
- LeViT: a Vision Transformer in ConvNet's Clothing for Faster Inference
 - LeViT: <https://arxiv.org/abs/2104.01136>

Thank you