

Data Science Homework 2 Report

311581010 陳彥儒

- torchsummary/torchinfo output (5%)

- See more below

```
=====
Total params: 85,956
Trainable params: 85,956
Non-trainable params: 0
-----
Input size (MB): 0.01
Forward/backward pass size (MB): 1.91
Params size (MB): 0.33
Estimated Total Size (MB): 2.25
-----
Number of remaining parameters: 85956
```

- Brief Explanation of Compression Methods (15%)

- Name, student_ID

- 陳彥儒 · 311581010

- Methods you used

- Regarding Knowledge Distillation, I adapted vanilla-KD under a smaller network (Resnet 14), i.e. simply adding an extra “KL divergence” loss on the original classification loss (CrossEntropy). Also, I’ve tried to tune the temperature used in KL loss and the weight between KL loss and classification loss.
 - Regarding model pruning, I adapted the package “torch_pruning” (see more in Reference). And also I chose “GroupNorm” as the normalization layer. To say more, I’ve pruned ~50% of the parameters of the network and got ~90000 parameters.

- Reference

- <https://github.com/DefangChen/SimKD>
 - <https://github.com/VainF/Torch-Pruning>

torchsummary

Layer (type)	Output Shape	Param #
Conv2d-1	[-1, 11, 28, 28]	297
BatchNorm2d-2	[-1, 11, 28, 28]	22
ReLU-3	[-1, 11, 28, 28]	0
Conv2d-4	[-1, 11, 28, 28]	1,089
BatchNorm2d-5	[-1, 11, 28, 28]	22
ReLU-6	[-1, 11, 28, 28]	0
Conv2d-7	[-1, 11, 28, 28]	1,089
BatchNorm2d-8	[-1, 11, 28, 28]	22
ReLU-9	[-1, 11, 28, 28]	0
BasicBlock-10	[-1, 11, 28, 28]	0
Conv2d-11	[-1, 11, 28, 28]	1,089
BatchNorm2d-12	[-1, 11, 28, 28]	22
ReLU-13	[-1, 11, 28, 28]	0
Conv2d-14	[-1, 11, 28, 28]	1,089
BatchNorm2d-15	[-1, 11, 28, 28]	22
ReLU-16	[-1, 11, 28, 28]	0
BasicBlock-17	[-1, 11, 28, 28]	0
Conv2d-18	[-1, 22, 14, 14]	2,178
BatchNorm2d-19	[-1, 22, 14, 14]	44
ReLU-20	[-1, 22, 14, 14]	0
Conv2d-21	[-1, 22, 14, 14]	4,356
BatchNorm2d-22	[-1, 22, 14, 14]	44
Conv2d-23	[-1, 22, 14, 14]	242
BatchNorm2d-24	[-1, 22, 14, 14]	44
ReLU-25	[-1, 22, 14, 14]	0
BasicBlock-26	[-1, 22, 14, 14]	0
Conv2d-27	[-1, 22, 14, 14]	4,356
BatchNorm2d-28	[-1, 22, 14, 14]	44
ReLU-29	[-1, 22, 14, 14]	0
Conv2d-30	[-1, 22, 14, 14]	4,356

BatchNorm2d-31	[-1, 22, 14, 14]	44
ReLU-32	[-1, 22, 14, 14]	0
BasicBlock-33	[-1, 22, 14, 14]	0
Conv2d-34	[-1, 45, 7, 7]	8,910
BatchNorm2d-35	[-1, 45, 7, 7]	90
ReLU-36	[-1, 45, 7, 7]	0
Conv2d-37	[-1, 45, 7, 7]	18,225
BatchNorm2d-38	[-1, 45, 7, 7]	90
Conv2d-39	[-1, 45, 7, 7]	990
BatchNorm2d-40	[-1, 45, 7, 7]	90
ReLU-41	[-1, 45, 7, 7]	0
BasicBlock-42	[-1, 45, 7, 7]	0
Conv2d-43	[-1, 45, 7, 7]	18,225
BatchNorm2d-44	[-1, 45, 7, 7]	90
ReLU-45	[-1, 45, 7, 7]	0
Conv2d-46	[-1, 45, 7, 7]	18,225
BatchNorm2d-47	[-1, 45, 7, 7]	90
ReLU-48	[-1, 45, 7, 7]	0
BasicBlock-49	[-1, 45, 7, 7]	0
AdaptiveAvgPool2d-50	[-1, 45, 1, 1]	0
Linear-51	[-1, 10]	460

Total params: 85,956

Trainable params: 85,956

Non-trainable params: 0

Input size (MB): 0.01

Forward/backward pass size (MB): 1.91

Params size (MB): 0.33

Estimated Total Size (MB): 2.25

Number of remaining parameters: 85956