LAB 4 Report

Introduction:

本次的 lab 分為兩個部分 4a 為練習利用 pytorch 做 model pruning，接下來為 intel 的 openvino tool 的使用練習

Experiment setup:

CPU: Intel(R) Xeon(R) W-2135 CPU @ 3.70GHz

GPU: Nvidia Titan RTX

System: Ubuntu 18.04.4 LTS

Implementation:

Lab4(a)

我們持續使用上次的 resnet18 架構，並保持上次的 layers, width, 及 input_size 來進行 model pruning。

```
prune.random_unstructured(module, name="weight", amount=0.3)
```

```
Conv2d(3, 64, kernel_size=(7, 7), stride=(2, 2), padding=(3, 3), bias=False)
```

```
model = ResNet(BasicBlock)

parameters_to_prune = (
    (model.conv1, 'weight'),
    (model.fc, 'weight'),
)


prune.global_unstructured(
    parameters_to_prune,
    pruning_method=prune.L1Unstructured,
    amount=0.5,
)
```

可以看到我們設定 pruning 的 amount 為 0.5

```
Sparsity in conv1.weight: 62.23%
Sparsity in fc.weight: 49.78%
Global sparsity: 50.00%
```

達成的 global sparsity 剛好為 50%

而 training 出來的結果如下圖所示

```
Test set: Top1 Accuracy: 1828/3347 (54 %) , Top3 Accuracy: 2755/3347 (82 %)
class 0 : 89/368 24 %
class 1 : 40/148 27 %
class 10 : 106/231 45 %
class 2 : 257/500 51 %
class 3 : 196/335 58 %
class 4 : 145/287 50 %
class 5 : 314/432 72 %
class 6 : 93/147 63 %
class 7 : 32/96 33 %
class 8 : 179/303 59 %
class 9 : 377/500 75 %
```

對比上次的結果 top1 accuracy 53%是持平的。可能我們的 accuracy 較低，因此 pruning 對於 model 精準度的影像並不顯著。

Lab4(a)-2

經過了多次的測試，我調整出 70%的 global sparsity

```
print(
    "Sparsity in conv1.weight: {:.2f}%".format(
        100. * float(torch.sum(model.conv1.weight == 0))
        / float(model.conv1.weight.nelement())
    )
)
```

```
Sparsity in conv1.weight: 78.09%
```

```
print(
    "Sparsity in fc.weight: {:.2f}%".format(
        100. * float(torch.sum(model.fc.weight == 0))
        / float(model.fc.weight.nelement())
    )
)
```

```
Sparsity in fc.weight: 69.85%
```

```
print(
    "Global sparsity: {:.2f}%".format(
        100. * float(
            torch.sum(model.conv1.weight == 0)
            + torch.sum(model.fc.weight == 0)
        )
        / float(
            model.conv1.weight.nelement()
            + model.fc.weight.nelement()
        )
    )
)
```

```
Global sparsity: 70.00%
```

而 accuracy 仍然持平於 55%並未受到影響

```
Epoch 24/24
----------
skewed_training Loss: 1.4189 Acc: 0.5215
validation Loss: 1.4500 Acc: 0.5507

Training complete in 59m 14s
Best val Acc: 0.557143
```

Lab4(b)

首先是將 Pytorch 的 model 轉換為 onnx 的形式。轉換的 code 如下

```
model.eval()

x = torch.FloatTensor( 512, 3, 244, 244)
print(x.shape)
torch.onnx.export(model, x,
          "Resnet_4.onnx",
          export_params = True,
          opset_version = 10,
          do_constant_folding=True,
          input_names=['input'], output_names=['output'])
```

轉換出的檔案如下圖:

```
(base) ubuntu@nctuws830:~/Andrew$ ls -l
total 2971608
drwxr-xr-x 2 ubuntu ubuntu      12288 Apr  7 10:02  configs
drwxr-xr-x 3 ubuntu ubuntu       4096 May 18 14:08  demo
drwxrwxr-x 6 ubuntu ubuntu       4096 May 18 08:22  food11re
-rw-rw-r-- 1 ubuntu ubuntu 2941072286 May 18 08:13  food11.zip
-rw-rw-r-- 1 ubuntu ubuntu      11591 May 18 11:20 'Lab 4b.py'
-rw-rw-r-- 1 ubuntu ubuntu   54980021 May 18 10:25  lab4_model.pth
-rw-rw-r-- 1 ubuntu ubuntu      12220 May 18 09:53 'Lab 4.py'
drwxr-xr-x 5 root   root         4096 Apr  7 10:18  model_optimizer
-rw-rw-r-- 1 ubuntu ubuntu   46808011 May 18 11:21  Resnet_4.onnx
-rw-rw-r-- 1 ubuntu ubuntu        894 May 18 14:54  resnet4.yml
drwxr-xr-x 5 root   root         4096 Apr  7 10:07  samples
```

產生 onnx 檔後，我在 ubuntu 的環境下使用助教釋出的 docker 安裝 openvino，使用其中的 model optimizer 轉換出 IR 如下所示:

```
ubuntu ubuntu   46719680 May 18 11:32 Resnet_4.bin
ubuntu ubuntu       9029 May 18 11:32 Resnet_4.mapping
ubuntu ubuntu   46808011 May 18 11:21 Resnet_4.onnx
ubuntu ubuntu      39045 May 18 11:32 Resnet_4.xml
```

但是再利用 inference engine 產生 output

```
[ INFO ] Creating Inference Engine
[ INFO ] Loading network files:
       /opt/intel/openvino/deployment_tools/open_model_zoo/tools/accuracy_checker/Resnet_4.xml
       /opt/intel/openvino/deployment_tools/open_model_zoo/tools/accuracy_checker/Resnet_4.bin
[ INFO ] Preparing input blobs
[ WARNING ] Image /opt/intel/openvino/deployment_tools/open_model_zoo/tools/accuracy_checker/evaluation/00/0_0.jpg is re
[ INFO ] Batch size is 1
[ INFO ] Loading model to the plugin
[ INFO ] Starting inference in synchronous mode
[ INFO ] Processing output blob
[ INFO ] Top 10 results:
Image /opt/intel/openvino/deployment_tools/open_model_zoo/tools/accuracy_checker/evaluation/00/0_0.jpg

classid probability
------- -----------
  282     358.0269165
  750     347.6483765
  498     318.5369568
  74      309.9864807
  370     299.2701416
  629     286.1187744
  653     283.3037109
  939     263.3056946
  393     262.1462402
  211     255.5621490
```

Lab4(c)

Accuracy Checker tool 是 OpenVINO 中的評分工具，我使用助教的 dataset config 產生如下:

```
models:
  - name: Resnet_4

    launchers:
      - framework: dlsdk
        device: CPU
        model:   /opt/intel/openvino/deployment_tools/open_model_zoo/tools/accuracy_checker/Resnet_4.xml
        weights:  /opt/intel/openvino/deployment_tools/open_model_zoo/tools/accuracy_checker/Resnet_4.bin
        adapter: classification
        batch: 1

    datasets:
      - name: food11_dataset
        data_source: /opt/intel/openvino/deployment_tools/open_model_zoo/tools/accuracy_checker/evaluation
        annotation_conversion:
          converter: food11
          data_dir: /opt/intel/openvino/deployment_tools/open_model_zoo/tools/accuracy_checker/evaluation
          labels_file: /opt/intel/openvino/deployment_tools/open_model_zoo/tools/accuracy_checker/label_map_food11.txt
        annotation: food11_eva_annotation.pickle
        metrics:
        - type: accuracy
          top_k: 1
```

但是發生 array size 無法對齊

```
root@a2405dba9ff6:/opt/intel/openvino/deployment_tools/open_model_zoo/tools/accuracy_checker# accur
Processing info:
model: resnet4
launcher: dlsdk
device: CPU
dataset: food11_dataset
OpenCV version: 4.3.0-openvino
IE version: 2.1.42025
Loaded CPU plugin version:
    CPU - MKLDNNPlugin: 2.1.42025
Input info:
        Layer name: input
        precision: FP32
        shape [1, 3, 244, 244]

Output info
        Layer name: Gemm_68
        precision: FP32
        shape: [1, 1000]

  0%|
14:41:00 accuracy_checker ERROR: cannot reshape array of size 786432 into shape (1,3,244,244)
```

Lab4(d)

Pot 函數無法使用?

報錯如下:

accuracy_checker.config.config_validator.ConfigError: Each model must specify name, launchers, datasets

可是 config 檔中皆已經包含這些內容