Report for the COSE474 project2 (Implementing ResNet-50)

2016320187 컴퓨터학과 홍성윤

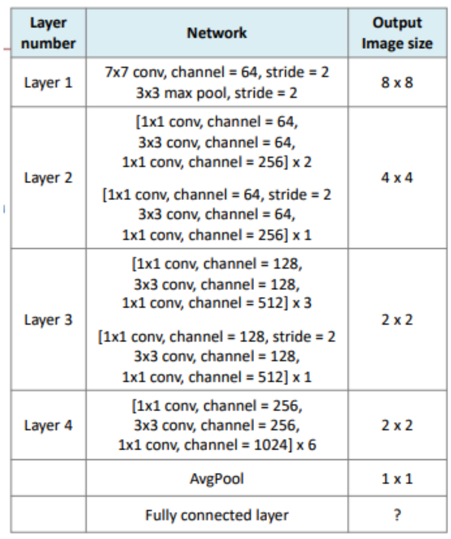
1. Code description (resnet50_skeleton.py)

(1) conv1x1, conv3x3 function

- each function applies 1 by 1 or 2 by 2 convolution filter and also applies batch normalization and ReLU non-linear function.

(2) ResidualBlock class

- According to next table, first and second conv layer in residual block have same output channel number. So, we can use in_channels parameter as input channel number of first conv layer and middle_channels as output channels number of first and second conv layer, and out_channels as output channels number of third conv layer.



Layer number	Network	Output Image size
Layer 1	7x7 conv, channel = 64, stride = 2 3x3 max pool, stride = 2	8 x 8
Layer 2	[1x1 conv, channel = 64, 3x3 conv, channel = 64, 1x1 conv, channel = 256] x 2 [1x1 conv, channel = 64, stride = 2 3x3 conv, channel = 64, 1x1 conv, channel = 256] x 1	4 x 4
Layer 3	[1x1 conv, channel = 128, 3x3 conv, channel = 128, 1x1 conv, channel = 512] x 3 [1x1 conv, channel = 128, stride = 2 3x3 conv, channel = 128, 1x1 conv, channel = 512] x 1	2 x 2
Layer 4	[1x1 conv, channel = 256, 3x3 conv, channel = 256, 1x1 conv, channel = 1024] x 6	2 x 2
	AvgPool	1 x 1
	Fully connected layer	?

- There are two types of ResidualBlock. One reduces activation map and the other doesn't. We can represent this by downsample parameter.
- If downsample is False, we should maintain activation map size. So, all strides of 3 conv layer should be 1 and padding of conv3x3 layer should be 1. Otherwise, activation map will be reduced. To apply residual connection, original and output size must be equal. Since downsample is False, size would not change. Use make_equal_channel function to make original and output channels number be equal.
- If downsample is True, stride of first conv1x1 layer should be 2 to reduce activation map. Since output size will be halved, we should reduce original x size to apply residual connection. We don't have to care about channel number during residual connection because downsample residual block always has same input and output channels. (From the table)

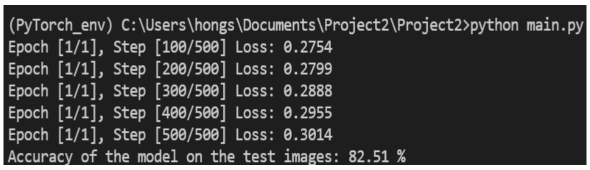
(3) layer1 in ResNet50_layer4 class

- We first apply 7x7 conv layer of which kernel size is 7 and stride 2. Input channel number will be 3 because we deal with RGB image data. Since the input image size is halved by this conv layer, padding should be 3.
- After applying batchNorm and ReLU, we apply maxpool(kernel size=3, stride=2). This maxpool must maintain input image size. So, 1 zero padding should be added.

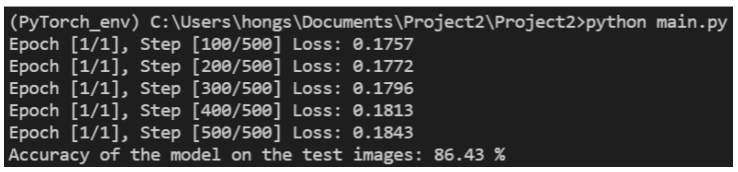
(4) The remainder of ResNet50_layer4 class

- Fill in the blank according to the table above. Each layer 2, 3, 4 has 3, 4, 6 residual blocks respectively. downsample parameter of last residual block of layer2 and layer3 should be True.
- Avgpool has kernel size 2, and stride 2, padding 0 to make input image size halved.
- Fc layer input size is 1x1x1024 (1024 is out_channels of the last residual block), and output size is 10 (the number of classes of our task).

2. Result and Discussions



```
(PyTorch_env) C:\Users\hongs\Documents\Project2\Project2>python main.py
Epoch [1/1], Step [100/500] Loss: 0.2754
Epoch [1/1], Step [200/500] Loss: 0.2799
Epoch [1/1], Step [300/500] Loss: 0.2888
Epoch [1/1], Step [400/500] Loss: 0.2955
Epoch [1/1], Step [500/500] Loss: 0.3014
Accuracy of the model on the test images: 82.51 %
```



```
(PyTorch_env) C:\Users\hongs\Documents\Project2\Project2>python main.py
Epoch [1/1], Step [100/500] Loss: 0.1757
Epoch [1/1], Step [200/500] Loss: 0.1772
Epoch [1/1], Step [300/500] Loss: 0.1796
Epoch [1/1], Step [400/500] Loss: 0.1813
Epoch [1/1], Step [500/500] Loss: 0.1843
Accuracy of the model on the test images: 86.43 %
```

- Result: Test accuracy of resnet-50(left), and vgg-16(right). The result came out as expected.
- Discussions: At first, I set the location of stride 2 in down-sampling res block wrongly by mistake. Intuitively, it seems this mistake will not affect the model, but the training loss was change almost 0.1. By this, I learned that this kind of parameter tuning is very sensitive and difficult job.
- It was quite interesting implementing the resnet architecture.