[COSE474] Deep Learning Project #2: CNN Architecture Implementation

2022320315 데이터과학과 김승주

Description of code

In the Question1, when Downsampling is needed(downsample=True), I use conv1x1 with stride=2 and padding=0, downsampling spatial dimensions. Next conv3x3 with stride=1, padding=1 and conv1x1 with stride=1, padding=0 keep spatial dimensions unchanged. When Downsampling is not needed(downsample=False), this part doesn't perform downsampling, maintaining spatial dimensions.

In the Question2, I applies a 7x7 convolution with a stride=2 and padding=3, reducing the spatial dimensions by half. Also, Max pooling with a kernel_size=3, stride=2, and padding=1 reduces the spatial dimensions by half. Output of the layer1 has a spatial dimension of 8x8 with channel=64. In layer 2 and 3, reduce an activation map in half by using "downsampling=True" at the last ResidualBlock. Output of the layer2 has a spatial dimension of 4x4 with channel=256 and output of the layer3 has a spatial dimension of 2x2 with channel=512. Layer4 keeps spatial dimension and channel=1024. The reason for having this FC layer is to convert the network's output into class scores for the classification task. The input dimension is set to 1024, which corresponds to the size of the feature maps from the last layer. The output dimension is the total number of classes to classify. AvgPool with stride=2 reduces the spatial dimensions by half.

Results

vgg16_full

```
Downloading https://www.cs.toronto.edu/~kriz/cifar-10-python.tar.gz to ../osproj/data/cifar-10-python.tar.gz
100%|██████████| 170498071/170498071 [00:03<00:00, 42904476.27it/s]
Extracting ../osproj/data/cifar-10-python.tar.gz to ../osproj/data/
Epoch [1/1], Step [100/500] Loss: 0.1792
Epoch [1/1], Step [200/500] Loss: 0.1848
Epoch [1/1], Step [300/500] Loss: 0.1838
Epoch [1/1], Step [400/500] Loss: 0.1848
Epoch [1/1], Step [500/500] Loss: 0.1884
Accuracy of the model on the test images: 85.82 %
```

resnet50

```
Epoch [1/1], Step [100/500] Loss: 0.2788
Epoch [1/1], Step [200/500] Loss: 0.2811
Epoch [1/1], Step [300/500] Loss: 0.2890
Epoch [1/1], Step [400/500] Loss: 0.2940
Epoch [1/1], Step [500/500] Loss: 0.2986
Accuracy of the model on the test images: 83.16 %
```

Discussions

I tried to keep in mind the changes in spatial dimensions and channels and I can understand the characteristics of downsampling and ResNet. Maybe I can employ several strategies to improve resnet50, like using more data, scheduling learning rate, tuning optimization algorithm etc. However, the effectiveness of each strategy may vary depending on the case. So I should try various strategies and make evaluation to find optimal performance.