Name: DO Thuy Trang

SID: 20549272

**Q1:**

**From the provided table, we can calculate the size of feature map according to the formula:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Layer | Kernel size | # of kernels | Stride | Padding | Size of feature map |
| Input | 1x32x32 image | | | | 32x32 |
| Conv | 3x3 | 32 | 1 | 1 | 32x32 |
| Conv | 3x3 | 32 | 1 | 1 | 32x32 |
| Max pooling | 2x2 |  | 2 | 0 | 16x16 |
| Conv | 3x3 | 64 | 1 | 1 | 16x16 |
| Conv | 3x3 | 128 | 1 | 1 | 16x16 |
| Conv | 3x3 | 256 | 1 | 1 | 16x16 |
| Conv | 3x3 | 512 | 1 | 1 | 16x16 |
| Average pooling | 16x16 |  | 1 | 0 | 1x1 |
| Output | 1x 512 vector | | | | |

Therefore, the correct kernel size in the average pooling layer is 16x16

**Q2:**

**Dropout in deep learning:** is a technique for regularization to avoid overfitting.

For each hidden unit or input of each training step, we will have probability ***p*** to temporarily dropped out this unit.

**Dropout works because:** when some units or inputs are randomly dropped out, the network cannot rely on just a few units, but it need to spread out the weights and rely on more other hidden units or inputs.

**Q3:**

When aggregation function is absolute difference, total number of trainable parameters is 1,842,465

**Q4:**

When aggregation function is concatenation, total number of trainable parameters is 2,104,609

**Q5:**

****

According to the plot, we should stop at about step 600th because at this state the loss of training and validation set are smallest

**Q6:**

Optimal threshold: 0.54

Corresponding validation accuracy: 0.7986

**Q7:**

**-** A photo application used to identify similar-looking family members: we should set low threshold because we may have about 5 people share the same result as family members. When we set the threshold to be low, the similarity could be low, therefore we get more people that are similar-looking family members.

**-** A photo authentication used for mobile banking: we should set high threshold because we only want 1 person for each identity information. When we get the threshold to be high, the similarity will be high, and we set less persons looking similar.

**Q8:**

1. **Optimization**

|  |  |  |
| --- | --- | --- |
| Setting | Accuracy |  |
| Learning rate: 0.0001  Optimizer: Adam  Training epochs: 30  Best threshold: 0.34 | 0.7629 | File name: PA2\_COMP4211\_Q8A\_Optimization\_1 |
| Learning rate: 0.001  Optimizer: SGD  Training epochs: 20  Best threshold: 0.46 | 0.5700 | File name: PA2\_COMP4211\_Q8A\_Optimization\_2 |
| Learning rate: 0.0001  Optimizer: Adam  Training epochs: 25  Best threshold: 0.44 | 0.7543 | File name: PA2\_COMP4211\_Q8A\_Optimization\_3 |

**D. Architecture:**

1. **VGG – 11**

**Convolutional Layers:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Layer | Kernel size | # of kernels | Stride | Padding | Size of feature map |
| Input | 3x64x64 image | | | | 64x64 |
| Conv | 3x3 | 32 | 1 | 1 | 32x32 |
| Conv | 3x3 | 32 | 1 | 1 | 32x32 |
| Max pooling | 2x2 |  | 2 | 0 | 16x16 |
| Conv | 3x3 | 64 | 1 | 1 | 16x16 |
| Conv | 3x3 | 128 | 1 | 1 | 16x16 |
| Conv | 3x3 | 256 | 1 | 1 | 16x16 |
| Conv | 3x3 | 512 | 1 | 1 | 16x16 |
| Average pooling | 16x16 |  | 1 | 0 | 1x1 |
| Output | 1x 512 vector | | | | |

**Aggregation Function:** takes in two input vectors of size 512 and outputs the element-wise absolute difference vector of size 512.

**Fully Connected Layers:**

|  |  |  |
| --- | --- | --- |
|  | Input features | Output Features |
| FC1 | 512 | 512 |
| Dropout p = 0.5 |  |  |
| FC2 | 512 | 512 |
| FC3 | 512 | 1 |

2**. ResNet 18:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Layer | Kernel size | # of kernels | Stride | Padding | Size of feature map |
| Input | 3x64x64 image | | | | 64x64 |
| Conv | 7x7 | 64 | 1 | 3 | 64x64 |
| MaxPool | 3x3 |  | 2 |  | 32x32 |
| Conv | 3x3 | 64 | 1 | 1 | 32x32 |
| Conv | 3x3 | 64 | 1 | 1 | 32x32 |
| Conv | 3x3 | 64 | 1 | 1 | 32x32 |
| Conv | 3x3 | 64 | 1 | 1 | 32x32 |
| Conv | 3x3 | 128 | 2 | 1 | 16x16 |
| Conv | 3x3 | 128 | 1 | 1 | 16x16 |
| Conv | 3x3 | 128 | 1 | 1 | 16x16 |
| Conv | 3x3 | 128 | 1 | 1 | 16x16 |
| Conv | 3x3 | 256 | 2 | 1 | 8x8 |
| Conv | 3x3 | 256 | 1 | 1 | 8x8 |
| Conv | 3x3 | 256 | 1 | 1 | 8x8 |
| Conv | 3x3 | 256 | 1 | 1 | 8x8 |
| Conv | 3x3 | 512 | 2 | 1 | 4x4 |
| Conv | 3x3 | 512 | 1 | 1 | 4x4 |
| Conv | 3x3 | 512 | 1 | 1 | 4x4 |
| Conv | 3x3 | 512 | 1 | 1 | 4x4 |
| AvgPool | 4x4 | 512 | 1 | 1 | 1x1 |
| Output | 1x 512 vector | | | | |

**Aggregation Function:** takes in two input vectors of size 512 and outputs the element-wise absolute difference vector of size 512.

**Fully Connected Layers:**

|  |  |  |
| --- | --- | --- |
|  | Input features | Output Features |
| FC1 | 512 | 512 |
| Dropout p = 0.5 |  |  |
| FC2 | 512 | 512 |
| FC3 | 512 | 1 |

3**. AlexNet:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Layer | Kernel size | # of kernels | Stride | Padding | Size of feature map |
| Input | 3x227x227 image | | | | 227x227 |
| Conv | 11x11 | 64 | 4 | 0 | 55x55 |
| MaxPool | 3x3 |  | 2 | 0 | 27x27 |
| Conv | 5x5 | 256 | 1 | 2 | 27x27 |
| MaxPool | 3x3 |  | 2 |  | 13x13 |
| Conv | 3x3 | 384 | 1 | 1 | 13x13 |
| Conv | 3x3 | 384 | 1 | 1 | 13x13 |
| MaxPool | 3x3 |  | 2 |  | 6x6 |
| AvgPool | 6x6 |  | 1 |  | 1x1 |
| Output | 1x 256 vector | | | | |

**Aggregation Function:** takes in two input vectors of size 256 and outputs the element-wise absolute difference vector of size 512.

**Fully Connected Layers:**

|  |  |  |
| --- | --- | --- |
|  | Input features | Output Features |
| FC1 | 256 | 512 |
| Dropout p = 0.5 |  |  |
| FC2 | 512 | 512 |
| FC3 | 512 | 1 |

**Comparison among 3 models:**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Setting | Accuracy |  |
| VGG-11 | Learning rate: 0.001  Optimizer: Adam  Training epochs: 20  Best threshold: 0.48 | 0.8186 | File name: PA2\_COMP4211\_Q8D\_ Architecture \_1\_VGG\_11 |
| ResNet-18 | Learning rate: 0.001  Optimizer: Adam  Training epochs: 20  Best threshold: 0.44 | 0.7300 | File name: PA2\_COMP4211\_Q8D\_ Architecture \_2\_ResNet\_18 |
| AlexNet | Learning rate: 0.001  Optimizer: Adam  Training epochs: 20  Best threshold: 0.42 | 0.8186 | File name: PA2\_COMP4211\_Q8D\_ Architecture \_3\_AlexNet |