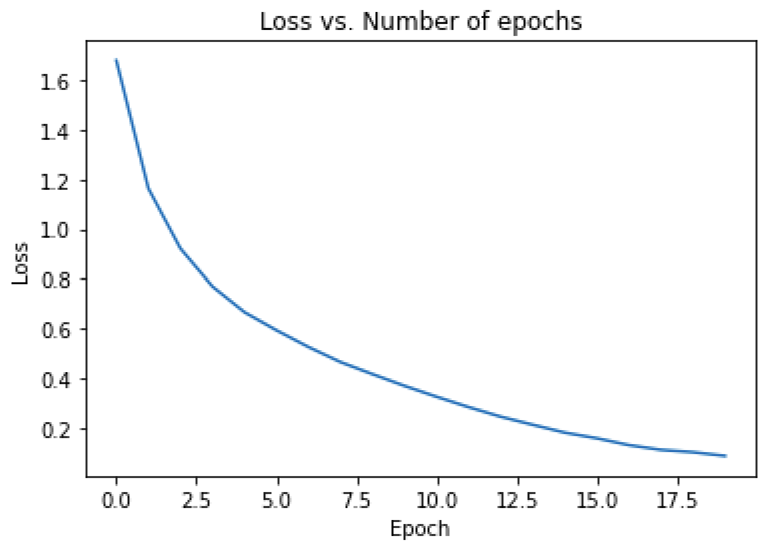
Exercise 3 - Report

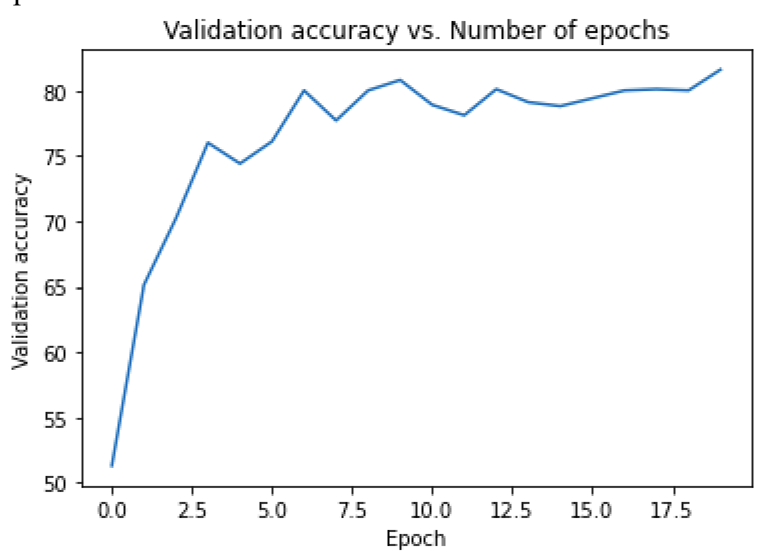
1. **Implement Convolutional Network (10 points)**

   a. In our architecture, there are five convolution blocks and in each block, the computations are in the order of convolution, pooling, ReLU. The first convolution block takes 3 input channels and produces 128 output channels and halves the feature size. Similarly, other convolution blocks are designed using given hyperparameters which give $512 \times 1 \times 1$ feature map in the end. At last a fully connected layer is used for classification.

   During training, loss starts decreasing from 1.6872 in the first epoch and decreases to 0.1183 in the final epoch.



   Validation accuracy starts improving from 51.3 % in the first epoch to 81.6 % in the final epoch.



Training loss decreases and validation accuracy improves with each epoch as follows:

```
Epoch [1/20], Step [100/245], Loss: 1.6872
Epoch [1/20], Step [200/245], Loss: 1.2512
Validataion accuracy is: 51.3 %
Epoch [2/20], Step [100/245], Loss: 1.1082
Epoch [2/20], Step [200/245], Loss: 0.9329
Validataion accuracy is: 65.1 %
Epoch [3/20], Step [100/245], Loss: 0.8105
Epoch [3/20], Step [200/245], Loss: 0.8719
Validataion accuracy is: 70.2 %
Epoch [4/20], Step [100/245], Loss: 0.6897
Epoch [4/20], Step [200/245], Loss: 0.7370
Validataion accuracy is: 76.0 %
Epoch [5/20], Step [100/245], Loss: 0.6915
Epoch [5/20], Step [200/245], Loss: 0.5746
Validataion accuracy is: 74.4 %
Epoch [6/20], Step [100/245], Loss: 0.5948
Epoch [6/20], Step [200/245], Loss: 0.5318
Validataion accuracy is: 76.1 %
Epoch [7/20], Step [100/245], Loss: 0.4429
Epoch [7/20], Step [200/245], Loss: 0.4901
Validataion accuracy is: 80.0 %
Epoch [8/20], Step [100/245], Loss: 0.4501
Epoch [8/20], Step [200/245], Loss: 0.4144
Validataion accuracy is: 77.7 %
Epoch [9/20], Step [100/245], Loss: 0.4455
Epoch [9/20], Step [200/245], Loss: 0.4448
Validataion accuracy is: 80.0 %
Epoch [10/20], Step [100/245], Loss: 0.3424
Epoch [10/20], Step [200/245], Loss: 0.3305
Validataion accuracy is: 80.8 %
Epoch [11/20], Step [100/245], Loss: 0.3408
Epoch [11/20], Step [200/245], Loss: 0.2741
Validataion accuracy is: 78.9 %
Epoch [12/20], Step [100/245], Loss: 0.1966
Epoch [12/20], Step [200/245], Loss: 0.3790
Validataion accuracy is: 78.1 %
Epoch [13/20], Step [100/245], Loss: 0.2063
Epoch [13/20], Step [200/245], Loss: 0.2077
Validataion accuracy is: 80.1 %
Epoch [14/20], Step [100/245], Loss: 0.2344
Epoch [14/20], Step [200/245], Loss: 0.2241
Validataion accuracy is: 79.1 %
Epoch [15/20], Step [100/245], Loss: 0.1934
Epoch [15/20], Step [200/245], Loss: 0.2430
Validataion accuracy is: 78.8 %
Epoch [16/20], Step [100/245], Loss: 0.1142
Epoch [16/20], Step [200/245], Loss: 0.1883
Validataion accuracy is: 79.4 %
Epoch [17/20], Step [100/245], Loss: 0.1039
Epoch [17/20], Step [200/245], Loss: 0.1223
Validataion accuracy is: 80.0 %
Epoch [18/20], Step [100/245], Loss: 0.0810
Epoch [18/20], Step [200/245], Loss: 0.1021
Validataion accuracy is: 80.1 %
Epoch [19/20], Step [100/245], Loss: 0.0729
Epoch [19/20], Step [200/245], Loss: 0.1240
Validataion accuracy is: 80.0 %
Epoch [20/20], Step [100/245], Loss: 0.0614
Epoch [20/20], Step [200/245], Loss: 0.1183
Validataion accuracy is: 81.6 %
```

b. The `number of trainable parameters in the model` are `7678474.`

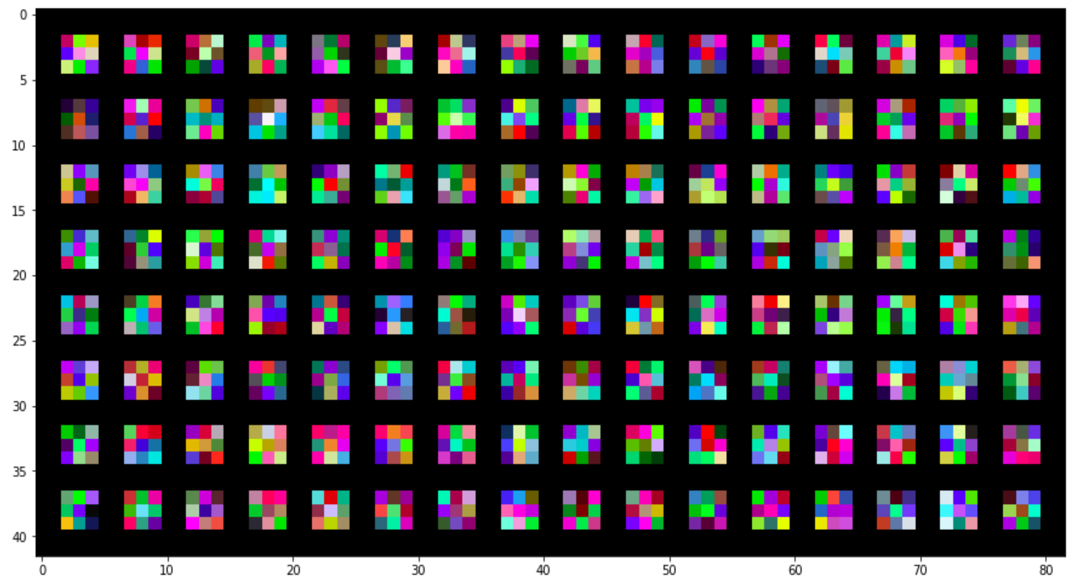
Zhang et al. (2017) in their paper “Understanding deep learning requires generalization”, shows that a simple two-layer neural network with 2n+d parameters is capable of perfectly fitting any dataset of n samples of dimension d.

However, deep neural networks have much more than 2n+d parameters, they do not necessarily overfit.

c. As it is a first convolution layer, it is very difficult to find any pattern in the filters before and after training. But we can notice a significant change in weights of filters after training.

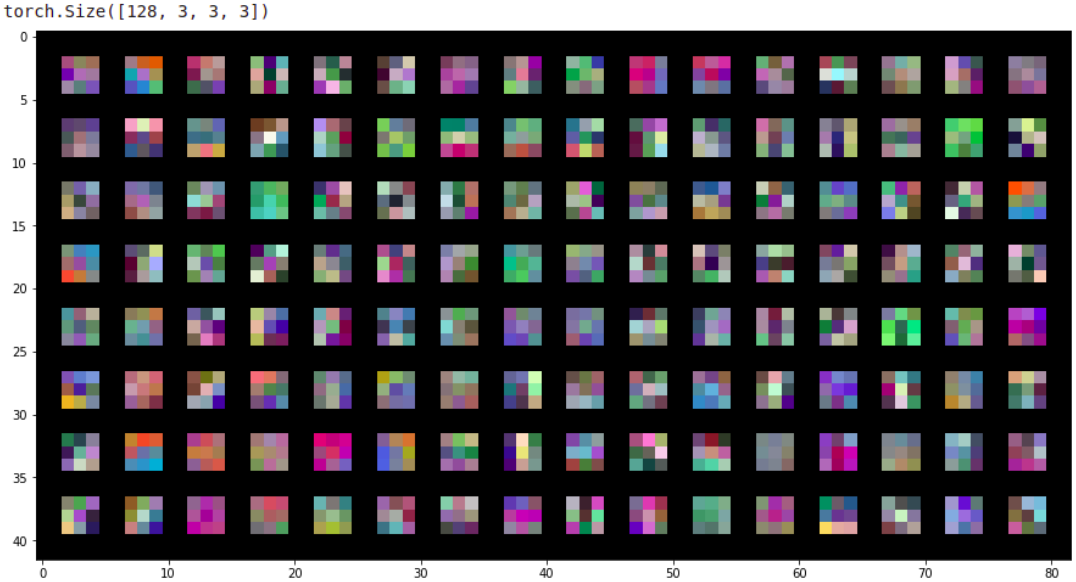
Before training :

`torch.Size([128, 3, 3, 3])`



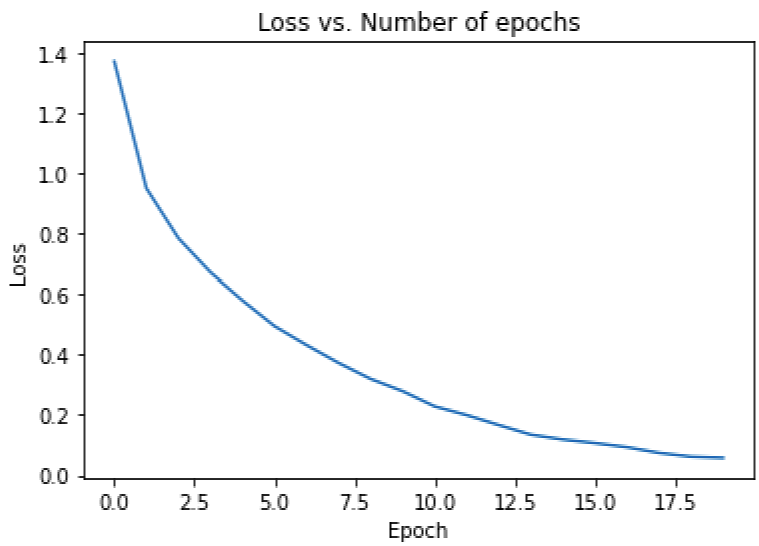
After training :

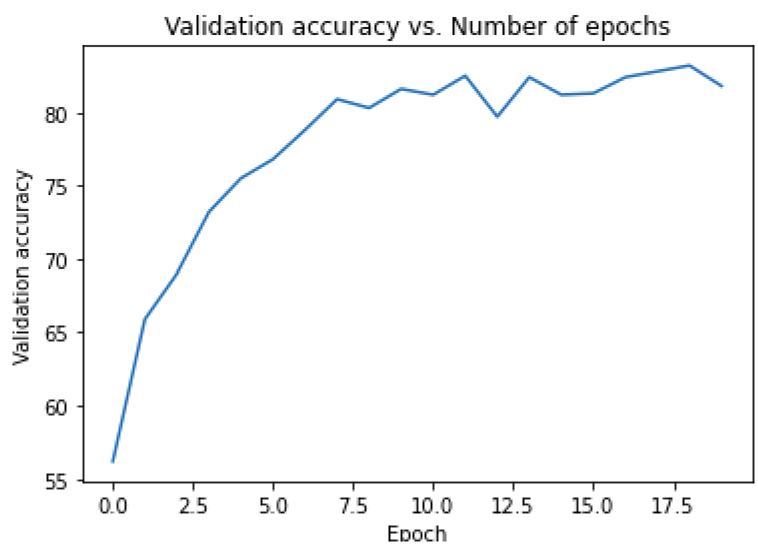
`torch.Size([128, 3, 3, 3])`



2. Improve training of Convolutional Networks (10 points)

a. Batch normalization forces earlier layers to not shift around much, because they're constrained to have the same mean and variance. This makes the later layers of the network learn more independently of other layers, and this speeds up the learning of the whole network. This is evident from following loss and validation accuracy curves on training data:





Loss decreases and validation accuracy increases significantly faster with batch normalization as compared to our implementation in Q1 - a - without batch normalization. With the model in Q2 - a, we get validation accuracy of more than 80 % without rapid changes in loss and accuracy in eighth epoch unlike the model in Q1 - a which tends to change drastically.