## Question 1.a)

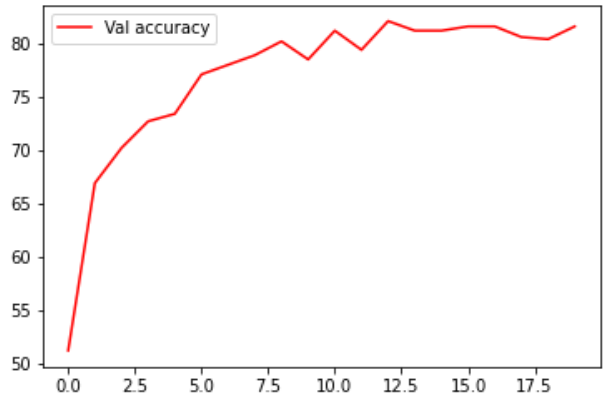
As we can see from the Loss and Accuracy figures there are clear signs of overfitting. 20 epochs are redundant for training. We can conclude that stopping the trainig at $10^{th}$-$13^{th}$ epoch can give better results.

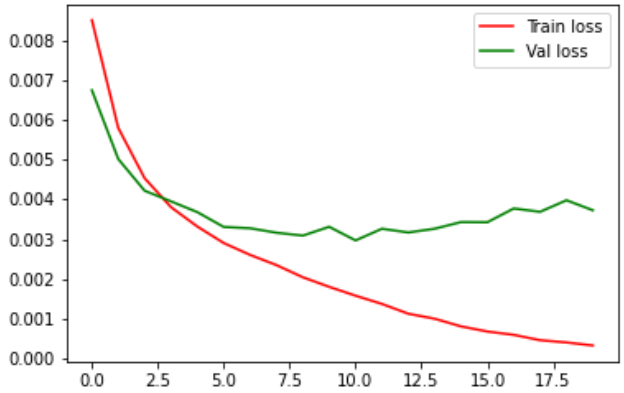
Epoch [1/20], Step [100/245], Loss: 1.6899
Epoch [1/20], Step [200/245], Loss: 1.2536
Validation accuracy is: 51.2 %

Epoch [13/20], Step [100/245], Loss: 0.2135
Epoch [13/20], Step [200/245], Loss: 0.1860
Validation accuracy is: 82.1 %

Epoch [20/20], Step [100/245], Loss: 0.0551
Epoch [20/20], Step [200/245], Loss: 0.0761
Validation accuracy is: 81.6 %

Accuracy of the network on the 1000 test images: 79.5 %




## Question 1.b)

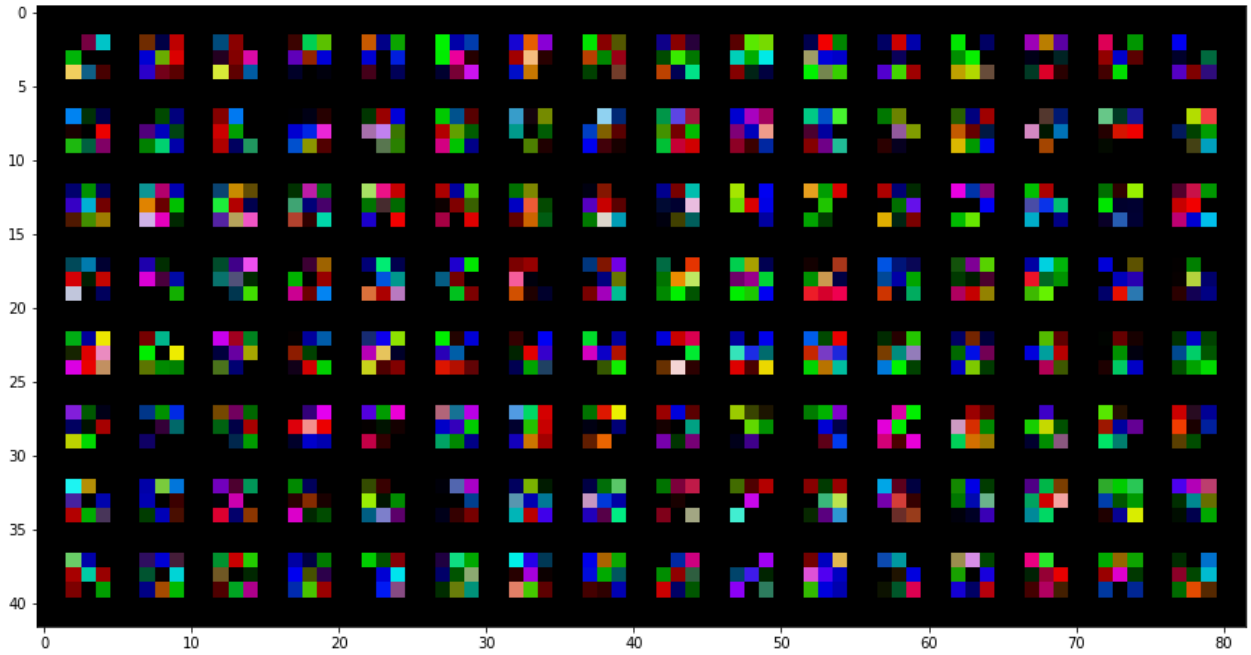
==========MODEL PARAMETERS==========

| | |
|---|---|
| Parameter name: layers.0.weight | Parameter shape: ([128, 3, 3, 3]) |
| Parameter name: layers.0.bias | Parameter shape: ([128]) |
| Parameter name: layers.3.weight | Parameter shape: ([512, 128, 3, 3]) |
| Parameter name: layers.3.bias | Parameter shape: ([512]) |
| Parameter name: layers.6.weight | Parameter shape: ([512, 512, 3, 3]) |
| Parameter name: layers.6.bias | Parameter shape: ([512]) |
| Parameter name: layers.9.weight | Parameter shape: ([512, 512, 3, 3]) |
| Parameter name: layers.9.bias | Parameter shape: ([512]) |
| Parameter name: layers.12.weight | Parameter shape: ([512, 512, 3, 3]) |
| Parameter name: layers.12.bias | Parameter shape: ([512]) |
| Parameter name: classifier.0.weight | Parameter shape: ([10, 512]) |
| Parameter name: classifier.0.bias | Parameter shape: ([10]) |

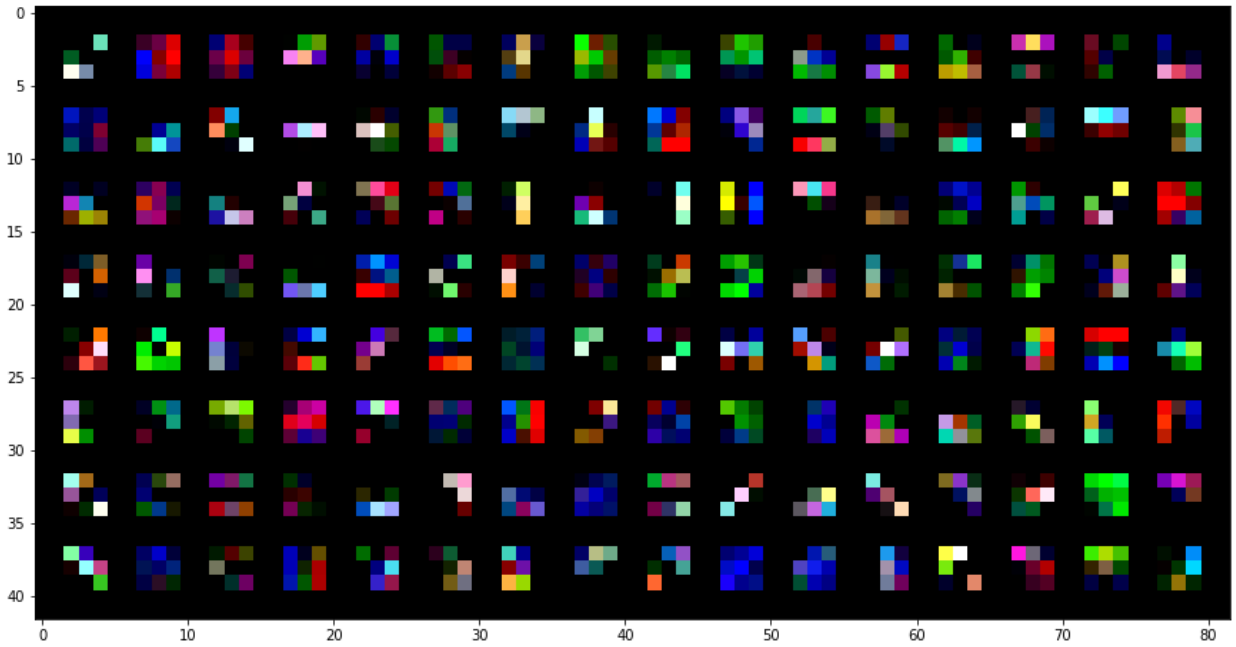
numberOfParameters: 7678474

Question 1.c)

Before:




After:




Here we visualized weights of the first convolutional layer before and after training. After training the model detects color patterns (mostly the green squares, but also blue) and diagonal[position: (8,1),(8,10),(5,5)], vertical [position: (3,7);(3,9);(3,10);(4,12)] and horizontal [position: (2,4),(2,11),(7,3),(7,9)] shapes. As we can see the model mostly retains the noisy view of the filters, which is another sign of overfitting.