CSC321 a2
Guanxiong Liu
liuguanx 1002077726

A.
1. There are six convolutional layers: downconv1, downconv2, rfconv, upconv1, upconv2, finalconv

	Filter size	Number of filters
Downconv1	3x3	32
Downconv2	3x3	64
Rfconv	3x3	64
Upconv1	3x3	32
Upconv2	3x3	3
Finalconv	3x3	3

2. The predictions of the images are not very good. The model favors a brownish color. Thus, all the predictions are more brownish. It predicts the RGB color poorly.

3. Since RGB only contains 3 channels and has no information about light intensity. If two pictures have the same RGB value but the light intensity is different. They should have great distance. However, if we use RGB, it contains no information about this. Therefore, using MSE on RGB color space would cause larger loss and would be more problematic.

4. Since the model in classification problem is more robust than in regression problem. The further you are away from the mean, the proportionally more you will be penalized. This is not what we want in colorization. We do not want to penalize severely for the points that are far away.

B.
1. Please see the code.

2. It is so much better than the previous model (regression). We can see some distinct colors in our predictions which were not shown in the previous model. However, there still some pixels desaturated.

CSC321 a2
Guanxiong Liu
liuguanx 1002077726

C.
1. Please see the code.

2. Please see the results attached.
CNN:

UNet:

3. The result is better than CNN. It predicts some colors a bit more accurately. For instance, in our second last set of images UNet predicted more detail on the ground than CNN and is closer

to the ground truth. CNN's prediction is blur. Also, UNet improves both validation loss and accuracy. UNet does a good job in improve the output qualitatively. The concatenation with the output of the same level layer will restore some original detail about the image and make the prediction more precisely. (i.e downconv1 and upconv2 are on the same level). Also, Unet does a good job in localization. This provides a significant segmentaion accuracy improvement compared to CNN. These are the reason Unet improves the performance of CNN. The below graphs also show that UNet's loss is less than CNN.
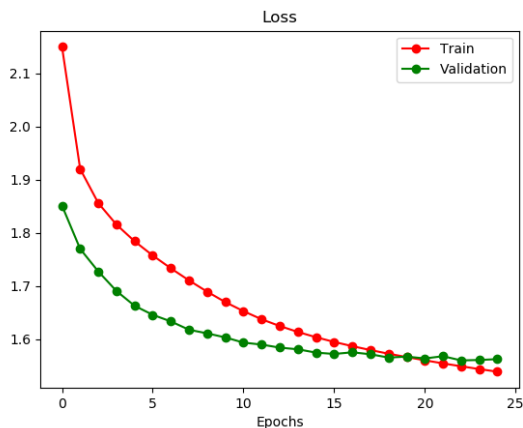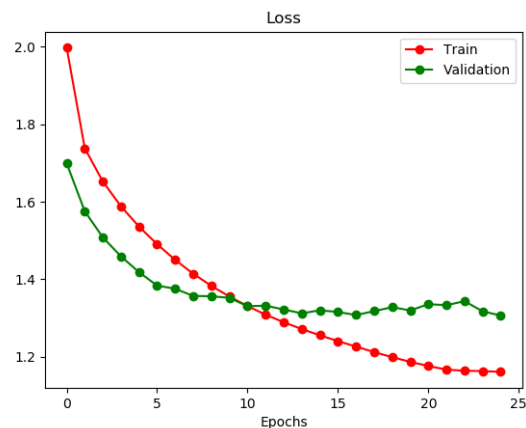
CNN.   UNet.

D.
1. (a) There are 3*3*C weights. The receptive field size is 3*3.
   (b) There are 5*5*C weights. The receptive field size is 5*5.
   (c) There are 3*3*C weights. The receptive field size is 5*5.
   Where C stands for number of channels.

2. Since other convolutional layer either do pooling or upsampling. Also, after rfconv layer, each upsampling's input has concatenation which makes the input bigger and dilation can help increase receptive field while remaining the number of weights the same. Thus, we replace middle layer with dilatedUNet.

E.
1. The first few layers have higher resolution and we can easily identify the objects from the first layer (dense). As we move from the first layer to the middle layer, the object becomes less identifiable (sparse). When we move from the middle layer to the last layer, the object again become more identifiable, but it does not have the same resolution as the first layer's output.

2. CNN has more noise and UNet has higher contrast ratio. It is not as smooth as UNet model (clear edge). Hence the object is easier to be identified in UNet's model.

F.

CSC321 a2

Guanxiong Liu

liuguanx 1002077726

1. (a) and (b) are helpful, because it has different orientations. It can help generate different feature maps from the previous maps. Thus, it can help train the model. (c) and (d) are not helpful, since the small translation is invariant to convolutional layer. (e) might be helpful, since it depends on what class do you choose. If you choose zebras to train horse model, it is acceptable since they have similar feature. If you choose pigs to train horse model, it will confuse the model.

2. Learning rate. Number of epochs. Batch size. Activation function. Number of hidden layers and units. Weight initialization.

G.

1.Please see the code.

2. There is no much improvement compared to UNet without dilation. Loss for UNet is 1.36, accuracy for UNet is 48.0%. For DUNet, the loss is 1.45, accuracy is 45.4%. So DUNet even increase the loss and decrease the accuracy. Dilation is used when you want to merge spatial information across the inputs much more aggressively with fewer layers.