CSC421 PA2

Yubo Wang 1002138377

A.

1. 6 layers

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

2. 25 epochs
3.

```
Epoch [10/10], Loss: 0.0102
Epoch [10/10], Val Loss: 0.0100
Epoch [50/50], Loss: 0.0066
Epoch [50/50], Val Loss: 0.0067
```

We can see that as epoch increases, loss will decreases

4. If we use RGB color space, the luminance/light level of the color cannot be displayed correctly based on the prediction. Besides, in RGB color space, even though square distance between color A and color C, color B and color C are same, color A and color B cannot be guaranteed to be same color, which may leads to a big problem.

5. When use regression model, we minimize the squared error to predict, which is inefficient and may cause overfitting of the model.

   When use classification, more color can be predicted based on training data even though they may be far from each other in color space.

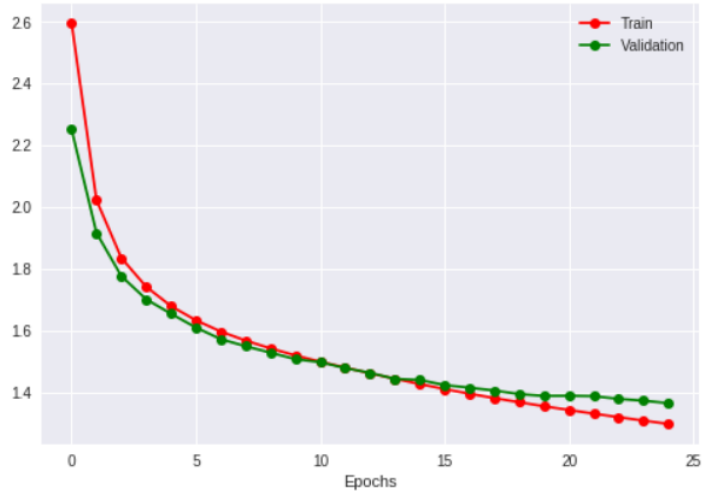
B.

Q2:

The color predicted is more accurate than the previous model. However, for some parts the are of the same color is predicted incorrectly.

C.

Q2:



The result compare to the previous model is better. Both training and validation loss decrease and accuracy increase. The color accuracy and shape are more accurate.

From my perspective, the result improves since

1. the image details are lost when there are too many layers in the network.
2. By skipping connections. The initial greyscale input image can be used as input to the final layer which provides more image details.

Q3:

As batch size increase, the loss increases, the accuracy decreases, the training time decreases and the quality of image decreases. The validation loss will be larger than training less when batch size is too large.

D.

Q1:

The resolution of the input is less than the resolution of the output.

Q2:

The result from CNN are more accurate in color and sharp compared with the bilinear interpolation result.

Reasons:

1. During bilinear interpolation, the image is blurred, which lead to bad image result.
2. Based on how bilinear interpolation is implemented, the input cell center close to output cell center will have more influence on the output value. Therefore, it should not be used on categorical data.

E.

Q1:

First few layers focus on details of images such as eyes, nose, mouth of the horse.

For the later layers, they will focus on big image pattern, like separating the horse from the background. The last layer will be the final prediction result.

Q2:

For activations in UNet, The first few layers do the same thing as activations in CNN do (focus on detail and specific parts). The last few layers also focus on detail too compare with CNN's since skip connections connects the output of first layer to back layers, which leads to the result that outputs from UNet are more detailed and sharp.

Q3:

The activations between these two are not so different. First few layers focus on detail and last few layers focus on details and the whole image. However, predictions of super-resolution is blurry since it is used to create high-resolution image by using low resolution input.

F.

Q1:

Number of filters, size of filters (number of colors), kernel size, learning rate, activation function.

Q2:

The output will not change after switching, but the running times will be different since max pooling reduce the size of input. We can save computation time by putting max pooling before ReLU.

Q3:

Using perceptual losses may lead to better result. The result produced by using per-pixel losses do not capture perceptual differences between output and ground-truth images. Perceptual losses measure image similarities more robustly than per –pixel. It can match over all pattern instead of comparing pixel and pixel, which produce more accurate image when assessed by human.

Q4:

We can either change number of filters to the size required to fit larger image, or we can downsample the input by applying a pooling layer.