CSC321 A2

February 2018

1 Part A

1.1 Q1

The model Regression CNN has 6 convolution layers.

Table 1: modelRegressionCNN

layer level	filter sizes	number of filters
downconv1	3	32
downconv2	3	64
rfconv	3	64
upconv1	3	32
upconv2	3	3
finalconv	3	3

1.2 Q2

No. Because the prediction pictures are not well coloured, which look pretty similar to the gray-scale ones.

1.3 Q3

Because RGB space stores three colours, under L2 distance, a color with high value in red, low in blue would be computed as close to a color with high value in blue, low in red, where value in blue is the same. i.e. Different colours might have the same L2 distance, which implies the same loss.

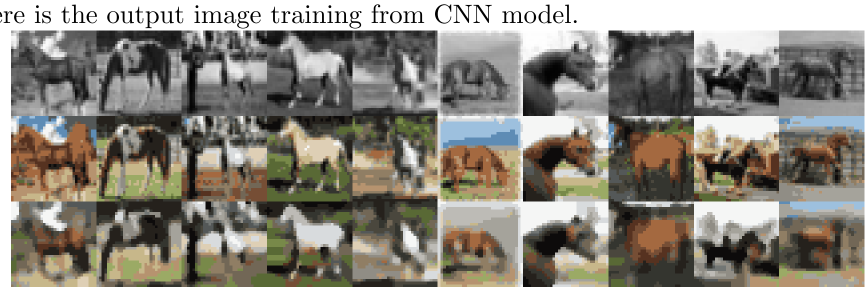
1.4 Q4

Because L2 distance encourages the sum of squared distance to be as small as possible, while using squared the distance between two colors, it does not tell apart positive or negative difference. But here in color scale, different numbers mean different colors. Therefore, it's better to frame as a classification problem because each pixel is either one color or the other, it's better to give a label to each pixel.

2 B

2.1 Q2.

Here is the output image training from CNN model.



It looks much better than the previous model. The horse is coloured properly in most cases.

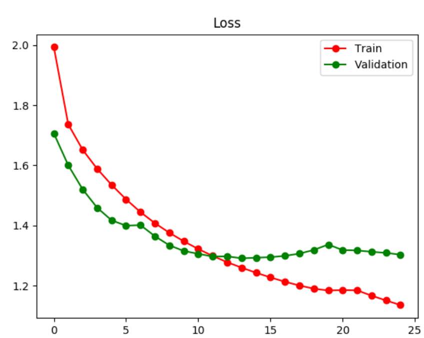
3 C

3.1 Q1

See code

3.2 Q2

Training curve with 10 batch size for 25 epochs:



3.3 Q3

Here is the output image from UNet model using pre-trained weights provided.

From the above results, we see that the picture now has sharper edges (clearer to see the boundary of horse) and more colours. Both validation loss and accuracy are improved.

```
Evaluating Model CNN: weights/cnn_k3_f32.pkl
Val Loss: 1.5881, Val Acc: 41.1%

Evaluating Model UNet: weights/unet_k3_f32.pkl
Val Loss: 1.3659, Val Acc: 48.0%
```

The skip connections did improve the output qualitatively, because 1) We lose details of image as the network adds more convolutional layers, making it harder to recover later. Adding the skip connection help feature maps obtain more image details from comparatively original stage. 2) From the perspective of backpropgation's speed, adding skip connections makes it easier to train a deeper network. For example, to back-propogate gradient, the last layer is directly connected to input layer and skipped all layers in between.

4 D

4.1 Q1

a) 9, 9 b) 25, 25,c) 9, 25 From the result, we see that a dilated convolution did increase the size of receptive without increasing the number of weights.

4.2 Q2

Because this convolution has the smallest input and output dimension(8*8 and 8*8), which implies that each pixel contains the most information from the original input compared to other layers, i.e. adding dilation on this layer allows small filter (3*3) to cover information over large regions of the original image.

5 E

5.1 Q1

From the first several activation layers of CNN, the number of feature maps increases and the resolution decreases along with dimension of each feature map, which becomes less clearer to see the shape of object.

From the later layers, the number of feature maps decreases and resolution increases along with dimension of each feature map.

5.2 Q2

downconv1, downconv2 and refconv looks pretty much the same as in the case of CNN model, but the later layers look better, as the output looks clearer than before.

6 F

6.1 Q1

(b), (c), (d) can be helpful when training set is small, because flipping image left to right and shifting pixel left/right or up/down can increase the effective size of the training set, and make it more likely that any given test example has a closely related training example. While it doesn't make much sense to flip it up-side-down as it's not normal to see a horse in that position.

6.2 Q2

Here are the five hyperparameters that can be tuned:

- padding size
- kernel size
- stride size
- max pooling kernel size
- upsampling scale factor

7 G

7.1 Q2

Here is the output image from the DUNet model using pre-trained weights provided, with self implemented forward method.

In this case, the model achieves a validation loss of 1.4353, and validation accuracy 45.8%, both of which are better than the CNN's result and worse than the UNet's result. So quantitatively, this model is better than CNN, but not UNet. Dilation might be more helpful in a low resolution picture.