A2 writeup

Josh Gelua 1002615805

February 26, 2020

1 Part A

1.1

1.2

These results are not effective at colouring, but at least they are recognizable.

The trained results are not coloured in, and they are also lower resolution than the original image.

1.3

The following table is based on the assumptions that the conv kernel is 3 * 3 and number of input channels is 1.

Layer	Weight	Output	Connection
1	3 * 3 * NF	32 * 32 * 1NF	9NF * 32 * 32
2	3 * 3 * $2NF^2$	16 * 16 * 2NF	$18NF^2 * 16 * 16$
3	3 * 3 * $4NF^2$	8 * 8 * 2NF	$36NF^2 * 8 * 8$
4	3 * 3 * $2NF^2$	8 * 8 * 1NF	$18NF^2 * 8 * 8$
5	3 * 3 * NF * NC	16 * 16 * 1NC	9NF*NC * 16 * 16
6	3 * 3 * NC^2	32 * 32 * 1NC	$9NC^2 * 32 * 32$

1.4

On inspection of replacing x with y = ax + b, with regards to scalar a, it training loss stays generally the same, validation loss however has a greater variance as scalar a increases, and with regards to b, similar results occur as increasing scalar a, however, the first Epoch generally has a much higher validation loss that skews the graph.

2 Part B

2.1

2.2

2.3

Results of this model is more accurate than the previous model. As shown, the validation loss and accuracy rates are much better than before. Qualitatively, I can see that the trained results are somewhat coloured in (compared to not at all) and the resolution of the image has sharpened and become more clear.

I can think of two reasons this may be the case: - Firstly, skip connections allow the information of the earlier layers to pass the final layers, as gradient information would otherwise be diluted as they are forwarded through each layer. - Secondly, they combine outputs from early layers and immediately previous layer that further strengthens correlated gradient information.

2.4

On keeping epochs=25, increasing batch size results in increase of train loss, validation loss, and lower validation accuracy, however, the time it takes to process these results decrease logarithmically.

Smaller batch sizes will become exponentially increasing in time to train, but will result in highest accuracy, lowest train and validation loss. The qualitative result is also staggering. Smaller batch sizes will dramatically produce clearer images and will colour in images that are either mostly accurate or a plausible guess (e.g. actual background is a grey sky, trained background is blue sky)

3 Part C

3.1

3.2

3.3

3.4

Assuming n is the number of layers, the memory complexity of fine-tuning the entire model is n times than fine-tuning only the last layer. The computational complexity of fine-tuning the entire model is also n times more than fine-tuning the last layer, but only w.r.t. back-propagation. Freezing other layers does not increase or decrease computational complexity w.r.t forwarding, so in this case it will be the same.

3.5

Since you are increasing two dimensions of the input image by 2, the memory complexity of fine-tuning should increase by 2x2=4. The number of parameters, however, will never change by changing height and width of input image.