**Atharva Joshi**

**FSRCNN results**

**hyperparameters:**

**learning rate:**

for the first 15 epochs: lr = 1e-3 for first two parts (feature extraction, shrinking, non linear mapping and expansion) and lr\*0.1 = 1e-4 for the last part (deconvolution)

for the next 10 epochs: lr = 0.5e-3

for the next 5 epochs: lr = 0.38e-3

**loss function:**

MSE perpixel loss + perceptual loss with vgg16 pretrained model with equal weightage

batch size = 16

gradient accumulation steps = 3

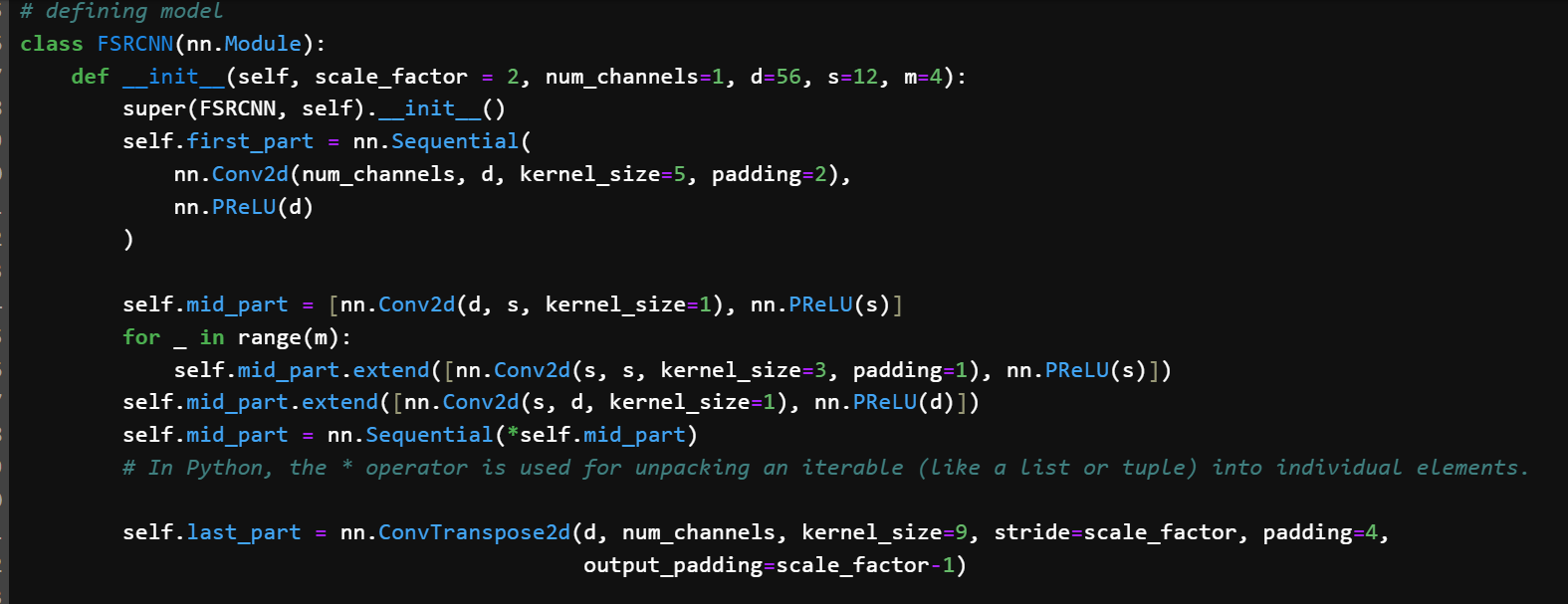
hence effective batch size = 48

used ycbcr channel separation and trained a y-only model

Optimiser used: Adam

**Average PSNR:** 24.8 dB

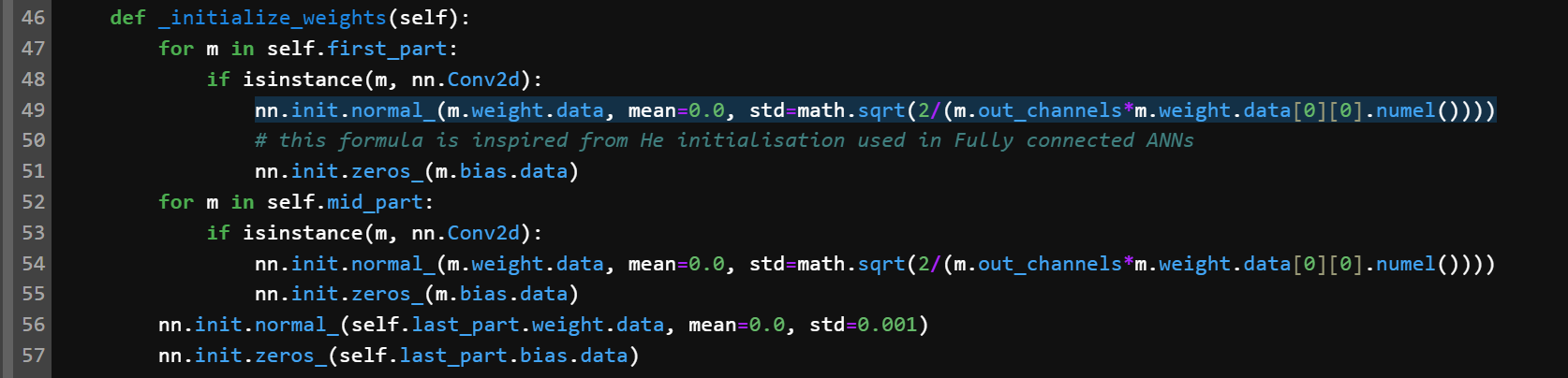
**Model architecture:**

****

**Initialisation step:**

Kaiming-He initialisation method for all layers except deconvolution layer

for deconvolution weights are intiialised from normal distribution with mean = 0 and std = 0.001

****

**Forward propagation:**

Downscalling performed to scale all pixel values to lie within [0, 1].

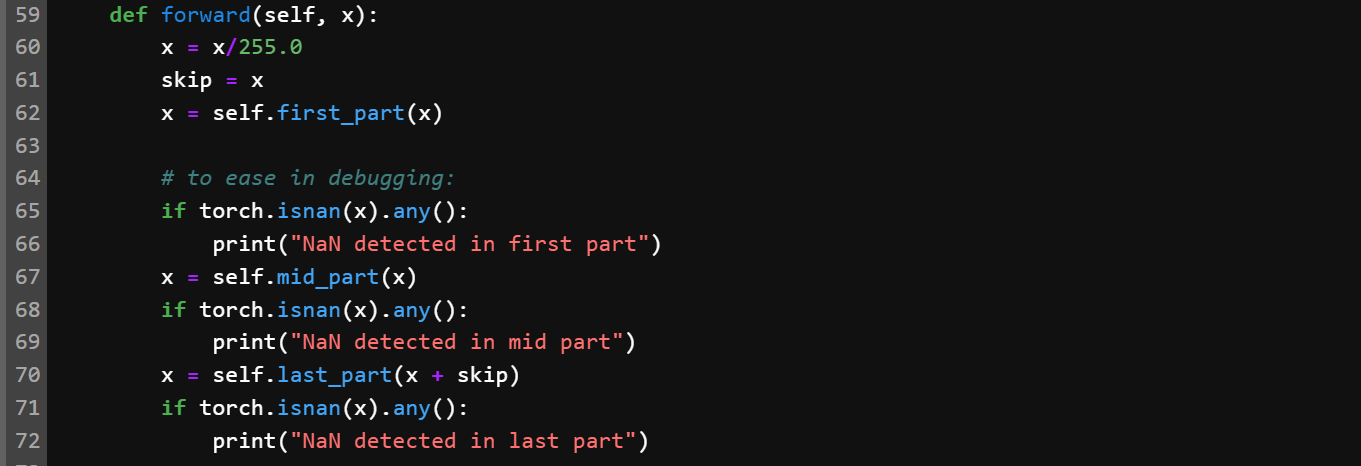
This prevents exploding gradients

As a further precaution, the max\_norm() function is used to put a limit to the maximum permissible gradient which is set to 2

Further, to debug for exploding gradients, the torch.isnan() method is used

**Skip connection** was introduced to the deconvolution layer to better learn a mapping close to the identity mapping. This was done because the nature of super-resolution task results in the image being similar in structure to the target image, with addition of new information. This mapping is hence semantically close to identity mapping.

Although theoretically sound, the use of skip connnection resulted in a decrease in performance for the same number of epochs (average PSNR 22.05 dB). A possible reason might be the increased ease of learning the exact identity mapping instead of a modified mapping close to but not exactly equal to the identity mapping.

****

**Min max normalisation step:**

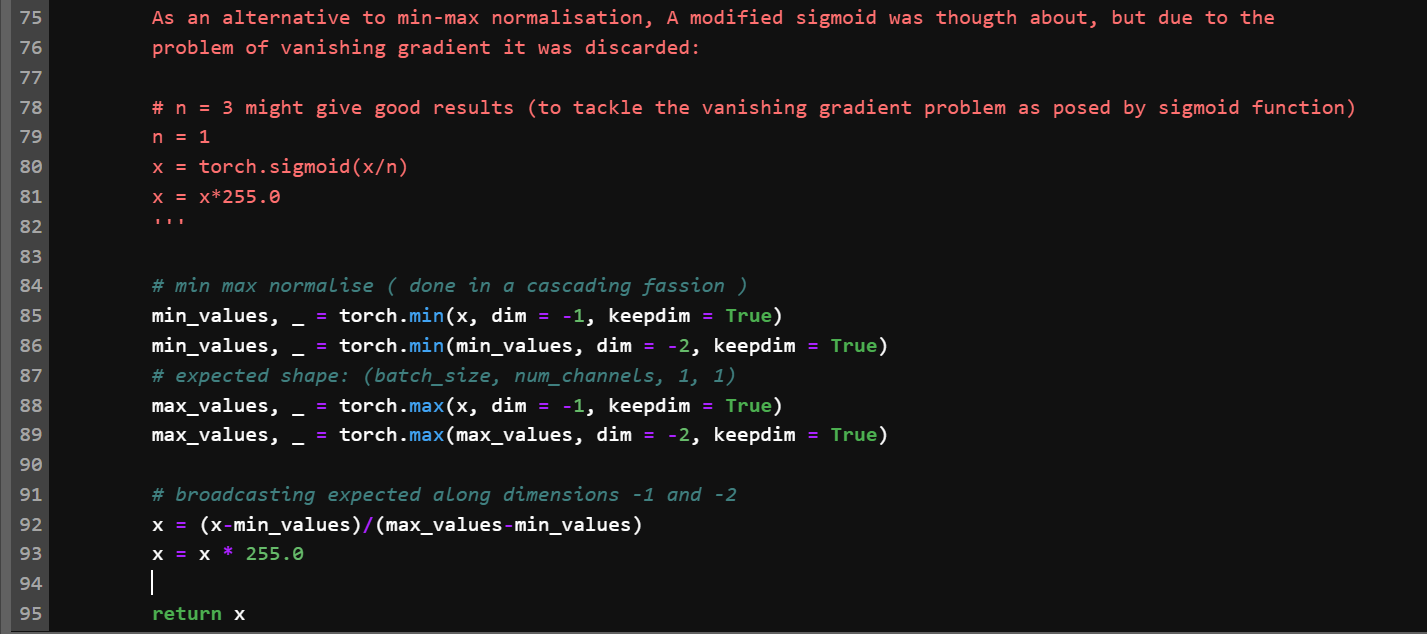
The min-max normalisation, which scales the output to the range [0, 1], followed by upscaling by 255 ensures that;  
1) All pixel values are within the permissible 8-bit range

2) The values exceeding 255 in the original model output are scalled linearly to fit within [0, 255] rather than clipping these values. In absence of this function, plain coloured patches were observed

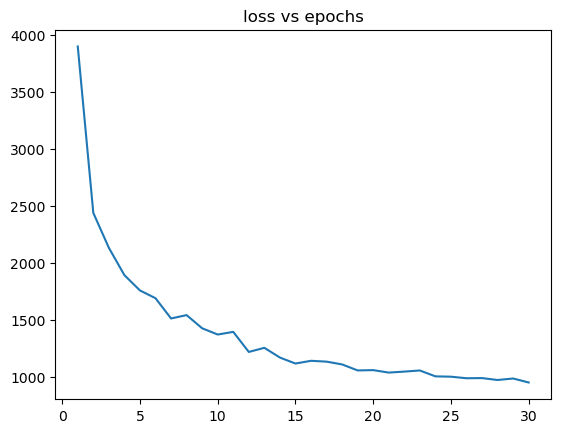
**Aliter:**

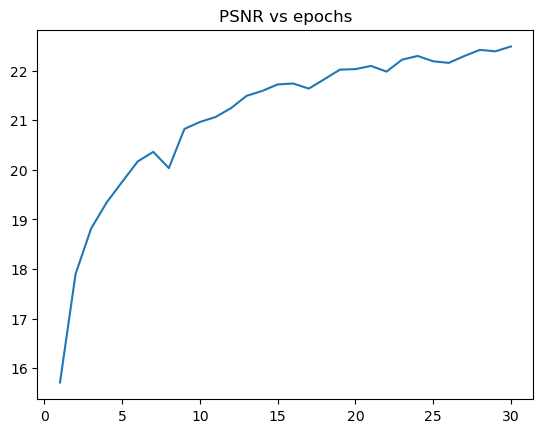
A modified sigmoid function was thought about:

with the intuition to ‘strech’ the graph of sigmoid to tackle vanishing gradient at large values of x. This did not work out. The modification does indeed lead to a relative increase in gradient value at the tail ends, it reduces the absolute value of gradient. Hence the idea was discarded and min-max normalisation was used instead.

****

**Training characteristics:**





**Results:**

