MRP - Week 8-9

This week we will be experimenting with using a variety of layers and α/β ratios

- The VGG-16 neural network allows us to extract the style and content of images separately
- The loss function is minimized during image synthesis and has 2 terms for content and style loss respectively, allowing us to control emphasis of one over the other.

Loss function

$$L_{total}(S, C, G) = \alpha L_{content}(C, G) + \beta L_{style}(S, G)$$

- The loss function needs to be fed a set of layers from the VGG-16 network
  - $L_{content}$: takes any 1 layer of the network
  - L_{style} : takes layers 'conv1 1', 'conv2 1', 'conv3 1', 'conv4 1' and 'conv5 1' (basically the 1st layer of each block of the VGG net to give a good variation of style details)
- The goal is to experiment generating an image using a variety of inputs layers for $L_{content}$ against a varying ratio of α/β

content   +   style   =   Generated image
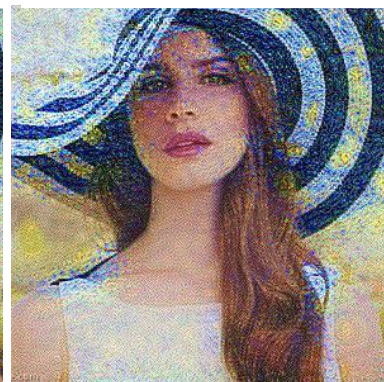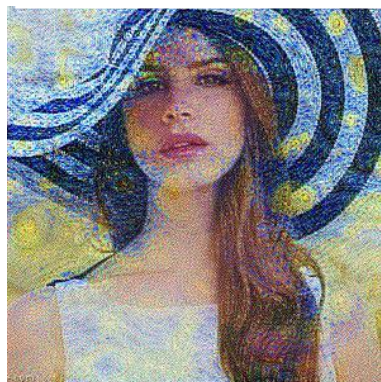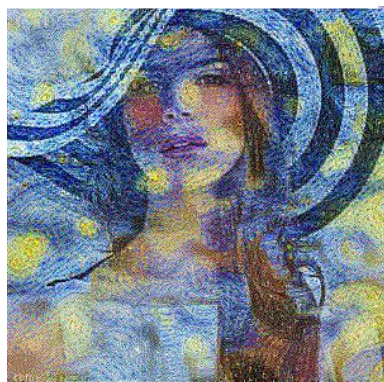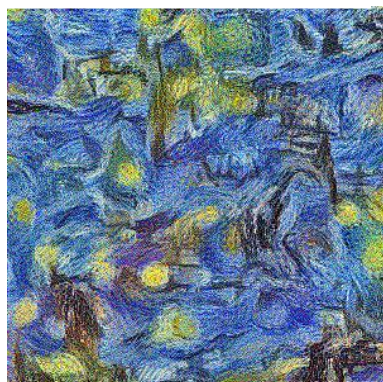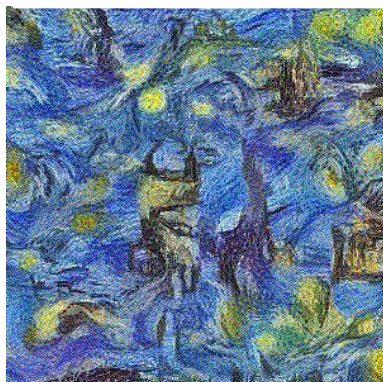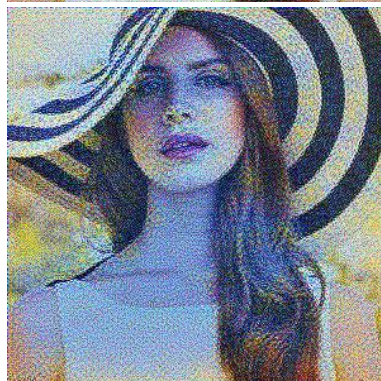
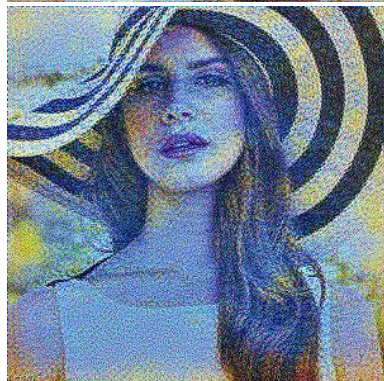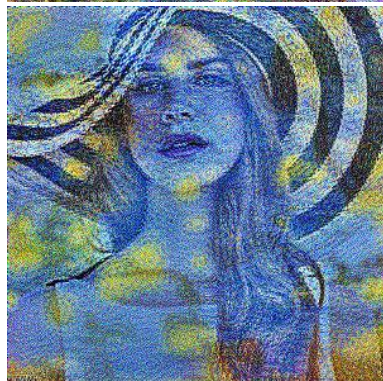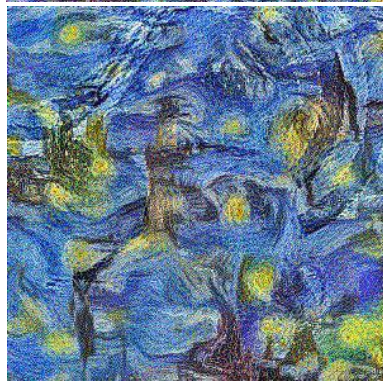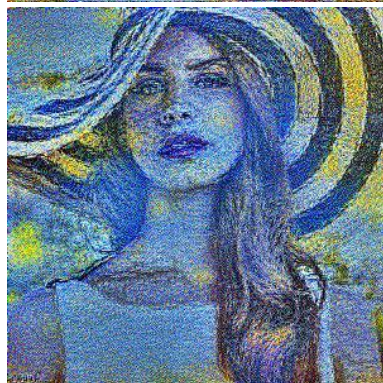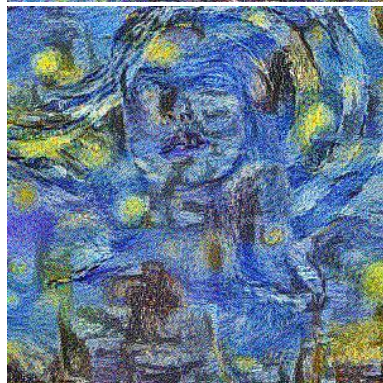|	$\alpha/\beta = 10^3$	$\alpha/\beta = 10^2$	$\alpha/\beta = 10^1$	$\alpha/\beta = 0.25$	$\alpha/\beta = 0.5$
Con1_1					
Con2_1					
Con3_1					
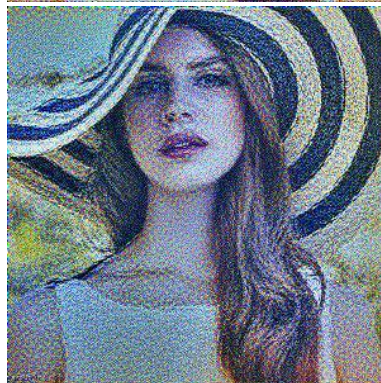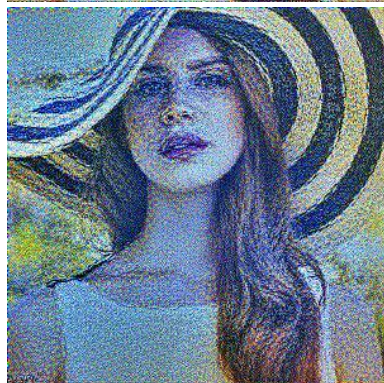
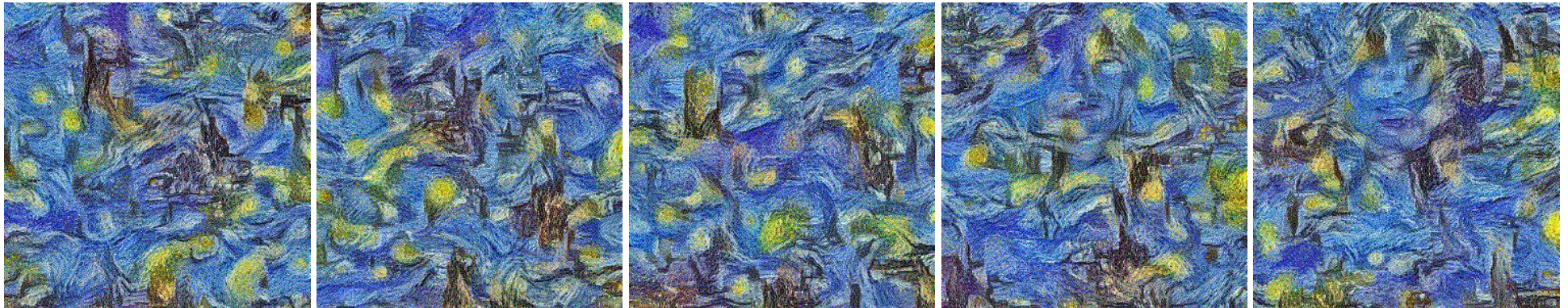$\alpha/\beta = 10^3$ | $\alpha/\beta = 10^2$ | $\alpha/\beta = 10^1$ | $\alpha/\beta = 0.25$ | $\alpha/\beta = 0.5$

Con4_1

Con5_1

- The grid of experiments above show how the generated image varies with respect to the layer used for content representation
- All experiments were run without adding the total variation loss factor in order to better see the effect of change

- Note:

  o Higher layers capture high-level content features. Not exact pixel value
  o Higher layers are more concerned with the shapes and arrangement of objects in the image
  o Progressively, the higher layers show less resemblance to the original content image

- The following trends can easily be observed in the grid of images above:

    o As α/β increases → more emphasis on the content image
    o As α/β decreases → more emphasis on the style image

- This grid can be used to visually optimize the generated results by allowing us to select how much detail we wish to retain
- Content images containing subjects like faces or other fine details may require that we use layers from block 1, 2, or 3. No higher
- My results are comparable to the work of Leon Gatys' in his paper , *"A neural Algorithm of Artistic Style" (see below)*

The column headings read the α/β ratio values as 10^{-5}, 10^{-4}, 10^{-3}, and 10^{-2}

The row headings read the layers used as con1_1, con2_1, con3_1, con4_1, con5_1. Basically the first layer of each block in the VGG network