NEURAL ARTISTIC STYLE

UNDERGRADUATE PROJECT-1

Deepak Kumar (15095026)

Ekansh Gupta (15095028)

Pramod Ghuge (15095029)

SUPERVISED BY:- DR. R. Dwivedi

Signature :- _____

ABSTRACT

Rendering the semantic content of an image in different styles is a difficult image processing task. Arguably, a major limiting factor for previous approaches has been the lack of image representations that explicitly represent semantic information and, thus, allow to separate image content from style. Here we use image representations derived from Convolutional Neural Networks optimised for object recognition, which make high level image information explicit.

We introduce A Neural Algorithm of Artistic Style that can separate and recombine the image content and style of natural images. The algorithm allows us to produce new images of high perceptual quality that combine the content of an arbitrary photograph with the appearance of numerous well known artworks. Our results provide new insights into the deep image representations learned by Convolutional Neural Networks and demonstrate their potential for high level image synthesis and manipulation.

INTRODUCTION

Transferring the style from one image onto another can be considered a problem of texture transfer. In texture transfer the goal is to synthesise a texture from a source image while constraining the texture synthesis in order to preserve the semantic content of a target image. For texture synthesis there exist a large range of powerful non-parametric algorithms that can synthesise photorealistic natural textures by resampling the pixels of a given source texture.

Although these algorithms achieve remarkable results, they all suffer from the same fundamental limitation: they use only low-level image features of the target image to inform the texture transfer. Ideally, however, a style transfer algorithm should be able to extract the semantic image content from the target image (e.g. the objects and the general scenery) and then inform a texture transfer procedure to render the semantic content of the target image in the style of the source image.

Therefore, a fundamental prerequisite is to find image representations that independently model variations in the semantic image content and the style in which it is presented.
The recent advance of Deep Convolutional Neural Networks has produced powerful computer vision systems that learn to extract high-level semantic information from natural images. It was shown that Convolutional Neural Networks trained with sufficient labelled data on specific tasks such as object recognition learn to extract high-level image content in generic feature representations that generalise across datasets and even to other visual information processing tasks, including texture recognition and artistic style classification.

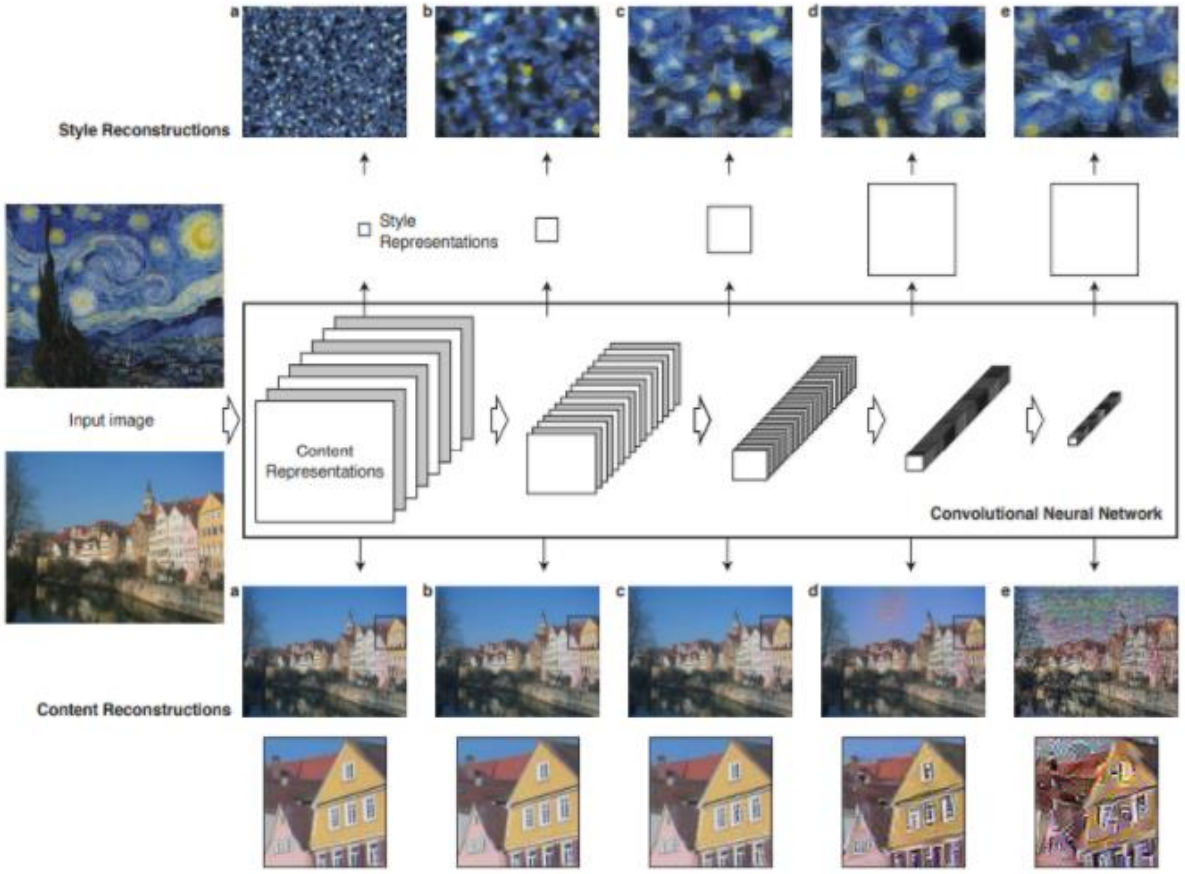


Figure 1. Image representations in a Convolutional Neural Network (CNN). A given input image is represented as a set of filtered images at each processing stage in the CNN. While the number of different filters increases along the processing hierarchy, the size of the filtered images is reduced by some down sampling mechanism (e.g. max-pooling) leading to a decrease in the total number of units per layer of the network.

Content Reconstructions. We can visualise the information at different processing stages in the CNN by reconstructing the input image from only knowing the network's responses in a particular layer. We reconstruct the input image from layers 'conv1 2' (a), 'conv2 2' (b), 'conv3 2' (c), 'conv4 2' (d) and 'conv5 2' (e) of the original VGG-Network. We find that reconstruction from lower layers is almost perfect (a–c). In higher layers of the network, detailed pixel information is lost while the high-level content of the image is preserved (d,e).

Style Reconstructions. On top of the original CNN activations we use a feature space that captures the texture information of an input image. The style representation computes correlations between the different features in different layers of the CNN. We reconstruct the style of the input image from a style representation built on different subsets of CNN layers ('conv1 1' (a), 'conv1 1' and 'conv2 1' (b), 'conv1 1', 'conv2 1' and 'conv3 1' (c), 'conv1 1', 'conv2 1', 'conv3 1' and 'conv4 1' (d), 'conv1 1', 'conv2 1', 'conv3 1', 'conv4 1' and 'conv5 1' (e). This creates images that match the style of a given image on an increasing scale while discarding information of the global arrangement of the scene.

DEEP IMAGE REPRESENTATION

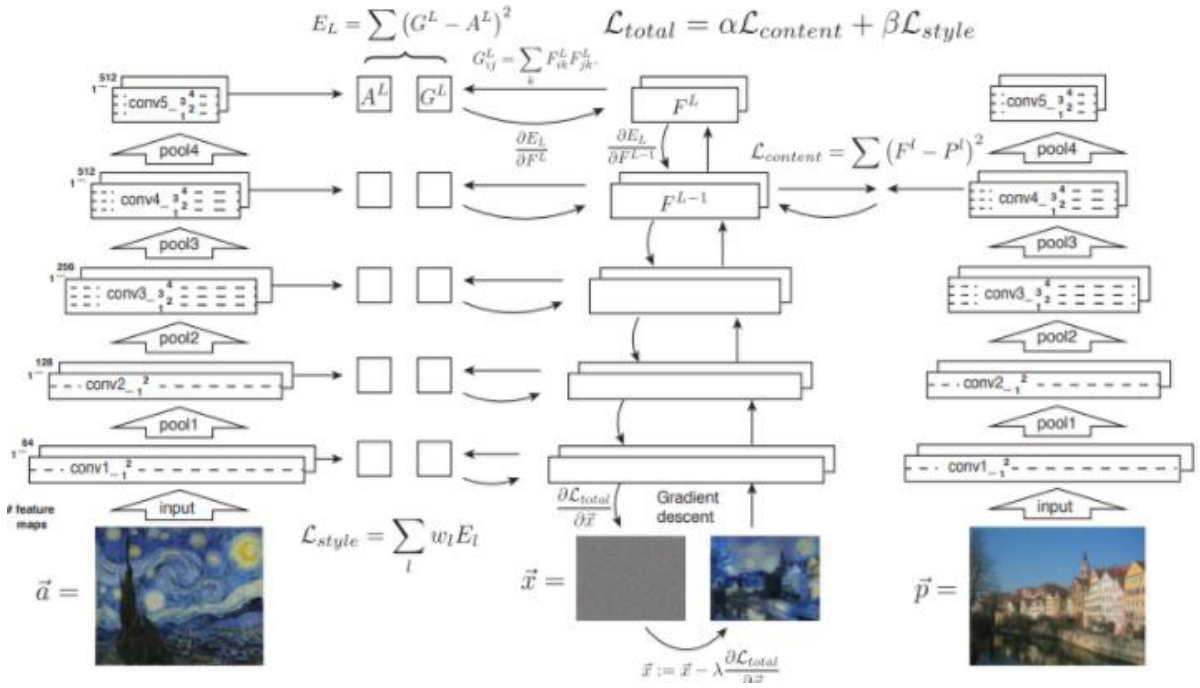
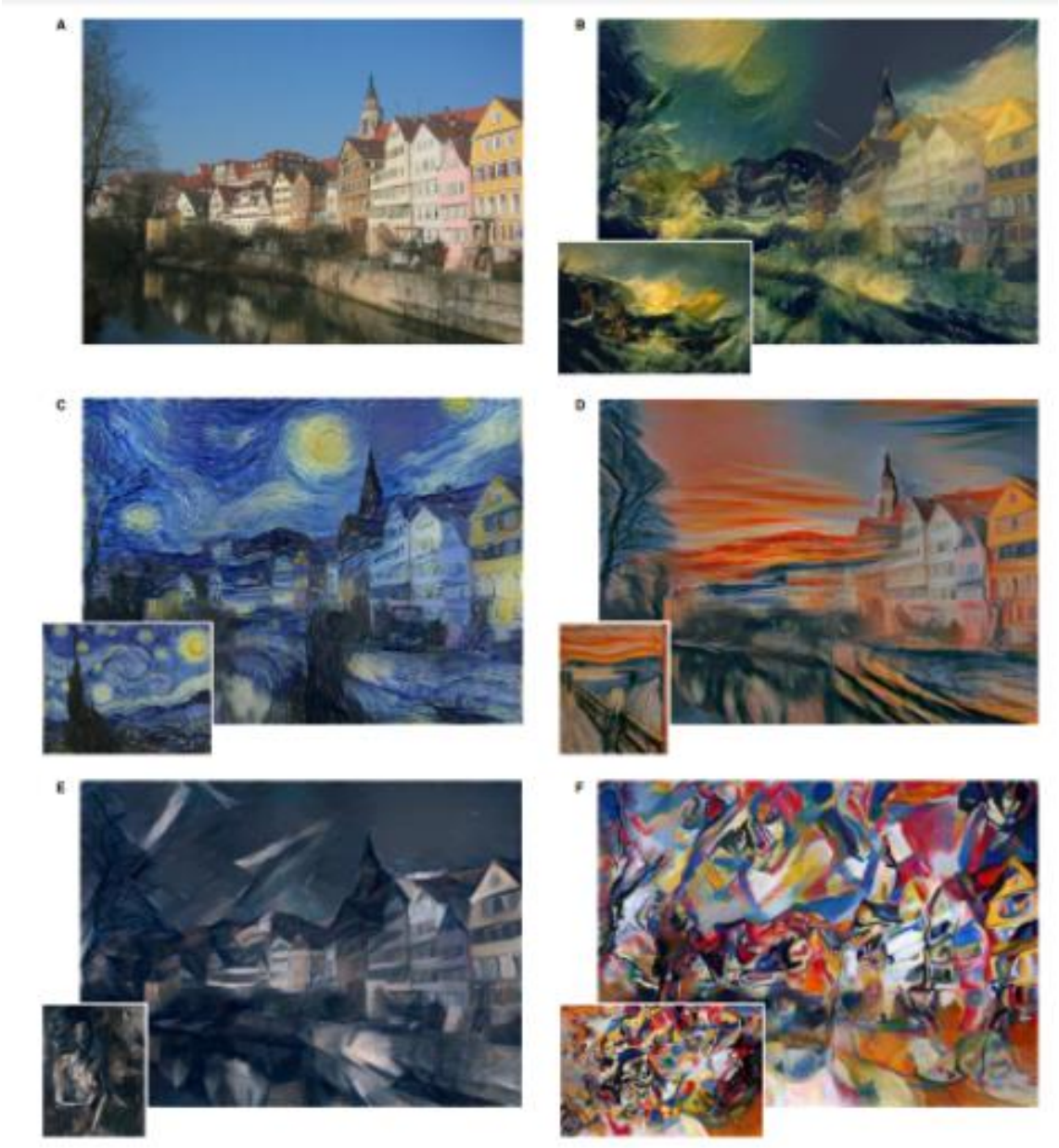


Figure 2. Style transfer algorithm. First content and style features are extracted and stored. The style image ***a*** is passed through the network and its style representation A_l on all layers included are computed and stored (left). The content image ***p*** is passed through the network and the content representation P_l in one layer is stored (right). Then a random white noise image ***x*** is passed through the network and its style features G_l and content features F_l are computed. On each layer included in the style representation, the element-wise mean squared difference between G_l and A_l is computed to give the style loss L_{style} (left). Also the mean squared difference between F_l and P_l is computed to give the content loss $L_{content}$ (right). The total loss L_{total} is then a linear combination between the content and the style loss. Its derivative with respect to the pixel values can be computed using error back-propagation (middle). This gradient is used to iteratively update the image ***x*** until it simultaneously matches the style features of the style image ***a*** and the content features of the content image ***p*** (middle, bottom).

RESULT

The key finding is that the representations of content and style in the Convolutional Neural Network are well separable. That is, we can manipulate both representations independently to produce new, perceptually meaningful images. To demonstrate this finding, we generate images that mix the content and style representation.



DISCUSSION

Here we demonstrate how to use feature representations from high-performing Convolutional Neural Networks to transfer image style between arbitrary images. While we are able to show results of high perceptual quality, there are still some technical limitations to the algorithm.

Probably the most limiting factor is the resolution of the synthesised images. Both, the dimensionality of the optimisation problem as well as the number of units in the Convolutional Neural Network grow linearly with the number of pixels. Therefore the speed of the synthesis procedure depends heavily on image resolution. Another issue is that synthesised images are sometimes subject to some low-level noise. While this is less of an issue in the artistic style transfer, the problem becomes more apparent when both, content and style images, are photographs and the photorealism of the synthesised image is affected. However, the noise is very characteristic and appears to resemble the filters of units in the network.

Thus it could be possible to construct efficient de-noising techniques to post-process the images after the optimisation procedure.
The separation of image content from style is not necessarily a well-defined problem. This is mostly because it is not clear what exactly defines the style of an image. It might be the brush strokes in a painting, the colour map, certain dominant forms and shapes, but also the composition of a scene and the choice of the subject of the image – and probably it is a mixture of all of them and many more. Therefore, it is generally not clear if image content and style can be completely separated at all.

FUTURE PLANNING

We are planning to apply semantic segmentation on the input image/frame using which we can try to implement multiple styles on a single content input.