

A Neural Algorithm of Artistic Style

Leon A. Gatys, Alexander S. Ecker, Matthias Bethge

<https://arxiv.org/abs/1508.06576>

<https://arxiv.org/pdf/1508.06576.pdf>

TL;DR

Artificial system based on a Deep Neural Network that creates artistic images of high perceptual quality. Uses neural representations to separate and recombine content and style of arbitrary images.

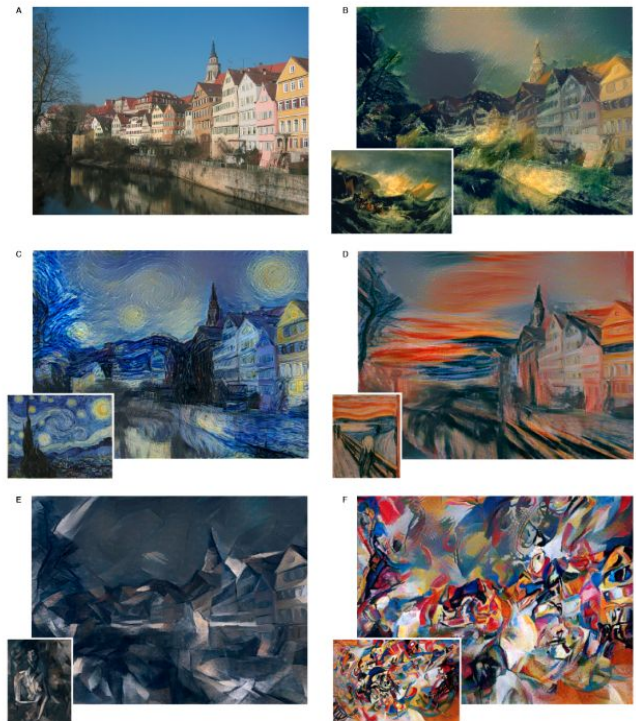
Paper

CNN's develop a representation of the image that makes object information increasingly explicit along the processing hierarchy. Therefore, along the processing hierarchy of the network, the input image is transformed into representations that increasingly care about the actual content of the image. The feature responses in higher layers of the network are the content representation.

To obtain a representation of the style of an image, we use a feature space originally designed to capture texture information. This feature space is built on top of the filter responses in each layer of the network. By including the feature correlations of multiple layers, we obtain a stationary, multi-scale representation of the input image, which captures its texture information but not the global arrangement.

Images are synthesized by finding an image that simultaneously matches the content representation and the style representation. The global arrangement of the original photo is preserved, and the colours, and texture, and local structures are provided by the style photo.

While synthesising an image that combines the content of one image with the style of another, there usually does not exist an image that perfectly matches both constraints at the same time. However, the loss function used contains two terms for content and style respectively. We can therefore smoothly regulate the emphasis on either reconstructing the content or the style.



Methods

The paper uses the feature space provided by the 16 convolutional and 5 pooling layers of VGG-19. They do not use any of the fully connected layers. For image synthesis, replacing the max-pooling operation by average pooling improves the gradient flow and on obtains slightly more appealing results, which is why the images shown were generated with average pooling.

In each layer of the network, a style representation is built that computes the correlations between the different filter responses, where the expectation is taken over the spatial extent of the input image. These feature correlations are given by the Gram matrix.