Implementation of paper A Neural Algorithm of Artistic Style

This paper is of great importance in the field of neural transfer, since it was the first one that described how you can separate content and style from an image with the same deep learning model. This is my interpretation and implementation of it.

The paper describes how, by getting the outputs of different feature maps alongside the deep neural network, we can get different representations of content and styles of an image. We then modify a white image by applying gradients via backpropagation to each pixel to reduce the distance between the white noise picture and the content and style representations of different pictures. The process to do this is described below.

Foto contenido foto blanco foto estilo

The neural style transfer architecture. I’ll describe each part in detail.

Foto todo el modelo

Content

The paper implies that the feature maps of higher layers in the network, have the content representation of the image. It describes what and where things are, but not the exact pixel value as the layers in lower levels. Given a feature map of the content image and a feature map of the white image, we minimize the mean square error between the two matrices via backpropagation, propagating the gradients to each pixel in the white image.

Foto blanc foto content del notebook

Com osacas el feature map the la content y la White image

Foto formula y foto formula código

Foto los calculada

Style

The style loss is very similar to the content. Instead of calculating the square mean error between feature maps, we calculate it between Gram Matrices. The concept of gram matrix is that, having two pictures similar to each other, each feature map in a specific layer will “see” similar things.

Foto pasto y lso features mapss

We can see how the first three feature maps have similar representation of the image. By calculating the dot product of each feature map with the others, we make a gram matrix of n\*n, where n is the number of feature maps in that specific layer. The diagonal is the dot product of each feature map with itself. By doing this we can observe that positions of high values in the gram matrix means that two feature maps dot product produce high value, therefore the two feature maps have high values. If we minimize the mean square error of the gram matrix of the style picture and the white image pixel, the gram matrix of the white image picture will try to assimilate to the gram matrix of the style picture, therefore capturing the style of this one.

Foto gram matrix

Fotos paper y fotos modelo

Content + Style loss:

If we sum up both the content loss and the style loss, we can make the white noise image become a combination of both pictures, with the content of the content style, and the gram matrix of the style picture, therefore with the style of it.

The paper uses the conv4\_2 layer for the content loss, and conv1\_1, conv2\_1, conv3\_1, conv4\_1 and conv5 1 for style loss. Each layer of the style loss represents 1/amount\_of\_layers\_used to the total style loss, so we multiply each style loss of each style layer by this amount. Also, the authors multiply each loss by a scalar to manipulate how much each loss weights to the final output. “The ratio α/β was either 1×10−3 (Fig 2 B,C,D) or 1 × 10−4 (Fig 2 E,F)”

Foto parte modelo que combina todo y backpropagation y toda la bola

Process y deprocess

Resultado final