

Friday 12.01.17

6 Likes Share

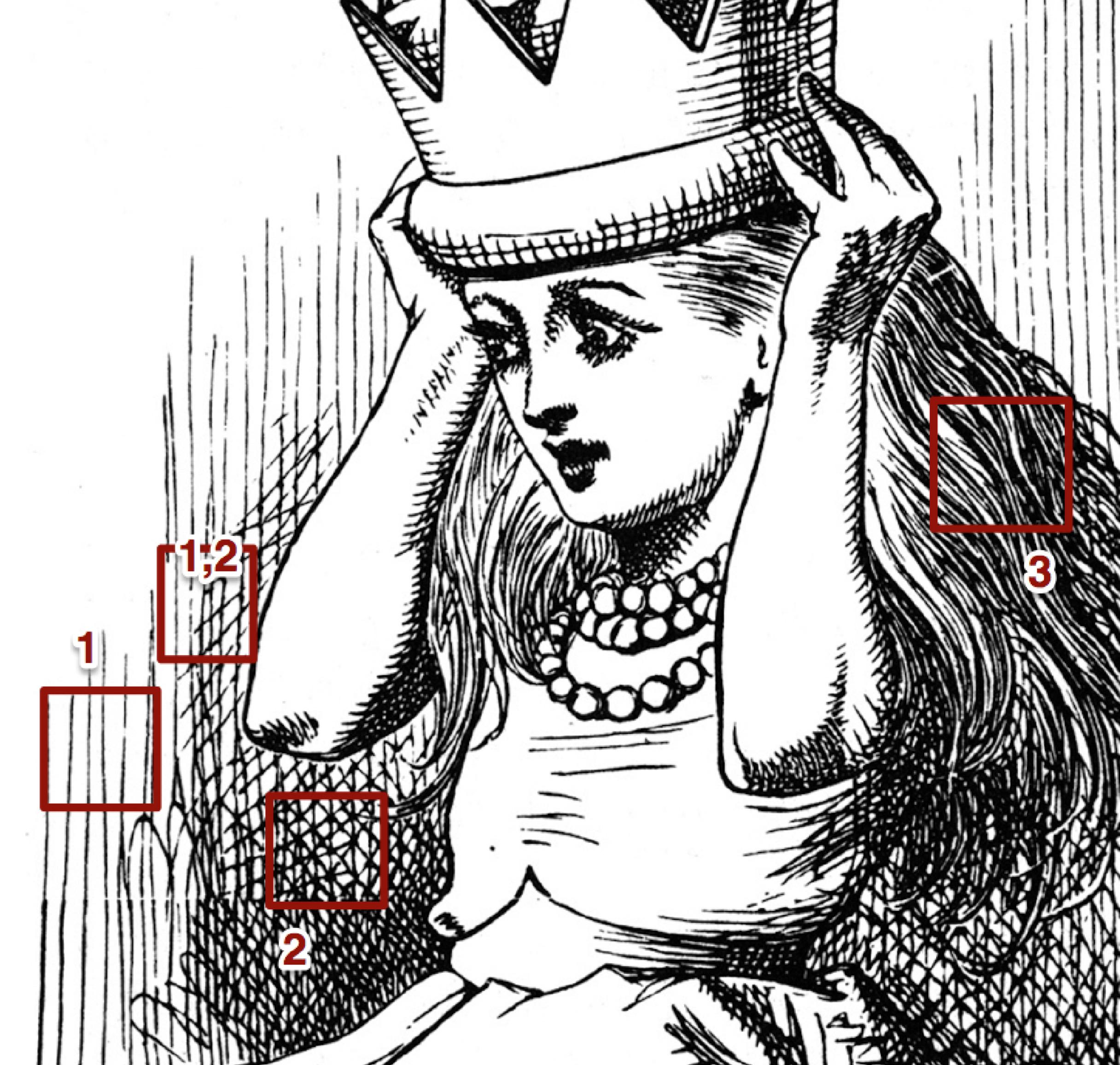
## How does Gram Matrix encode the Style of an Image?

In non-photoreal renderings, capturing the style of an image is an extremely tricky and difficult issue. Recently there has been a resurgence in this field with the application of Deep Learning to this problem. Specifically, in Gatys' paper, the stylization problem is posed as an optimization problem where two cost functions capture the content distance to one image and style distance to another image.

Powered by squarespace

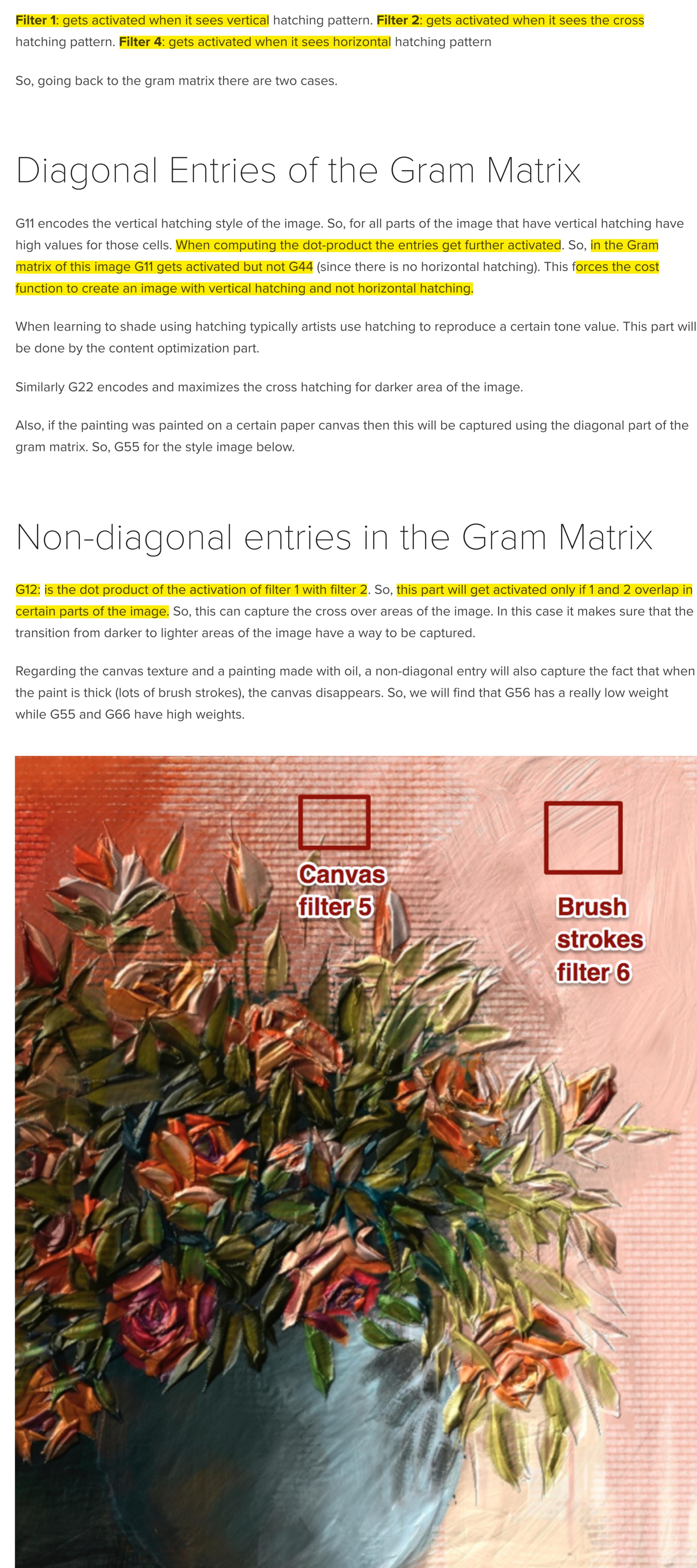
### The Genius is in the Details (Again)

If a hidden layer at a certain point has width ( $w$ ) and height ( $h$ ) and if we shape the activation at a certain layer as a matrix with each row capturing the activation for a particular filter then the activation matrix on the left results (see Figure below).



In this case the gram matrix is defined as the matrix multiplication of the activation matrix and its transpose.

We will investigate why optimizing the gram matrix preserves the style in the context of the following style image by John Tenniel from Alice in Wonderland.



Two Filters that may capture the low level style of the image are denoted by squares 1 and 2 (see figure above).

**Filter 1:** gets activated when it sees vertical hatching pattern. **Filter 2:** gets activated when it sees the cross hatching pattern. **Filter 4:** gets activated when it sees horizontal hatching pattern

So, going back to the gram matrix there are two cases.

### Diagonal Entries of the Gram Matrix

G11 encodes the vertical hatching style of the image. So, for all parts of the image that have vertical hatching have high values for those cells. When computing the dot-product the entries get further activated. So, in the Gram matrix of this image G11 gets activated but not G44 (since there is no horizontal hatching). This forces the cost function to create an image with vertical hatching and not horizontal hatching.

When learning to shade using hatching typically artists use hatching to reproduce a certain tone value. This part will be done by the content optimization part.

Similarly G22 encodes and maximizes the cross hatching for darker area of the image.

Also, if the painting was painted on a certain paper canvas then this will be captured using the diagonal part of the gram matrix. So, G55 for the style image below.

### Non-diagonal entries in the Gram Matrix

G12 is the dot product of the activation of filter 1 with filter 2. So, this part will get activated only if 1 and 2 overlap in certain parts of the image. So, this can capture the cross over areas of the image. In this case it makes sure that the transition from darker to lighter areas of the image have a way to be captured.

Regarding the canvas texture and a painting made with oil, a non-diagonal entry will also capture the fact that when the paint is thick (lots of brush strokes), the canvas disappears. So, we will find that G56 has a really low weight while G55 and G66 have high weights.

**Canvas filter 5** and **Brush strokes filter 6** are highlighted in red in the image below.

**Assumption:** That VGG-19 and VGG-16 capture the right filters that an artist may draw. Since human beings have no difficulty recognizing art, this is a valid assumption assuming that VGG captures the human visual system.

Newer / Older