

A Neural Algorithm of Artistic Style

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

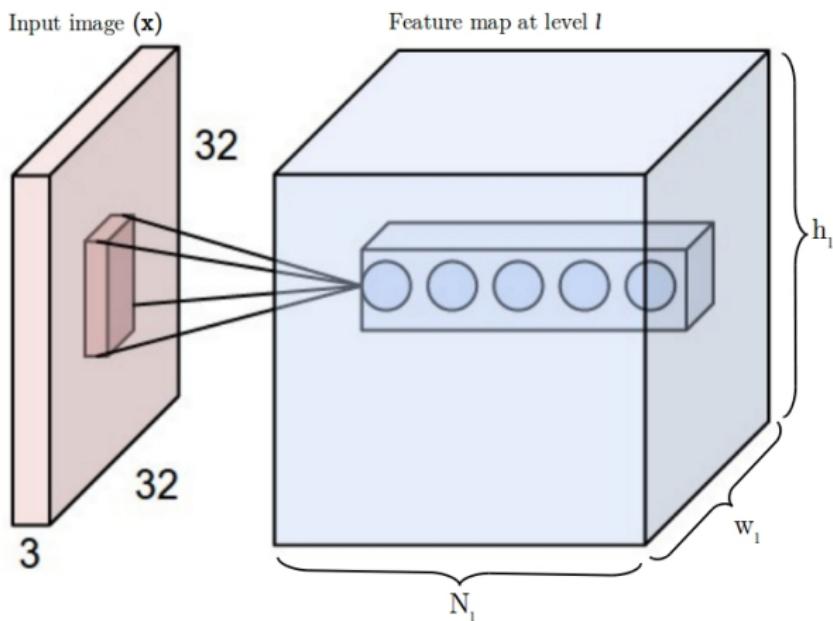
Behnam Rasoolian, Christian Kauten

Auburn University

Notations

- \mathbf{p} : The original, content image
- \mathbf{a} : The original, artwork image
- \mathbf{x} : The image to be generated. It is initiated as a random noise image.
- F^l : **Feature Map** at level l , is the result of applying filters at level l . If N_l filters are applied at level l , then this feature map has a depth of N_l .
- N_l : The number of filters applied at level l . This is the same as the depth of the feature map at level l .
- M_l : the dimension of the feature map at level l , which is equal to $N_l \times M_l$.
- \mathbf{F}^l : The feature map at level l . it is an $N_l \times M_l$ matrix.

Notations



Content Representation

- Gradient descent is performed on a white noise image (\mathbf{x}) and a content image (\mathbf{p})
- F^l and P^l : Respective feature maps of the noise image and the original image
- Goal: Reduce the squared-error loss between F^l and P^l .

$$\mathcal{L}(\mathbf{p}, \mathbf{x}, l) = \frac{1}{2} \sum_{i=1}^{N_l} \sum_{j=1}^{M_l} (F_{ij}^l - P_{ij}^l)^2 \quad (1)$$

Content Representation

The gradient of this loss with respect to activations in I can be easily calculated:

$$\frac{\partial \mathcal{L}_{content}}{\partial F_{ij}^I} = \begin{cases} (F^I - P^I)_{ij} & \iff F_{ij}^I > 0 \\ 0 & \iff F_{ij}^I < 0 \end{cases} \quad (2)$$

Content Reconstruction

Figure: white noise image \mathbf{x}



Figure: content image \mathbf{p}



Content Reconstruction

Figure: Block 1 Conv 1



Content Reconstruction

Figure: Block 2 Conv 1



Content Reconstruction

Figure: Block 3 Conv 1



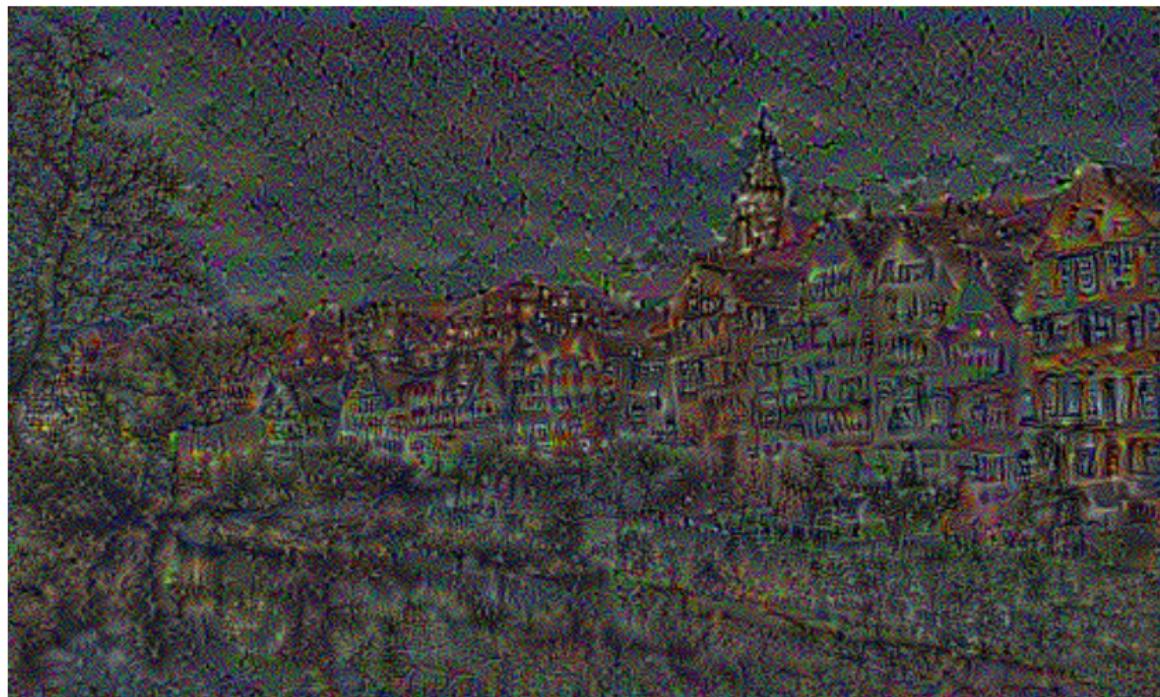
Content Reconstruction

Figure: Block 4 Conv 1



Content Reconstruction

Figure: Block 5 Conv 1



Style Representation

Style representation is achieved via the “Gram Matrix” G . Gram matrix is an $N_l \times N_l$ matrix which calculates the correlations between different filter responses.

$$\mathbf{G}^l_{ij} = \mathbf{F}^{l\top}_i \times \mathbf{F}^l_j = (\mathbf{F}^{l\top} \times \mathbf{F}^l)_{ij} \quad (3)$$

Style Representation

Given G^l and A^l as respective Gram matrices of the noise image and the original image, our goal is to reduce the overall difference between G^l and A^l . In this sense, Contribution of layer l to the total loss is

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_i^{N_l} \sum_j^{N_l} (G_{ij}^l - A_{ij}^l)^2 = \mathbf{1}^T (\mathbf{G} - \mathbf{A}) (\mathbf{G} - \mathbf{A})^T \quad (4)$$

Style Representation

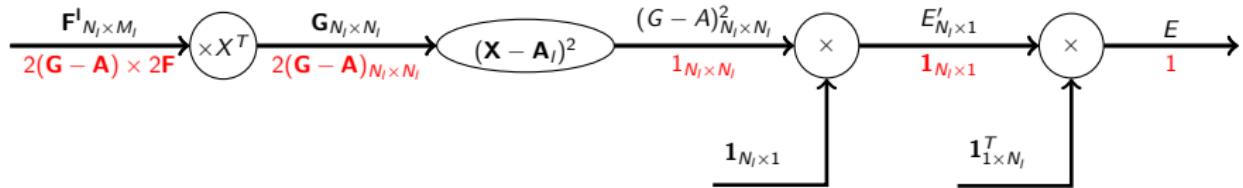
The total loss is:

$$\mathcal{L}_{style}(\mathbf{a}, \mathbf{x}) = \sum_{l=0}^L w_l E_l \quad (5)$$

$$\frac{\partial \mathcal{L}_{style}}{\partial F_{ij}^l} = \frac{\partial E_l}{\partial F_{ij}^l} = (4(\mathbf{G}^l - \mathbf{A}^l) \times \mathbf{F}^l)_{ij} \quad (6)$$

Style Representation

$$\frac{\partial \mathcal{L}_{style}}{\partial F_{ij}^I} = \frac{\partial E_I}{\partial F_{ij}^I} = (4(\mathbf{G}^I - \mathbf{A}^I) \times \mathbf{F}^I)_{ij} \quad (7)$$



Style Reconstruction

Figure: white noise image x

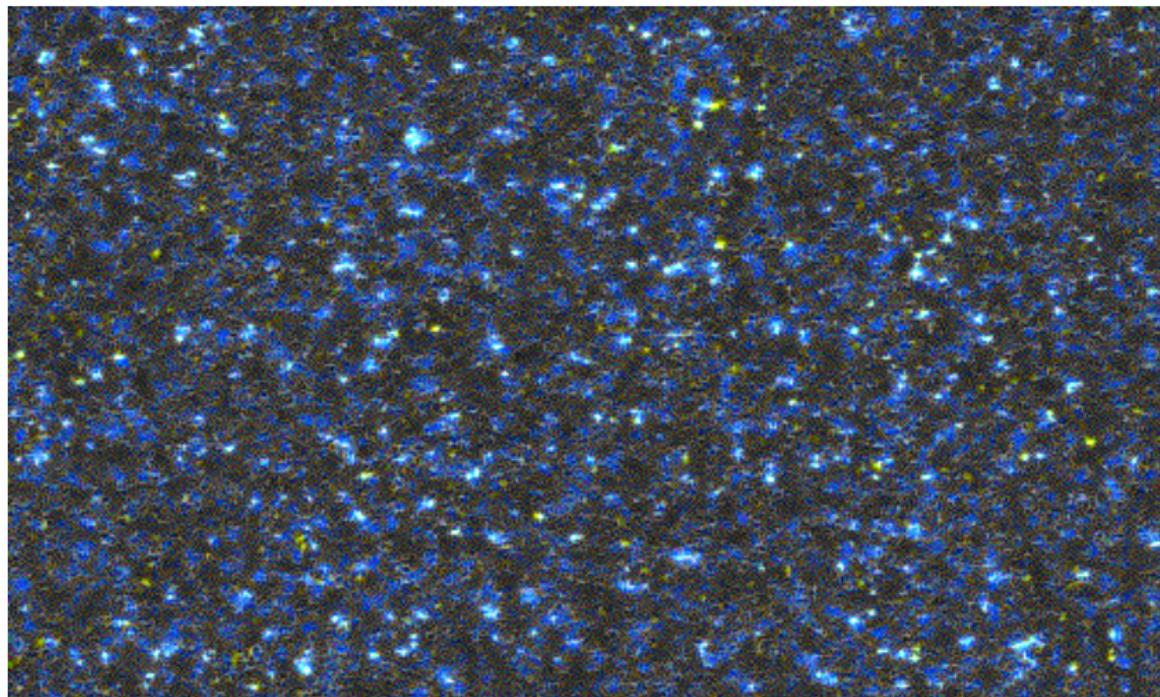


Figure: artwork image a



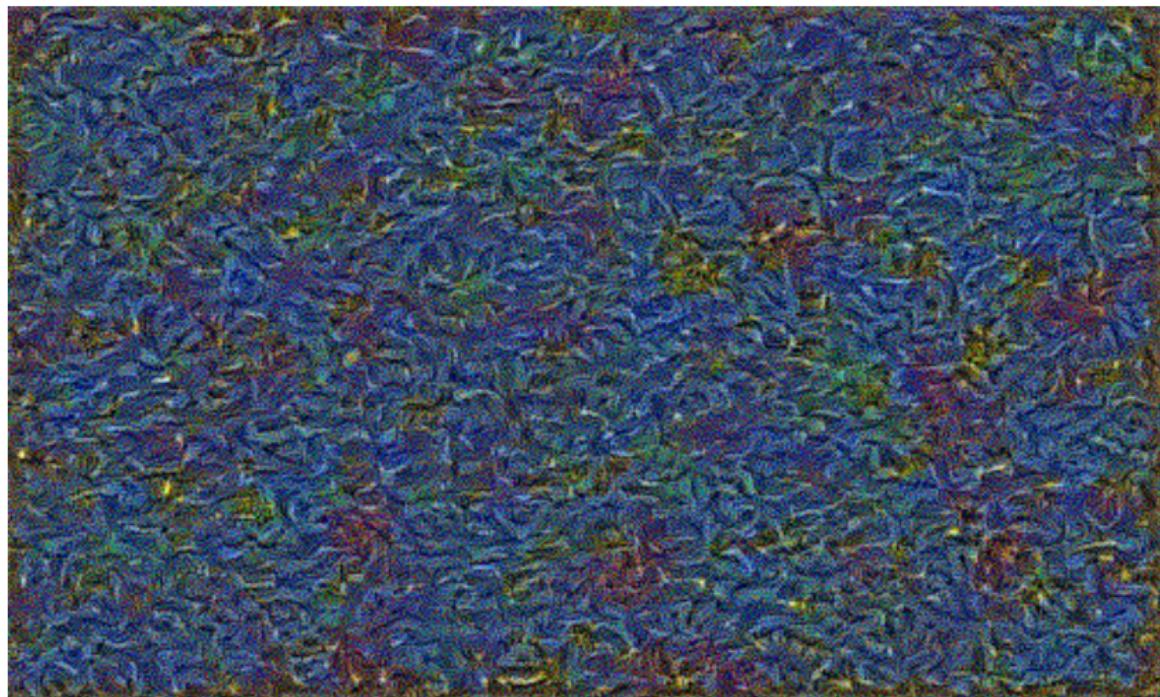
Style Reconstruction

Figure: Conv 1 of Block 1



Style Reconstruction

Figure: Conv 1 of Block 1, 2



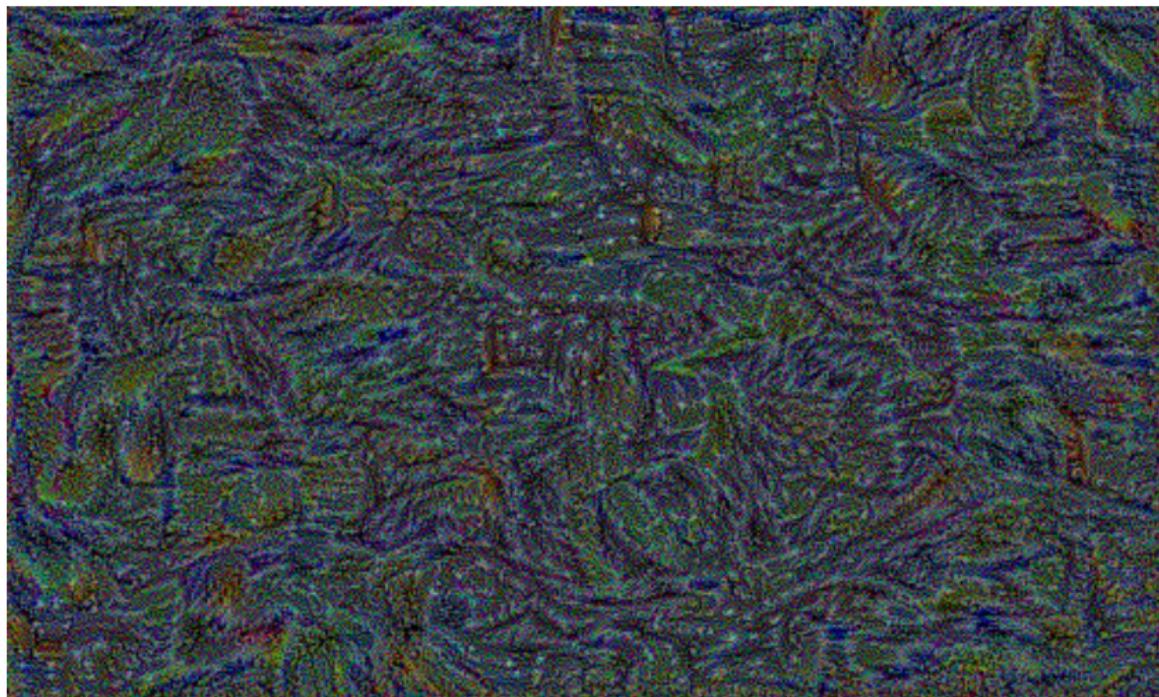
Style Reconstruction

Figure: Conv 1 of Block 1, 2, 3



Style Reconstruction

Figure: Conv 1 of Block 1, 2, 3, 4



Style Reconstruction

Figure: Conv 1 of Block 1, 2, 3, 4, 5

