A Neural Algorithm of Artistic Style

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

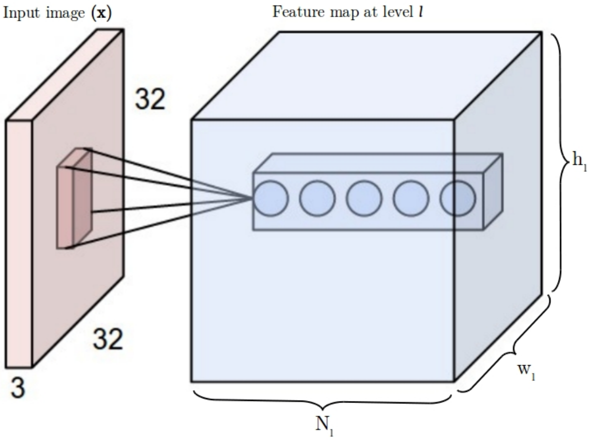
Behnam Rasoolian, Christian Kauten

Auburn University

Notations

- $\overrightarrow{p}$: The original, content image
- $\overrightarrow{a}$: The original, artwork image
- $\overrightarrow{x}$: The image to be generated. It is initiated as a random noise image.
- F^l: **Feature Map** at level l, is the result of appling filters at level l. If N_l filters are applier at level l, then this feature map has a depth of N_l.
- N_l: The number of filters applier at level l. This is the same as the depts 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 ($\overrightarrow{x}$) and a content image ($\overrightarrow{p}$)
- 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.

$$\mathcal{L}(\overrightarrow{p}, \overrightarrow{x}, l) = \frac{1}{2} \sum_{i,j} (F^l_{ij} - P^l_{ij})^2 \tag{1}$$

Content Representation

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

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

Content Reconstruction

Figure: white noise image $\vec{x}$

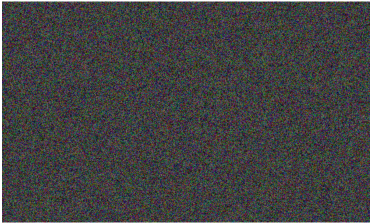


Figure: content image $\vec{p}$

Content Reconstruction

Figure: Block 1 Conv 1

Content Reconstruction

Figure: Block 2 Conv 1

Content Reconstruction

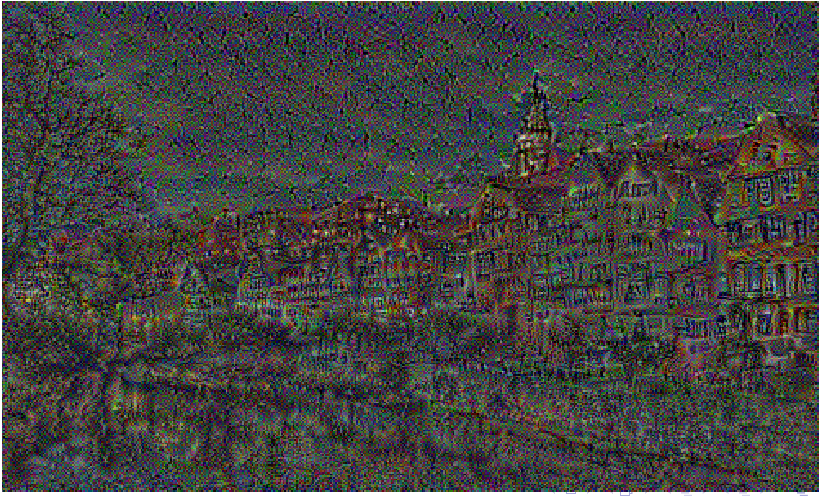
Figure: Block 3 Conv 1

Content Reconstruction

Figure: Block 4 Conv 1

Content Reconstruction

Figure: Block 5 Conv 1



Content Reconstruction

Figure: Block 4 Conv 2

Style Representation

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

$$\mathbf{G^l}_{ij} = \mathbf{F^l}^{\mathbf{T}}_i \times \mathbf{F^l}_j = (\mathbf{F^l}^{\mathbf{T}} \times \mathbf{F^l})_{ij} \tag{3}$$

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 overal difference between G^l and A^l. In ths sense, Contribution of layer l to the total loss is

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

Style Representation

The total loss is:

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

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

Style Representation

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