



Notations

- **a**: The original image
- **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.

Content representation

In order to come up with an image that has the same content as the input image, gradient descent is performed on a white noise image (\mathbf{x}). At each level l , given F^l and P^l as respective feature maps of the noise image and the original image, our goal is to reduce the overall difference between F^l and P^l . Therefore, the loss function should look like minimizing the square error:

$$\mathcal{L}(p, x) = \sum_{l=1}^L E_l \quad (1)$$

where E_l is

$$E_l = \frac{1}{2} \sum_{i=1}^{N_l} \sum_{j=1}^{M_l} (F_{ij}^l - P_{ij}^l)^2 \quad (2)$$

thus the gradient can be easily calculated:

$$\frac{\partial \mathcal{L}_{style}}{\partial F_{ij}^l} = \frac{\partial E_l}{\partial F_{ij}^l} = (\mathbf{F}^l - \mathbf{P}^l)_{ij} \quad (3)$$

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.

$$\mathbf{G}_{ij}^l = \mathbf{F}^{lT}{}_i \times \mathbf{F}^l{}_j = (\mathbf{F}^{lT} \times \mathbf{F}^l)_{ij} \quad (4)$$

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

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_i^{N_l} \sum_j^{N_l} (G_{ij}^l - A_{ij}^l)^2 = \mathbf{1}^T (\mathbf{G} - \mathbf{A})(\mathbf{G} - \mathbf{A})^T \quad (5)$$

The total loss is:

$$\mathcal{L}_{style}(\mathbf{a}, \mathbf{x}) = \sum_{l=0}^L w_l E_l \quad (6)$$

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

