Image Style Transfer
Using Convolutional Neural Networks

L. Gatys, S. Ecker, M. Bethge

October 19, 2018

- Create new images with content of arbitrary photograph and appearance of artworks: $I = \textbf{Content}(I_1) + \textbf{Style}(I_2)$
- Usually only the low-level features are considered; ideally the semantic features of target image must be rendered with the required style.
- **Content Representation**: Higher (later) layers of the NN concern the content in the image and arrangement, but not the exact pixel values. The lower layers can get exact representation (paper used `conv4_1`).
- To reconstruct image x as p from l^{th} layer, perform gradient descent on white noise p with loss

$$\mathcal{L}_{\textbf{content}} = \frac{1}{2}\Sigma_{i,j}(X^l_{ij} - P^l_{ij})^2$$

.

- Gram matrix of a layer l is defined as (each computation is the inner product)

$$G^l_{ij} = \Sigma_k F^l_{ik} F^l_{jk}$$

- **Style Representation**: The style of the image is represented by the Gram matrices of various layers in NN (paper used `conv1_1`, `conv2_1`, `conv3_1`, `conv4_1`, `conv5_1`).
- To estimate loss on style, with same notation as above:

$$\mathcal{L}_{\textbf{style}} = \Sigma_{l=0}^{L} w_l E_l, \text{ where } E_l = \frac{1}{4N_l^2 M_l^2}\Sigma_{i,j}(Y^l_{ij} - P^2_{ij})^2$$

- Finally, the total loss on synthesized image p from content image x and style image y is

$$\mathcal{L}_{\textbf{total}} = \alpha\mathcal{L}_{\textbf{content}} + \beta\mathcal{L}_{\textbf{style}}$$

- *Topic in literature is called Non-Photorealistic Rendering.*

Figure 1: Whole process in a diagram