A Neural Algorithm of Artistic Style

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

Presented by Weidi Xie (1st Oct 2015)
What does the paper do?

• Create artistic images of high perceptual quality.

• The key observation: the representations of content and style of an image in the Convolutional Neural Networks are separable.

• Use a pre-trained VGG-19 (average pooling) to find another image simultaneously matches the content of a photograph and the style of a piece of art work.

(Simonyan & Zisserman, 2015)
Related Papers:

   - Match four distinct families of statistics, but only use **Wavelet-based** pyramid decomposition of the image.
   - Joint (**Correlation**) statistics were introduced.

   - Style representation.

   - Content representation.
Style of an art work (Textures) ?

- **Style representation**: **correlations** between the different filter responses over the spatial extent of feature maps.
  - Provide colours and local structures.
- Synthesize texture by matching correlation matrices calculated from different layers.
- **Key equations**: (Check paper for notation)

  \[ G^l = F^l (F^l)^T \]
  \[ E_l = \frac{1}{Norm} \sum_{i,j} (G^l_{i,j} - A^l_{i,j})^2 \]
  \[ Loss_{style} = \sum_{l=0}^L w_l E_l \]

  **Correlation matrix**
  **Cost for style reconstruction**
  **Accumulate cost for lower layers**

Portilla & Simoncelli, 2000
Gatys, et al. 2015
More Results:

Original

Up to Conv1-1 layer

Up to Pool1 layer

Up to Pool2 layer

Up to Pool3 layer

Up to Pool4 layer
More Results:

Original

Up to Pool4 layer

Original (My result)

Up to Pool4 layer (only Conv layers)

Up to Pool5 layer (only Conv layers)
• Refer to the feature responses in higher layers of the network as the *content representation*.

  — Higher layers in the network capture the **high-level content** in terms of objects and their arrangement.

  — Reconstructions based on high-layer feature maps preserve global arrangement, but does **not** constrain the exact pixel values.

Reconstruction based on feature responses of different layers.
Content and Style
• Loss function is defined as: (Check the paper for notations)

\[
\text{Loss}_{\text{style}} = \sum_{l=0}^{L} w_l E_l
\]

\[
\text{Loss}_{\text{content}} = \frac{1}{2} \sum_{i,j} (F_{ij} - P_{ij})^2
\]

\[
\text{Loss}_{\text{total}} = \alpha L_{\text{content}} + \beta L_{\text{style}}
\]

\(\alpha\) and \(\beta\) are the weighting factors for content and style reconstruction.

\[
\frac{\partial L_{\text{content}}}{\partial F_{ij}^l} = \begin{cases} (F^l - P^l)_{ij} & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0 \end{cases}
\]

\[
\frac{\partial E_l}{\partial F_{ij}^l} = \begin{cases} \frac{1}{N_i^2 M_j^2} (F^T (G^l - A^l))_{ji} & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0 \end{cases}
\]
Combine content and style (My result)

Used for Content

\[ \alpha / \beta = 10^{-3} \]

Used for Style

\[ \alpha / \beta = 10^{-4} \]

\[ \alpha / \beta = 10^{-5} \]
Combine content and style

Decrease $\alpha/\beta$

Used for \textit{Content}

Used for \textit{Style}
Multiple style
Appendix:


a. Original Image.

b. Spectrally matched noise images. Only analyzed with linear filters and energy filters, tuned to different orientations, spatial frequencies and spatial positions.

c. Naturalistic texture images. **Correlations** are computed by taking products of linear and energy filter responses across different **orientations, spatial frequencies** and **positions**
• Wavelet-based pyramid decomposition of the image.

• Four distinct families of statistics:
  
  — Pixel statistics.
  
  Representation of the distribution of the raw intensities present in the original stimulus.

  — Correlation coefficients
  
  Periodicity, or the degree to which a pattern repeats.

  — Magnitude Correlations
  
  Consider low layers, the correlation matrix records edge co-occurrence across:
    1. **Position** (Does a particular edge continue in a straight line or a corner?)
    2. **Orientation** (Do vertical edges tend to co-occur with horizontal edges?)
    3. **Scale** (Is a small-scale edge part of a larger structure at a coarser scale?)

  — Relative phase
  
  Measure the relative phase of wavelet features between neighbouring spatial scales within the pyramid decomposition of the target image.