A Neural Algorithm of Artistic Style

Leon A. Gatys et. al
The Question

Can we apply *any* style to *any* content?

**MASP Art School campaign done by DDB Brazil**

**Google search on “images”**
“Previous Work”

Print by James Hance

http://thirddime.com/blog/10_awesome_lego_versions_of_famous_paintings/

http://www.artfido.com/blog/artist-photoshops-her-fat-cat-into-famous-artworks/
Previous Work: Learning Styles

“We also show that style is highly content-dependent.”

Contributions

- Learn best pairing between content and style

Contributions: Visual Results

https://youtu.be/cB84sgqIkR4?t=27

https://www.youtube.com/watch?v=g9BxlwIqWlc
Methodology

● Leverage CNNs
  ○ Trained for object recognition

● Jointly learn content and style
  ○ Texture synthesis captures style
  ○ Separate representations of content and style
  ○ Recombinations of content and style based on loss functions
Methodology

Very Deep ConvNets (VGG)

- **Key factors**: small kernels, stride of 1, ReLU, deeper depths

![VGG Team ILSVRC Progress](http://www.robots.ox.ac.uk/~karen/pdf/ILSVRC_2014.pdf)
Methodology

Very Deep ConvNets (VGG)

- **Key factors**: small kernels, stride of 1, ReLU, deeper depths

Slight modifications:

- Average pooling
- Maxpool
- Conv-64
- Conv-128
- Conv-256
- Conv-512
- Fully connected layers
Methodology

- More flexibility!
- Content and style trained separately (for the most part)
  - Cannot have perfect synthesis → loss functions with its parameters
Methodology

Methodology

Methodology

Loss Function: Content

\[ \mathcal{L}_{content}(p, x, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2 \]

- \( p \) - original image
- \( x \) - generated image
- \( F \) - responses stored in matrix
- \( P \) - feature representation of original image
- \( i,j \) - ith position, jth filter

Loss Function: Content

\[ \mathcal{L}_{\text{content}}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2 \]

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

Methodology: Texture Synthesis

- As previously seen:
  - Create textures from feature representations
  - Discriminative
  - Captures salient features
  - Also uses VGG architecture

Loss Function: Style

\[
E_l = \frac{1}{4N_l^2M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2
\]

\[
\mathcal{L}_{style}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} w_l E_l
\]

- \( G \) - content representation
- \( A \) - style representation
- \( a \) - original art image
- \( N, M \) - \( N \) feature maps of size \( M \)
- \( w \) - weight factors of layer \( l \)
- \( E \) - style loss at layer \( l \)

Loss Function: Style

\[ E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} \left( G_{ij}^l - A_{ij}^l \right)^2 \]

\[ \frac{\partial E_l}{\partial F_{ij}^l} = \begin{cases} \frac{1}{N_l^2 M_l^2} \left( (F^l)^T (G^l - A^l) \right)_{ji} & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0 \end{cases} \]

- \( G \) - content representation
- \( A \) - style representation
- \( a \) - original art image
- \( N, M \) - features map of size \( M \)
- \( w \) - weight factors of layer \( l \)
- \( E \) - style loss at layer \( l \)

Loss Function: Style

\[ E_l = \frac{1}{4N_l^2M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2 \]

\[ \mathcal{L}_{style}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} \omega_l E_l \]

- \( \omega_l = 1/\text{numActiveLayers} \)
- with non-zero loss weight \( \omega_l \)

Loss Function: All together

\[ \mathcal{L}_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{content}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{style}(\vec{a}, \vec{x}) \]

- \( \alpha \) and \( \beta \) are parameters to control regularization
Loss Function: All together

\[ \mathcal{L}_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{content}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{style}(\vec{a}, \vec{x}) \]

- \( \alpha \) and \( \beta \) are parameters to control regularization
- alpha / beta

Content/Style Representations

10^{-5}

Conv1_1

10^{-2}

Conv5_1

Content/Style Representations

- Content represented in lower layers

Content/Style Representations

- Style represented in feature space
  - local arrangements, textural information

Conclusions

- Learn best pairing between **content** and **style**
  - (Mostly) separable