Image Style Transfer Using Convolutional Neural Networks

L. Gatys, A. Ecker, M. Bethge

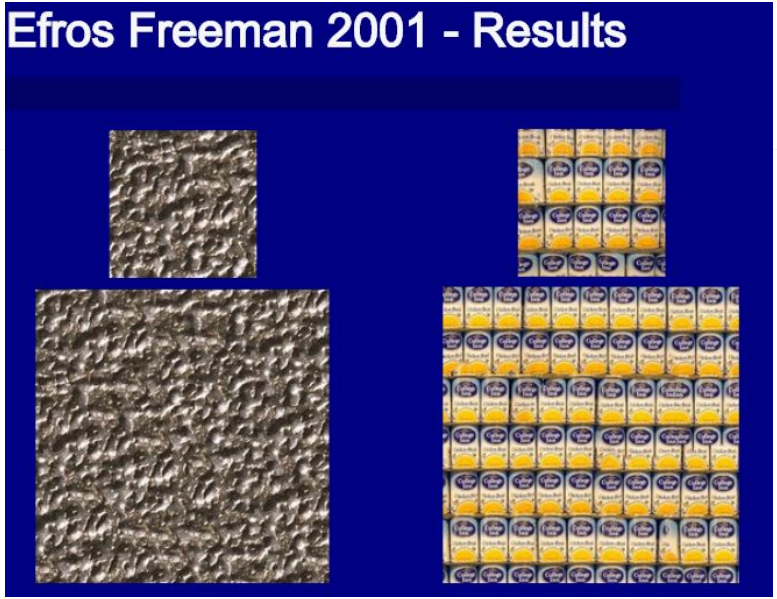
Zach Harris, Felix Portillo

UCDAVIS
UNIVERSITY OF CALIFORNIA

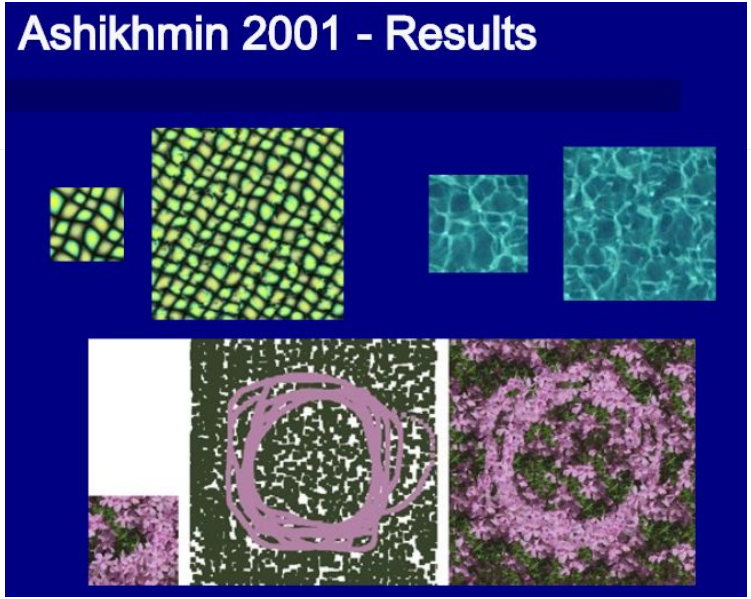
Image Style Transfer

UCDAVIS

Older Methods of Texture Generation



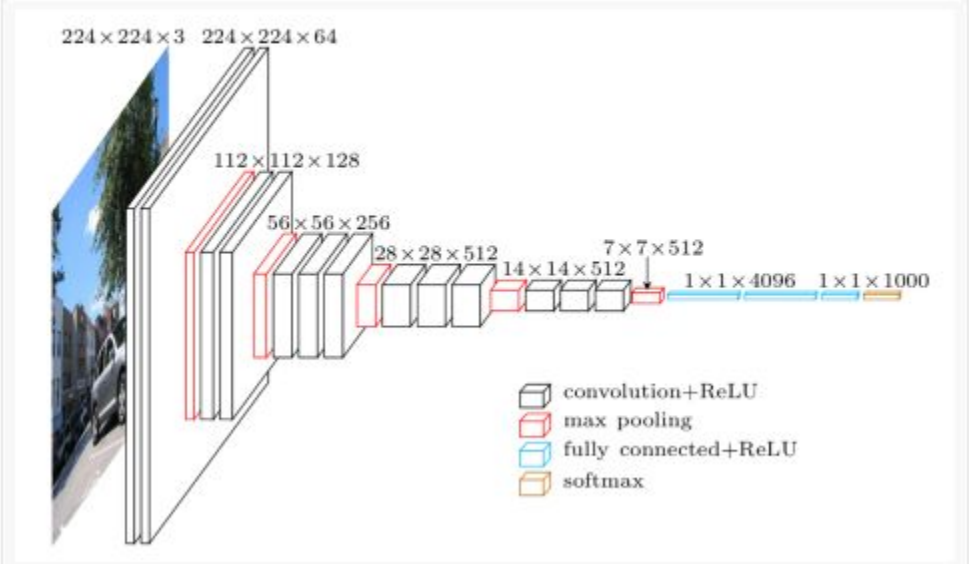
A.Efros, W. Freeman, Image Quilting for Texture Synthesis and Transfer, 2001



M. Ashikhmin, Synthesizing Natural Textures, 2001

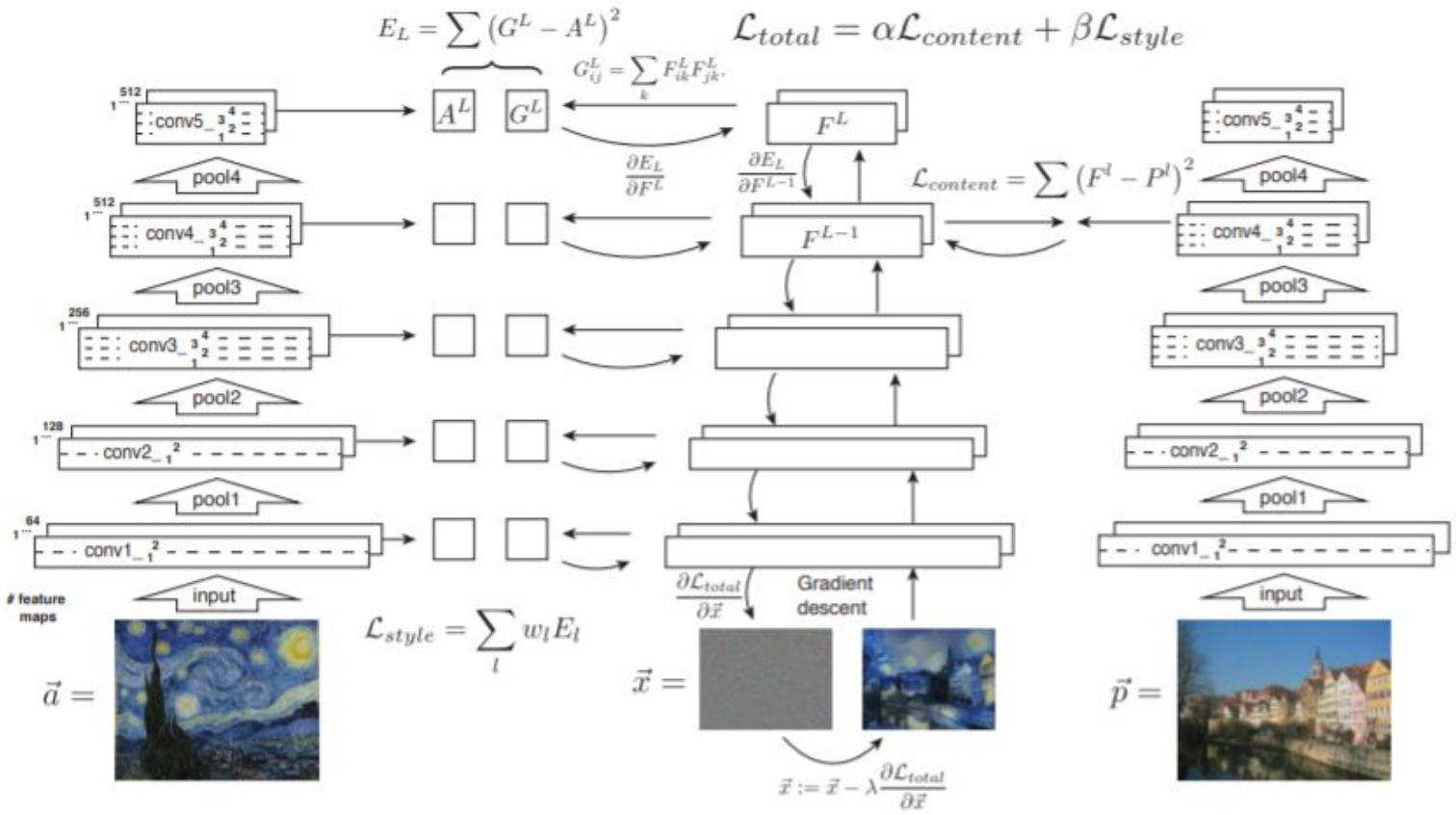
A VGG Network Was Used to Obtain Results

VGG16 and VGG19

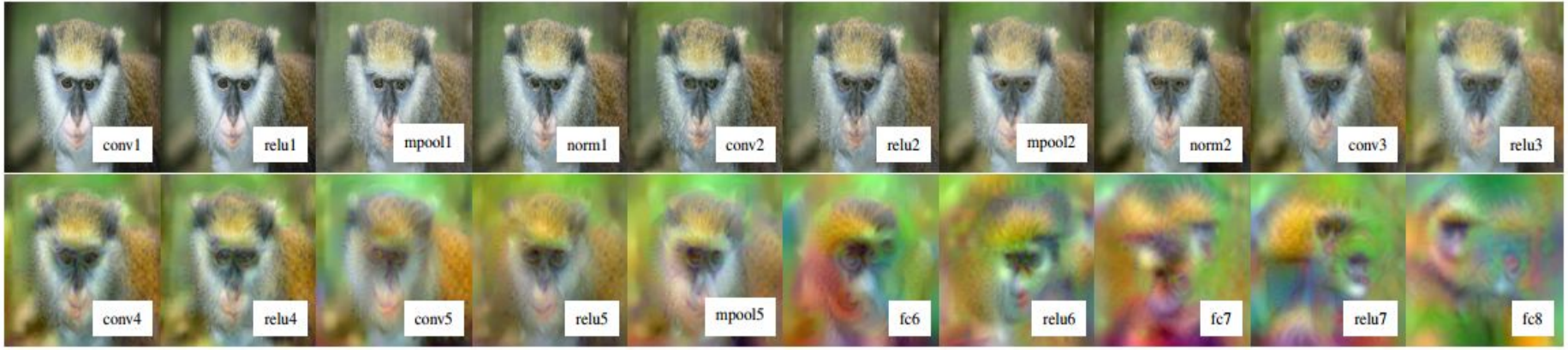


https://www.cs.toronto.edu/~frossard/post/vgg16/

Image Style Transfer Algorithm



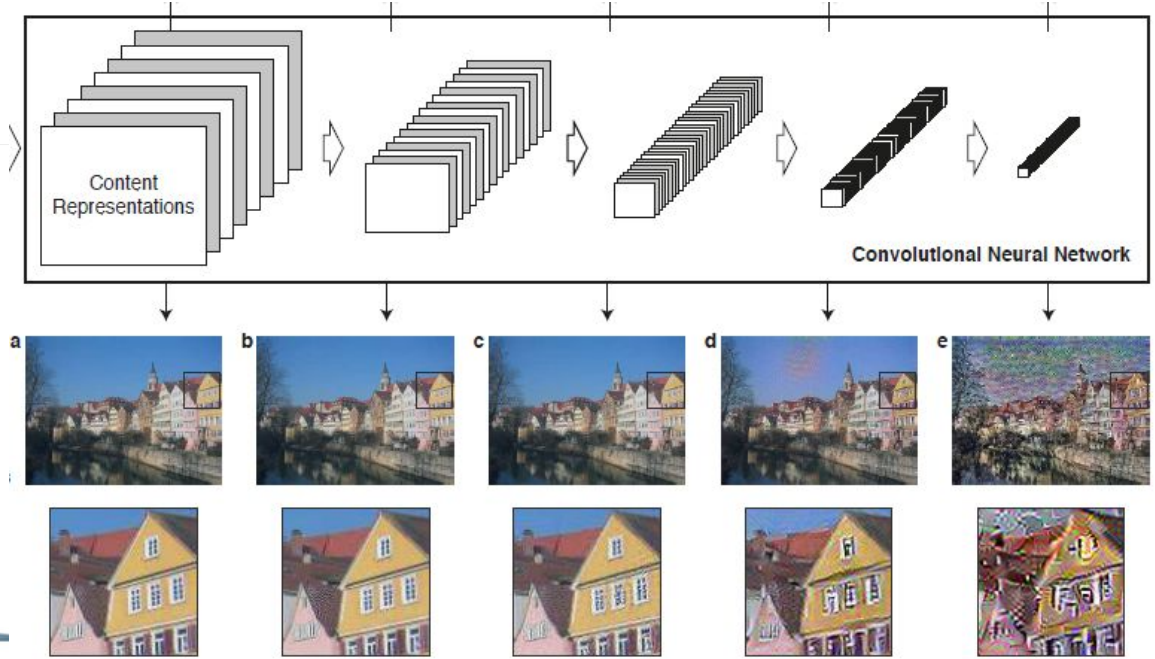
Deep Image Representation



A. Mahendran and A. Vedaldi, "Understanding deep image representations by inverting them," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 5188-5196.
doi: 10.1109/CVPR.2015.7299155

UCDAVIS

Higher Levels of the Network Preserve Semantic Content



UCDAVIS

Content Representation

To visualize the image information at each layer, a gradient descent is performed on a white noise image to find another image that matches the feature response of the original.

Loss Function for Image Content

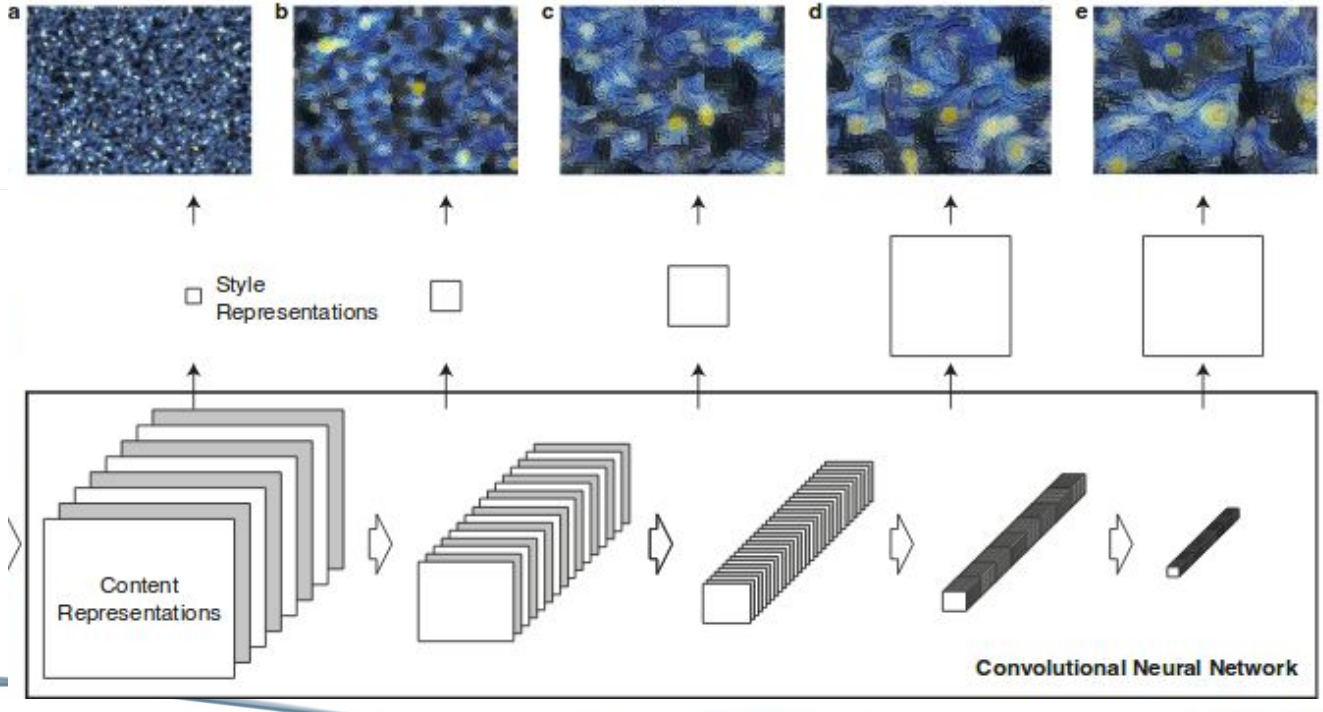
$$\mathcal{L}_{\text{content}}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} \left(F_{ij}^{l} - P_{ij}^{l}\right)^2$$

Derivative with Respect to Activations

$$\frac{\partial \mathcal{L}_{\text{content}}}{\partial F_{ij}^{l}} = \begin{cases} \left(F^{l} - P^{l}\right)_{ij} & \text{if } F_{ij}^{l} > 0 \\ 0 & \text{if } F_{ij}^{l} < 0 \,, \end{cases}$$

UCDAVIS

Style Representation



UCDAVIS

Style Representation

- Compute activations @ desired layers
- Calculate Gram Matrices of activations

$$G(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \begin{pmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \cdots & \langle \mathbf{v}_1, \mathbf{v}_k \rangle \\ \vdots & & \vdots \\ \langle \mathbf{v}_k, \mathbf{v}_1 \rangle & \cdots & \langle \mathbf{v}_k, \mathbf{v}_k \rangle \end{pmatrix}.$$

- Style Loss function:

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

- Total Style Loss:

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

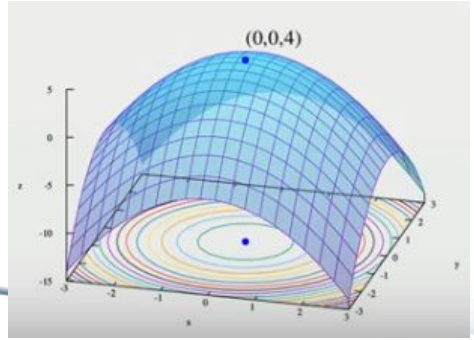
UCDAVIS

Training the Network

- Total Loss:

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

- L-BFGS Optimization

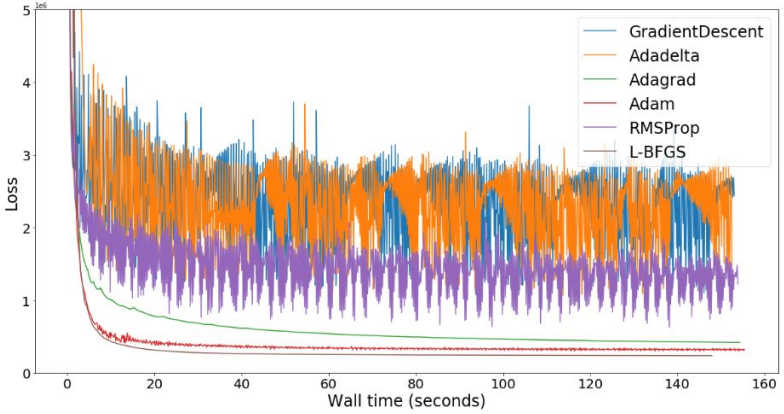


https://www.youtube.com/watch?v=1A6l0vcMNdc

Why L-BFGS?

Content image

Style image



Gradient Descent

Adadelta

RMSprop

Adam

Adagrad

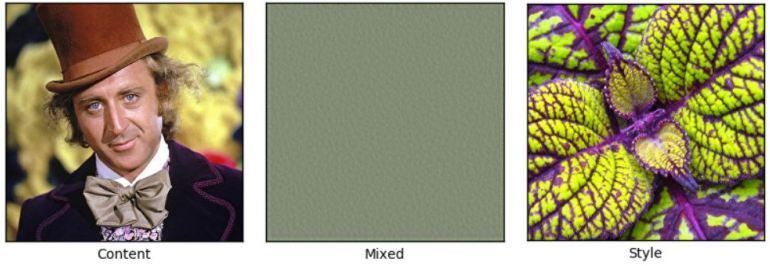
L-BFGS

UCDAVIS

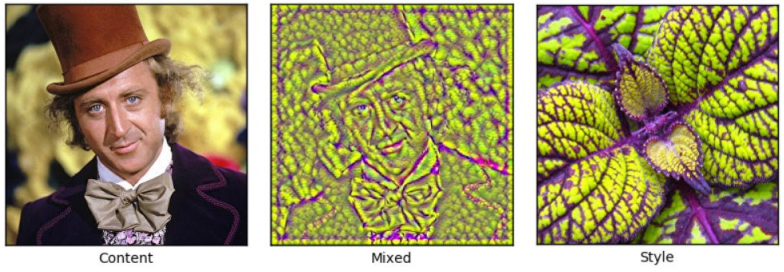
https://blog.slavv.com/picking-an-optimizer-for-style-transfer-86e7b8cba84b

Example

Iteration: 0
Weight Adj. for Content: 5.18e-11, Style: 2.14e-29, Denoise: 5.61e-06



Iteration: 10
Weight Adj. for Content: 2.79e-11, Style: 4.13e-28, Denoise: 1.25e-07



Iteration: 50
Weight Adj. for Content: 2.75e-11, Style: 1.12e-27, Denoise: 1.24e-07

Final image:

https://github.com/Hvass-Labs/TensorFlow-Tutorials/blob/master/15_Style_Transfer.ipynb

Style Matching vs. Content

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

UCDAVIS

Effect of Different Layers for Content

- Lower layers vs. Higher Layers
- Fine Structure vs. Style Blending

$(\alpha/\beta = 1 \times 10^{-3})$

UCDAVIS

Discussion

Strengths

- New process of image generation combining features of 2 images from different layers of a CNN
- “I don’t care what it was designed to do, I care about what it can do”

Weaknesses

- System is not perfect, photorealism is still a hard area to handle
- Can be used to create fake content
- Resolution is limited (better in more recent examples)

Photorealistic Image Style Transfers

Original photo | Reference photo | Result

https://phys.org/news/2017-05-cornell-cis-adobe-collaboration-artificial.html

UCDAVIS

Future Research

- Removing noise to achieve photorealistic image transfers
- Artistic outlets for creativity

UCDAVIS

Thank You

UCDAVIS