Deep Style Transfer

Andrea Battistello, Fabio Chiusano, Samuele Conti

Politecnico di Milano

2018

Table of Contents

Table of Contents

The starting point

Transferring the style of an image onto another image is a challenging task.

In recent years, Gatys et al. demonstrated the power of **Convolutional Neural Networks (CNN)** in combining the content of an image using the style of another image.

Why CNNs?

Remark

CNNs greatly improved the performance of style transfer algorithms thanks to their **higher-level representation** of the images.

When CNNs are trained on object recognition, they develop a high-level representation of an image leveraging on simple pixel values.

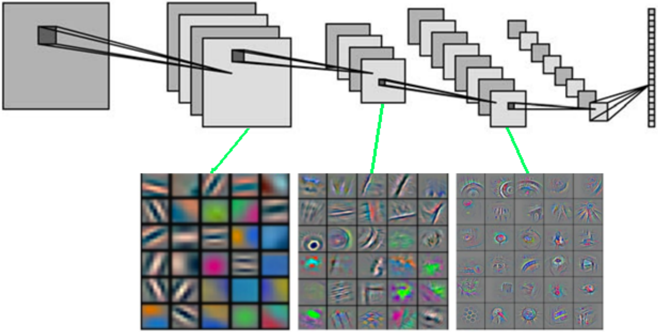


Table of Contents

Style Transfer with CNN

Reproducing painting styles on natural images, Gatys et al. found that the representations of image content and style were separable.

Remark

Image = Content + Style

Style Transfer with CNN

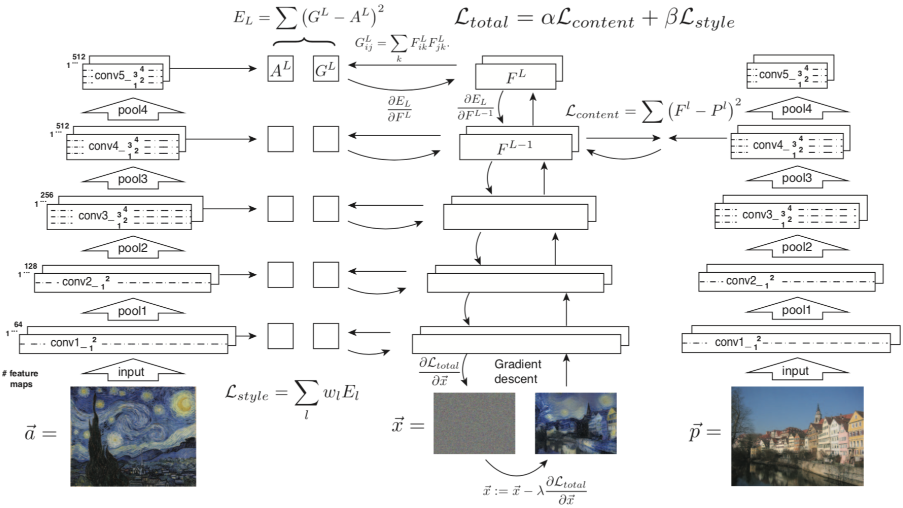
The algorithm starts from random noise as the initial result and then change the values of pixels iteratively.

In order to transfer the style from an image onto another, we can define two constraints: the **content loss** (the final image must resemble the content image) and the **style loss** (the final image must have the style of the style image).

Remark

$$\mathcal{L}_{Gatys} = \alpha \mathcal{L}_c + \beta \mathcal{L}_s$$

Style Transfer with CNN



Style Transfer with CNN

Content loss:

Remark

$$\mathcal{L}_c = \sum_{\ell=1}^{L} \frac{w_{c,\ell}}{2N_\ell D_\ell} \sum_{i=1}^{N_\ell} \sum_{p=1}^{D_\ell} (F_\ell[O] - F_\ell[I])_{ip}^2$$

Style loss:

Remark

$$\mathcal{L}_s = \sum_{\ell=1}^{L} \frac{w_{s,\ell}}{2N_\ell^2} \sum_{i=1}^{N_\ell} \sum_{j=1}^{N_\ell} (G_\ell[O] - G_\ell[S])_{ij}^2$$

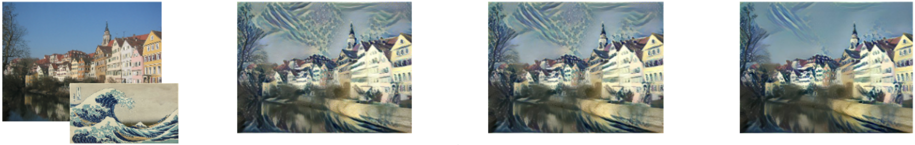
Table of Contents

Total variation loss

Johnson et al.[2] showed that adding a **total variation loss** produces smoother outputs thus improving style transfer result:

Total Variation Loss

$$\mathcal{L}_{tv}(O) = \sum_{x,y} (O_{x,y} - O_{x,y-1})^2 + (O_{x,y} - O_{x-1,y})^2$$



(a) Content and Style (b) $w_{TV} = 0$ (c) $w_{TV} = 0.001$ (d) $w_{TV} = 0.01$

Histogram Loss

The standard Gatys algorithm does not provide guarantees that the mean or variance of the texture is preserved. This problem is solved with **histogram loss**, introduced by Wilmot et al [7]:

Histogram Loss

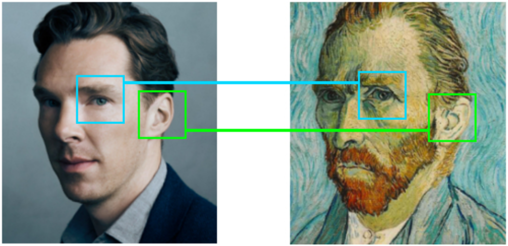
$$\mathcal{L}_{hist} = \sum_{\ell=1}^{L} \gamma_\ell \sum_{i=1}^{N_\ell} \sum_{p=1}^{D_\ell} (F_\ell[O] - R_\ell[O])^2_{ip}$$

(a) Content and Style (b) $w_{HIST} = 0$ (c) $w_{HIST} = 100$ (d) $w_{HIST} = 500$

Table of Contents

Transfer Style From Semantically Similar Patches

The original Gatys approach is based on the Gram matrix, which considers the whole style image. Better results can be obtained by taking the style from the patch in the style image **whose content is the most similar** to the one in the current patch in the content image. Li and Wand [3] built a nearest neighbors Markov Random Field (MRF) to match content patches to the most semantically similar style patch.

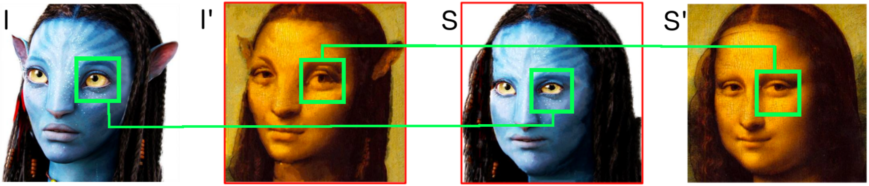


Transfer style from semantically similar patches

Example of bad style transfer due to wrong style selection:

Transfer style from semantically similar patches

Liao et al. improve the nearest neighbor field NNF [4] by considering bidirectional correspondence between the content and the style image.



Transfer style from semantically similar patches

[5] improves the matching of patches by adding a bonus whenever contiguous patches are mapped to contiguous patches.

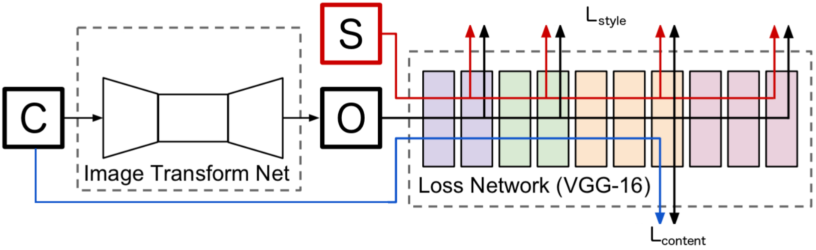
Table of Contents

Style transfer challenges

- **Fast style transfer**
- **Semantic augmentations**
- **Video**

Fast style transfer

[2] train a neural network on a specific style, so that it transforms the input image to a new image with same content but with the style the network was trained on. Therefore, time is spent on training the neural network so that style transfer can be applied with a single forward pass.



Semantic augmentations

In [1] semantic annotations are proposed to augment the CNNs architecture for style transfer, in order to obtain more control over the final outcome and increase its quality.

Video

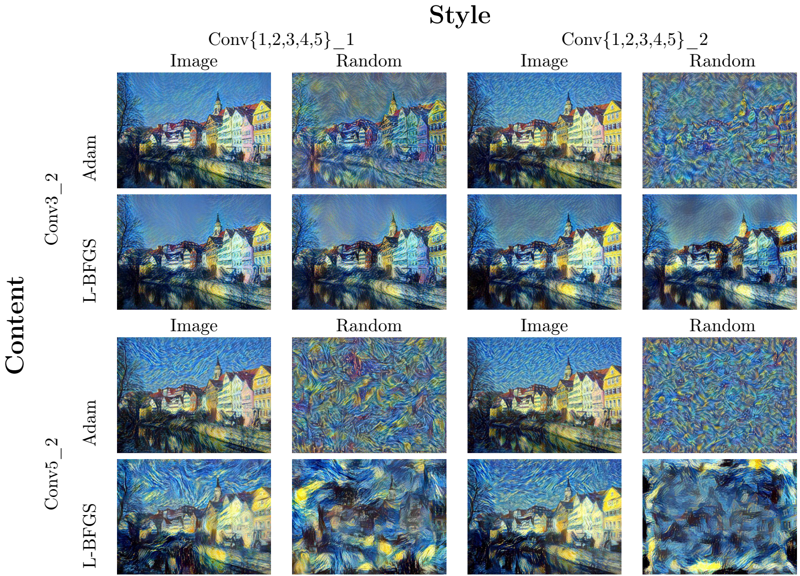
As showed in [6], style transfer can be applied to video too, but special attention must be taken in enforcing consistency between adjacent frames.

Table of Contents

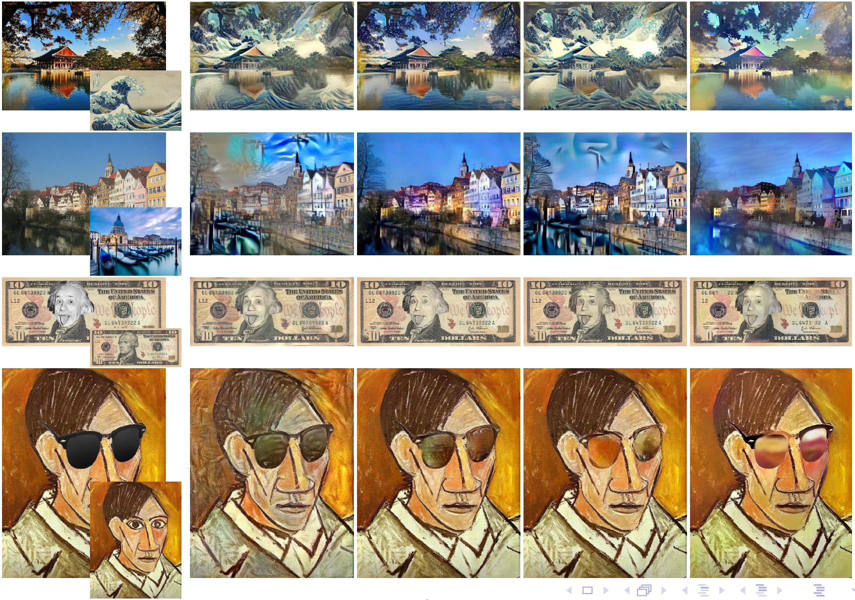
Experiments

We tested state-of-the-art algorithms for style transfer on a set of common used images for content and style. The first experiment uses an implementation of Gatys et al. and investigates the parameter space that motivated the choices made by the authors of the paper. The second experiment, instead, aims to compare the results of several algorithms on several content and style choices.

Experiment 1: Gatys parameters exploration



Experiment 2: Algorithms output comparison



Y.-L. Chen and C.-T. Hsu.
Towards deep style transfer: A content-aware perspective.
pages 8.1–8.11, 01 2016.

J. Johnson, A. Alahi, and L. Fei-Fei.
Perceptual losses for real-time style transfer and super-resolution.
In *European Conference on Computer Vision*, 2016.

C. Li and M. Wand.
Combining markov random fields and convolutional neural networks for image synthesis.
CoRR, abs/1601.04589, 2016.

J. Liao, Y. Yao, L. Yuan, G. Hua, and S. B. Kang.
Visual attribute transfer through deep image analogy.
CoRR, abs/1705.01088, 2017.

F. Luan, S. Paris, E. Shechtman, and K. Bala.
Deep painterly harmonization.
arXiv preprint arXiv:1804.03189, 2018.

M. Ruder, A. Dosovitskiy, and T. Brox.

Artistic style transfer for videos.

In B. Rosenhahn and B. Andres, editors, *Pattern Recognition*, pages 26–36, Cham, 2016. Springer International Publishing.

P. Wilmot, E. Risser, and C. Barnes.

Stable and controllable neural texture synthesis and style transfer using histogram losses.

CoRR, abs/1701.08893, 2017.