Controlling Perceptual Factors in Neural Style Transfer

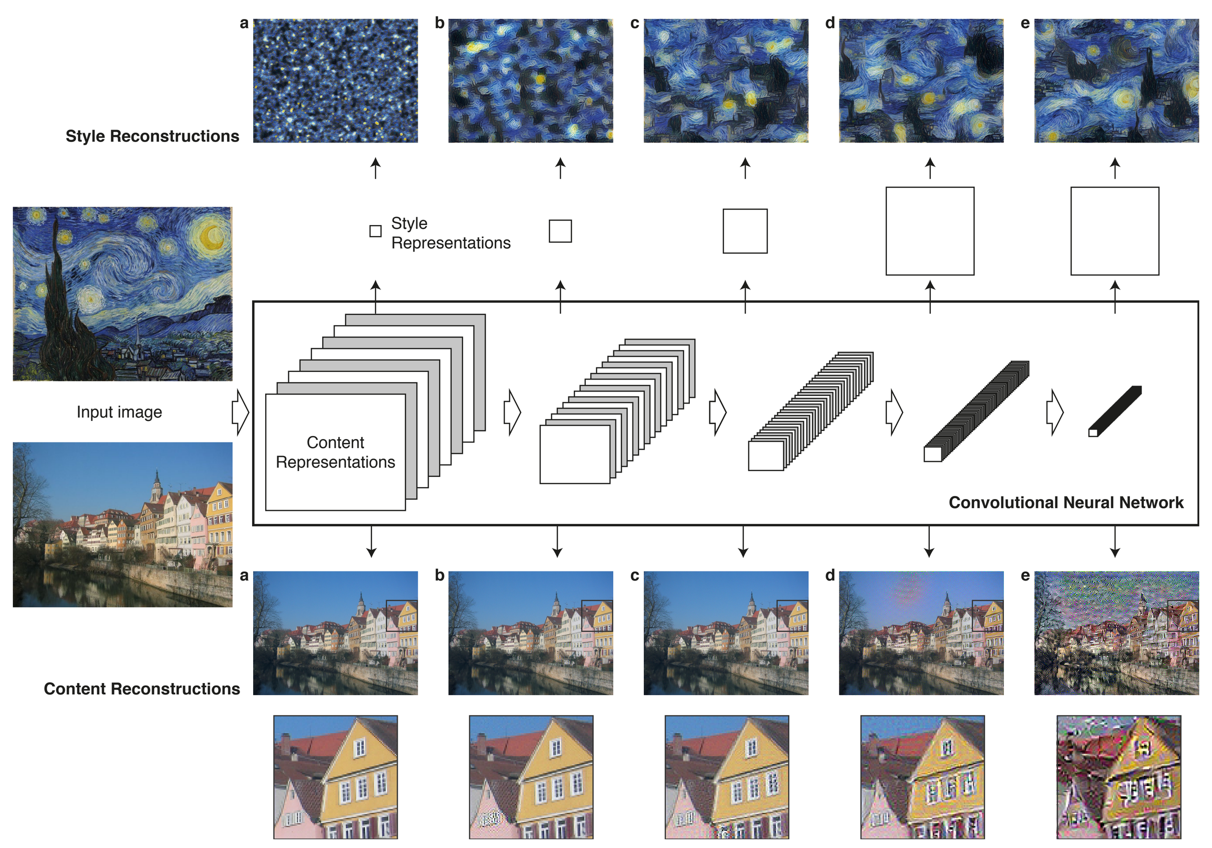
Leon A. Gatys[1-3], Alexander S. Ecker[1,2,4,5], Matthias Bethge[1,2,4], Aaron Hertzmann[6], Eli Shechtman[6]

[1]Center for Integrative Neuroscience and Institute for Theoretical Physics, University of Tuebingen, [2]Bernstein Center for Computational Neuroscience, Tuebingen, [3]Graduate School of Neural Information Processing, University of Tuebingen, [4]Max Planck Institute for Biological Cybernetics, Tuebingen, [5]Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA, [6]Adobe Research, USA

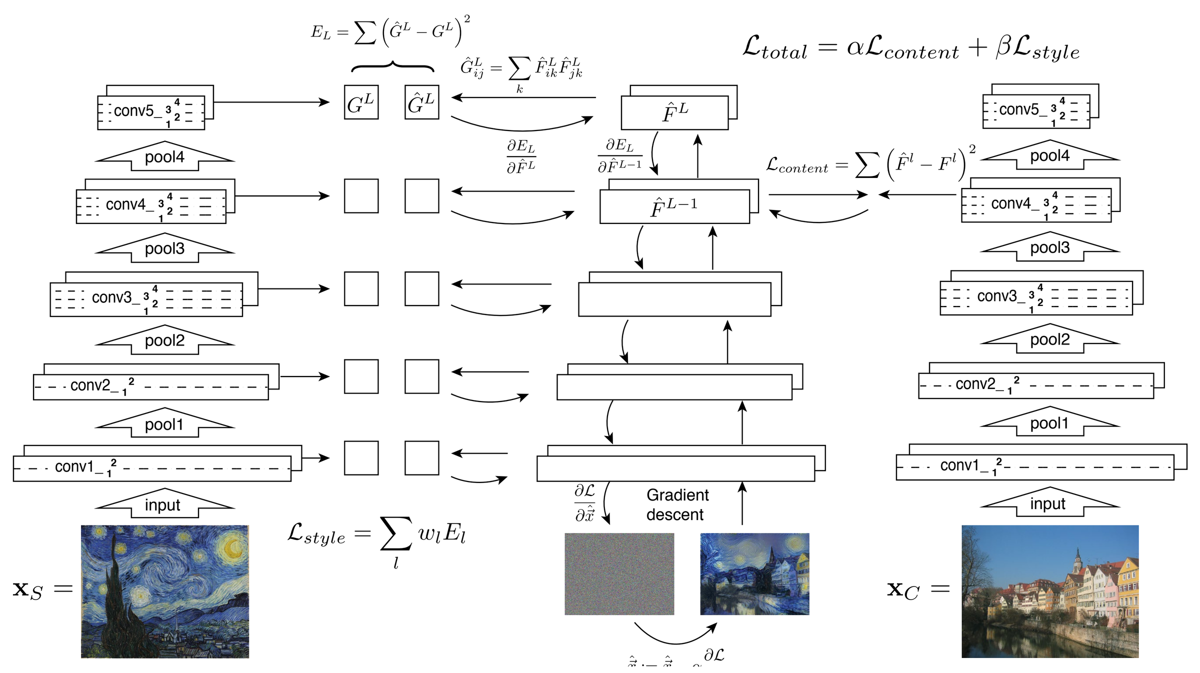
Introduction

Neural Style Transfer has shown exciting results enabling new forms of image manipulation. Here we extend the existing method beyond the paradigm of transferring global style information between pairs of images. In particular, we introduce control over spatial location, colour information and across spatial scale. We demonstrate how this enhances the method by allowing high-resolution controlled stylisation and helps to alleviate common failure cases such as applying ground textures to sky regions. Furthermore, by decomposing style into these perceptual factors we enable the combination of style information from multiple sources to generate new, perceptually appealing styles from existing ones. Finally we show how the introduced control measures can be applied in recent methods for Fast Neural Style Transfer.

CNN Image Representations



Neural Style Transfer



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

$$\mathcal{L}_{content} = \frac{1}{N_{\ell_c} M_{\ell_c}(\mathbf{x}_C)} \sum_{ij} (\mathbf{F}_{\ell_c}(\hat{\mathbf{x}}) - \mathbf{F}_{\ell_c}(\mathbf{x}_C))^2_{ij}$$

$$\mathcal{L}_{style} = \sum_{\ell} w_\ell E_\ell$$

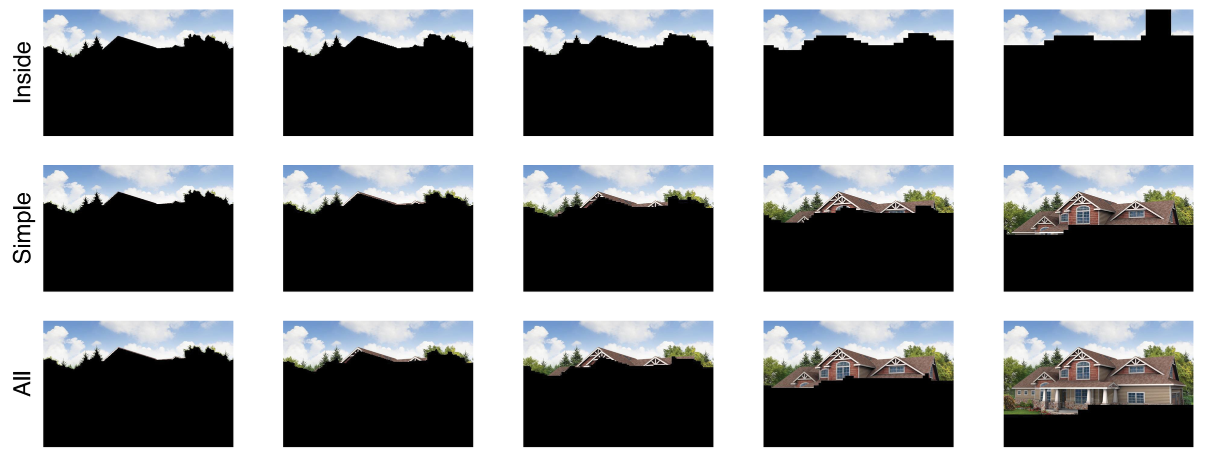
$$E_\ell = \frac{1}{4N_\ell^2} \sum_{ij} (\mathbf{G}_\ell(\hat{\mathbf{x}}) - \mathbf{G}_\ell(\mathbf{x}_S))^2_{ij} \qquad \mathbf{G}_\ell(\mathbf{x}) = \frac{1}{M_\ell(\mathbf{x})} \mathbf{F}_\ell(\mathbf{x})^T \mathbf{F}_\ell(\mathbf{x})$$

Spatial Control

Content | Style I | Output without spatial control

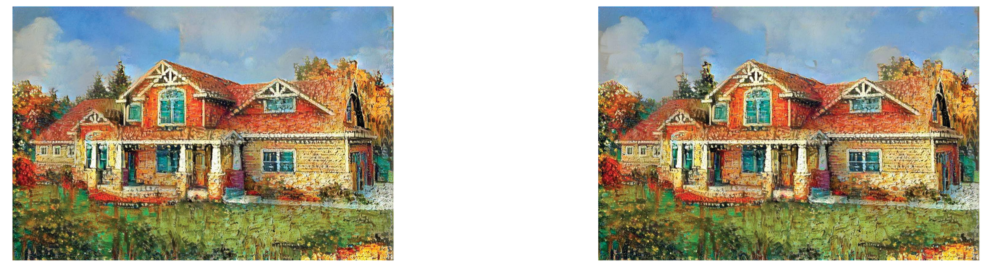
Guidance Propagation

Receptive fields on image



Conv1_1 | Conv2_1 | Conv3_1 | Conv4_1 | Conv5_1

'Inside' guidance | 'Simple' guidance | 'All' guidance



Inside' guidance, dilation: 3 | 'Simple' guidance, erosion: 1

Guidance Method

Mix styles with guided Gram Matrices | Mix styles with guided means | Style II

Guided Gram Matrices

- Compute one Gram Matrix for each spatial region. Optionally combine with global Gram Matrix: $E_\ell = \frac{\lambda_{global}}{4N_\ell^2} \sum_{ij} (\mathbf{G}_\ell(\hat{\mathbf{x}}) - \mathbf{G}_\ell(\mathbf{x}_S))^2_{ij} + \frac{1}{4N_\ell^2} \sum_{r=1}^{R} \lambda_r \sum_{ij} (\mathbf{G}^r_\ell(\hat{\mathbf{x}}) - \mathbf{G}^r_\ell(\mathbf{x}_S))^2_{ij}$

Guided feature map means

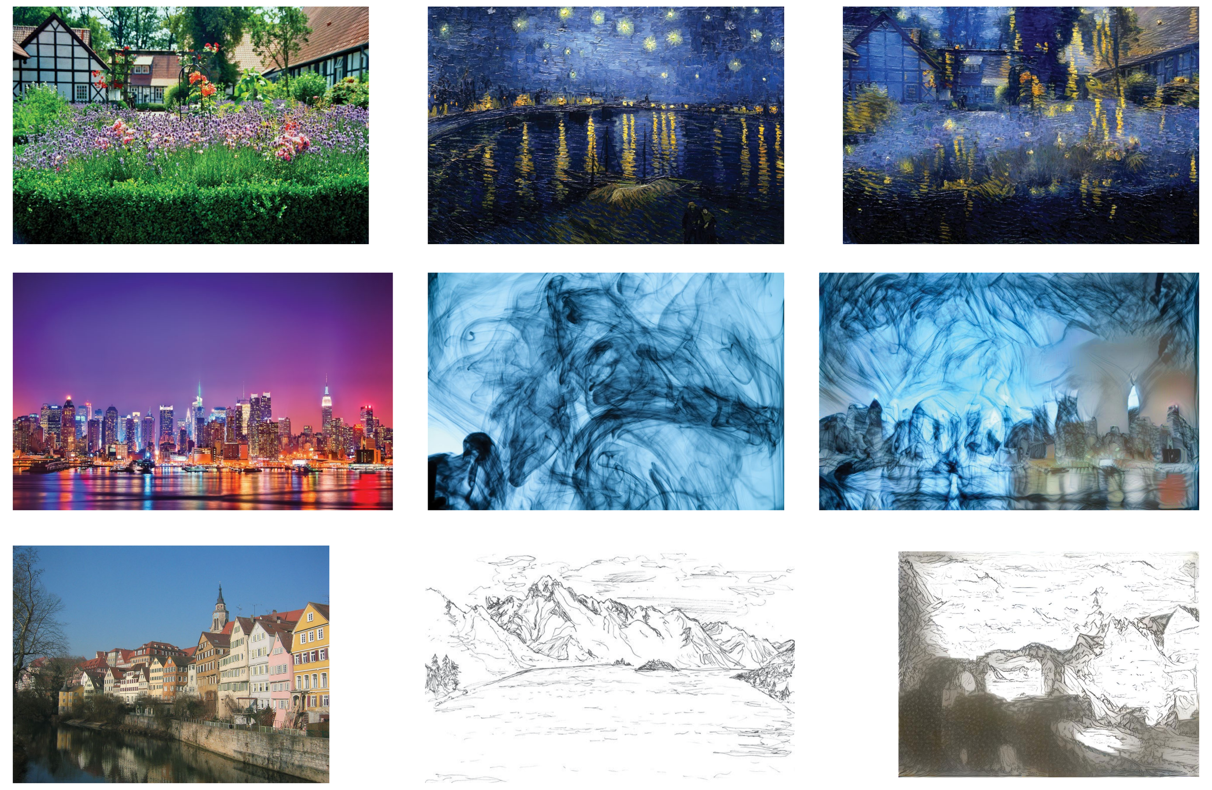
- Stack guidance channels with feature maps before computing the Gram Matrix. Equivalent to combining the global Gram Matrix with the local mean of the feature maps in each spatial region:

$$E_\ell = \frac{\lambda_{global}}{4N_\ell^2} \sum_{ij} (\mathbf{G}_\ell(\hat{\mathbf{x}}) - \mathbf{G}_\ell(\mathbf{x}_S))^2_{ij} + \frac{1}{2N_\ell} \sum_{r=1}^{R} \lambda_r \sum_{i} (\langle \mathbf{F}^r_\ell(\hat{\mathbf{x}}) \rangle - \langle \mathbf{F}^r_\ell(\mathbf{x}_S) \rangle)^2_{i}$$

with: $\mathbf{T}^r_\ell$ is a vectorised guidance channel and

$$\mathbf{F}^r_\ell(\mathbf{x})_{[:,i]} = \mathbf{T}^r_\ell \circ \mathbf{F}_\ell(\mathbf{x})_{[:,i]} \qquad \mathbf{G}^r_\ell(\mathbf{x}) = \mathbf{F}^r_\ell(\mathbf{x})^T \mathbf{F}^r_\ell(\mathbf{x})$$

Colour Control



Content I/II | Style I/II | Outputs without colour control

Colour Preservation

Output with luminance-only style transfer | Output with colour histogram matching

- Method I: Transfer style only in luminance channel and combine outcome with the colour channels of the content image.
- Method II: Transfer colour map from content to style image and then use result as the style image in Neural Style Transfer.

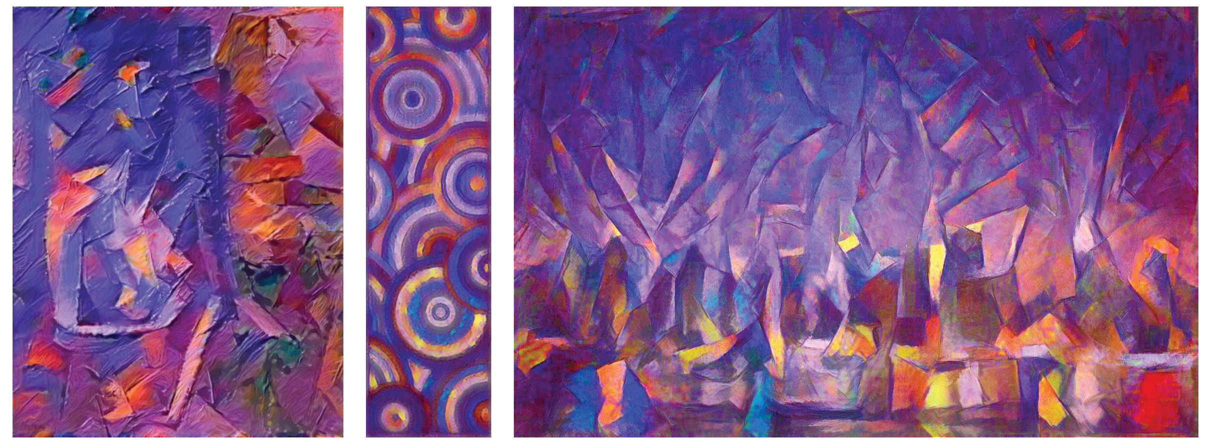
Improved Colour Transfer in Neural Style Transfer

Colour-matched content | Output using colour-matched content

- Transfer colour map from style to content image and then use result as the content image in Neural Style Transfer.

Scale Control

Content | Style I/II/III



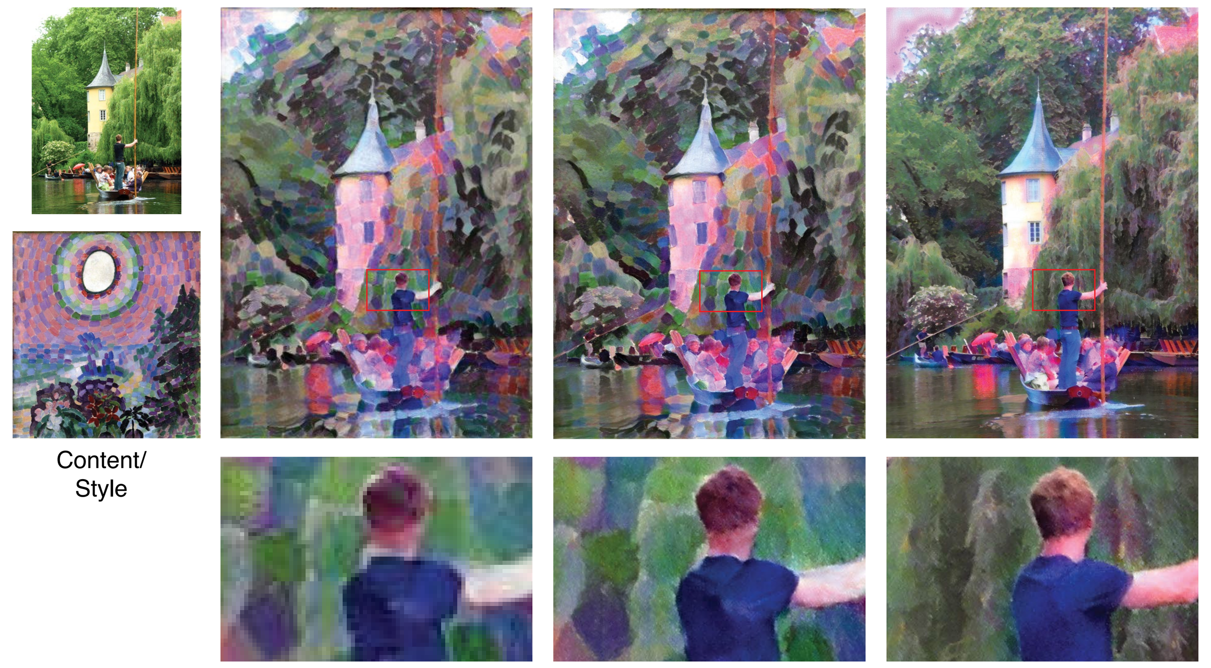
Style IV/V (new) | Output with style II

Output with style IV | Output with style V

Painting style factorises along spatial scale:

- Fine scale: material properties, brushstrokes. Coarse scale: round and angular shapes, swirls
- Combine fine and coarse scales from different images to generate new styles.
- Method: Transfer fine-scale style (only using low-layer Gram Matrices) from one style image onto another to generate new style image. Then use new style image in Neural Style Transfer.
- This works because image structures that are much larger than the receptive field size of the filters included in the style features remain unchanged during optimisation-based Neural Style Transfer.

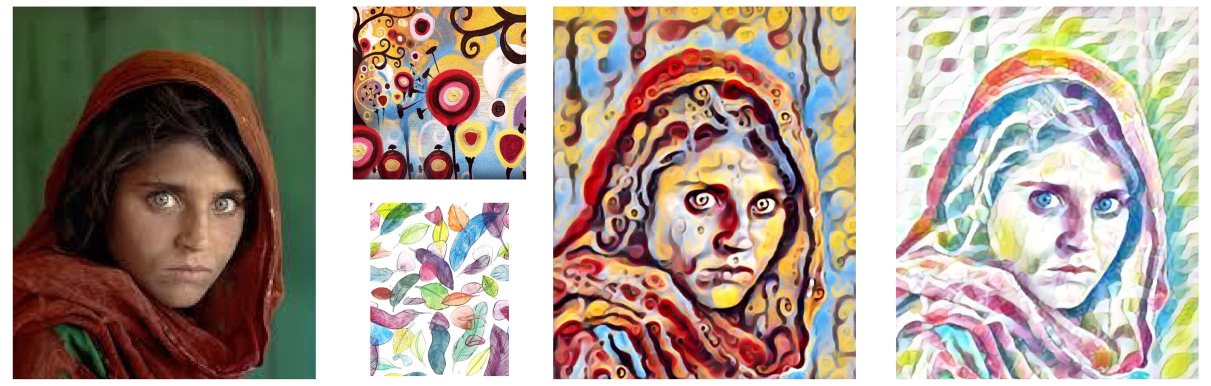
Scale Control for High Resolution



Content/ Style | Low-res | High-res (ctf) | High-res

- Naive stylisation in high-resolution fails because Network filters are too small compared to image size.
- Instead use coarse-to-fine procedure: Perform stylisation in low-resolution and use the upsampled result as initialisation for high-resolution stylisation.

Controlling Fast Neural Style Transfer



Content | Style I/II | Output with style I/II

Spatial Control

Vertical mask | Horizontal mask | Automated Person/Background masks

- Train network that applies different styles to different regions.
- Input: Content image and spatial mask, output: region-specific stylised image.
- Surprising: Training with fixed mask generalises to arbitrary masks.

Colour Preservation

Content | Original network | Luminance network

- Method I: Combine luminance channel of standard output with colour channels of the content image
- Method II: Train network only on luminance images and combine output with colour channels of the content image

Conclusions

- Perceptual factors of style: spatial location, spatial scale, colour and luminance information.
- Introduced practical ways to control these factors in Neural Style Transfer
- Recombining factors from several images allows creation of perceptually appealing new styles
- Control measures transfer to recent Fast Style Transfer methods

For more Neural Style check out: www.deepart.io

Code online at: github.com/leongatys/NeuralImageSynthesis

Leon Gatys: leon.gatys@bethgelab.org

This work was funded by the German National Academic Foundation (L.A.G.), the Bernstein Center for Computational Neuroscience (FKZ 01GQ1002) and the German Excellency Initiative through the Centre for Integrative Neuroscience Tuebingen (EXC307)(M.B., A.S.E, L.A.G.). Declaration of conflict of interest: The L.A.G, A.S.E and M.B have co-founded the company DeepArt UG (haftungsbeschraenkt) that runs the website www.deepart.io.