Intro to Deep Learning

Chapter: Style Transfer

Prof. Noam Koenigstein

Style Transfer

A Neural Algorithm of Artistic Style

Leon A. Gatys,[1,2,3*] Alexander S. Ecker,[1,2,4,5] Matthias Bethge[1,2,4]

[1]Werner Reichardt Centre for Integrative Neuroscience
and Institute of Theoretical Physics, University of Tübingen, Germany
[2]Bernstein Center for Computational Neuroscience, Tübingen, Germany
[3]Graduate School for Neural Information Processing, Tübingen, Germany
[4]Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[5]Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA
[*]To whom correspondence should be addressed; E-mail: leon.gatys@bethgelab.org

A Neural Algorithm of Artistic Style

Leon A. Gatys,[1,2,3*] Alexander S. Ecker,[1,2,4,5] Matthias Bethge[1,2,4]

- *"In fine art, especially painting, humans have mastered the skill to create unique visual experiences through composing a complex interplay between the **content** and **style** of an image."*

- *"We introduce an artificial system based on a Deep Neural Network that creates **artistic** images of high perceptual quality."*

- *"The key finding of this paper is that the representations of **content** and **style** in the Convolutional Neural Network are **separable**. Therefore, we can manipulate both representations independently to produce new, perceptually meaningful images."*

The original photograph depicting the Neckarfront in Tubingen, Germany

The Shipwreck of the Minotaur by J.M.W. Turner, 1805

The Starry Night by Vincent van Gogh, 1889.

The Scream by Edvard Munch, 1893.

Femme nue assise by Pablo Picasso, 1910.

Composition VII by Wassily Kandinsky, 1913

A Neural Algorithm of Artistic Style

Leon A. Gatys,[1,2,3*] Alexander S. Ecker,[1,2,4,5] Matthias Bethge[1,2,4]

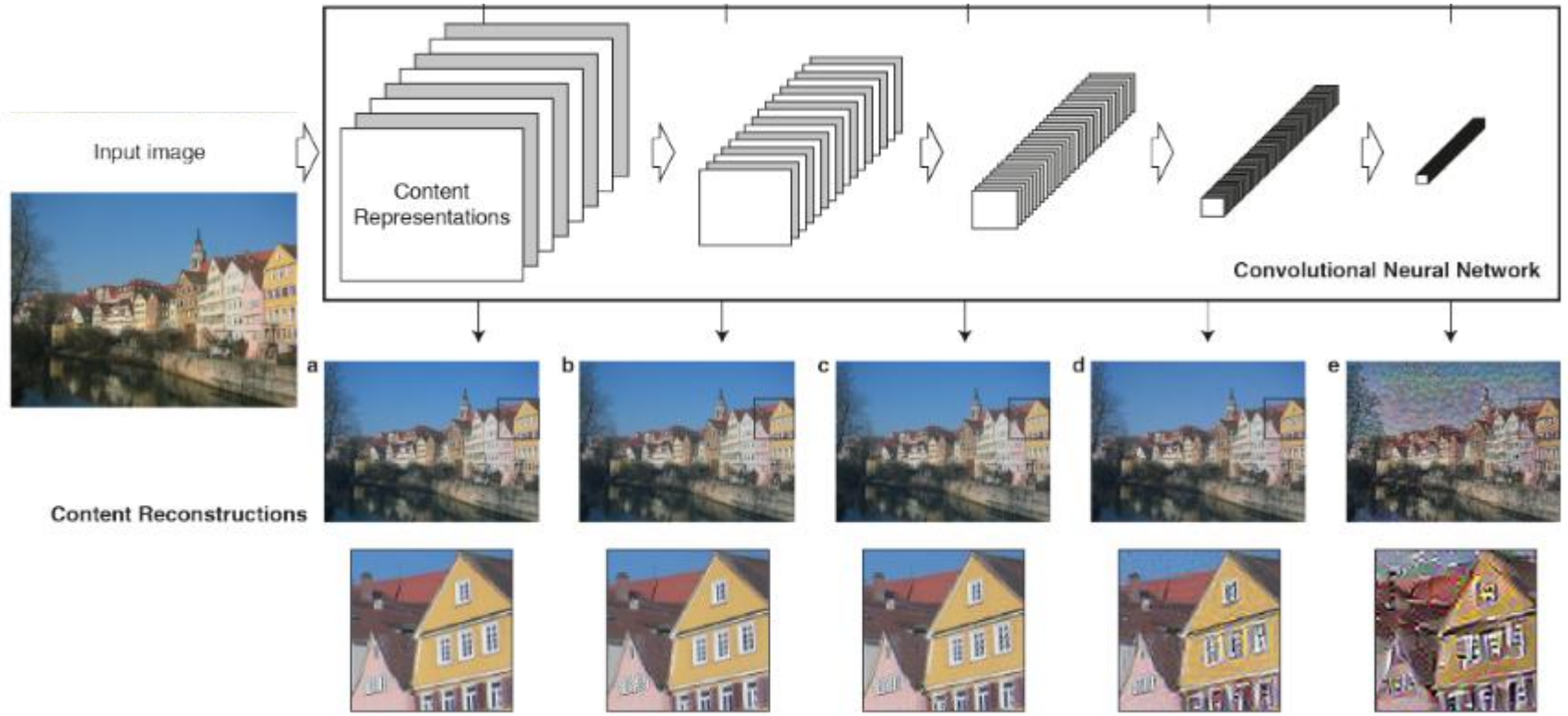
- *"When Convolutional Neural Networks are trained on object recognition, they develop a **representation** of the image that makes object information **increasingly explicit** along the processing hierarchy."*

- *"Therefore, along the processing hierarchy of the network, the input image is transformed into **representations** that **increasingly care about** the actual **content** of the image compared to its detailed pixel values."*

Content Representations

- *"Higher layers in the network capture the **high-level content** in terms of objects and their arrangement in the input image but **do not constrain the exact pixel values** of the reconstruction."*

- *"In contrast, reconstructions from the **lower layers** simply reproduce the **exact pixel values** of the original image."*



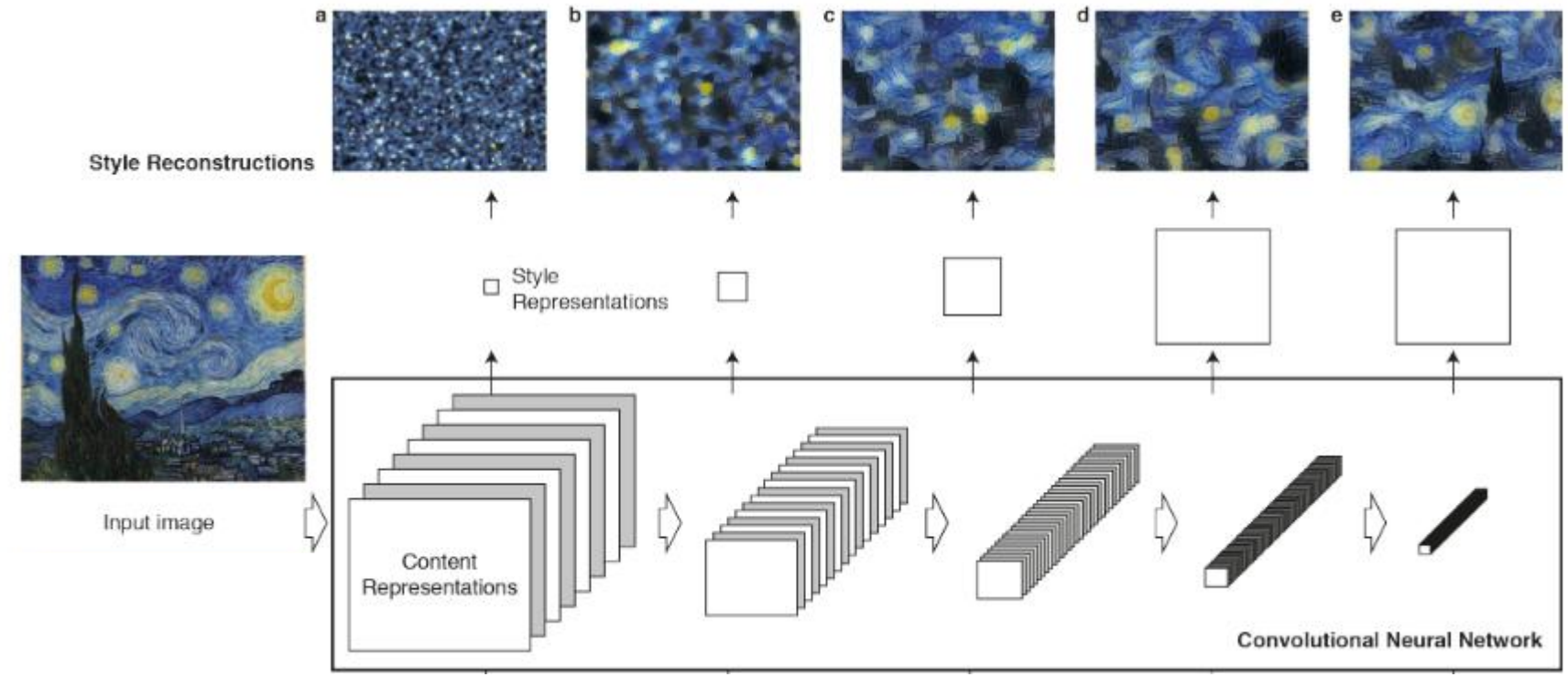
<u>Content Reconstructions</u>: We can visualize the information at different processing stages in the CNN by reconstructing the input image from only knowing the network's responses in a particular layer. We reconstruct the input image from layers 'conv1 1' (a), 'conv2 1' (b), 'conv3 1' (c), 'conv4 1' (d) and 'conv5 1' (e) of the original VGG-Network. We find that reconstruction from lower layers is almost perfect (a,b,c). In higher layers of the network, detailed pixel information is lost while the high-level content of the image is preserved (d,e).

Style Representations

- *"To obtain a representation of the **style** of an input image, we use a feature space designed to capture **texture information**."*

- *"The feature space is built on top of the filter responses in each layer of the network and consists of the **correlations between the different filter responses** over the spatial extent of the feature maps."*

- *"By including the feature correlations of multiple layers, we obtain a multi-scale representation of the input image, which captures its **texture** information but not the global arrangement."*


a
b
c
d
e
Style Reconstructions
Style
Representations
Input image
Content
Representations
Convolutional Neural Network

Method – content loss

Let $\vec{p}$ and $\vec{x}$ be the original image and the image that is generated and P^l and F^l their respective feature representation in layer l.

The content loss term between two feature representations is given by:

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

Method – style loss

On top of the CNN responses in each layer of the network we built a style representation that computes the correlations between the different filter responses, where the expectation is taken over the spatial extend of the input image.

These ***feature correlations*** *are given by the Gram matrix* $G^l \in \mathbb{R}^{N_l \times N_l}$*,* *where* $\boldsymbol{G_{ij}}$ ***is the inner product between the vectorized feature map*** $\boldsymbol{i}$ ***and*** $\boldsymbol{j}$ ***in layer*** $\boldsymbol{l}$*:*

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l.$$

Method – style loss cont.

Let $\vec{a}$ and $\vec{x}$ be the original image and the image that is generated and A^l and G^l their respective style representations in layer l. The contribution of that layer to the total loss is then:

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

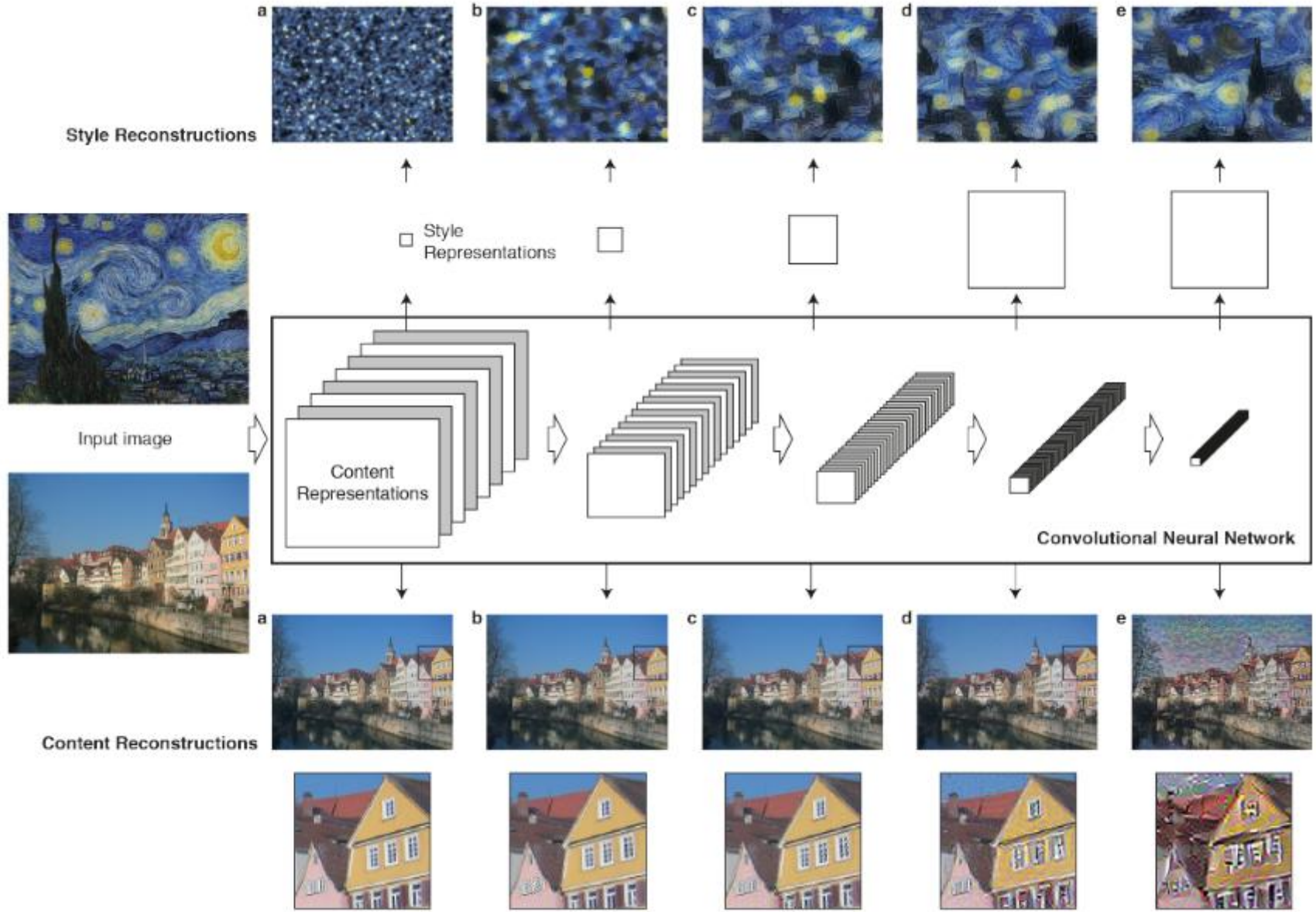
and the total loss is

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

where w_l are weighting factors of the contribution of each layer to the total loss.

The **rows** show the result of matching the style representation of increasing subsets of the CNN layers. We find that the local image structures captured by the style representation increase in size and complexity when including style features from higher layers of the network. This can be explained by the increasing receptive field sizes and feature complexity along the network's processing hierarchy.

The **columns** show different relative weightings between the **content** and **style** reconstruction. The number above each column indicates the ratio α/β between the emphasis on matching the **content** of the photograph and the **style** of the artwork. (α- content loss coefficient. B- style loss coefficient).


a
b
c
d
e
Style Reconstructions
Style
Representations
Input image
Content
Representations
Convolutional Neural Network
a
b
c
d
e
Content Reconstructions


A
B
C
D
E
F

A Neural Algorithm of Artistic Style

Leon A. Gatys,[1,2,3*] Alexander S. Ecker,[1,2,4,5] Matthias Bethge[1,2,4]

- *"It is fascinating that a neural system, which is trained to perform **object recognition**, one of the core computational tasks of biological vision, automatically learns image representations that allow the separation of image content from style."*

- *"The explanation could be that when learning object recognition, the network has to become invariant to all image variation that preserves object identity."*

- *"Thus, **our ability to separate content from style**, and therefore **our ability to create and enjoy art**, may **result from the way our visual system processes information**!"*