Image Style Transfer

NST

Definition

Image-to-Image Translation

- Transformation between different domains.
- shared features : content, domain-specific features : style

Style Transfer

- Preserving the content of an image while applying the style of a reference image.

Recent Trend

- Both fields have evolved toward disentangling content and style for more flexible and controllable generation.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

Image Style Transfer
NST

CVPR

2016

Image Style Transfer Using Convolutional Neural Networks

Leon A. Gatys
Centre for Integrative Neuroscience, University of Tübingen, Germany
Bernstein Center for Computational Neuroscience, Tübingen, Germany
Graduate School of Neural Information Processing, University of Tübingen, Germany
leon.gatys@bethgelab.org

Alexander S. Ecker
Centre for Integrative Neuroscience, University of Tübingen, Germany
Bernstein Center for Computational Neuroscience, Tübingen, Germany
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
Baylor College of Medicine, Houston, TX, USA

Matthias Bethge
Centre for Integrative Neuroscience, University of Tübingen, Germany
Bernstein Center for Computational Neuroscience, Tübingen, Germany
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Abstract

Rendering the semantic content of an image in different styles is a difficult image processing task. Arguably, a major limiting factor for previous approaches has been the lack of image representations that explicitly represent semantic information and, thus, allow to separate image content from style. Here we use image representations derived from Convolutional Neural Networks optimised for object recognition, which make high level image information explicit. We introduce A Neural Algorithm of Artistic Style *that can sep-*

there exist a large range of powerful non-parametric algorithms that can synthesise photorealistic natural textures by resampling the pixels of a given source texture [7, 30, 8, 20]. Most previous texture transfer algorithms rely on these non-parametric methods for texture synthesis while using different ways to preserve the structure of the target image. For instance, Efros and Freeman introduce a correspondence map that includes features of the target image such as image intensity to constrain the texture synthesis procedure [8]. Hertzman et al. use image analogies to transfer the texture from an already stylised image onto a target image[13].

Image Style Transfer

NST

Background & Goal

Previous research limitation

- Separate content from style in natural images is still a difficult problem

Goal

- Separate the content and style of the image and apply a new style

Style **Content**

Fig 1. Style transfer results

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

Image Style Transfer

NST

Network Architecture

VGG Network

- Basis of the 19 layers VGG network

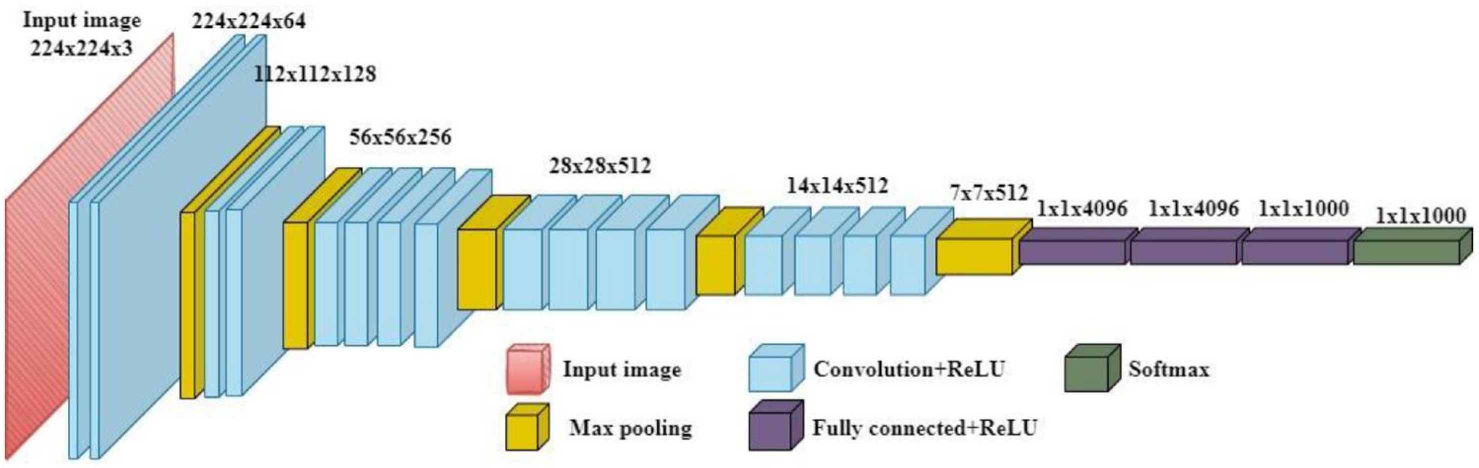


Fig 2. VGG19 architecture

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

NST

Feature Representations

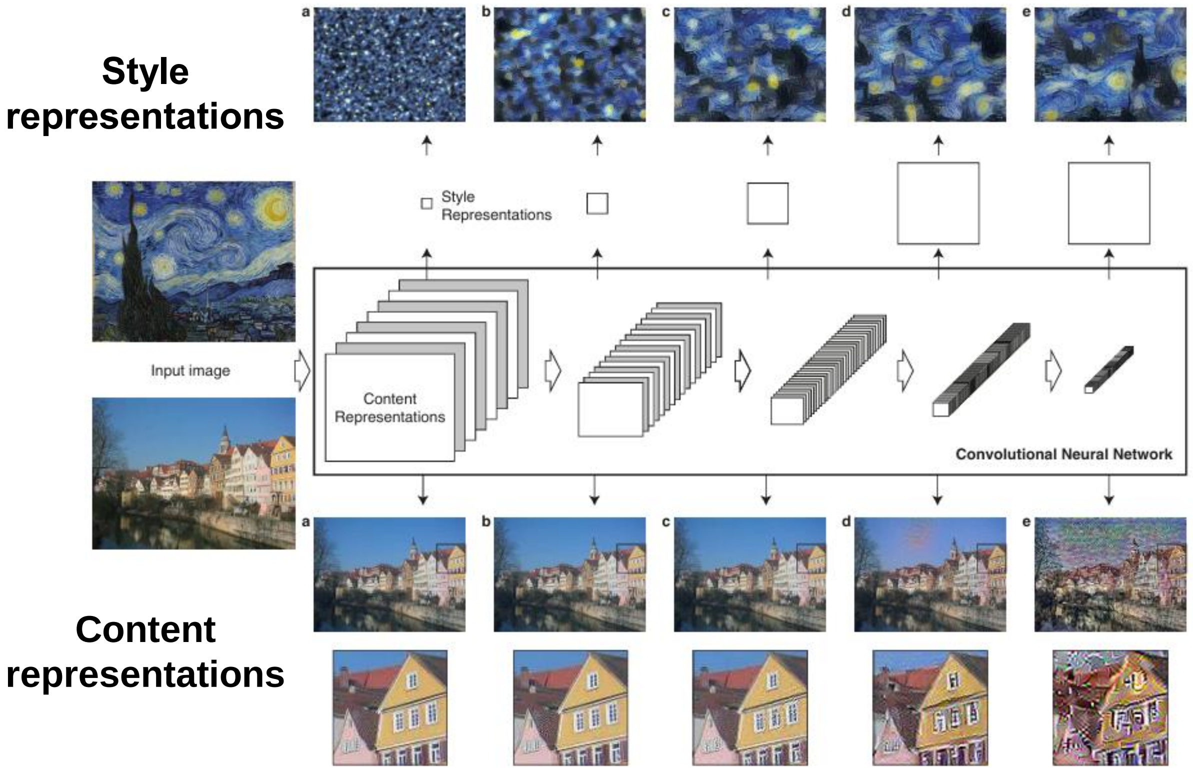
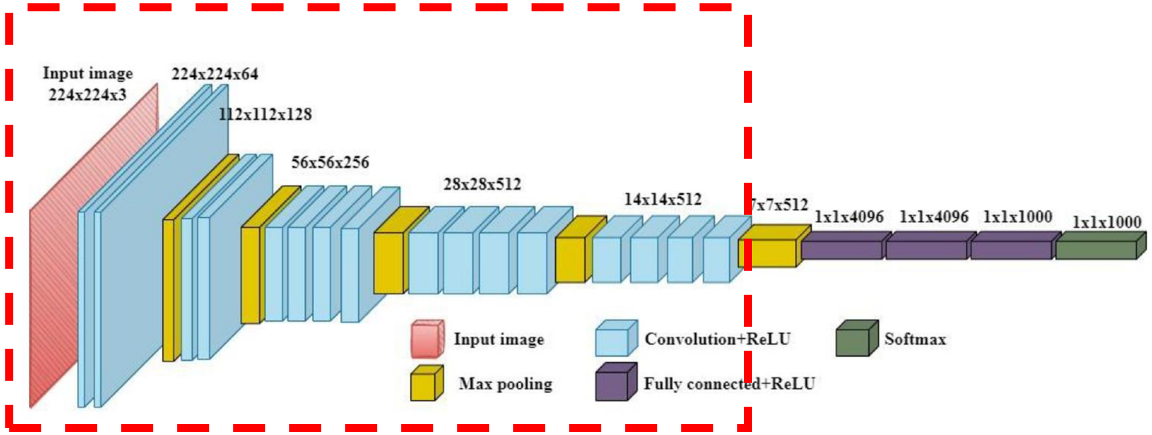


Fig 3. Image representations in a CNN



Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

Image Style Transfer

NST

Feature Representations

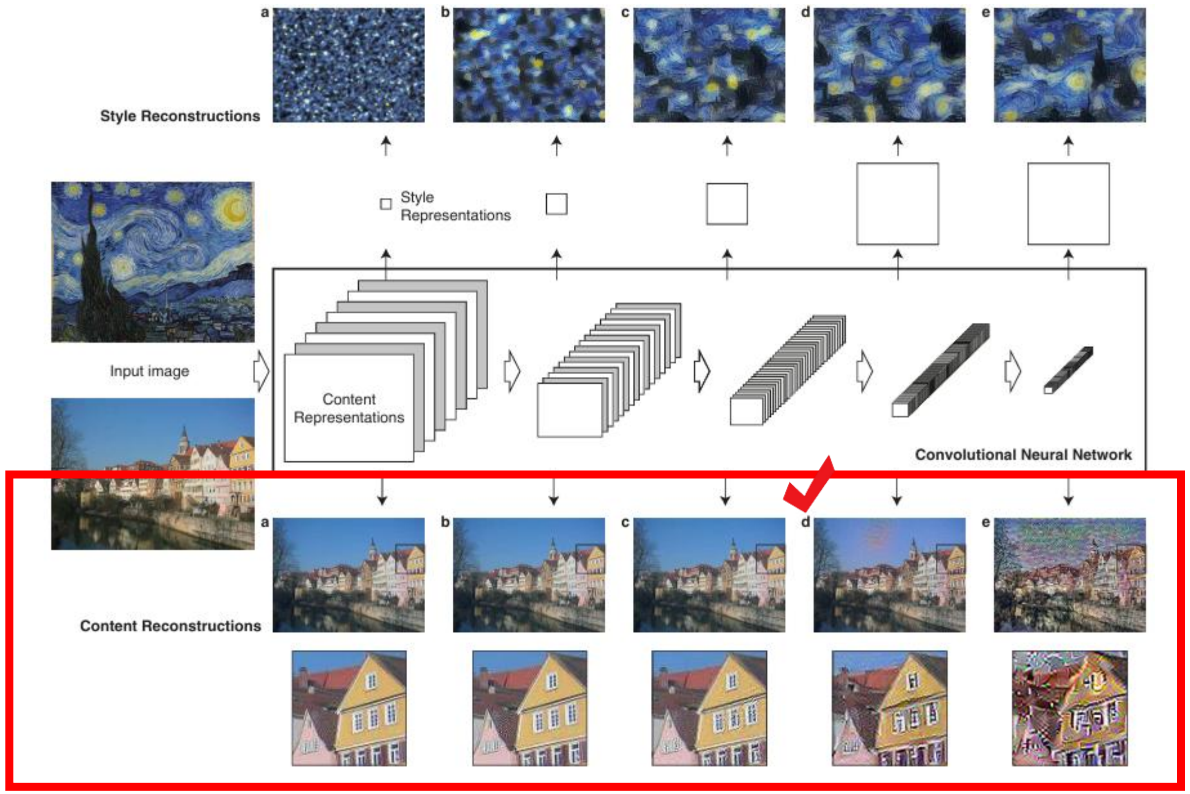


Fig 3. Image representations in a CNN

Content representations

- The higher layers (conv4_2) of a pre-trained VGG network capture high-level content that represents the semantic meaning of an image.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

Image Style Transfer
NST

Feature Representations

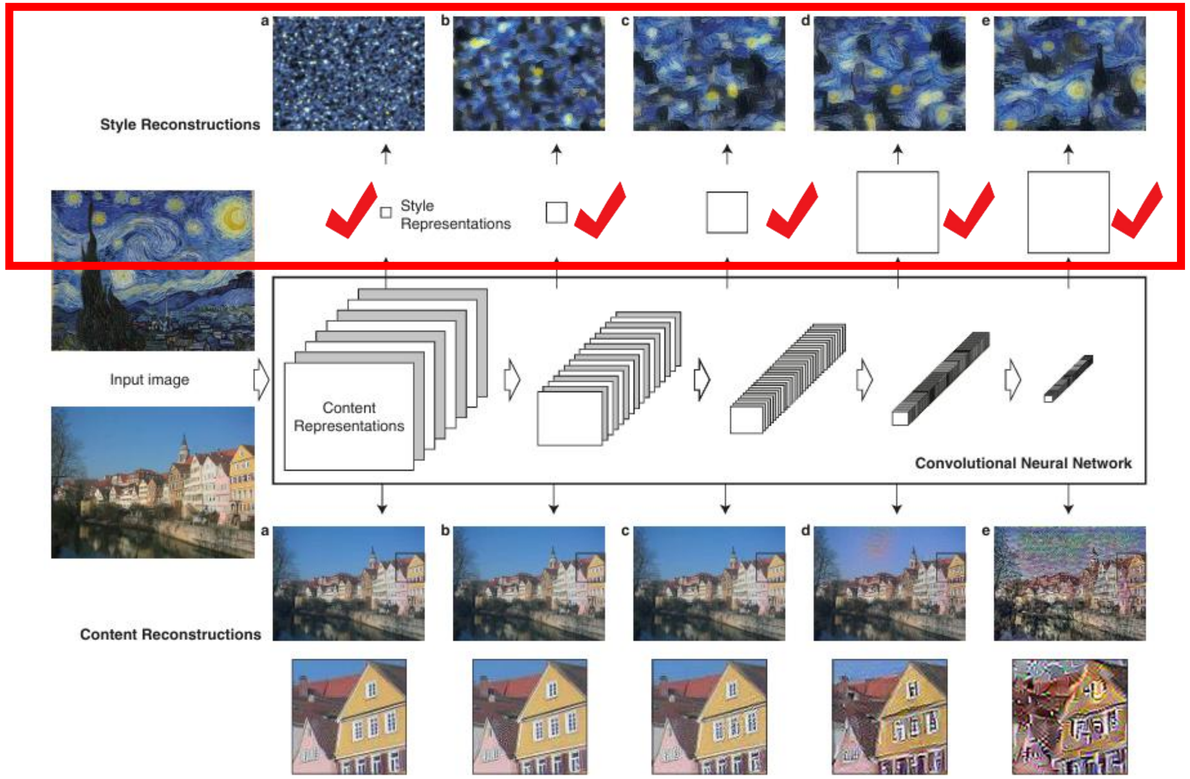


Fig 3. Image representations in a CNN

Style representations

- Style is the overall texture, color distribution, and pat tern of the image.
- Expressed as a relationship between entire channel s (features), not individual pixels
  => Gram matrix

$$\mathbf{F} \in \mathbb{R}^{C \times H \times W} \qquad \mathbf{F}_{\text{flat}} \in \mathbb{R}^{C \times (H \times W)}$$

$$\mathbf{G} = \mathbf{F}_{\text{flat}} \mathbf{F}_{\text{flat}}^T \in \mathbb{R}^{C \times C}$$

$$G_{ij} = \sum_k F_{ik} F_{jk}$$

- Gram matrix calculates the correlations between var ious feature maps while disregarding spatial informa tion

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

Image Style Transfer

NST

Feature Representations

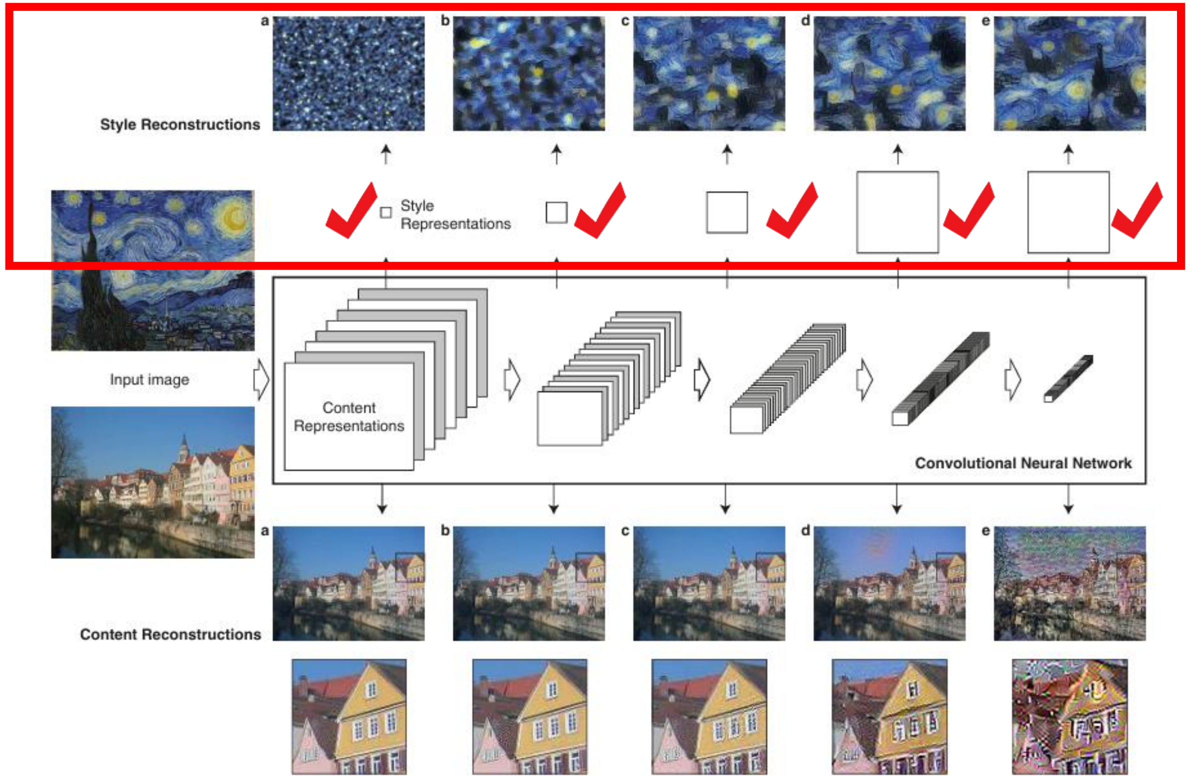


Fig 3. Image representations in a CNN

Style representations

- The lower layers (conv1_1) of a pre-trained VGG network capture fine-grained texture patterns and colors, while the higher layers (conv4_1, conv5_1) represent more abstract style properties such as global patterns.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

Image Style Transfer

NST

Loss Functions

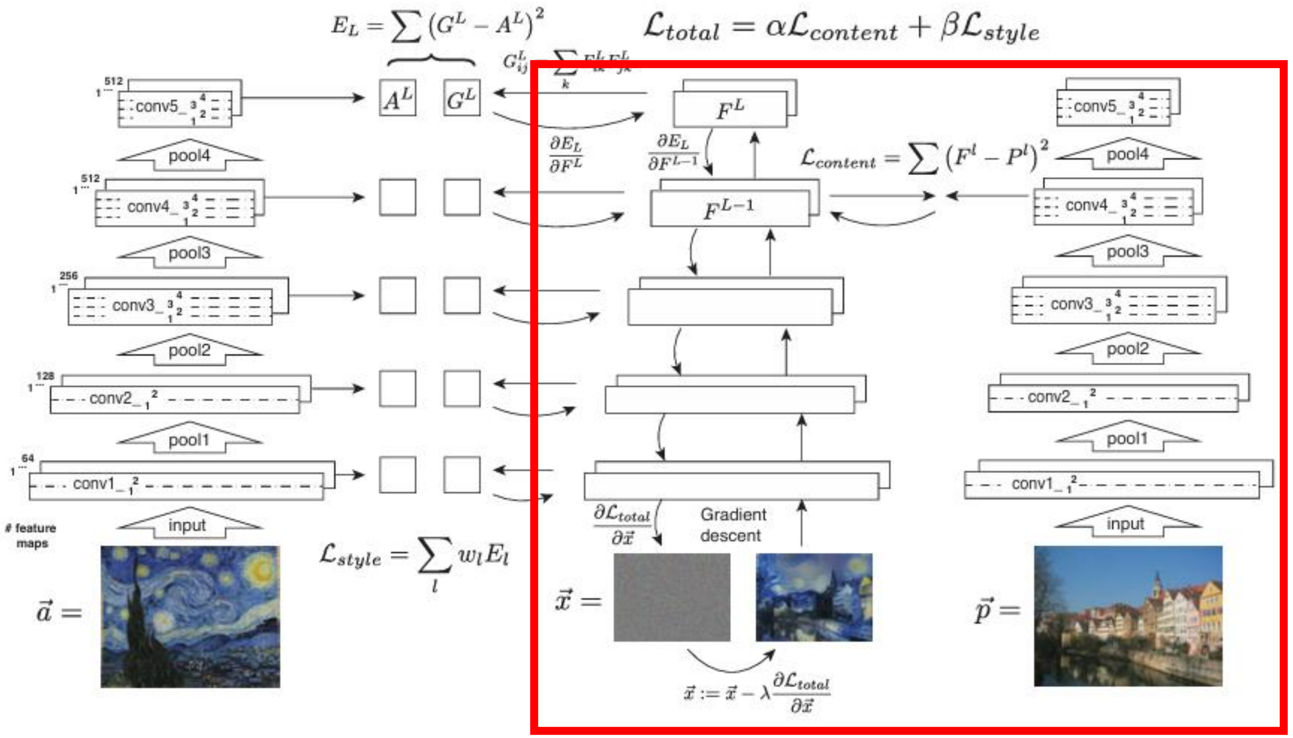


Fig 4. Style transfer Algorithm

Content loss

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

- $\vec{p}$, $\vec{x}$: Original & Generate image
- P^l, F^l: Feature representations in layer l
- $F_{i,j}^l$: Activation of the i^{th} filter at position j in layer l

Eq1. Content loss

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

Image Style Transfer

NST

Loss Functions

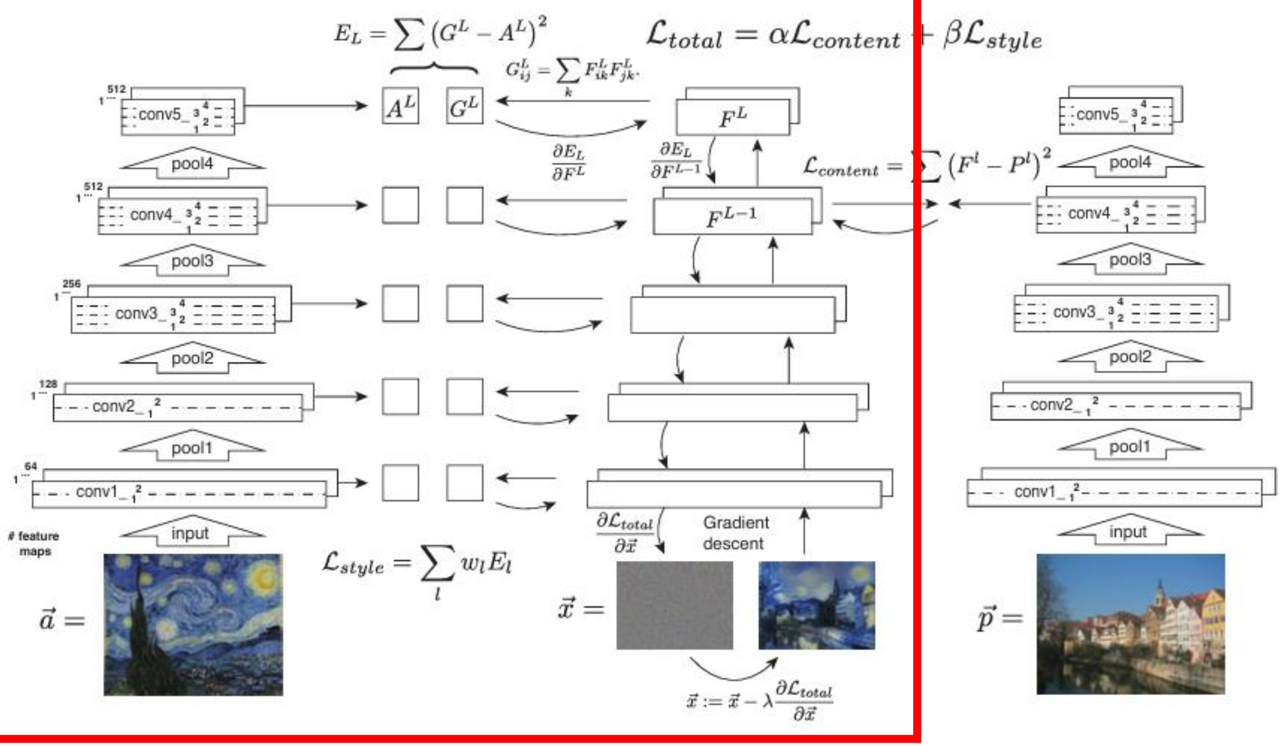


Fig 4. Style transfer Algorithm

Style loss

$$E_L = \sum (G^L - A^L)^2 \quad G^l_{ij} = \sum_k F^l_{ik} F^l_{jk}$$

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

- $\vec{a}, \vec{x}, w_l$: Original, Generate image, weight factor
- A^l, G^l: Gram Matrix in layer l
- $G^l_{i,j}$: Inner between the vectorized feature maps

Eq2. Style loss

Image Style Transfer

NST

Loss Functions

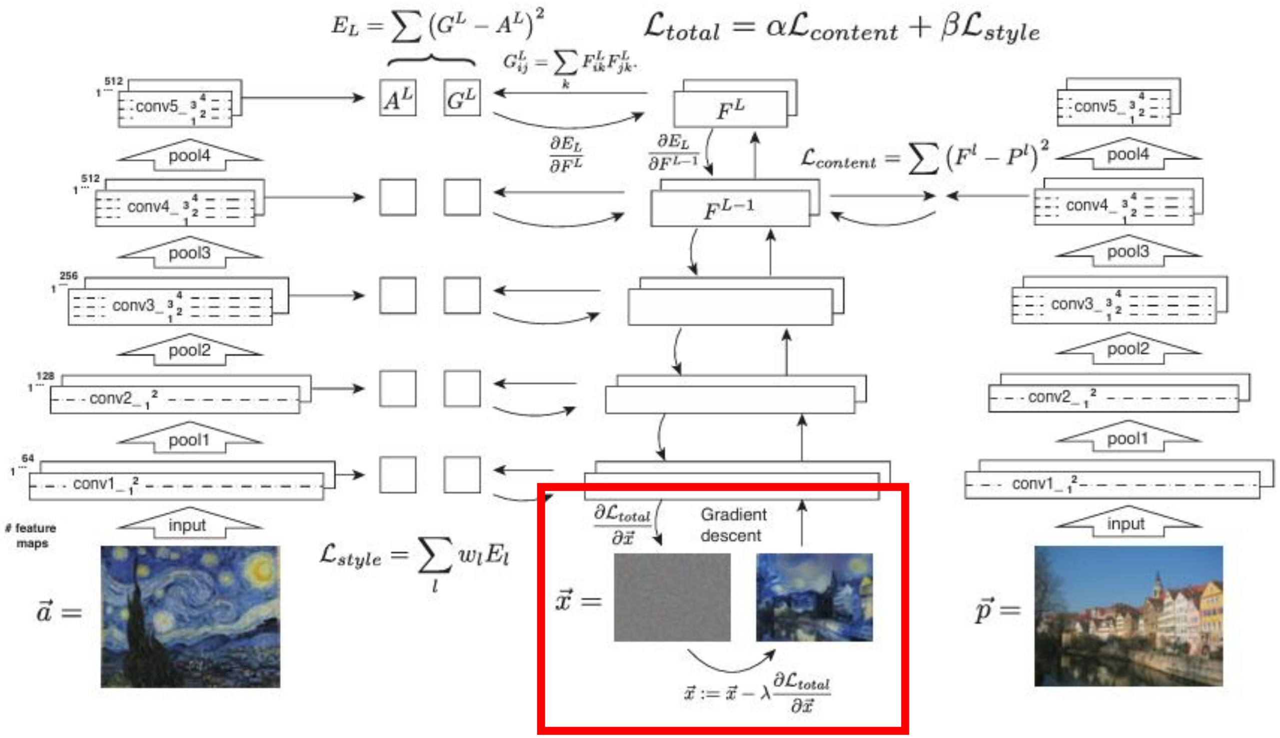


Fig 4. Style transfer Algorithm

Transfer

- Perform gradient descent on a white noise image
- Synthesis a new image that simultaneously matches the content representation of $\vec{p}$ and the style representation of $\vec{a}$

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

Eq 3. Total loss function

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

Image Style Transfer

NST

Experiment Results

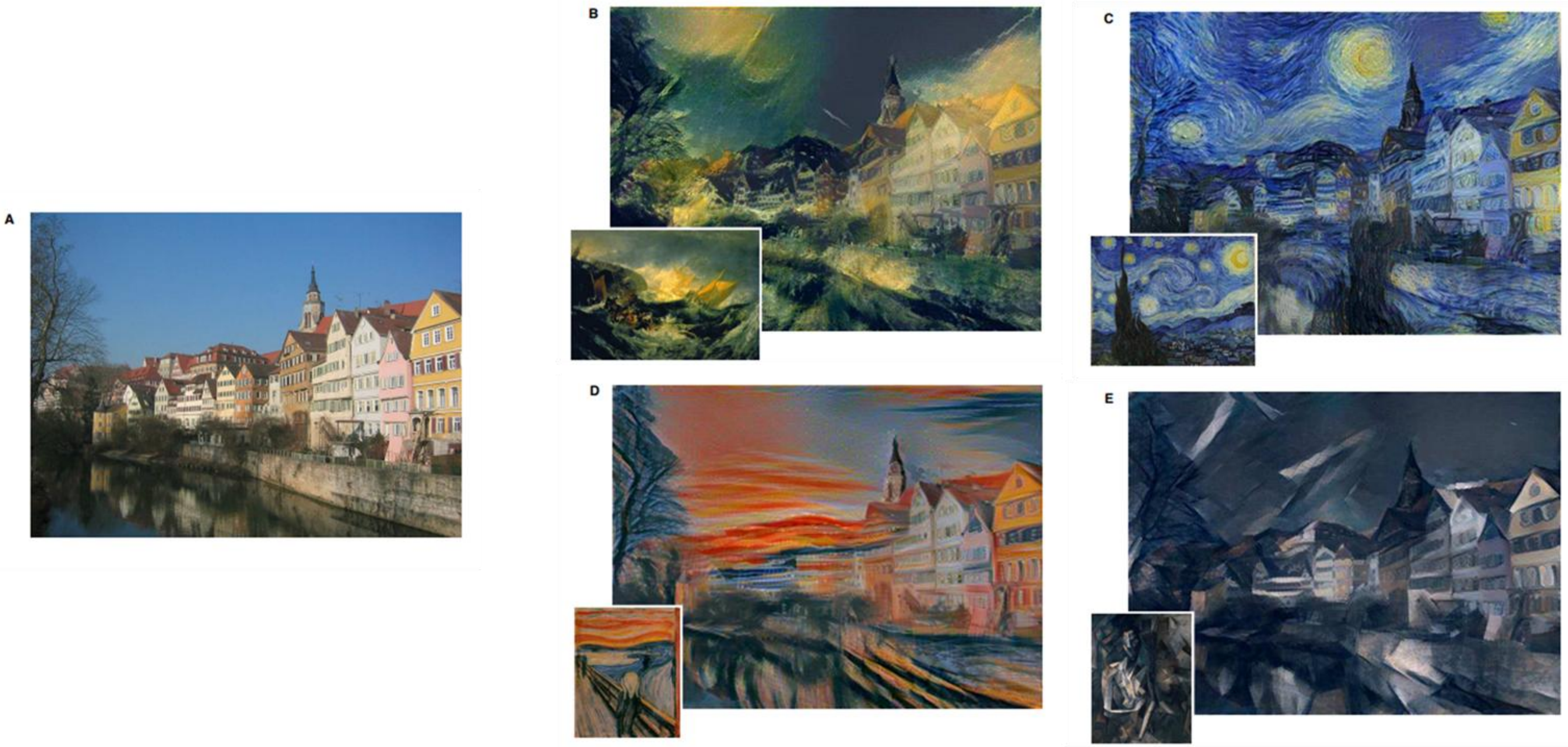


Fig 5. Images that combine the content of a photograph with the style

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

NST

Experiment Results

Why are content and style entangled in the synthesized image?

- VGG Network was not originally trained by separating content and style.
- Content and style are mixed in the feature map extracted from VGG.
- Gram matrix captures texture and global statistics but is agnostic to spatial information.
- Interpret and utilize the features extracted by a pre-trained VGG network in a meaningful way.

Image Style Transfer

NST

Experiment Results

Why are content and style entangled in the synthesized image?

- VGG Network was not originally trained by separating content and style.
- Content and style are mixed in the feature map extracted from VGG.
- Gram matrix captures texture and global statistics but is agnostic to spatial information.
- Interpret and utilize the features extracted by a pre-trained VGG network in a meaningful way.

Recent Trend

- To better disentangle content and style
- Adaptive Instance Normalization (AdaIN) separate style information through normalization techniques.
- Disentangled representation learning, content and style are explicitly separated into different latent spaces.

Image Style Transfer

NST

Additional Experiments

Trade-off between content and style matching

- A linear combination between the loss functions for content and style
- Strong emphasis on style: produces a texturized version of the style image
- Strong emphasis on content: produces an image with only little stylization

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

Fig 6. Images generated by α / β ratio

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

Image Style Transfer

NST

Additional Experiments

- **Effect of different layers**

Fig 7. The effect of matching the content representation in different layers of the network.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

Image Style Transfer

NST

Additional Experiments

- **Initialization of gradient descent**

Content Style Random noise

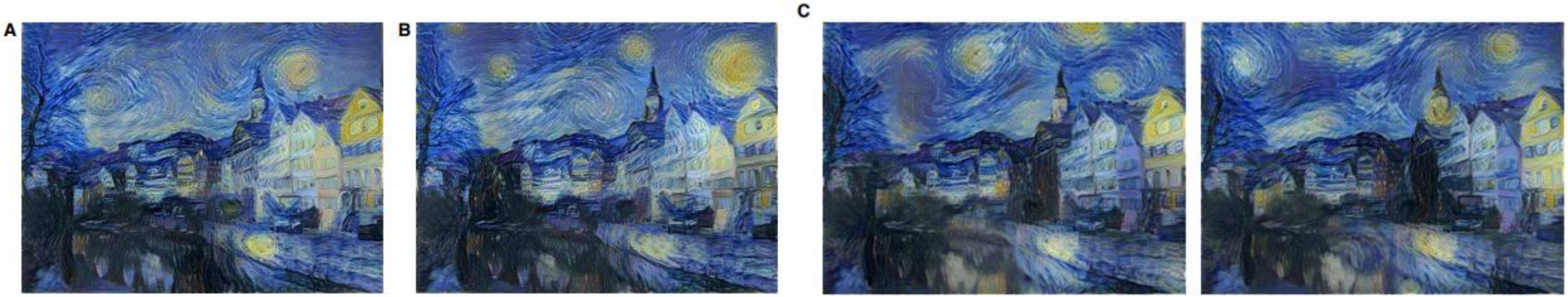


Fig 8. Results for initialization image

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

Image Style Transfer

NST

Implications & Limitations

Implications

- To achieve style transfer between arbitrary images using feature representations extracted from Convolutional Neur al Network

Limitations

- Optimization-based approach => Not suitable for real-time processing

- Difficulty in separating content and style