Neural Style Transfer

ICFP M2 Machine Learning – Final Presentation

Ömer Ünlüsoy, Jonathan Menssen, Jorge Cerrada

2024 - 2025

WHAT IS STYLE TRANSFER?

- Computer Vision subfield
- Extraction of the artistic style and its transfer into another image.
    - While preserving the semantic content [1].

- **Evaluation Criterion**: human inspection
- **Main Challenge**: find image representations that independently model content and style.

WHAT IS STYLE TRANSFER?

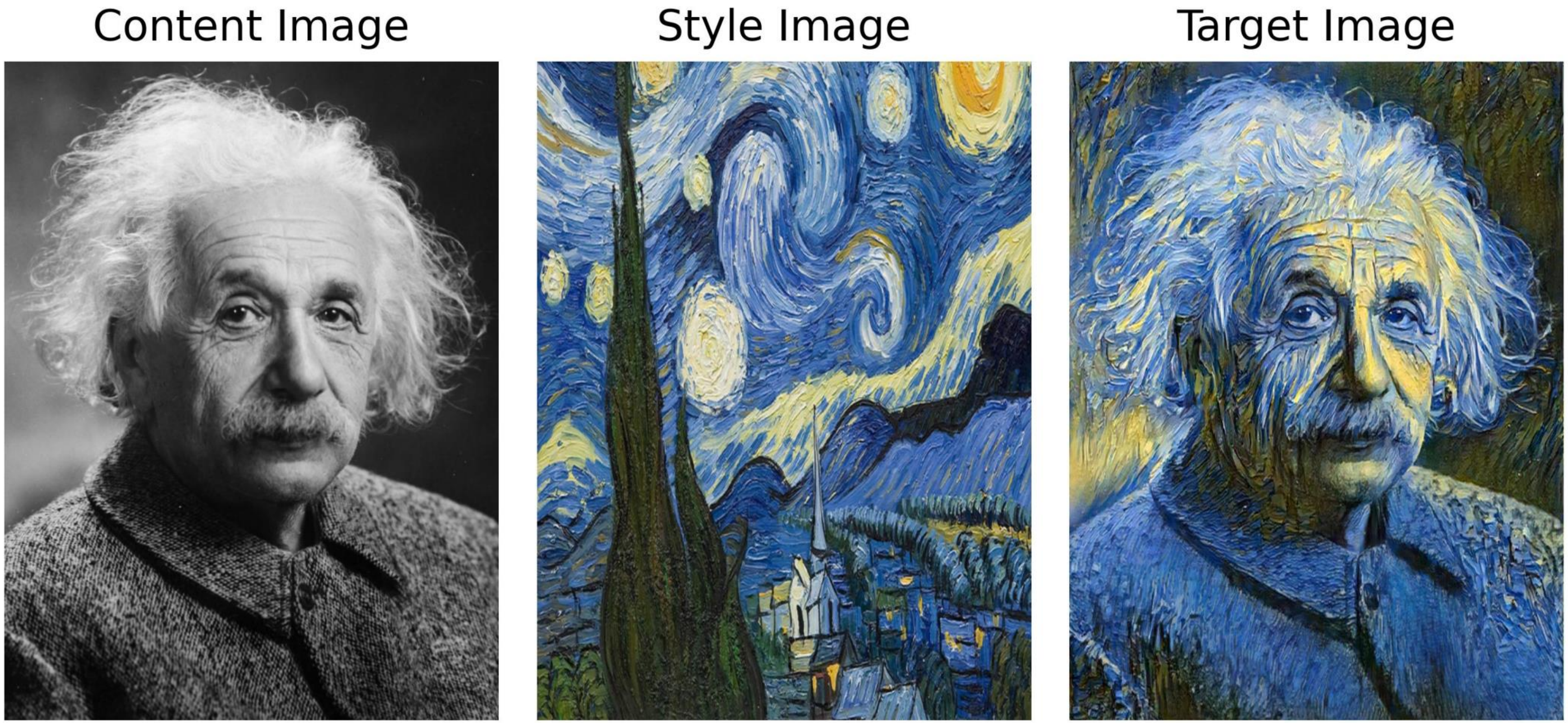


Figure 1: Neural Style Transfer with 'A Neural Algorithm of Artistic Style' algorithm (our implementation)

BEFORE CNN'S

- Large range of powerful **non-parametric algorithms** [2-5].
- Limitation: only low-level content extraction
- Poor separation of content from the artistic style.
- Quality highly dependent on the content image.

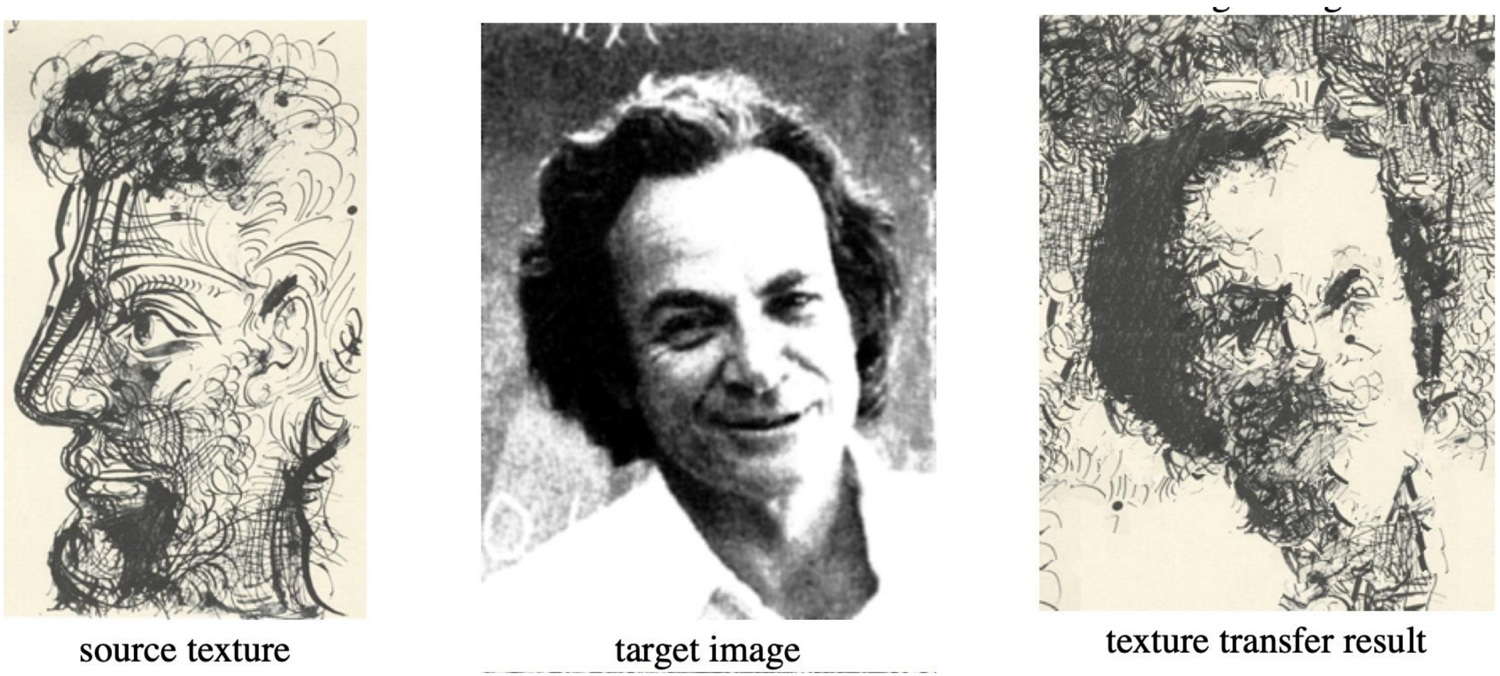


Figure 2: Texture Transfer with Image Quilting [3]

NEURAL STYLE TRANSFER

- DNNs to extract high-level semantic information.
- What type of DNN?
  - ✓ Specifically designed for Vision
  - ✓ Understanding for object recognition and artistic style classification
  - ✓ pre-trained on sufficient labeled data

- **Answer**: Convolutional Neural Networks (CNNs)
  - VGG-19 architecture pre-trained on ImageNet dataset

VGG-19 ARCHITECTURE

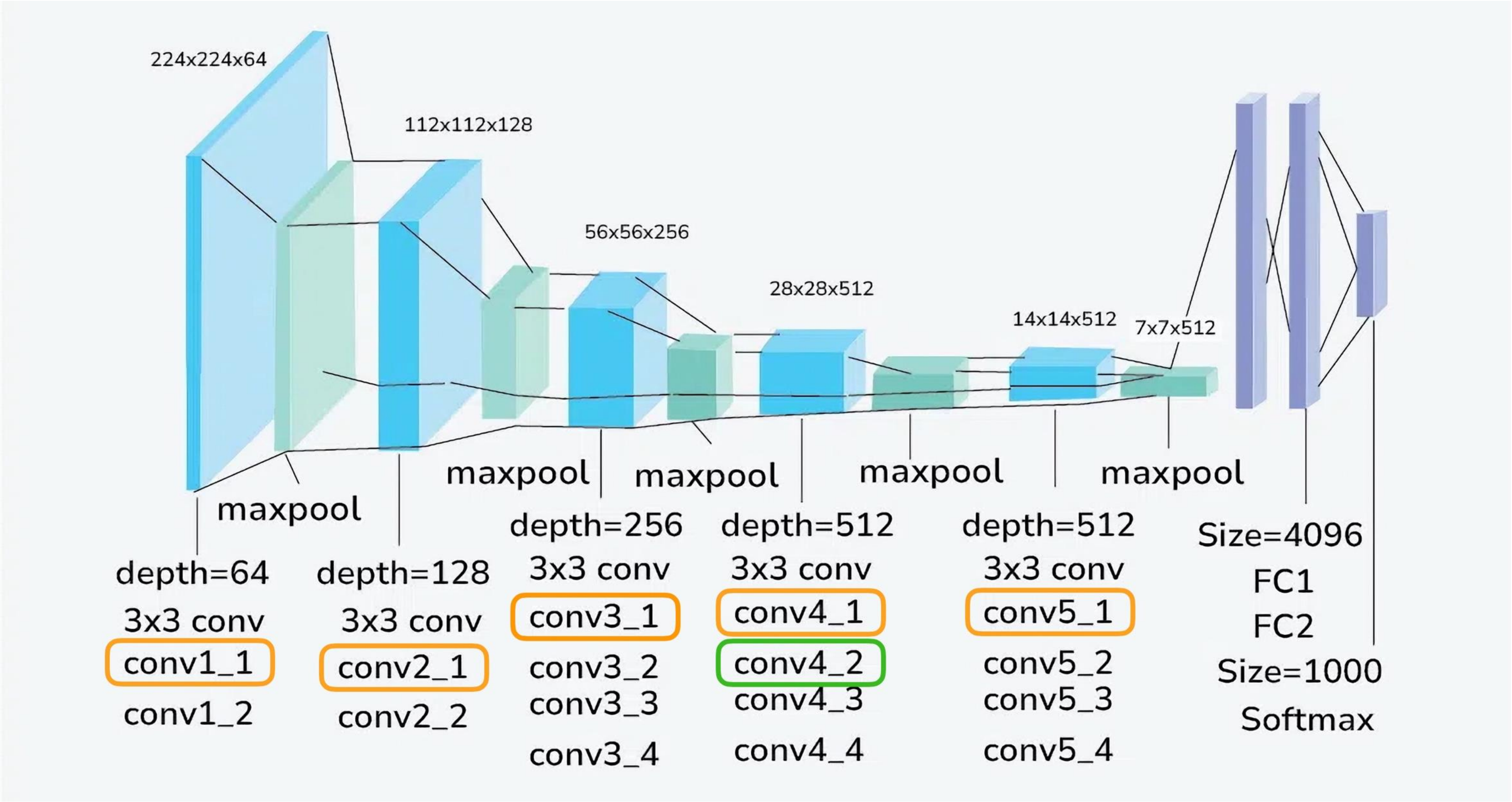


Figure 3: VGG-19 Architecture [6-7]

A NEURAL ALGORITHM OF ARTISTIC STYLE

- proposed by Gatys et al. in their seminal work [1]
- Style Transfer algorithm based on
  - VGG-19 architecture (pre-trained on ImageNet)
  - LBFGS Optimizer
  - Training on the Target Image (rather than on the model)
  - Loss function is the weighted sum of Content Loss $\mathcal{L}_{content}$ and Style Loss $\mathcal{L}_{style}$.

LOSS FUNCTION

Content Loss: $$\mathcal{L}_{content}(\vec{x}, \vec{p}, \vec{v}) = \sum_l v^l (F^l - P^l)^2$$

where l runs through the selected Content Layers [1].

$\vec{x}$: target image
$\vec{p}$: content image

F^l : Content Features of target image
P^l : Content Features of content image

v^l : Content weights

LOSS FUNCTION

Style Loss:

$$\mathcal{L}_{style}(\vec{x}, \vec{a}, \vec{w}) = \sum_{L} w^L (G(F^L) - G(A^L))^2$$

where L runs through the selected Style Layers [1].

$\vec{x}$: target image
$\vec{a}$: style image

F^L : Style Features of target image
A^L : Style Features of style image
$G(X)$: Gram matrix of X

w^L : Style weights

LOSS FUNCTION

Total Loss: $$\mathcal{L}_{total}(\alpha, \beta) = \alpha\,\mathcal{L}_{content} + \beta\,\mathcal{L}_{style}$$

α : Content weight

- A high emphasis on the content produces an image with only little stylization [1].

β : Style weight

- A high emphasis on the style effectively produces a texturized version of the style image [1].

A NEURAL ALGORITHM OF ARTISTIC STYLE

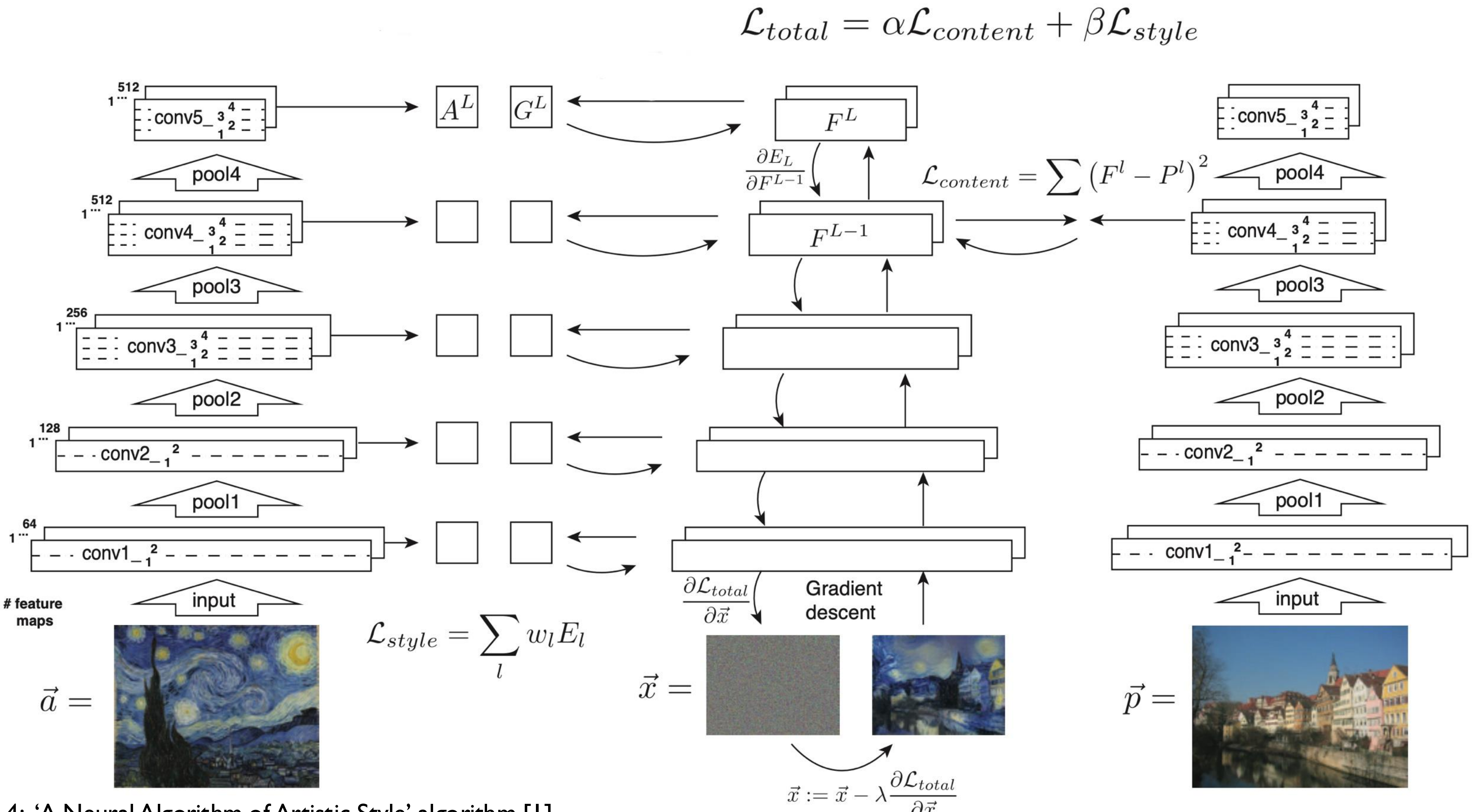


Figure 4: 'A Neural Algorithm of Artistic Style' algorithm [1]

HYPERPARAMETERS

- **Weight ratio (α / β)** : we fixed the content weight α and run an experiment with different style weights β.
  - β = [1e-2, 1e-1, 1e0, 1e1, 1e2, 1e4]

- **Learning rate** : Our experiments suggest an optimal value of 0.1 - 0.3 for LBFGS Optimizer.

- **Epoch** : Our experiment suggests an optimal value of 200 - 300 for LBFGS based on the the images.

HYPERPARAMETERS

- **Selected Content layers** : Gatys et al. paper proposes 'conv4_2’ [1].

- **Content weights** : Gatys et al. paper proposes 1 [1].

- **Selected Style layers** : Gatys et al. paper proposes 'conv1_1', 'conv2_1', 'conv3_1', 'conv4_1', 'conv5_1’ [1].

- **Style weights** : Gatys et al. paper proposes [1e3/n**2 for n in [64, 128, 256, 512, 512]] [1].

RATIO EXPERIMENT

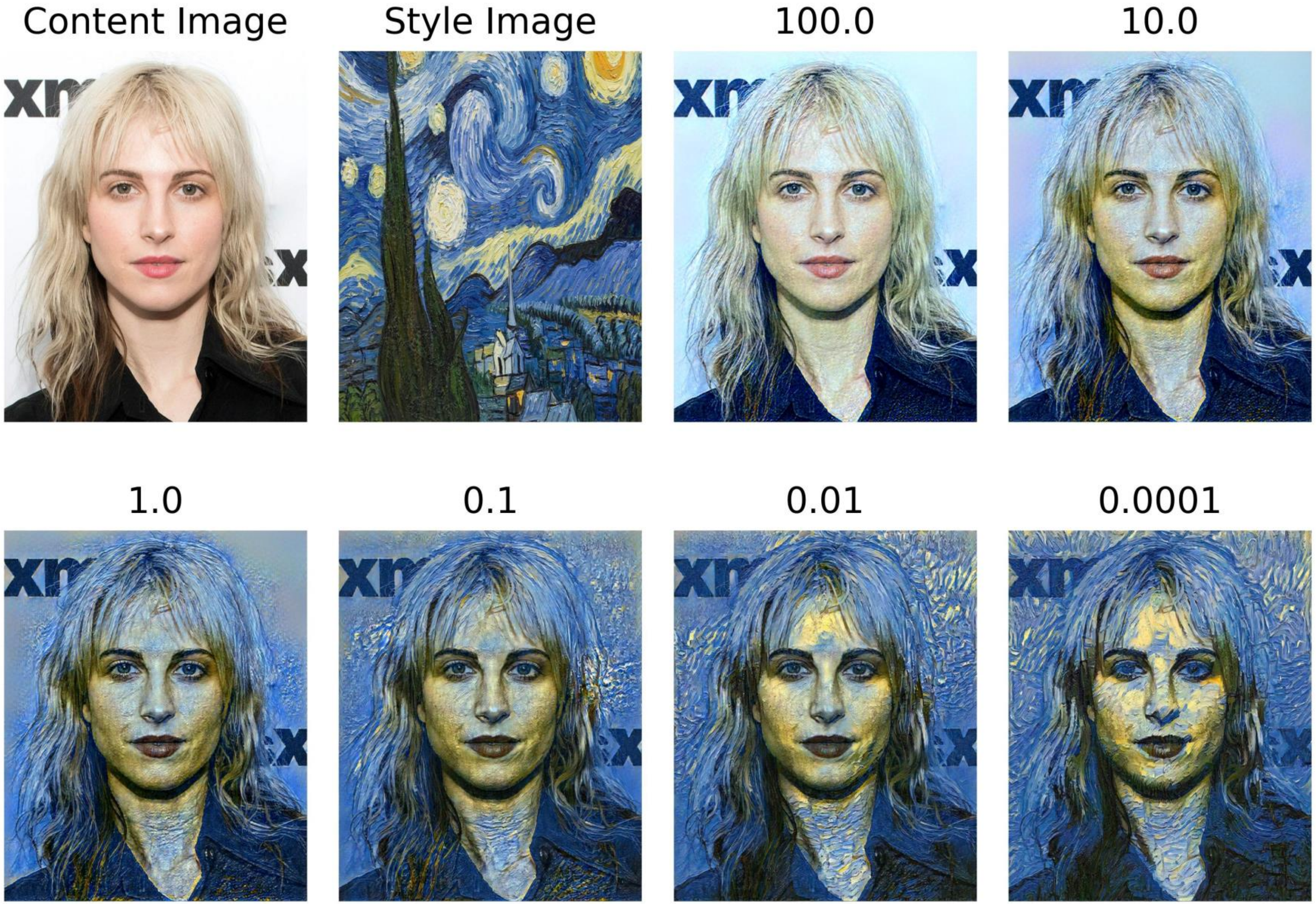


Figure 5: α / β weight ratio test with our implementation

EXAMPLE RUNS

Figure 6: Albert Einstein and Gallery of Cubism by Pablo Picasso [8] (LBFGS epoch = 200, learning rate = 0.2, style weight = 100.0)

EXAMPLE RUNS

Figure 7: Oak tree and The Starry Night by Vincent van Gogh (LBFGS epoch = 200, learning rate = 0.2, style weight = 100.0)

EXAMPLE RUNS

Figure 8: Mona Lisa by Leonardo da Vinci and Gallery of Cubism by Pablo Picasso [8] (LBFGS epoch = 200, learning rate = 0.2, style weight = 100.0)

EXAMPLE RUNS

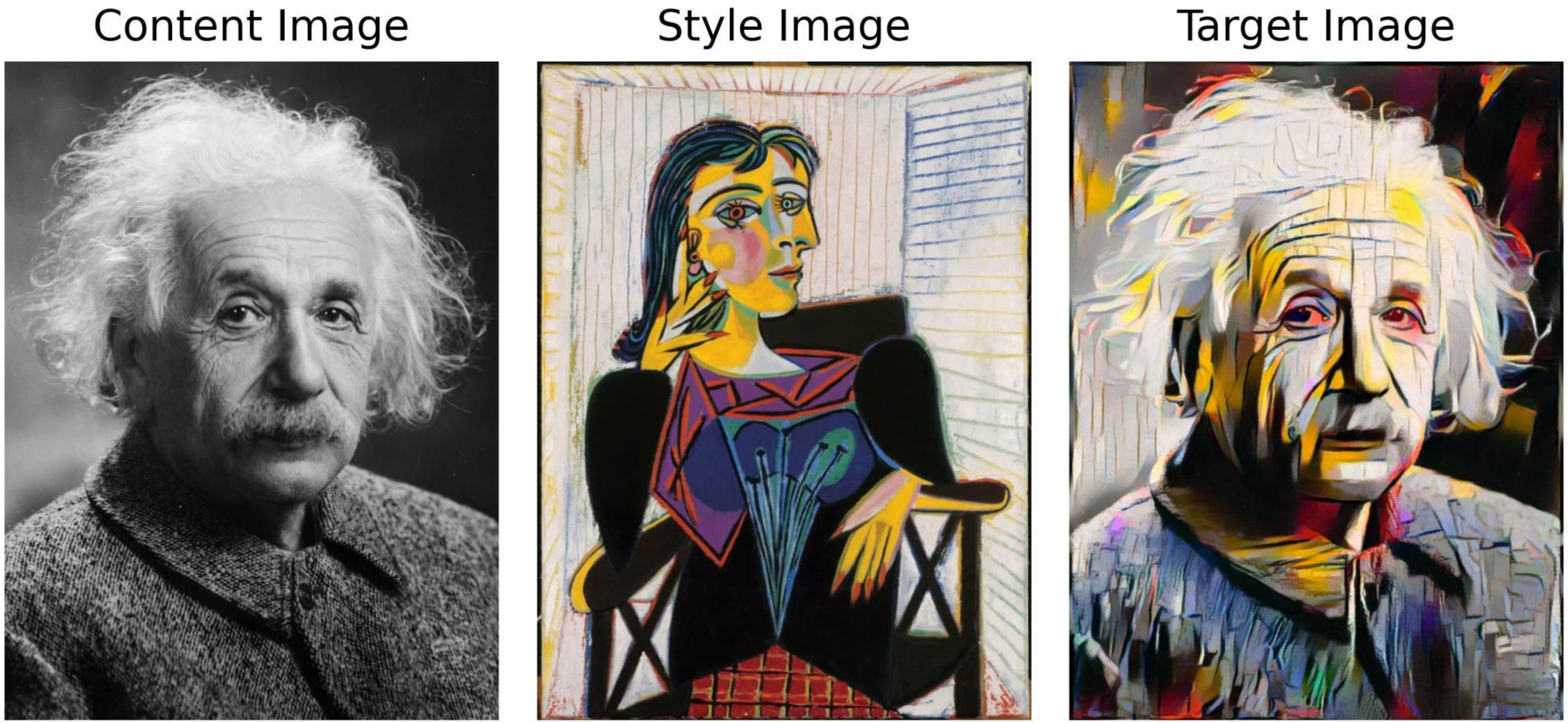
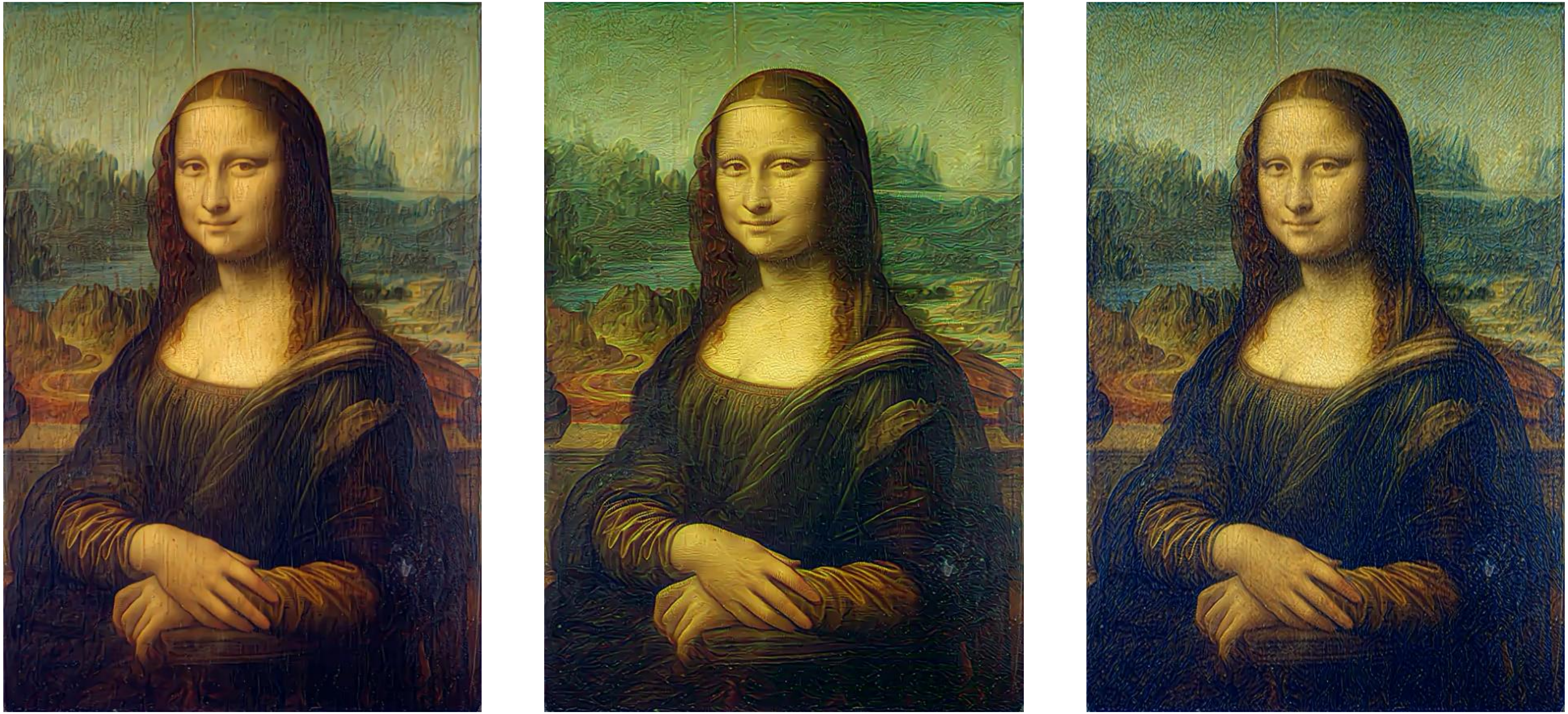


Figure 9: Albert Einstein and Portrait of Dora Maar by Pablo Picasso [9] (LBFGS epoch = 200, learning rate = 0.2, style weight = 100.0)

EXAMPLE RUNS



Video 10: Mona Lisa training videos (LBFGS epoch = 200, learning rate = 0.2, style weight = 100.0)

RESULTS

- **Key Finding**: Representations of content and style in the CNNs are well separable [1].

- **Trade-off** exists between content and style matching.
  - requires fine adjustment to weight ratio

- Results may be highly **subjective**.
  - human inspection as the sole evaluation criteria

- Fascinating to see CNNs, designed for object classification, can understand the **semantic content independently from artistic style**.

REFERENCES

[1] L. A. Gatys, A. S. Ecker and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2414-2423, doi: 10.1109/CVPR.2016.265.

[2] A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, volume 2, pages 1033–1038. IEEE, 1999.

[3] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pages 341–346. ACM, 2001.

[4] V. Kwatra, A. Sch¨ odl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: image and video synthesis using graph cuts. In ACM Transactions on Graphics (ToG), volume 22, pages 277–286. ACM, 2003.

[5] L. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pages 479–488. ACM Press/Addison-Wesley Publishing Co., 2000.

REFERENCES

[6] Zheng, Yufeng & Yang, Clifford & Merkulov, Aleksey. (2018). Breast cancer screening using convolutional neural network and follow-up digital mammography. 4. 10.1117/12.2304564.

[7] Simonyan, Karen and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” CoRR abs/1409.1556 (2014).

[8] “Gallery of Cubism by Pablo Picasso.” LatAm ARTE, www.latamarte.com/en/galleries/3kKJ/. Accessed 12 Mar. 2025.

[9] Wikipedia contributors. "Portrait of Dora Maar." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 19 Feb. 2024. Web. 12 Mar. 2025.

THANKS FOR LISTENING