Synthesis: Style Transfer

Danna Gurari
University of Colorado Boulder
Fall 2021

https://home.cs.colorado.edu/~DrG/Courses/RecentAdvancesInComputerVision/AboutCourse.html

Review

- Last lecture topic:
  - Computer vision with self-supervised learning
- Assignments (Canvas)
  - Final project outline due Wednesday
- Questions?

Style Transfer: Today's Topics

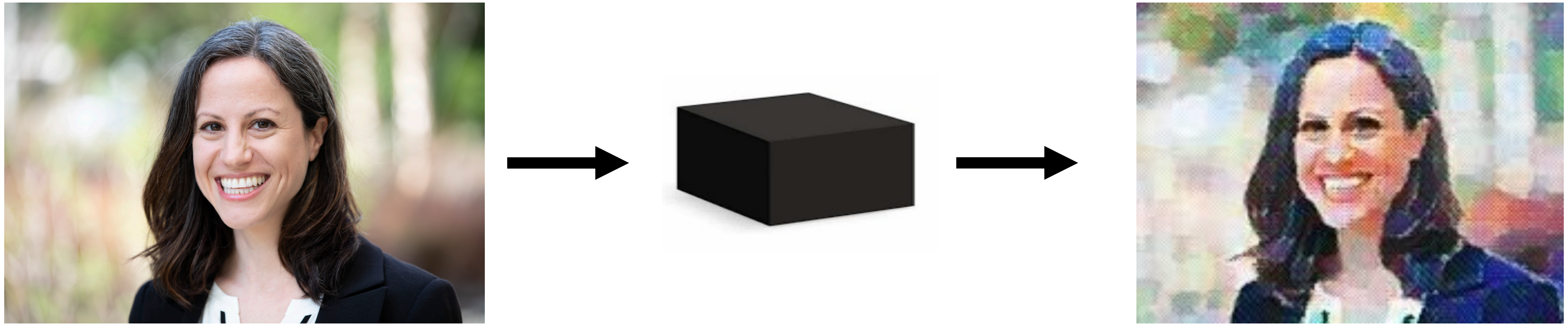
- Problem
- Applications
- Computer vision models
- Evaluation metrics

Style Transfer: Today’s Topics

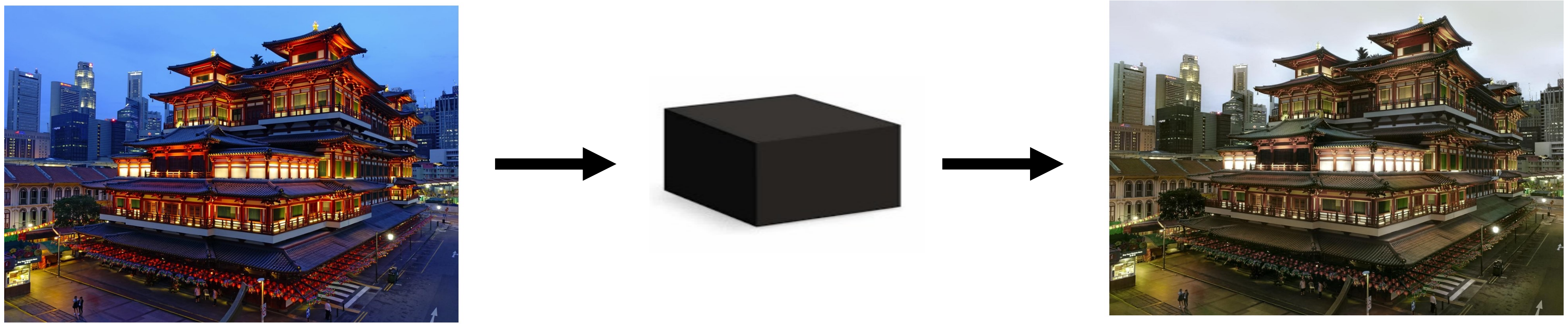
- Problem
- Applications
- Computer vision models
- Evaluation metrics

An Image Transformation Problem:
Transform **Content** of Image into a New **Style**

Artistic:



Photorealistic:



An Image Transformation Problem: Transform **Content** of Image into a New **Style**

How would you define “content”?

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

An Image Transformation Problem: Transform **Content** of Image into a New **Style**

How would you define “style”?

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Style Transfer: Today's Topics

- Problem
- Applications
- Computer vision models
- Evaluation metrics

Entertainment (Mobile Phone Applications)

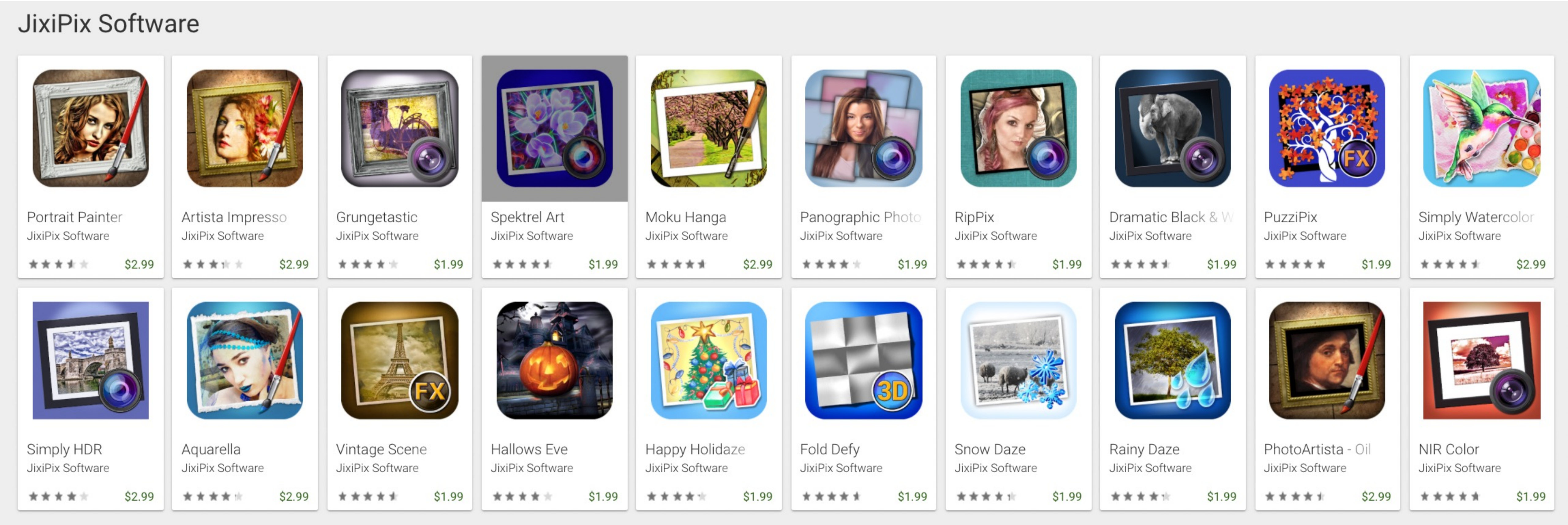
Browser demo: https://reiinakano.com/arbitrary-image-stylization-tfjs/

Entertainment (Mobile Phone Applications)

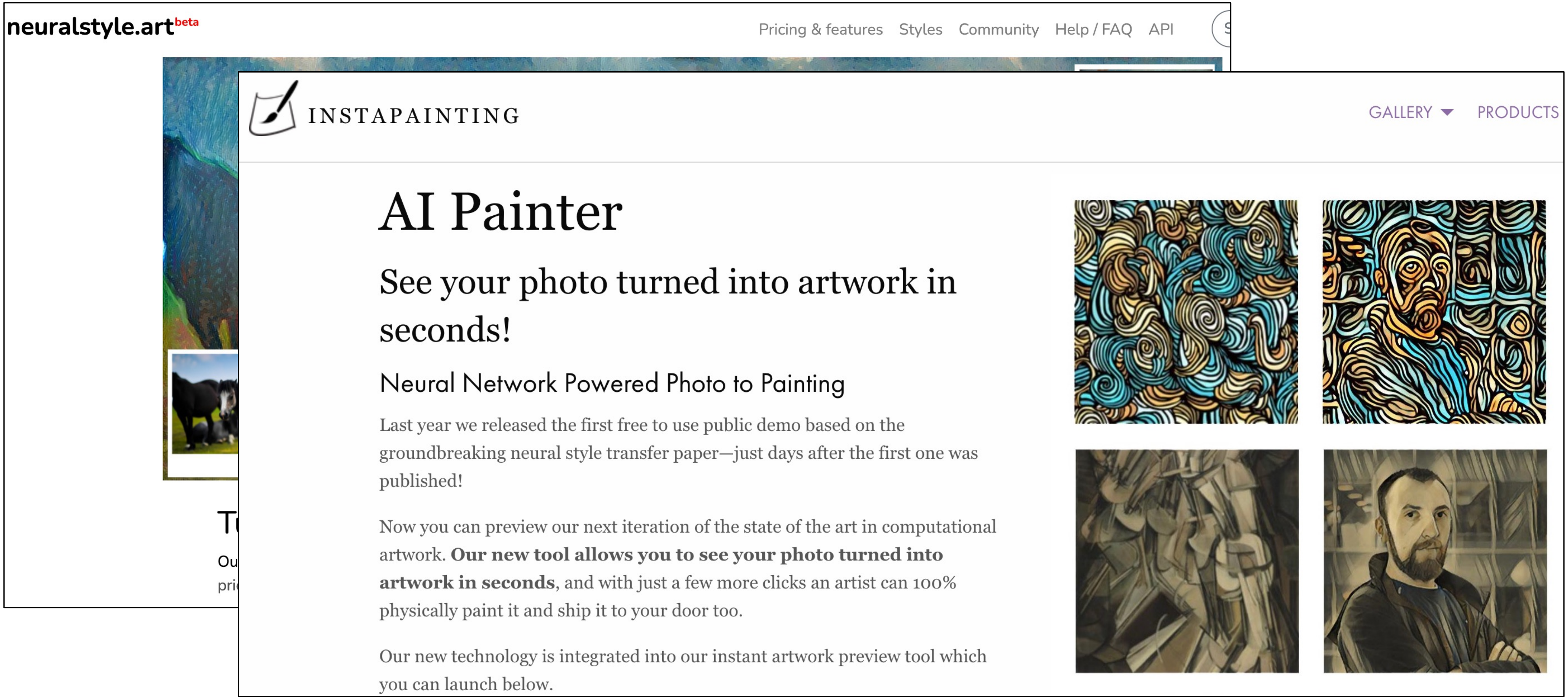
ARTIFY

PicsArt

Entertainment (Mobile Phone Applications)



Commercial Art



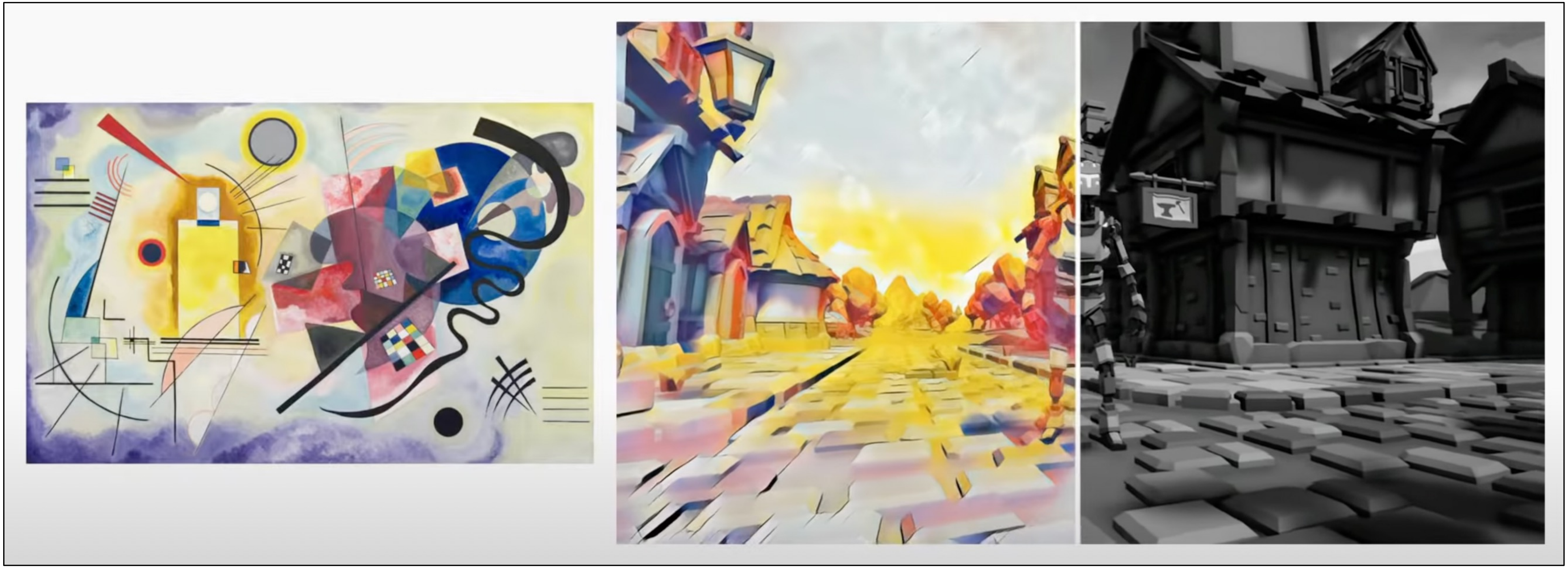
Virtual and Augmented Reality

Demo: https://youtu.be/Rz4J3T1uYYo

Virtual and Augmented Reality

Demo: https://youtu.be/Rz4J3T1uYYo

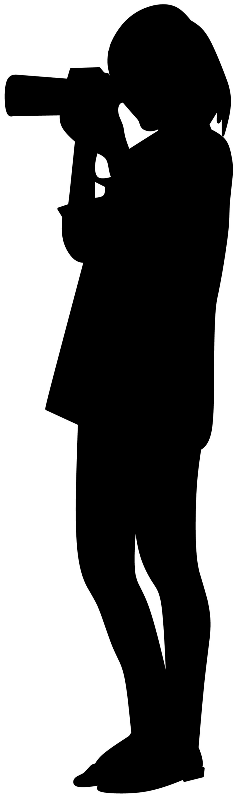
Gaming (e.g., Stadia from Google)

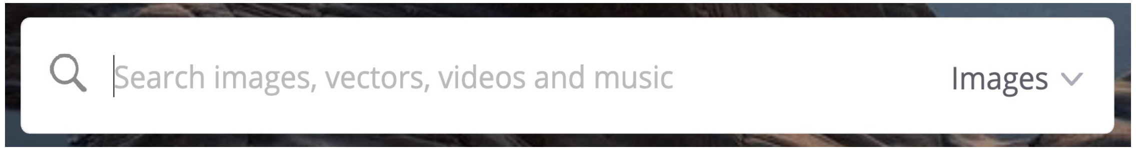


Demo: https://www.youtube.com/watch?v=yF1bZiH-wJQ

Improved Messaging via Visual Content

- Marketing
- Artwork
- Presentations
- Blogs
- Websites





Potential sources:

Photographer
(self or hired)

Stock photos

For what other applications might style transfer be useful?

Are there any applications that you can imagine that would be unethical uses of style transfer?

Style Transfer: Today's Topics

- Problem
- Applications
- Computer vision models
- Evaluation metrics

Key Challenges

- How to computationally isolate the content of the content image?
- How to computationally isolate the style of the style image?
- How to blend the content and style?
- How to support any arbitrary style?
- How to do all these fast?

Neural Style Transfer (NST): Addresses…

- How to computationally isolate the content of the content image?
- How to computationally isolate the style of the style image?
- How to blend the content and style?
- How to support any arbitrary style?
- How to do all these fast?

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Neural Style Transfer (NST): Key Insight

"The representations of content and style in the Convolutional Neural Network are well separable."

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Neural Style Transfer (NST)

Approach: iteratively modify a random image guided by the content image and style image

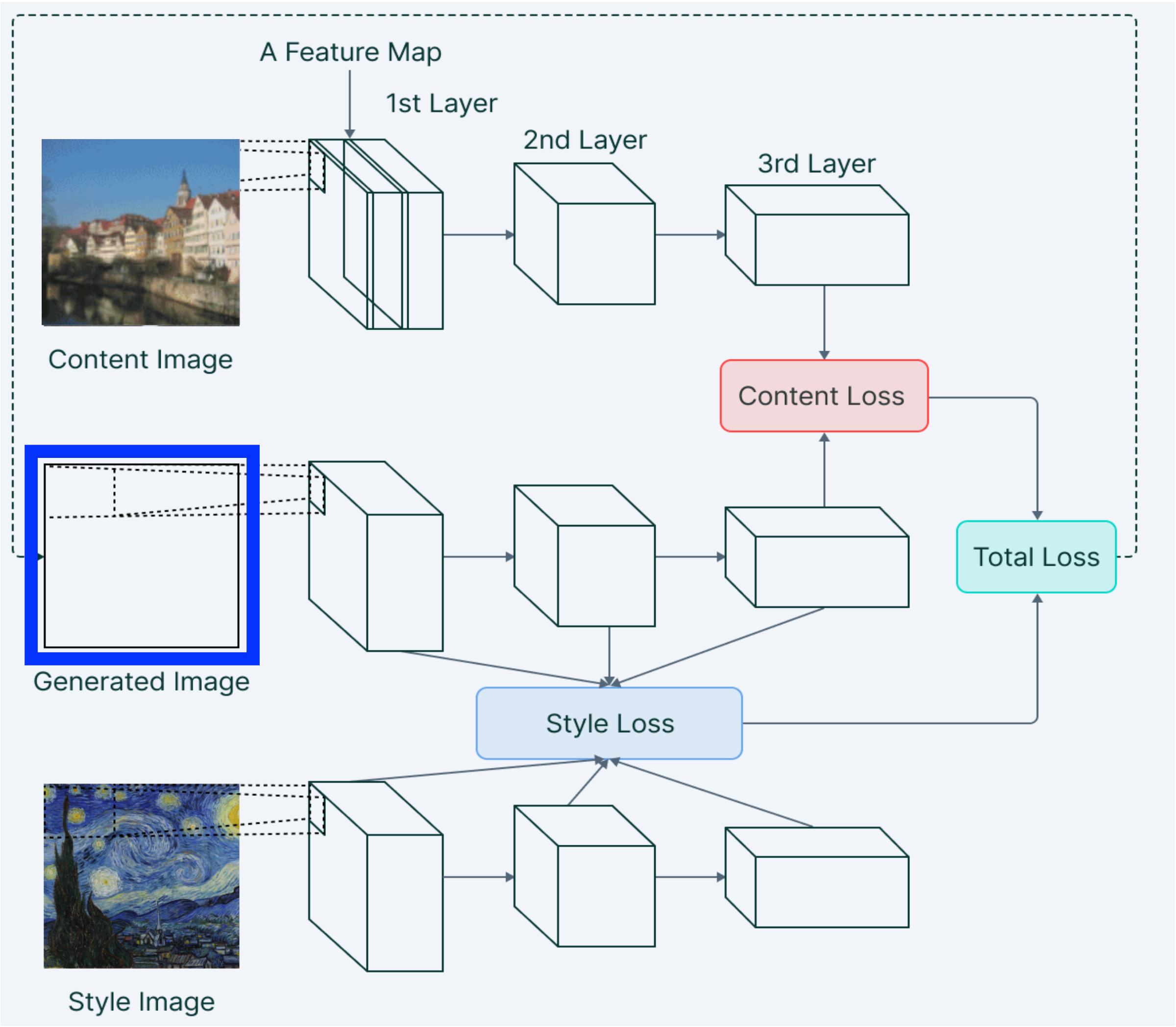


Figure Sources: https://ndres.me/images/style-transfer.gif; https://www.v7labs.com/blog/neural-style-transfer

Neural Style Transfer (NST)

Approach: iteratively modify a random image guided by the content image and style image

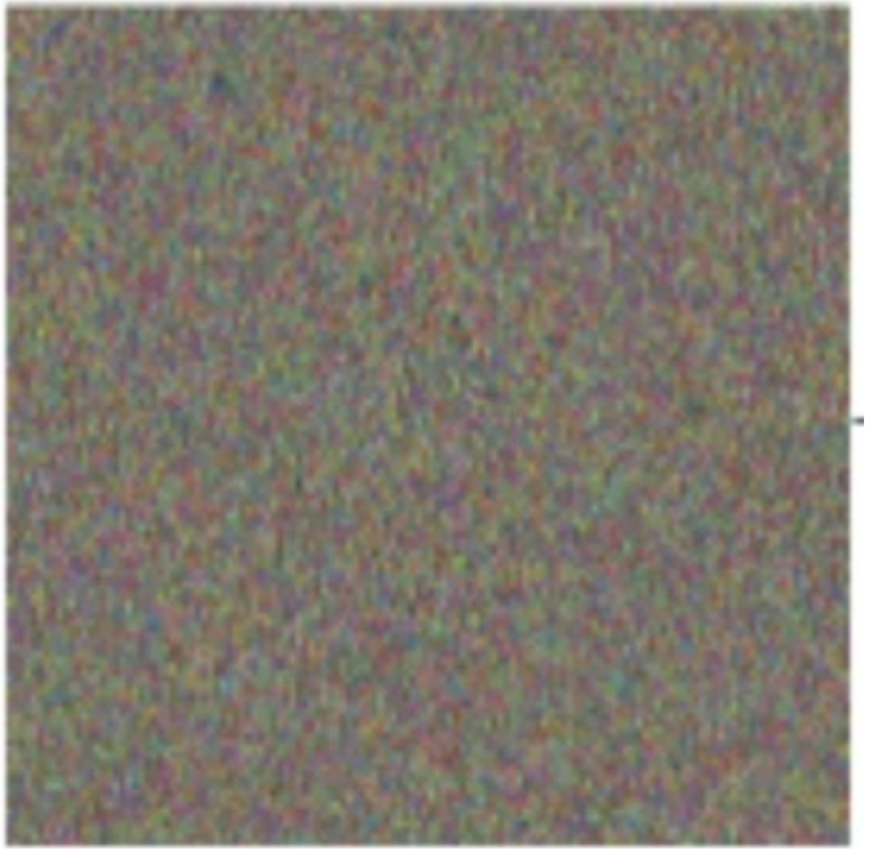


Figure Source: https://towardsdatascience.com/a-brief-introduction-to-neural-style-transfer-d05d0403901d

Neural Style Transfer (NST)

Approach: iteratively modify a random image guided by the content image and style image

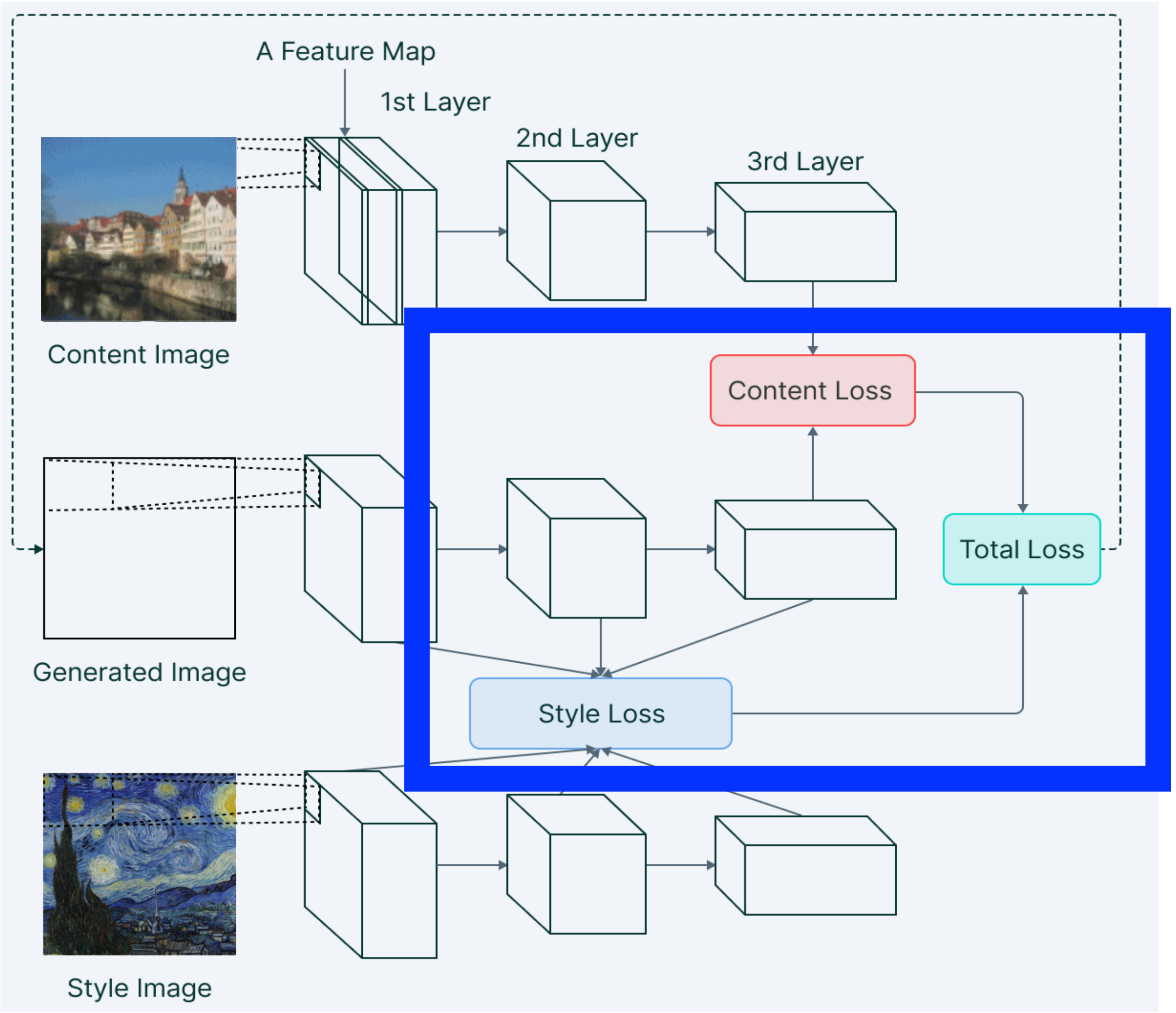


Figure Sources: https://ndres.me/images/style-transfer.gif; https://www.v7labs.com/blog/neural-style-transfer

Neural Style Transfer (NST)

Approach: iteratively modify a random image guided by the content image and style image

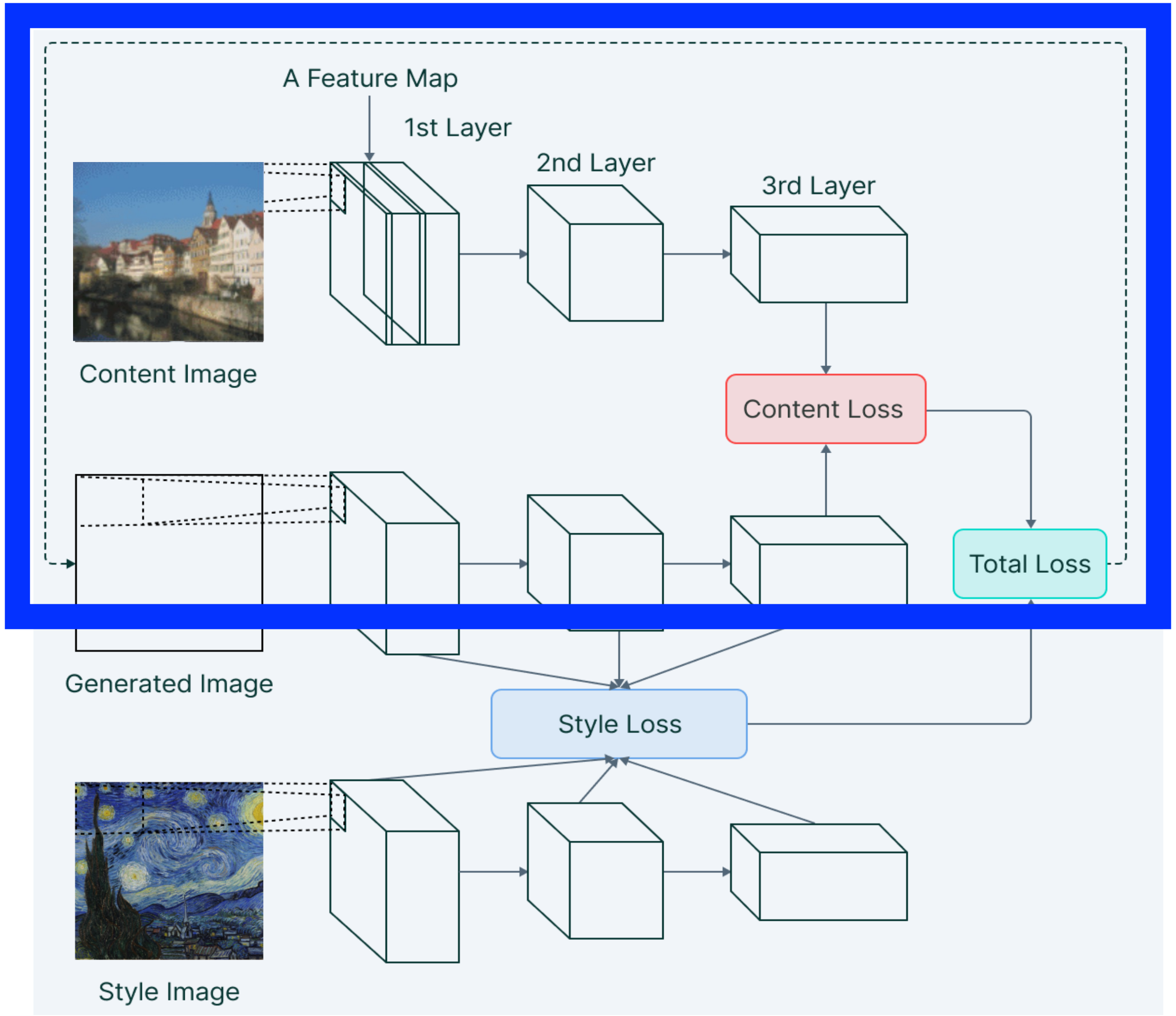


Figure Sources: https://ndres.me/images/style-transfer.gif; https://www.v7labs.com/blog/neural-style-transfer

Neural Style Transfer (NST)

- How to computationally isolate the content of an image?
  - Recall, what CNNs typically learn:

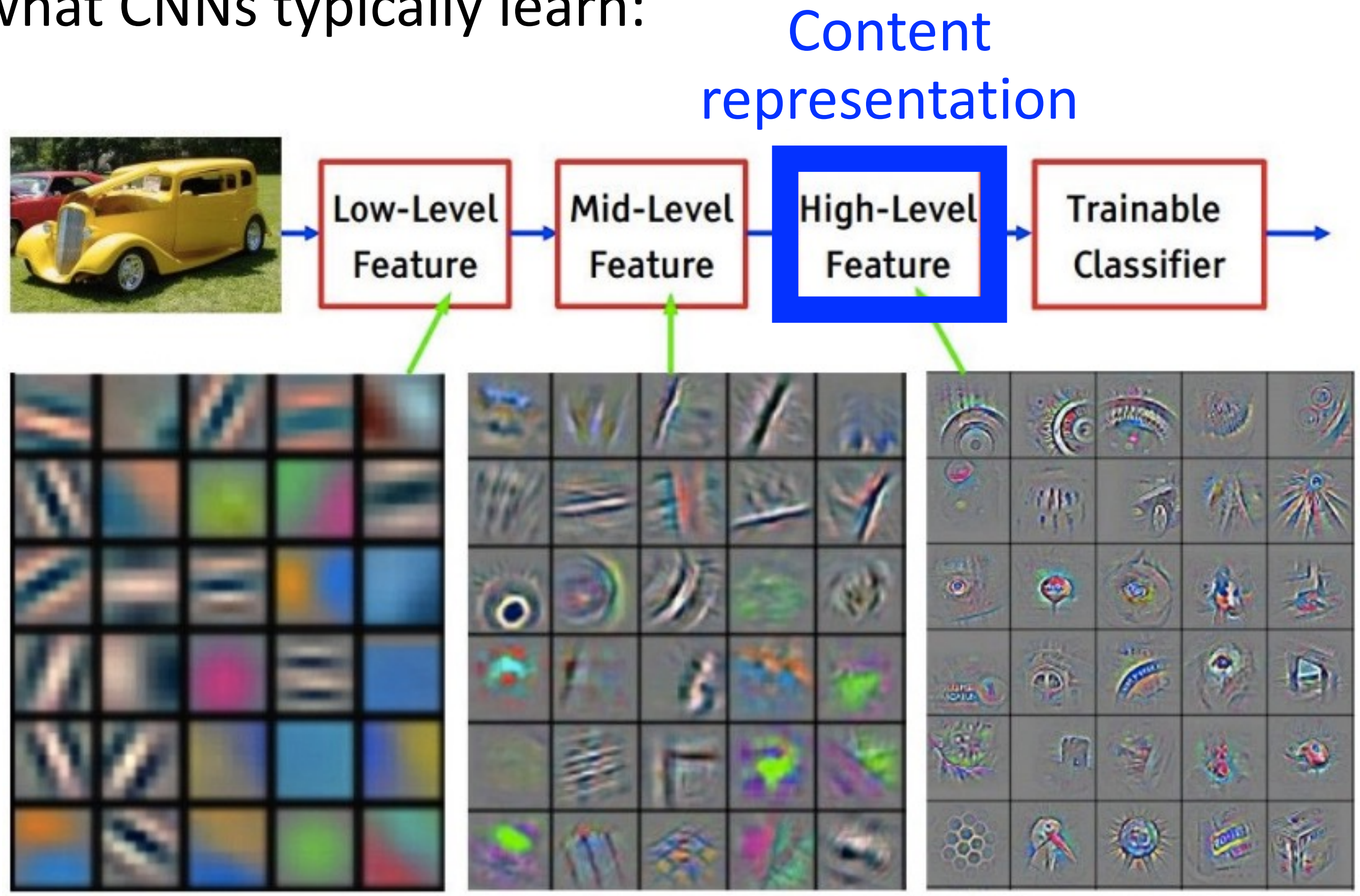


Figure Credit: Yann LeCun

Neural Style Transfer (NST)

Iteratively adjust the generated image until its high level features match the high level features of the content image

Layer of the network

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

Index of feature map generated by the *i*-th filter in layer *l* of the network

Position to look at in the feature map

Approach: iteratively modify a random image guided by the content image and style image

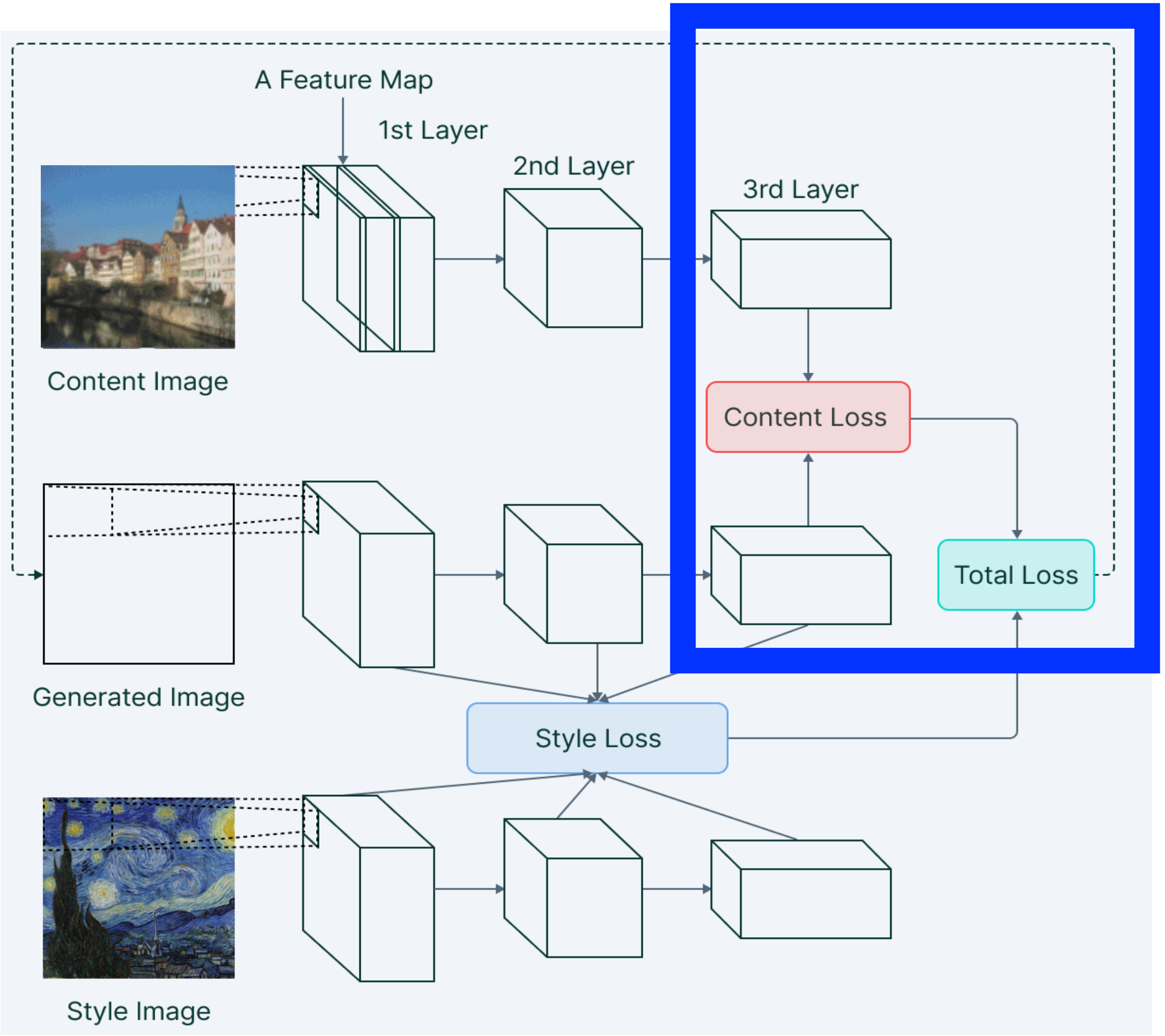


Figure Sources: https://ndres.me/images/style-transfer.gif; https://www.v7labs.com/blog/neural-style-transfer

Neural Style Transfer (NST)

Iteratively adjust the generated image until its feature correlations match the feature correlations of the style image

Approach: iteratively modify a random image guided by the content image and style image

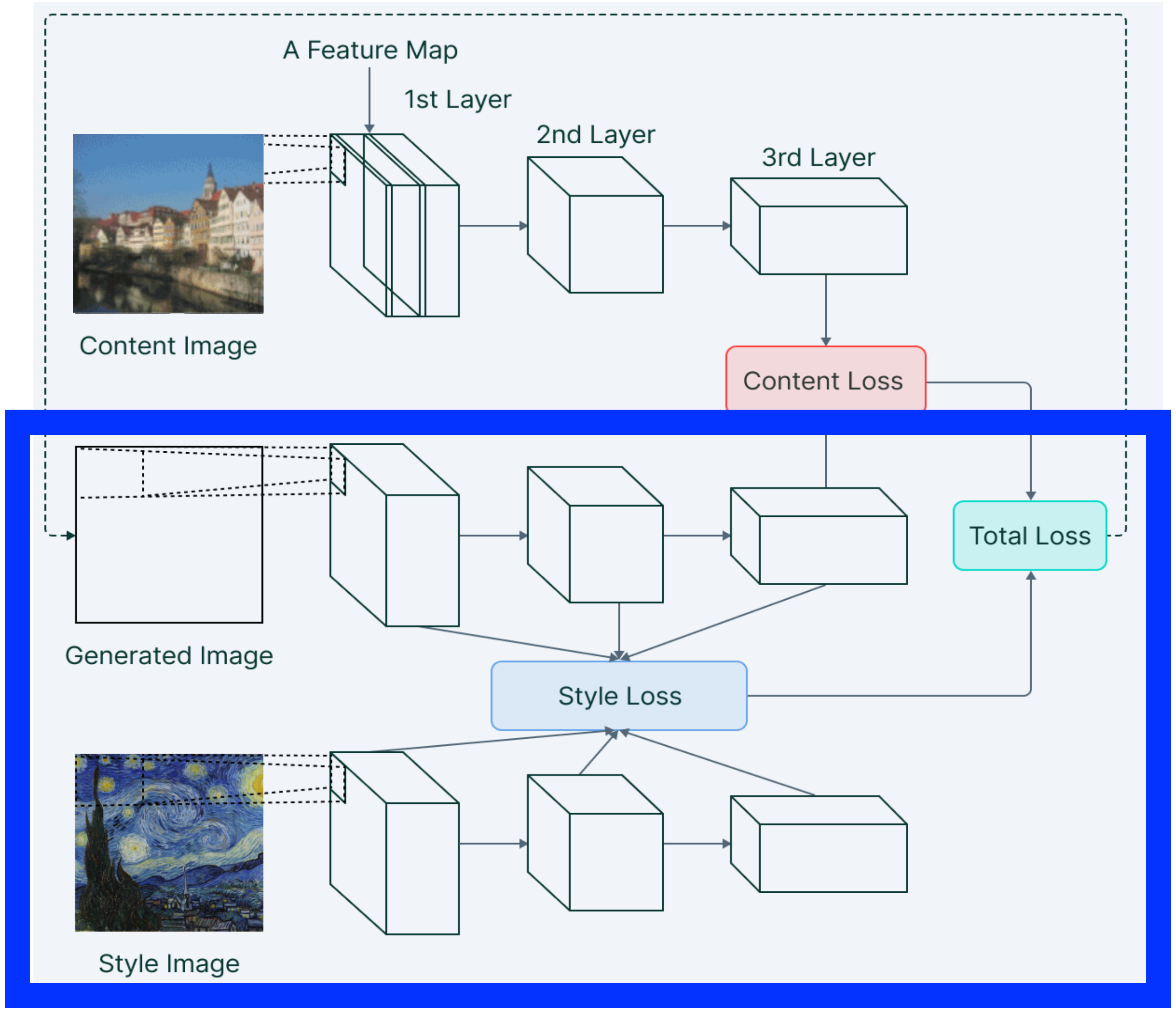


Figure Sources: https://ndres.me/images/style-transfer.gif; https://www.v7labs.com/blog/neural-style-transfer

Neural Style Transfer (NST)

- How to computationally isolate the style of an image?
  - Recall, what CNNs typically learn:

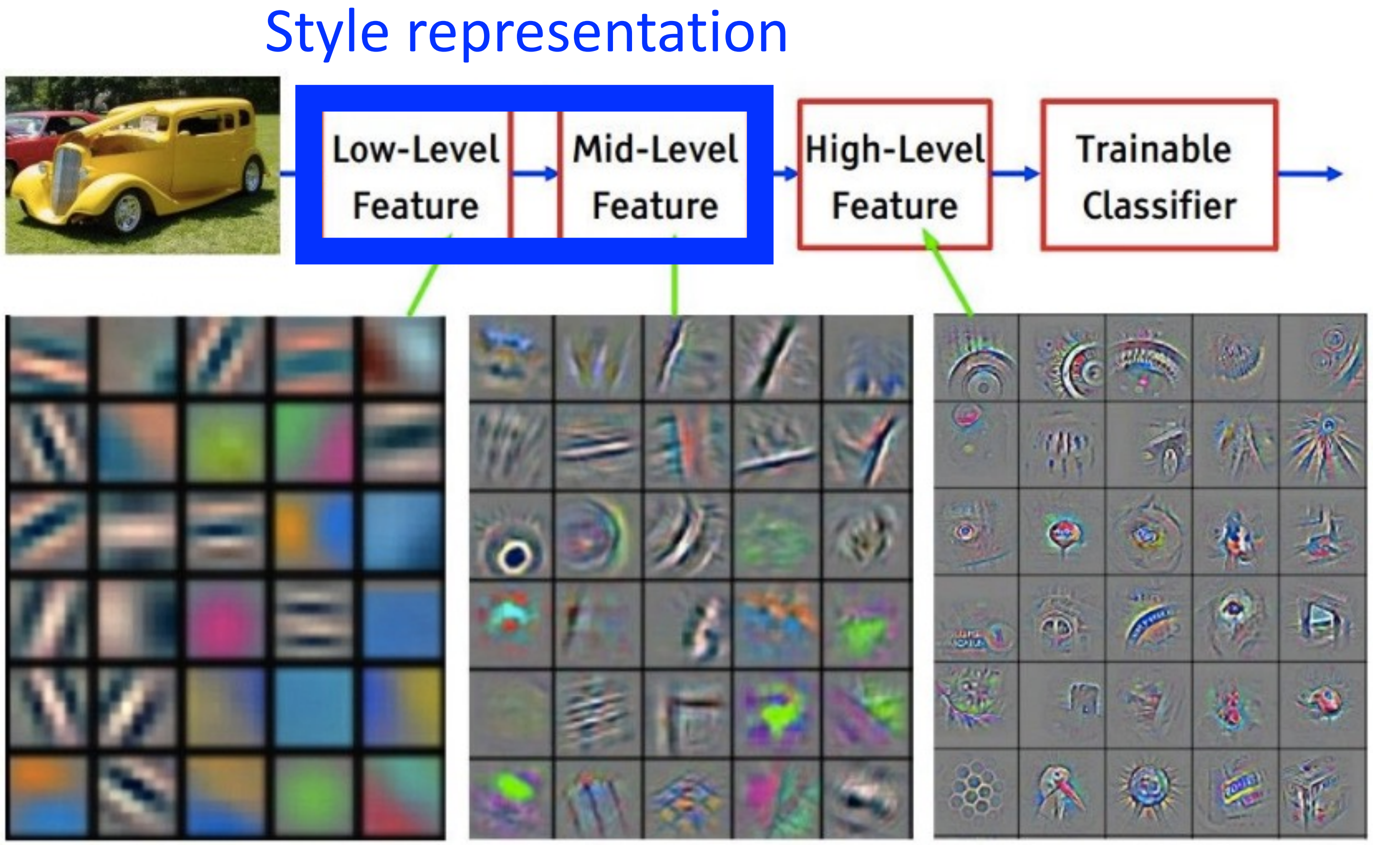


Figure Credit: Yann LeCun

Neural Style Transfer (NST)

Iteratively adjust the generated image until its feature correlations match the feature correlations of the style image

Gram matrix measures correlation of feature maps *i* and *j* (for layer *l*)

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

Approach: iteratively modify a random image guided by the content image and style image

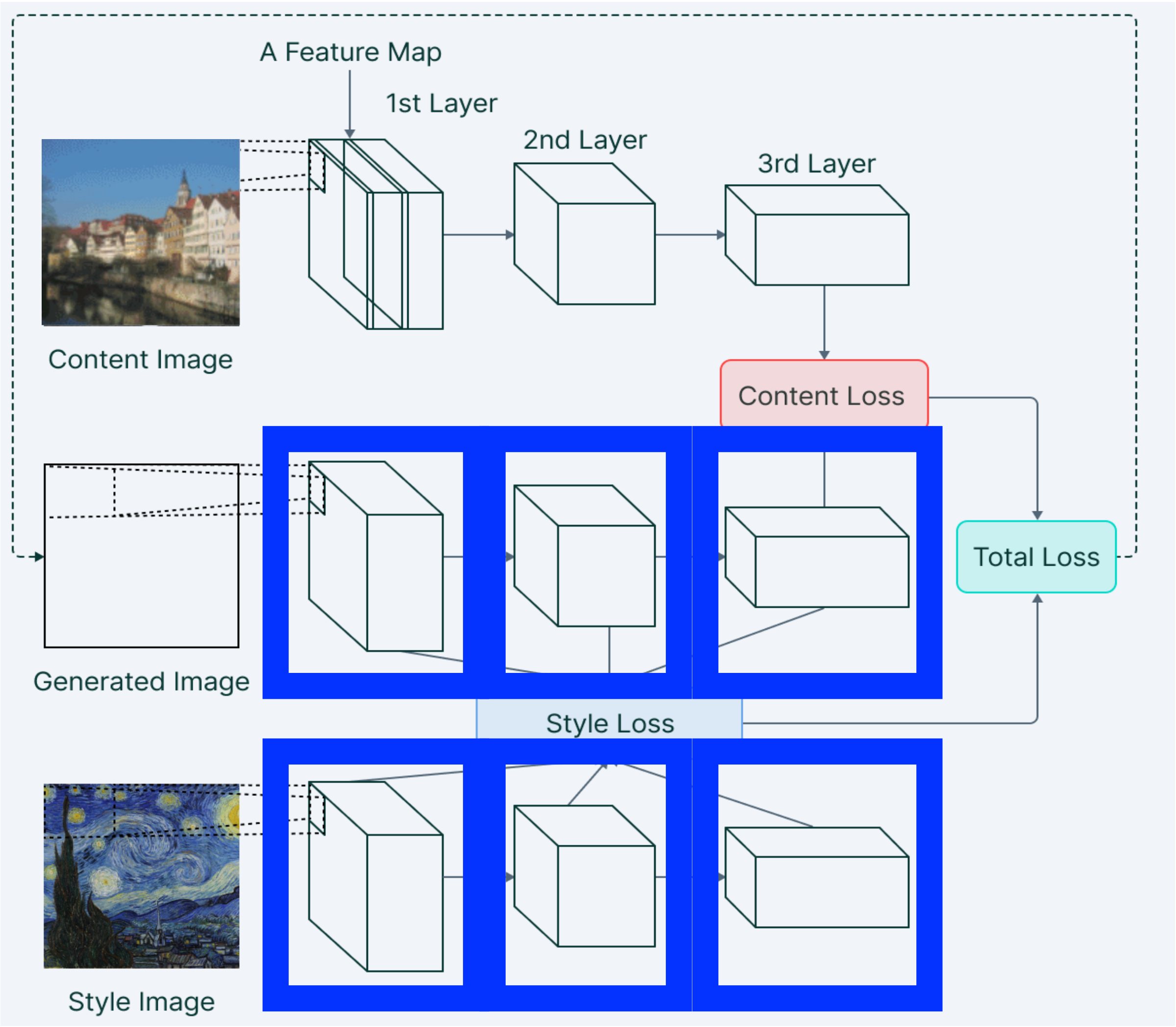


Figure Sources: https://ndres.me/images/style-transfer.gif; https://www.v7labs.com/blog/neural-style-transfer

Neural Style Transfer (NST)

Iteratively adjust the generated image until its feature correlations match the feature correlations of the style image

Loss for 1 layer:

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

Layer of the network

Number of feature maps x each map's size in layer *l*

Index of feature map from *i*-th filter

Position in feature map

Approach: iteratively modify a random image guided by the content image and style image

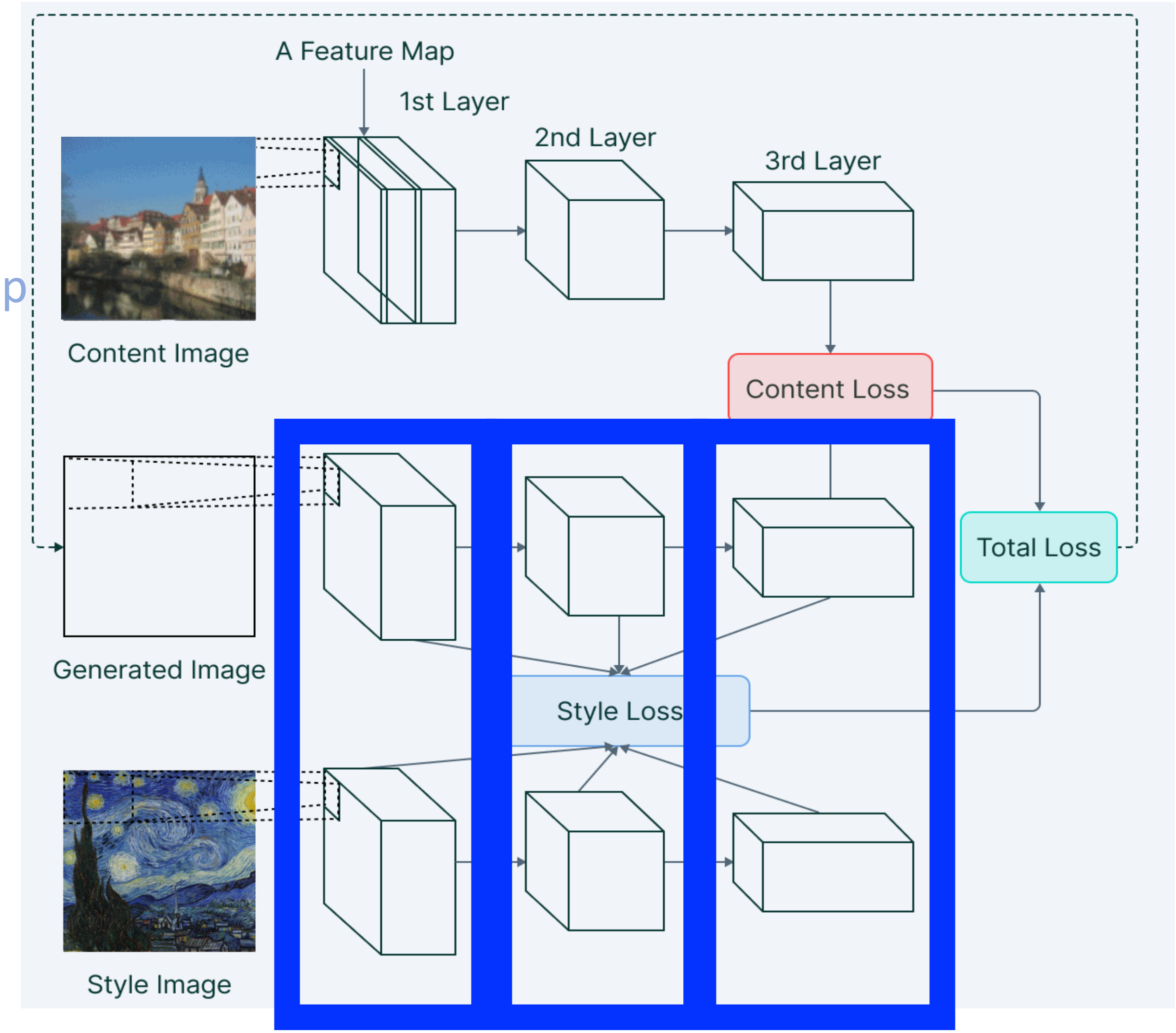
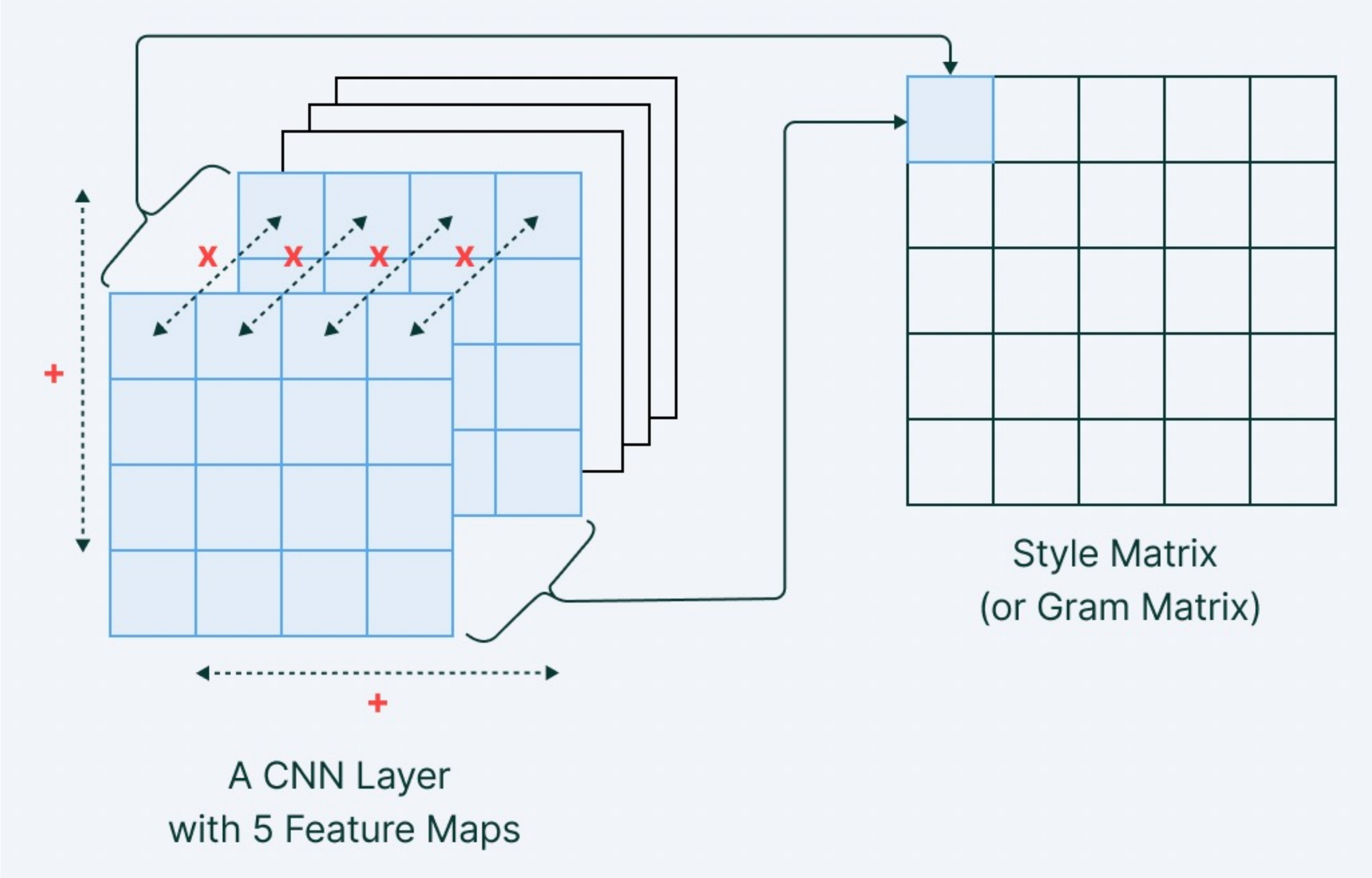


Figure Sources: https://ndres.me/images/style-transfer.gif; https://www.v7labs.com/blog/neural-style-transfer

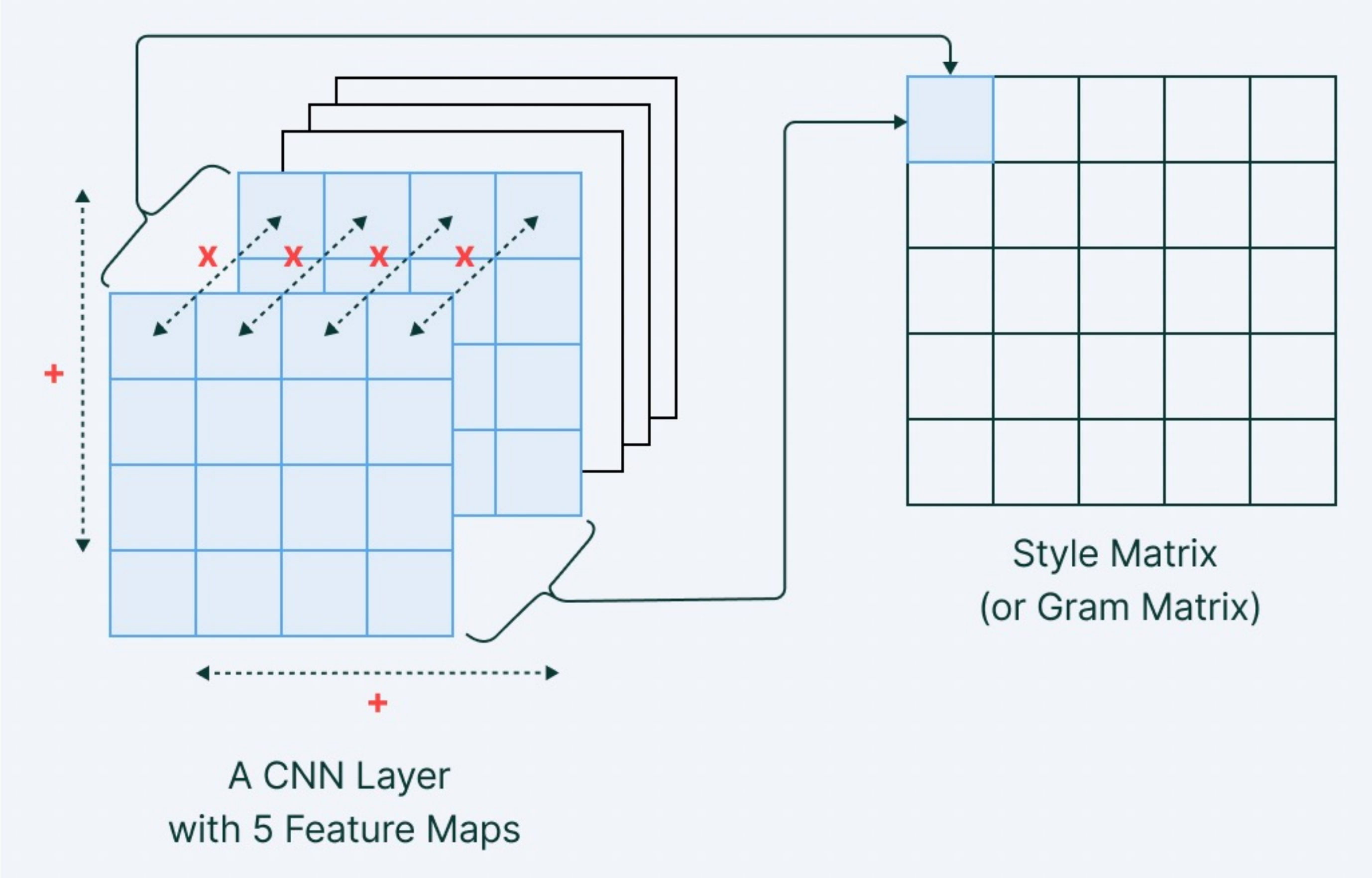
Neural Style Transfer (NST): Gram Matrix



Shows distribution of features in a layer by computing amount of correlation between features maps in that layer.

Figure Source: https://www.v7labs.com/blog/neural-style-transfer

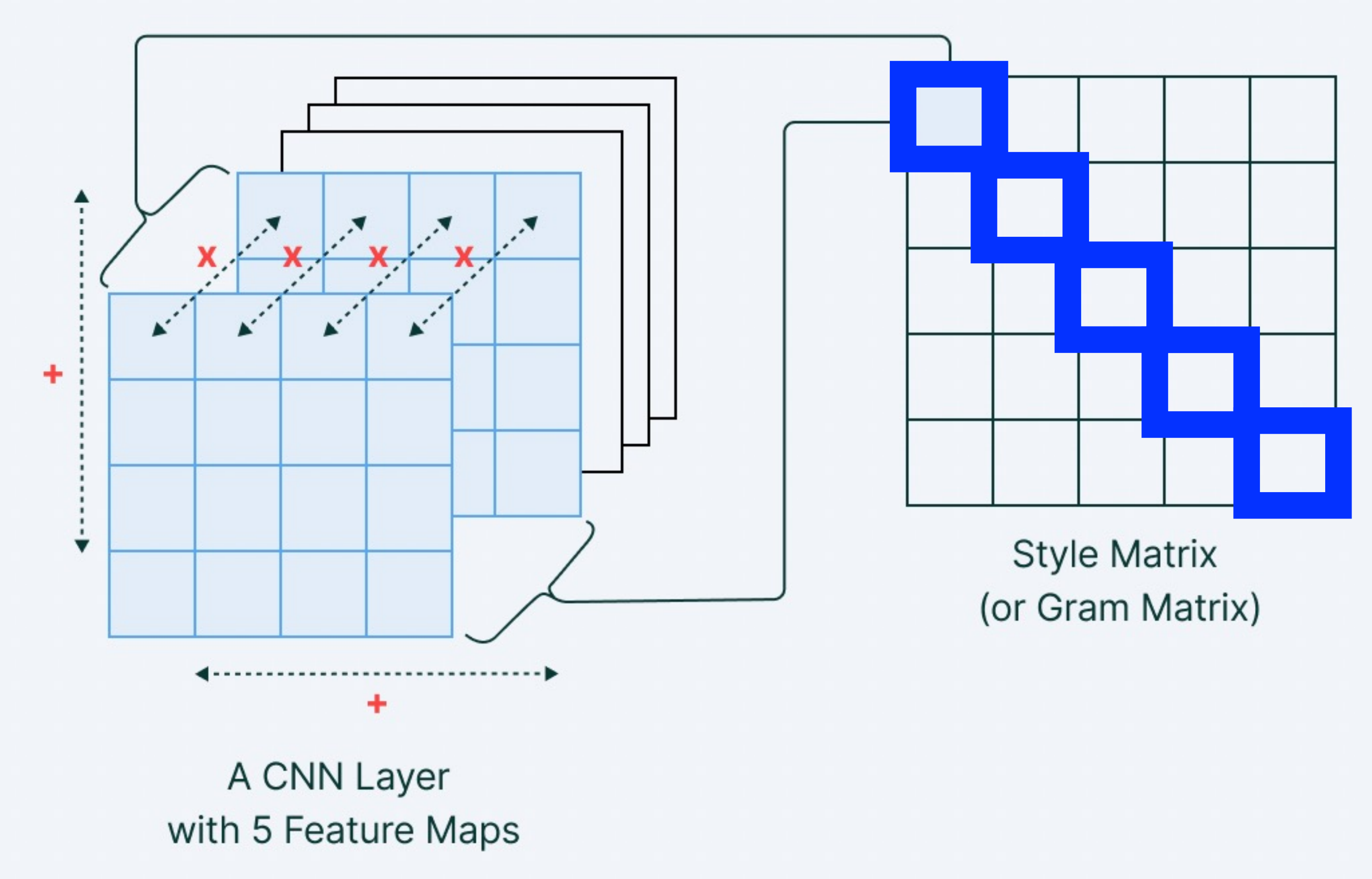
Neural Style Transfer (NST): Gram Matrix



We know we start with *N* feature maps each containing *M* values. What will be the dimension of the Gram matrix?

Figure Source: https://www.v7labs.com/blog/neural-style-transfer

Neural Style Transfer (NST): Gram Matrix



What should be the values on the diagonal of the Gram matrix?

Figure Source: https://www.v7labs.com/blog/neural-style-transfer

Neural Style Transfer (NST)

Iteratively adjust the generated image until its feature correlations match the feature correlations of the style image

Total loss is the weighted sum of correlation differences across all layers

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

Approach: iteratively modify a random image guided by the content image and style image

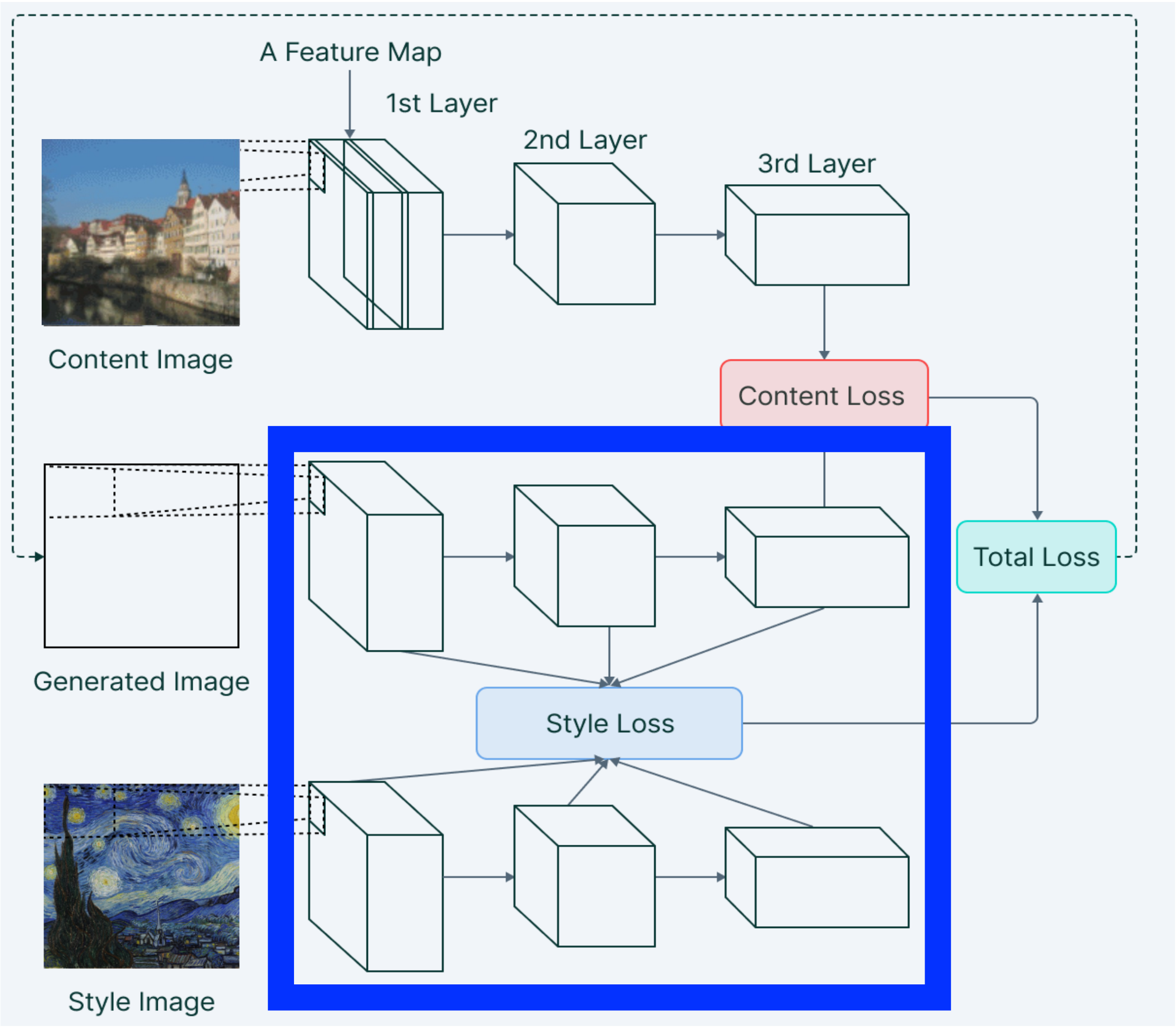
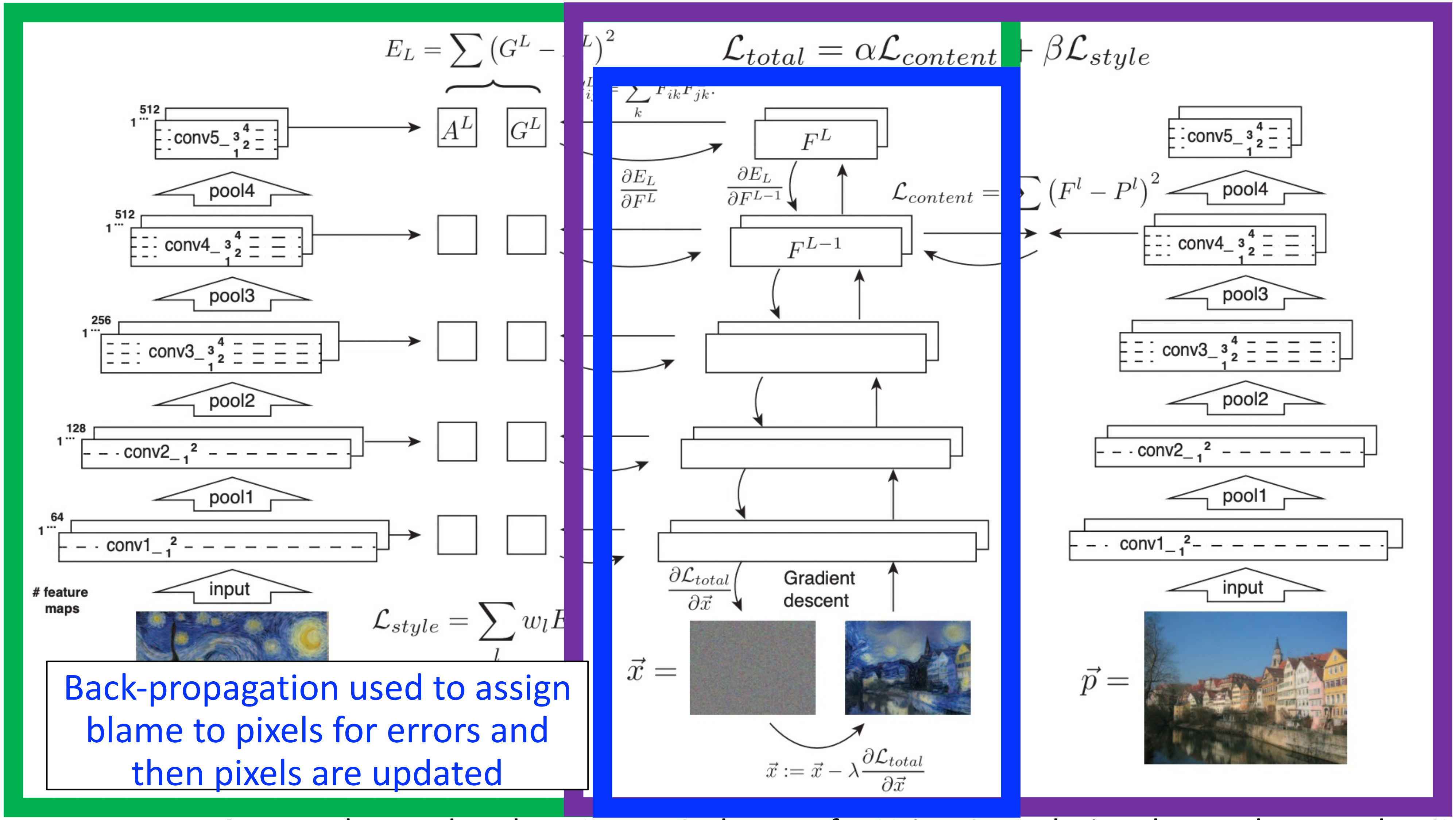


Figure Sources: https://ndres.me/images/style-transfer.gif; https://www.v7labs.com/blog/neural-style-transfer

Neural Style Transfer (NST): Algorithm

Compute content loss based on feature maps from 1 layer

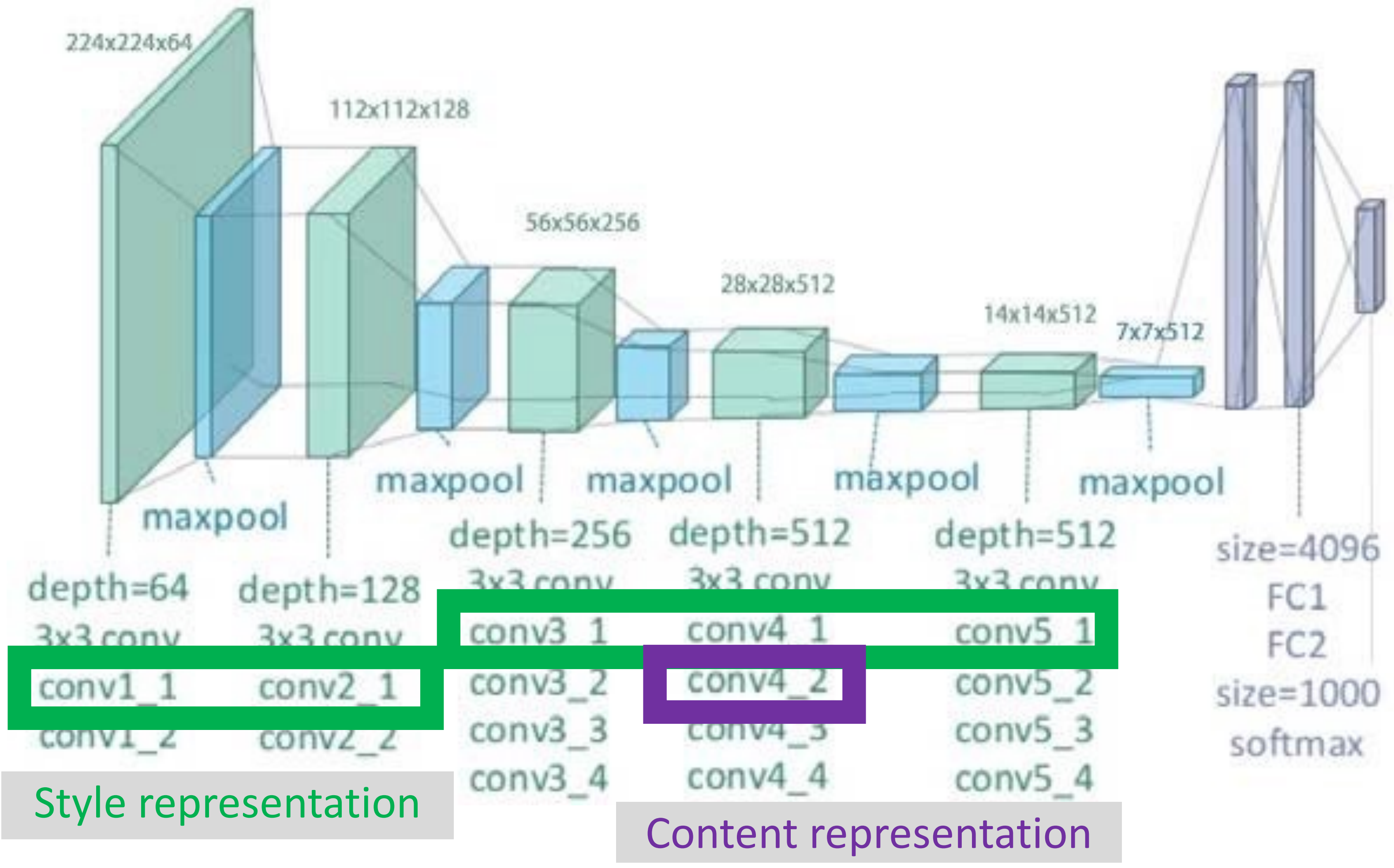
Compute style loss based on feature maps from 5 layers



Back-propagation used to assign blame to pixels for errors and then pixels are updated

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Neural Style Transfer (NST): Implementation



Uses VGG-19 for feature extraction

Figure Source: https://towardsdatascience.com/making-deep-learning-your-artist-with-style-transfer-

Neural Style Transfer (NST): Influence of Different CNN Layers in Representing Content

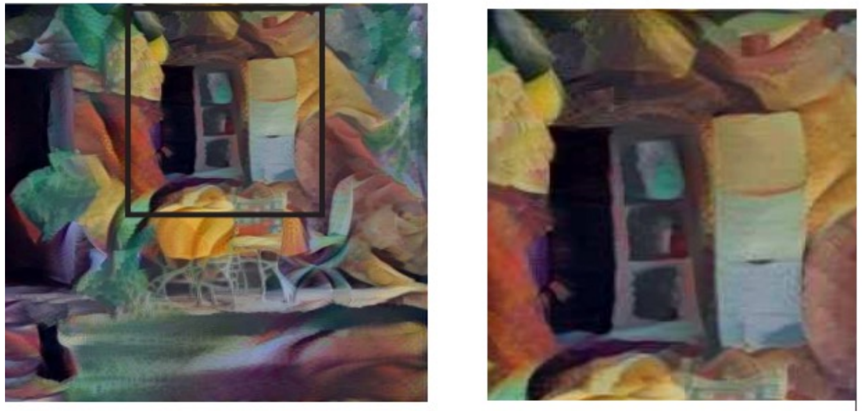
Content image

Style image

What are the differences in the stylized results?

2^{nd} convolutional layer of VGG-19

4^{th} convolutional layer of VGG-19



Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

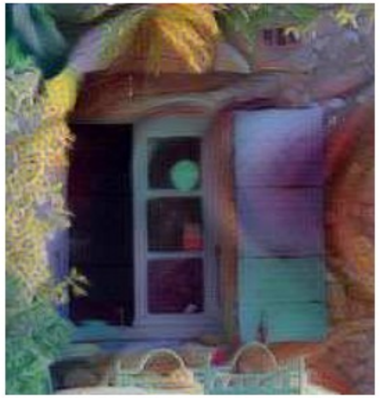
Neural Style Transfer (NST): Influence of Different CNN Layers in Representing Content

Content image

Style image

Which result do you prefer for artistic style transfer?

2nd convolutional layer of VGG-19



4th convolutional layer of VGG-19

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Neural Style Transfer (NST): Influence of Different CNN Layers in Representing Content

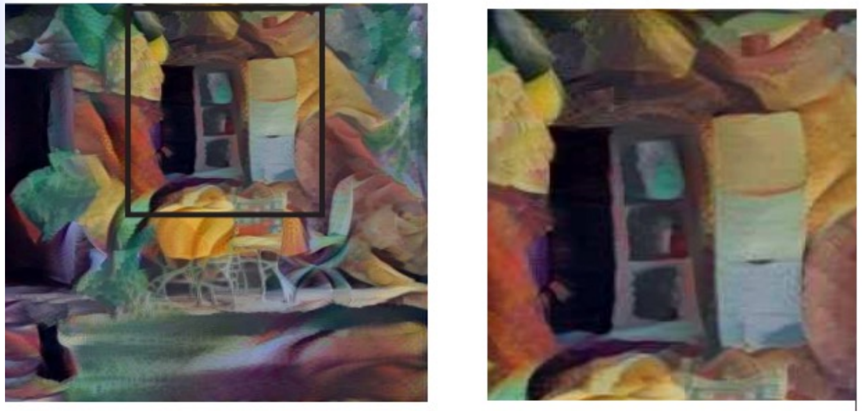
Content image

Style image

Higher layer features lead to different colors and edges that reflect the style of the artwork without requiring rendered pixels to match those in the content image

2nd convolutional layer of VGG-19

4th convolutional layer of VGG-19



Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Neural Style Transfer (NST): Trade-off When Optimizing for Style Loss vs Content Loss

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

Greater focus on minimizing content loss

$$\alpha / \beta$$

Greater focus on minimizing style loss

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Neural Style Transfer (NST): Trade-off When Optimizing for Style Loss vs Content Loss

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

What visual qualities arise from this style/content trade-off?

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Neural Style Transfer (NST): Trade-off When Optimizing for Style Loss vs Content Loss

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

What ratio should be used to balance style and content?

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Neural Style Transfer (NST): Intuition Behind Findings

Can separate the representation of content with a CNN because, when the CNN trains for the object recognition task, it learns to ignore image variations that can occur when recognizing an object.

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Neural Style Transfer (NST): Intuition Behind Findings

More concisely, a representation learned for discrimination can be useful for generation

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Neural Style Transfer (NST): Limitation

Slow; for example, synthesizing a 512x512 image takes ~1 hour

(it requires *iterative* optimization)

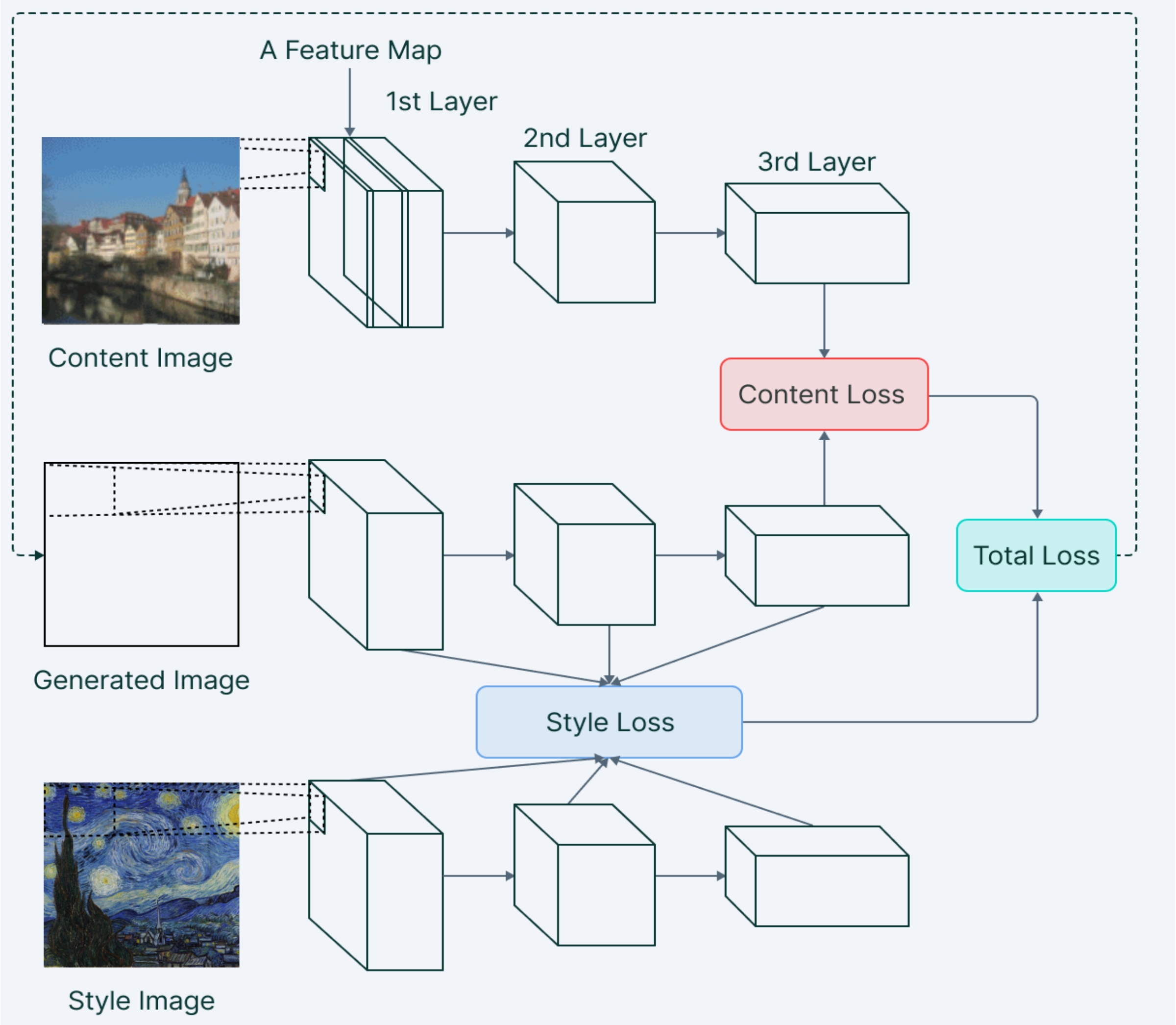


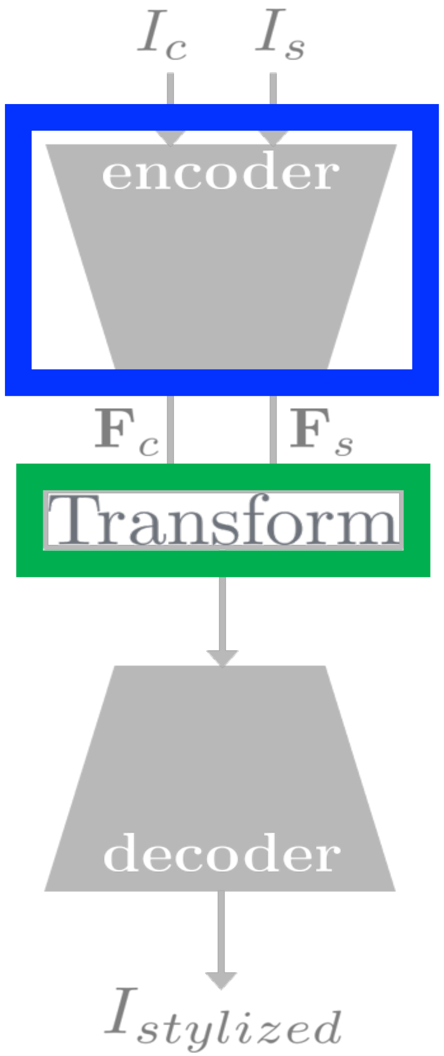
Figure Sources: https://ndres.me/images/style-transfer.gif; https://www.v7labs.com/blog/neural-style-transfer

Autoencoders

- How to computationally isolate the content of the content image?
- How to computationally isolate the style of the style image?
- How to blend the content and style?
- How to support any arbitrary style?
- How to do all these fast?

Gatys, Ecker, and Bethge. Image Style Transfer Using Convolutional Neural Networks. CVPR 2016.

Autoencoder: Basic Architecture



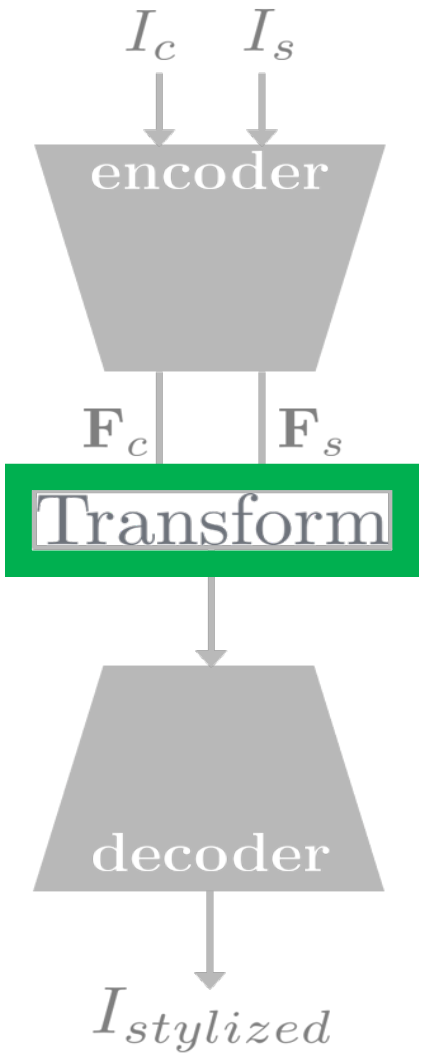
Typically uses VGG-19 to encode input content and style images

Key idea: directly transform features describing the content to match the statistics of the features describing the style image

(rather than matching statistics of the synthesized image to the style image)

Figure Source: Chiu and Gurari. ECCV 2020

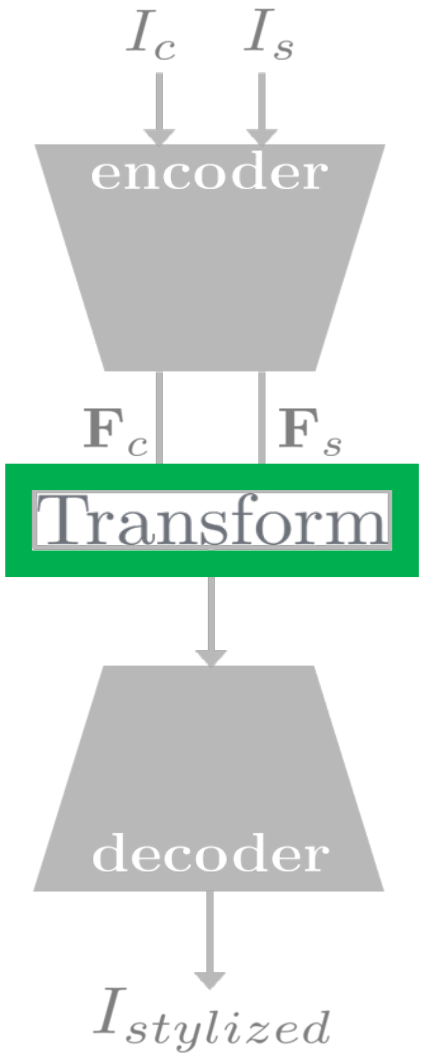
Autoencoder: WCT Transformation



e.g., adjusts the covariance of the content features to match that of the style image (through “whitening” and then “coloring”)

Yijun Li et al. Universal Style Transfer via Feature Transforms. Neurips 2017.

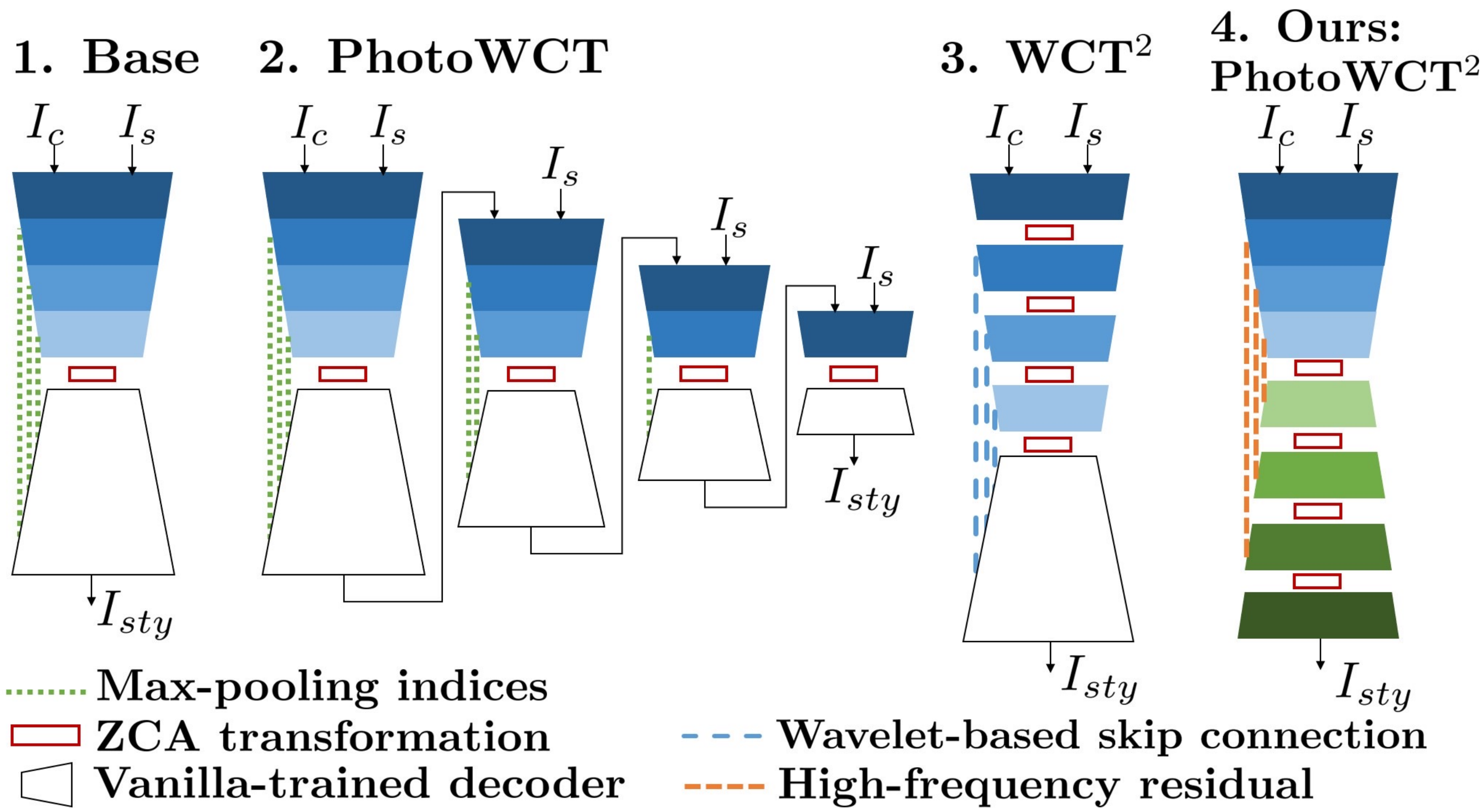
Autoencoder: AdaIn Transformation



e.g., adjusts the mean and variance of the content features to match those of the style features

Xun Huang and Serge Belongie. Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization. ICCV 2017.

Autoencoder: PhotoWCT and variants



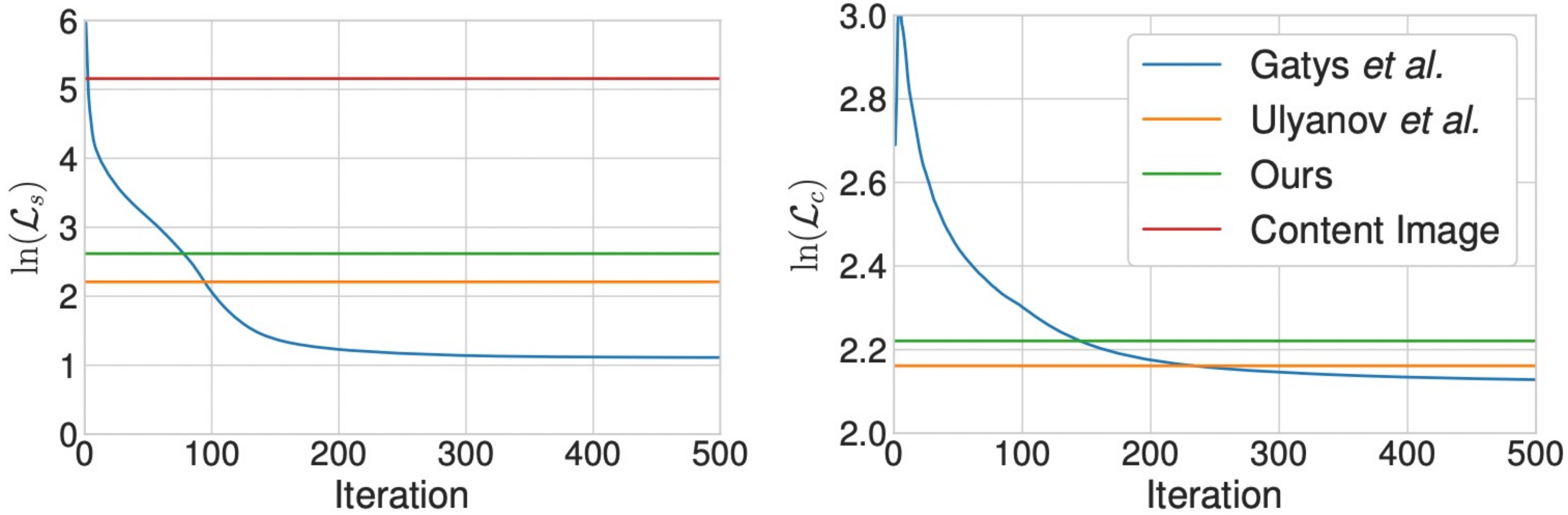
Often multiple layers of a network are used in order to capture both coarse and fine style features such that stronger style effects are achieved

[Chiu and Gurari. WACV 2022]

Style Transfer: Today's Topics

- Problem
- Applications
- Computer vision models
- Evaluation metrics

Content Loss and Style Loss



(a) Style Loss (b) Content Loss

Figure 3. Quantitative comparison of different methods in terms of style and content loss. Numbers are averaged over 10 style images and 50 content images randomly chosen from our test set.

Want to arrive at lower losses in as few iterations as possible

Xun Huang and Serge Belongie. Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization. ICCV 2017.

Speed (and sometimes size)

Model	(a) Size		(b) Speed performance				
	# par	# layer	1024×512	HD 1280×720	FHD 1920×1080	QHD 2560×1440	4K 3840×2160
PhNAS	40.24M	35	0.23	OOM	OOM	OOM	OOM
WCT^2	10.12M	**24**	0.30	0.43	0.80	OOM	OOM
PhWCT	8.35M	48	0.21+0.03	0.32+0.06	0.61+0.14	1.01+0.23	OOM
Ours (E2E) Ours (BT)	**7.05M**	**24**	**0.18+0.03**	**0.24+0.06**	**0.39+0.14**	**0.59+0.23**	**1.22+0.54**

Want model to run faster across many resolutions
(and so typically have fewer parameters)

[Chiu and Gurari. WACV 2022]

User Study: Which better carries the style?

A

B

C

D

E

Yijun Li et al. Universal Style Transfer via Feature Transforms. Neurips 2017.

User Study: Which is your favorite for a style?

A

B

C

D

E

Yijun Li et al. Universal Style Transfer via Feature Transforms. Neurips 2017.

User Study: Which looks more like a real photo?

A

B

[Chiu and Gurari. WACV 2022]

User Study: Which looks more like a real photo?

A

B

[Chiu and Gurari. WACV 2022]

User Study: Which looks more like a real photo?

A

B

[Chiu and Gurari. WACV 2022]

Style Transfer: Today’s Topics

- Problem
- Applications
- Computer vision models
- Evaluation metrics


The End