Image Style Transfer using Neural Methods

Xide Xia
xidexia@bu.edu

Agenda

- Motivation
- History
- NST Literature Review
- Empirical Comparison
- Future Direction
- Conclusion

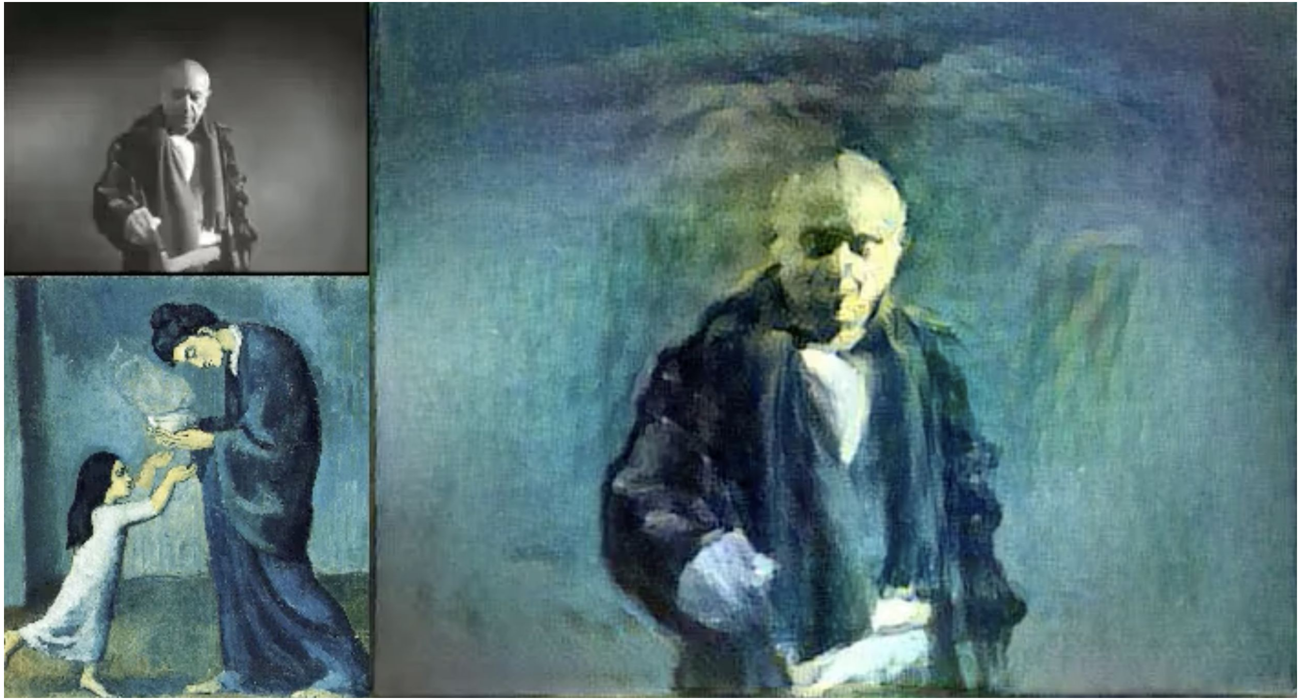
Agenda

- **Motivation**
- History
- NST Literature Review
- Empirical Comparison
- Future Direction
- Conclusion

What’s style transfer?

content

style



stylized
image

Agenda

- Motivation
- **History**
- Literature Review
- Empirical Comparison
- Future Direction
- Conclusion

Before NST

Before the appearance of NST, the related researches have expanded into an area called **non-photorealistic rendering (NPR)**.

Before NST

Before the appearance of NST, the related researches have expanded into an area called **non-photorealistic rendering (NPR)**.

- **Image-based artistic rendering (IB-AR)**: [Rosin *et al.*, 2012], [Kyprianidis *et al.*, 2013]
- **Stroke-Based Rendering (SBR)**: [Hertzmann *et al.*, 2012]
- **Example-Based Rendering (EBR)**. [Hertzmann *et al.*, 2001], [Zhao *et al.*, 2011]

Before NST

Before the appearance of NST, the related researches have expanded into an area called **non-photorealistic rendering (NPR)**.

- **Image-based artistic rendering (IB-AR)**: [Rosin *et al.*, 2012], [Kyprianidis *et al.*, 2013]
- **Stroke-Based Rendering (SBR)**: [Hertzmann *et al.*, 2012]
- **Example-Based Rendering (EBR)**. [Hertzmann *et al.*, 2001], [Zhao *et al.*, 2011]

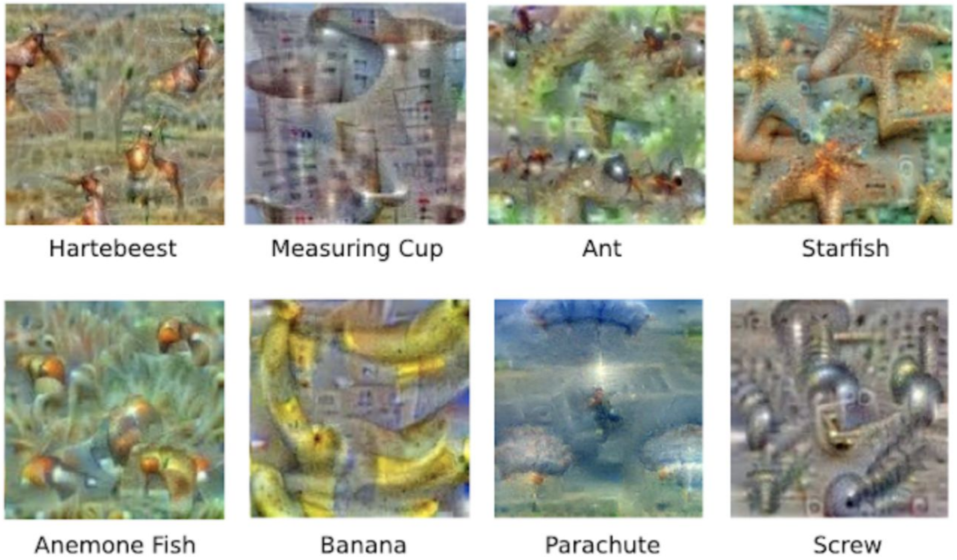
However, most of these NPR stylisation algorithms are designed for particular artistic styles and cannot be easily extended to other styles.

Agenda

- Motivation
- History
- **NST Literature Review**
- Empirical Comparison
- Future Direction
- Conclusion

DeepDream

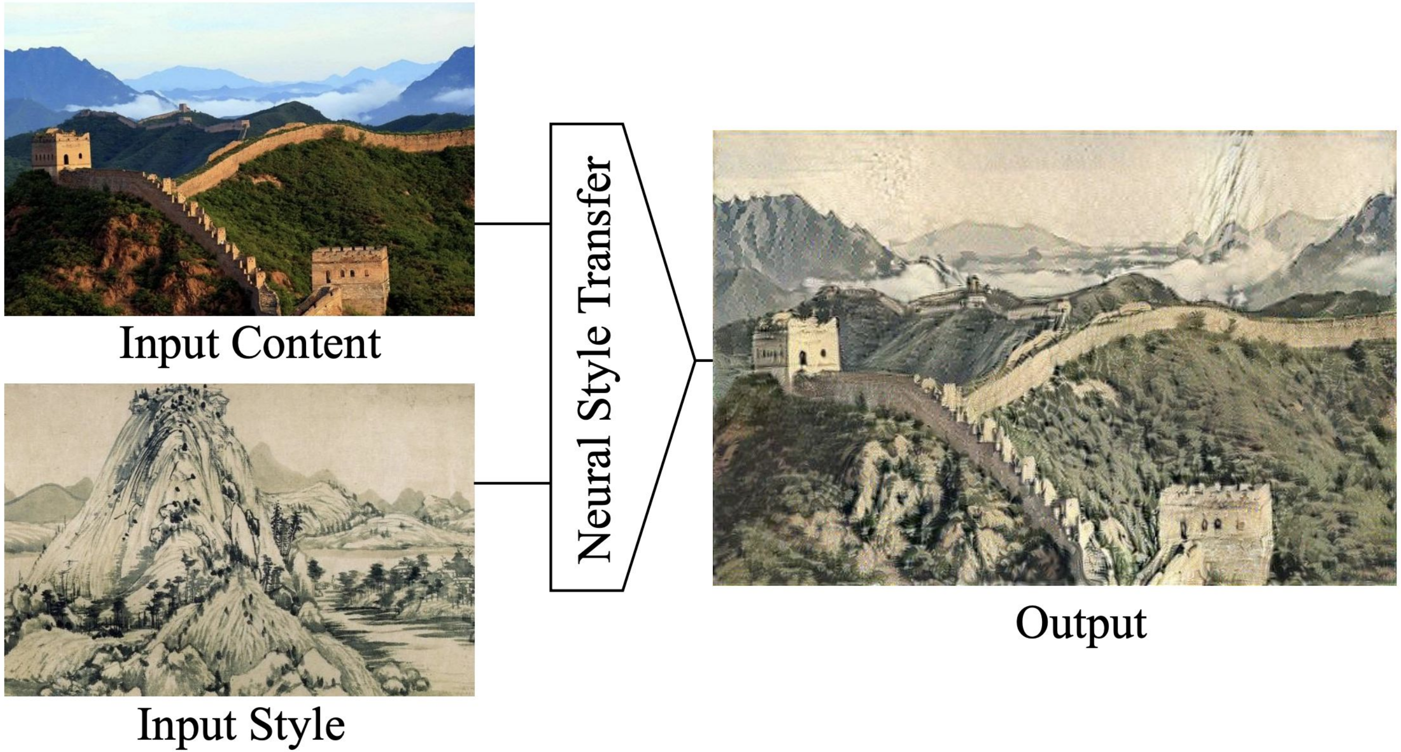
Google's DeepDream is the first attempt to produce artistic images by reversing CNN representations with IOB-IR techniques.



NST Literature Reviews

	Image-Optimisation-Based (IOB)	**Model-Optimisation-Based (MOB)**
Artistic Style Transfer	**[Gatys *et al.*, CVPR’16] (NST)** [Li *et al.*, IJCAI’17]	**[Johnson *et al.*, ECCV’16] (Perceptual Loss)** **[Huang *et al.*, ICCV’17] (AdaIN)** **[Li *et al.*, NIPS’17] (WCT)** [Li *et al.*, CVPR’19]
Photorealistic Style Transfer	**[Luan *et al.*, CVPR’17] (DPST)** [Mechrez *et al.*, BMVC’17]	**[Li *et al.*, ECCV’18] (PhotoWCT)** [Yoo *et al.*, ICCV’19] (WCT2)

Artistic Style Transfer



Photorealistic Style Transfer

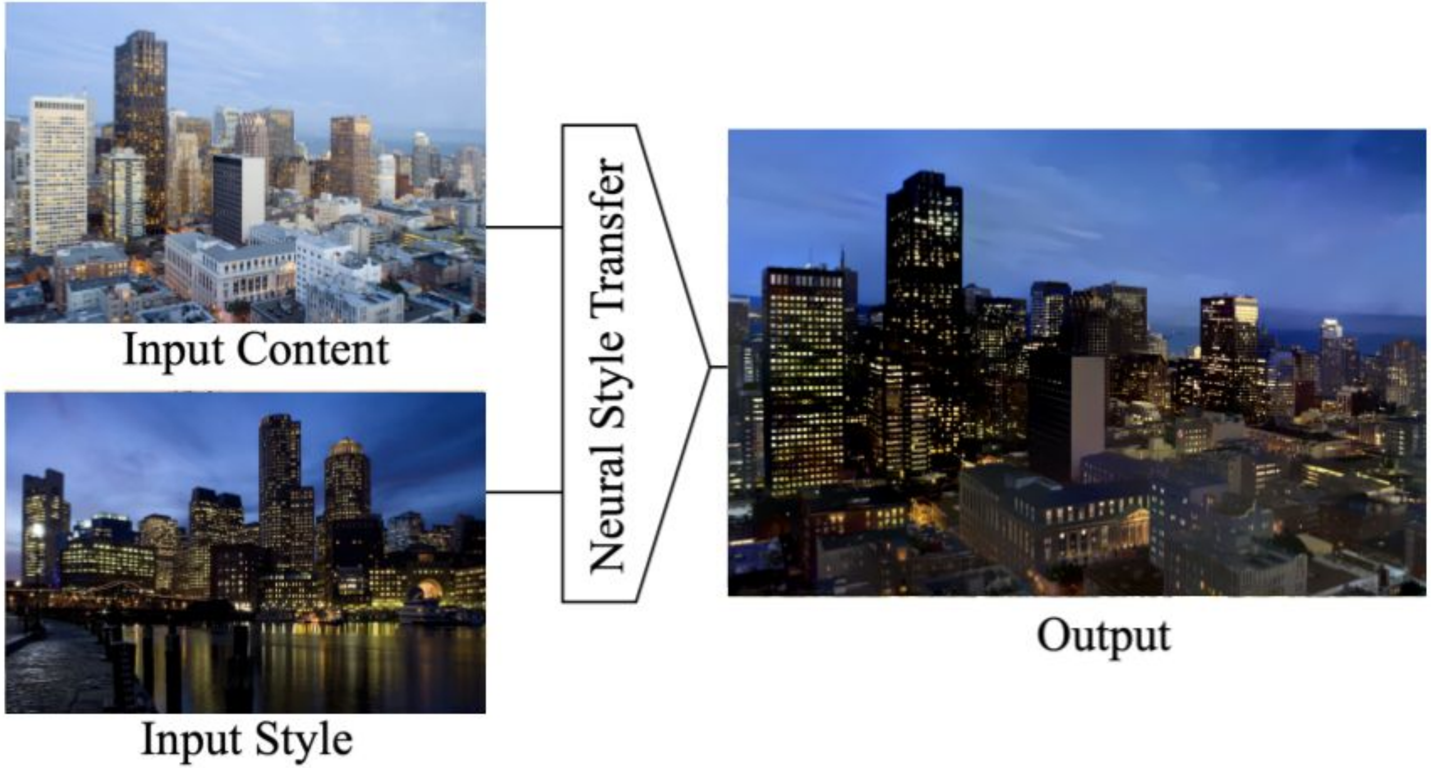
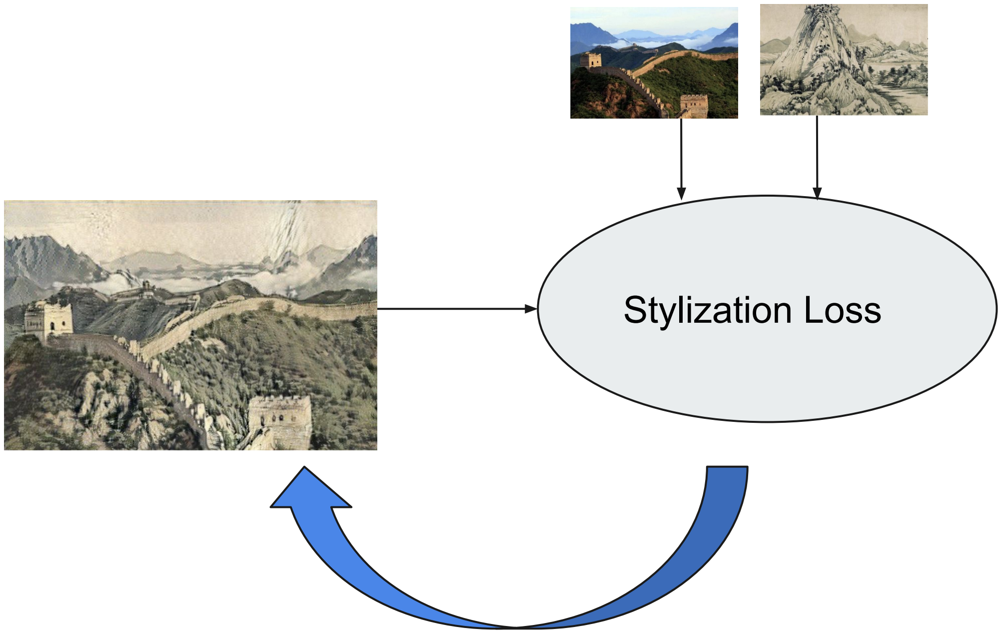


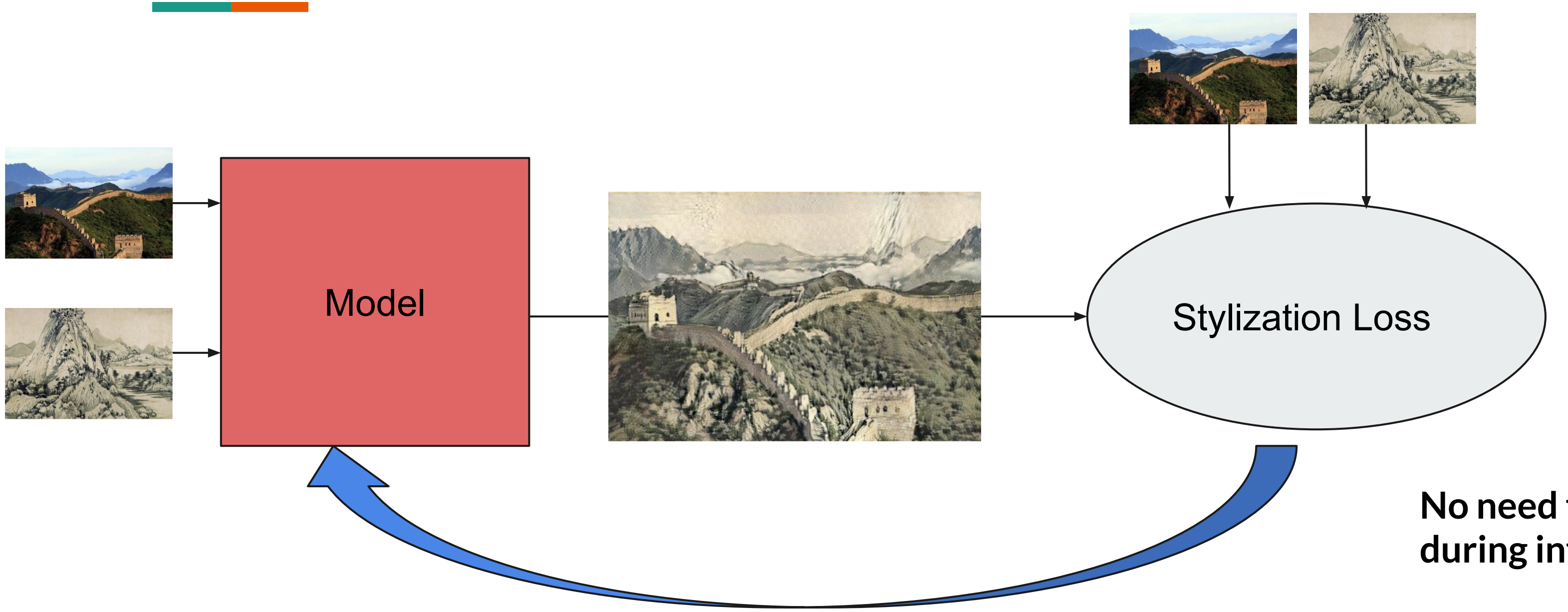
Image-Optimization-based (IOB)



Need to run optimization processing everytime!

: back-prop

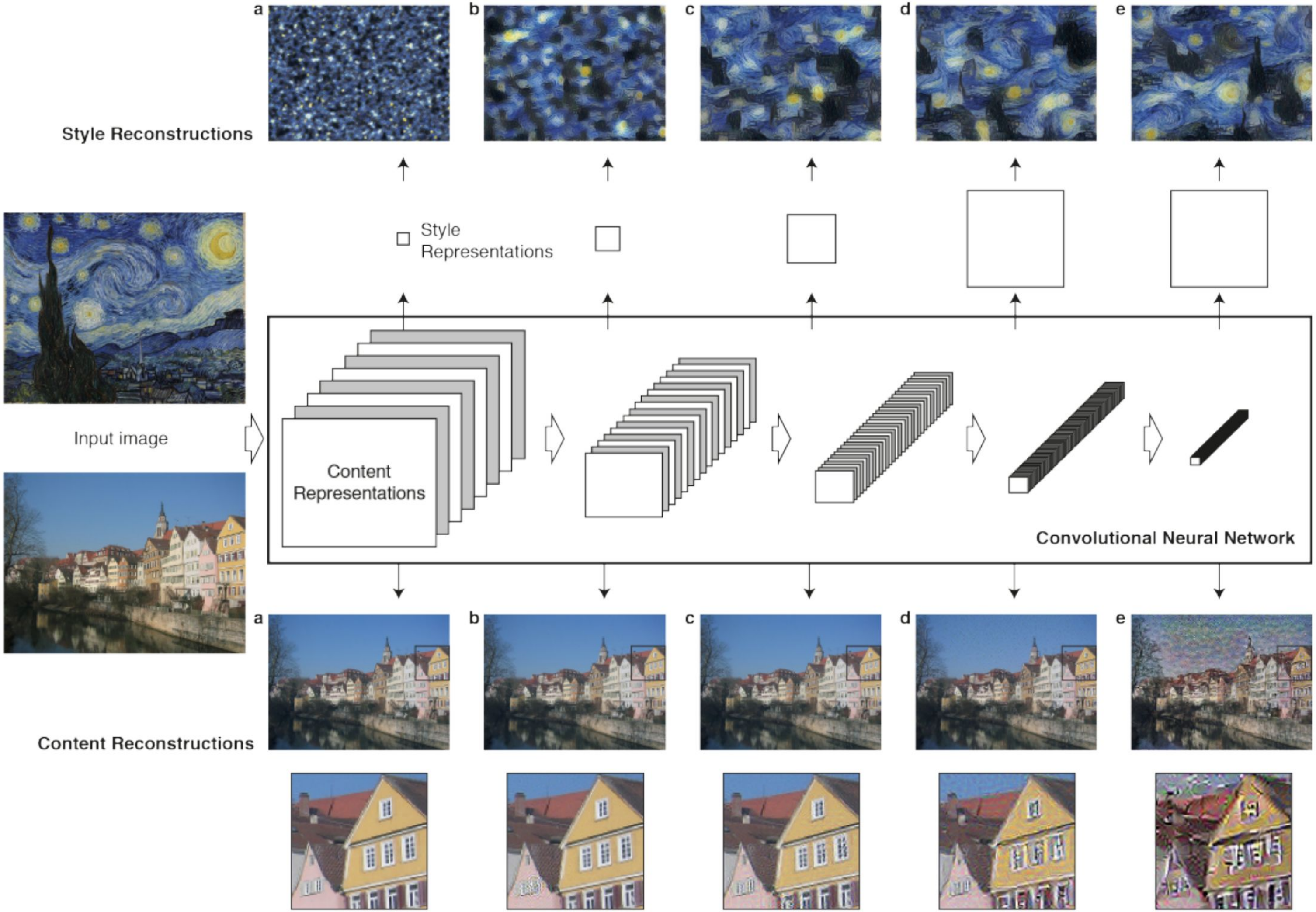
Model-Optimization-based (MOB)



No need to optimized during inference.

NST

A Neural Algorithm of Artistic Style
[Gatys *et al.*, CVPR'16]


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

NST

Given a content image ***c*** and a style image ***s***, the NST algorithm tries to to seek a stylized image that minimises the following objective:

$$
\begin{aligned}
I^* &= \arg\min_I \mathcal{L}_{total}(I_c, I_s, I) \\
&= \arg\min_I \; \alpha \mathcal{L}_c(I_c, I) + \beta \mathcal{L}_s(I_s, I),
\end{aligned}
$$

content loss

style loss

NST

Let **p** and **x** be the original image and the image that is generated and **P** and **F** their respective feature representation in current layer.

content loss

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

Selected content layers in NST: [conv4_2]

NST

Let **p** and **x** be the original image and the image that is generated and **P** and **F** their respective feature representation in current layer. And let **A** and **G** be their respective style representations (Gram Matrix).

Gram matrix is the inner product between the vectorised feature maps:

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

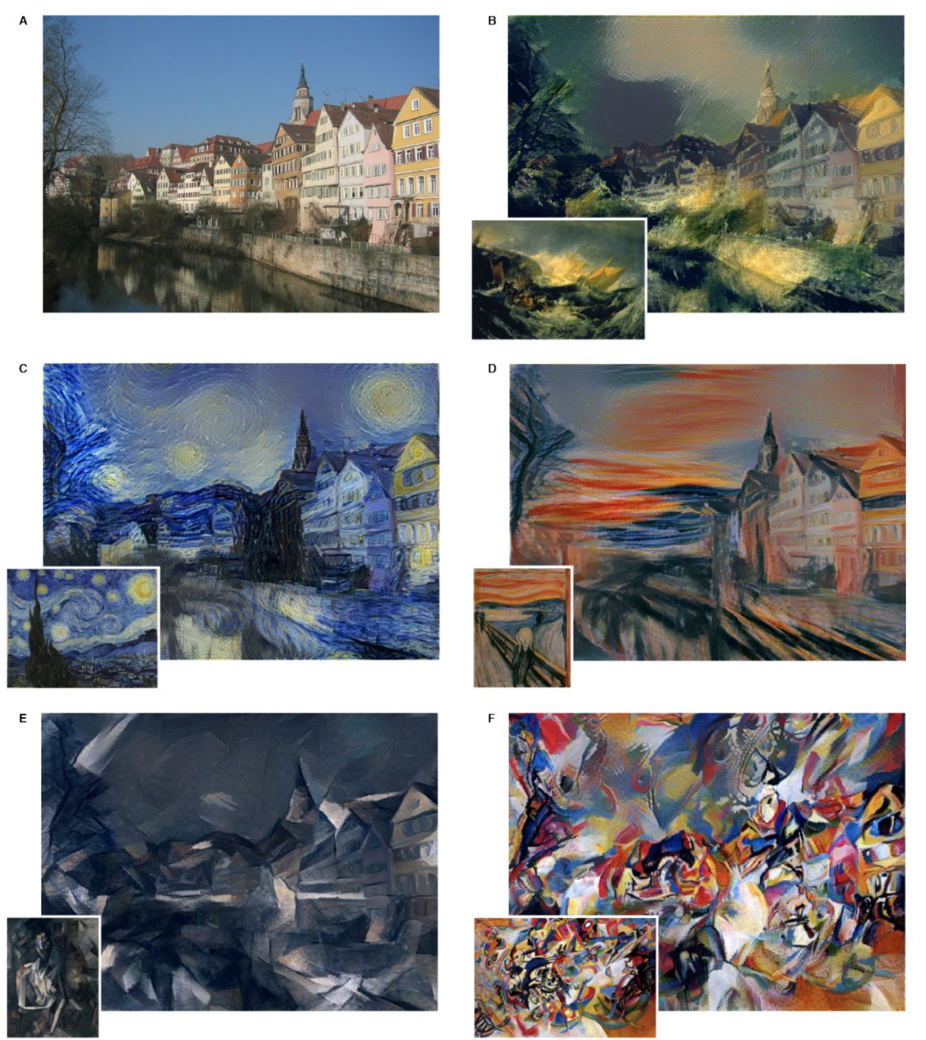
style loss

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

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

Style layers:
[conv1_1, conv2_1, conv3_1, conv4_1, conv5_1]
with equal weights.

wc/ws ratio
deeper layer



Content layers: [conv4_2]

Style layers: [conv1_1, conv2_1, conv3_1, conv4_1, conv5_1] with equal weights.

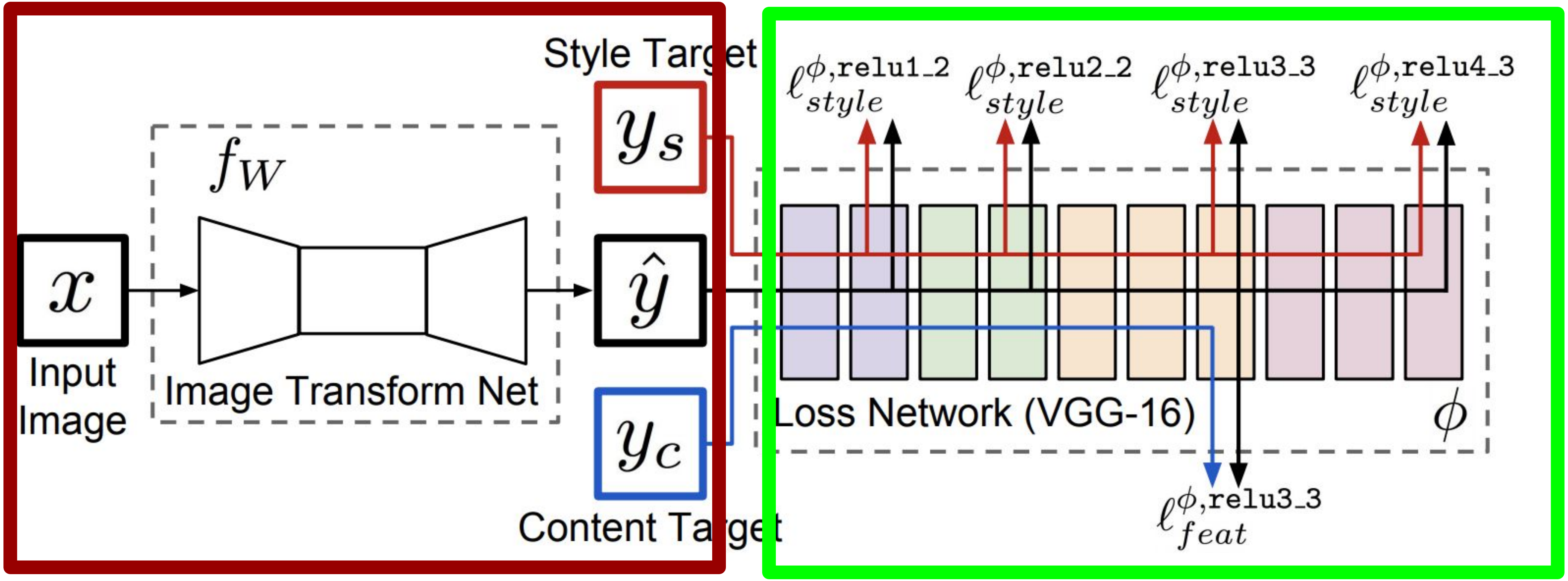
Analysis

	Task Type	Model Type	Arbitrary style?	Speed
NST	Artistic	IOB	Y	Slow

Perceptual Losses for Real-Time Style Transfer and Super-Resolution [Johnson *et al.*, ECCV'16]

Perceptual Loss

PSPM



encoder-decoder

Perceptual loss

Style
Composition VII, Wassily Kandinsky, 1913

Style
The Great Wave off Kanagawa, Hokusai, 1829-1832

content | NST | perceptual loss | content | NST | perceptual loss

Perceptual Loss

Image Size	NST 100	300	500	perceptual loss	Speedup 100	300	500
256×256	3.17	9.52s	15.86s	**0.015s**	212x	636x	**1060x**
512×512	10.97	32.91s	54.85s	**0.05s**	205x	615x	**1026x**
1024×1024	42.89	128.66s	214.44s	**0.21s**	208x	625x	**1042x**

Runtime comparison.

Analysis

	Task type	Model type	Arbitrary style?	Speed
NST	Artistic	IOB	Y	Slow
Perceptual Loss	Artistic	MOB	N (PSPM)	Fast

K styles → K models

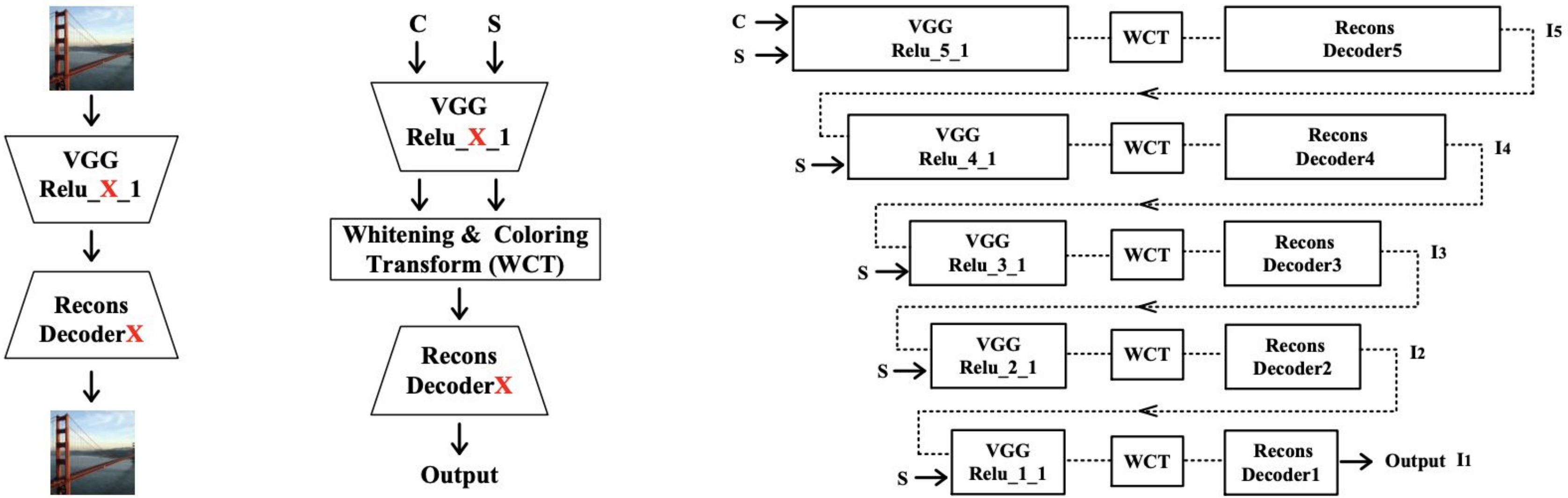
~~unseen style?~~

⋮

WCT

Universal Style Transfer via Feature Transforms
[Li *et al.*, NIPS'17]

WCT



(a) Reconstruction (b) Single-level stylization (c) Multi-level stylization

WCT

Whitening Transform.

$$\hat{f}_c = E_c \, D_c^{-\frac{1}{2}} \, E_c^\top \, f_c \,,$$

where D_c is a diagonal matrix with the eigenvalues of the covariance matrix $f_c \, f_c^\top \in \Re^{C \times C}$, and E_c is the corresponding orthogonal matrix of eigenvectors, satisfying $f_c \, f_c^\top = E_c D_c E_c^\top$.

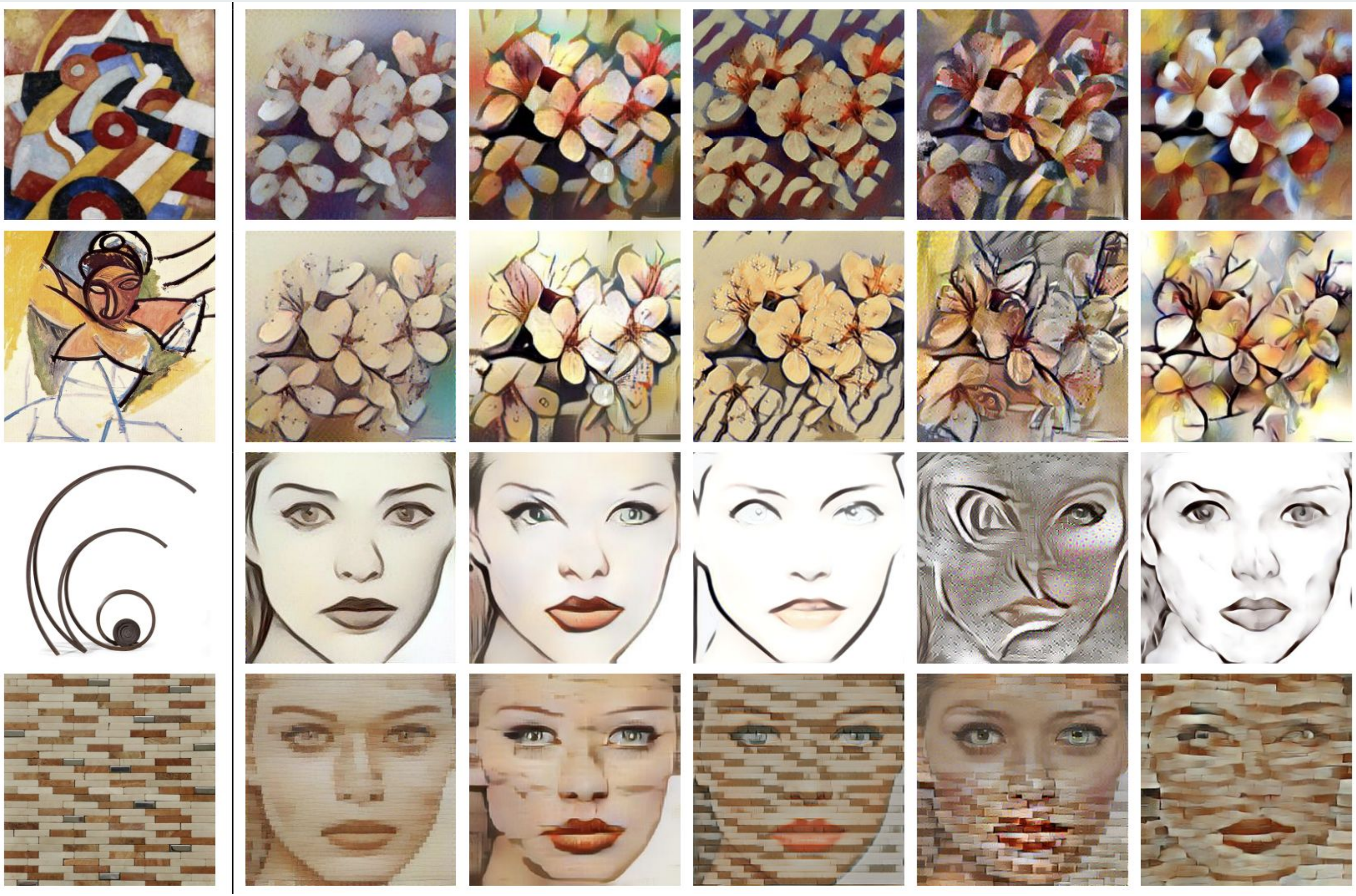
WCT

Coloring Transform.

$$\hat{f}_{cs} = E_s \, D_s^{\frac{1}{2}} \, E_s^\top \, \hat{f}_c \; ,$$

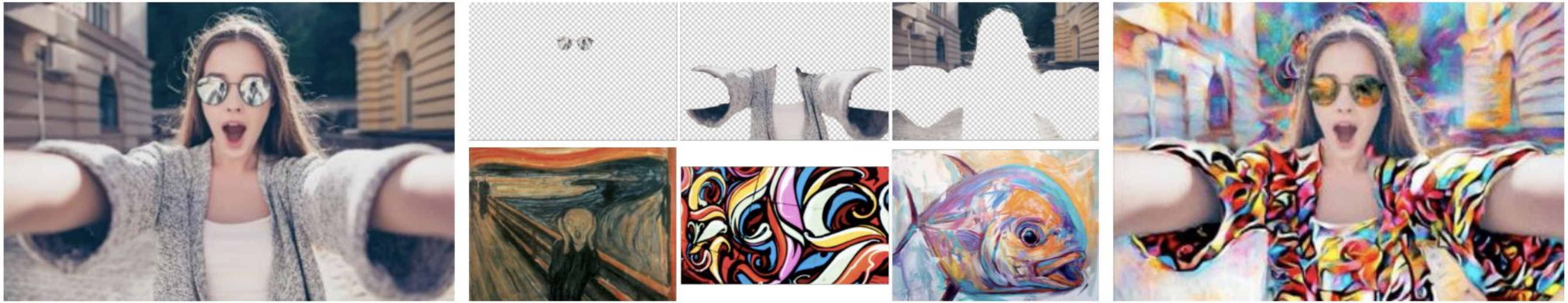
where D_s is a diagonal matrix with the eigenvalues of the covariance matrix $f_s \, f_s^\top \in \Re^{C \times C}$, and E_s is the corresponding orthogonal matrix of eigenvectors.

(a) Style (b) Content (c) HM (d) WCT (e) Style (f) Content (g) HM (h) WCT



Style [Chen *et al.*, Arxiv'16] AdaIN [Ulyanov *et al.*, ICML'16] Gatys WCT

WCT



(a) Content (b) Different masks and styles (c) WCT result

Analysis

	Task type	Model type	Arbitrary style?	Speed
NST	Artistic	IOB	Y	Slow
Perceptual Loss	Artistic	MOB	N (PSPM)	Fast
WCT	Artistic	MOB	Y(ASPM)	Fast

training-free

Photorealistic: DPST

Deep Photo Style Transfer
[Luan *et al.*, CVPR'17]

DPST


content


style

?

DPST

content

style

NST (non-photorealistic)

DPST


content


style


DPST (photorealistic)

DPST

Given a content image ***c*** and a style image ***s***, the NST algorithm tries to to seek a stylized image that minimises the following objective:

NST loss:

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_s^\ell$$

content loss

style loss

NST result

DPST

Given a content image ***c*** and a style image ***s***, the NST algorithm tries to to seek a stylized image that minimises the following objective:

NST loss:

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_s^\ell$$

content loss

style loss

NST result

DPST loss:

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_{s+}^\ell + \boxed{\lambda \mathcal{L}_m}$$

photorealistic regularization

DPST

Given a content image ***c*** and a style image ***s***, the NST algorithm tries to to seek a stylized image that minimises the following objective:

NST loss:

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_s^\ell$$

content loss

style loss

DPST loss:

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_{s+}^\ell + \boxed{\lambda \mathcal{L}_m}$$

photorealistic regularization

DPST result

DPST

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_{s+}^\ell + \boxed{\lambda \mathcal{L}_m}$$

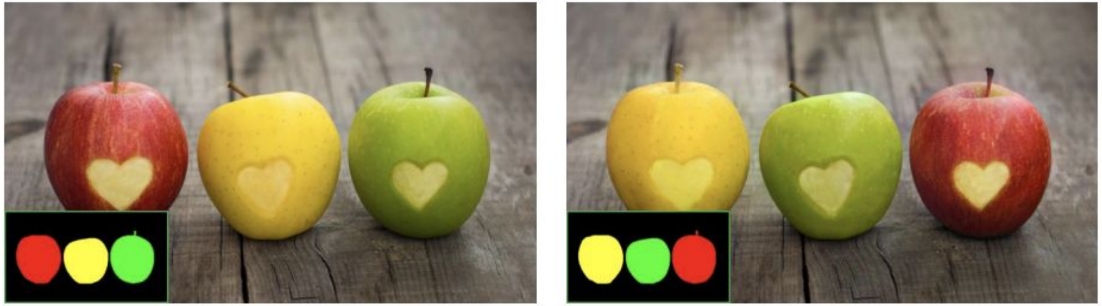
photorealistic regularization

$$\mathcal{L}_m = \sum_{c=1}^{3} V_c[O]^T \mathcal{M}_I V_c[O]$$

(a) Input and Style (b) $\lambda = 1$ (c) $\lambda = 10^2$ (d) $\lambda = 10^4$ (e) $\lambda = 10^6$ (f) $\lambda = 10^8$

DPST

(a) Input (b) Reference style (c) DPST



(d) Input (e) DPST

Compare to other NST methods:

(a) Input image (b) Reference style image (c) Neural Style (d) CNNMRF (e) DPST

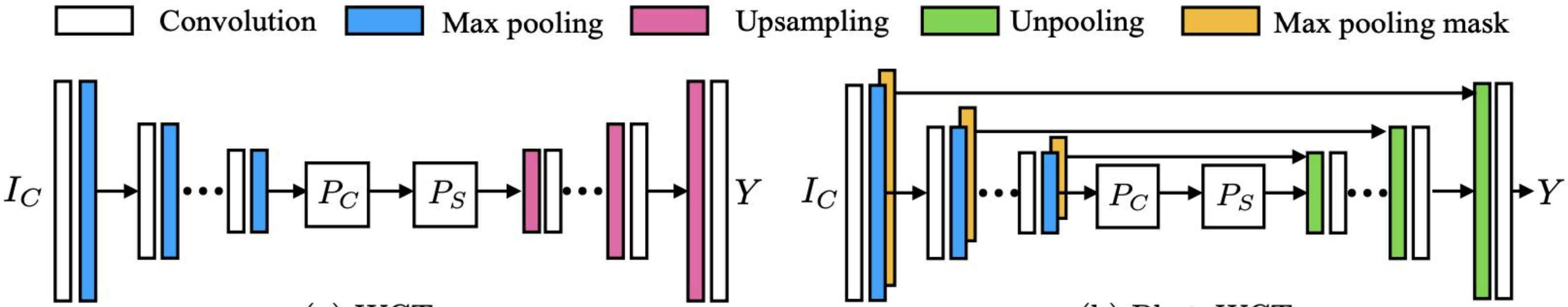
Analysis

	Task type	Model type	Arbitrary style?	Speed
DPST	Photorealistic	IOB	Y	Slow

PhotoWCT

A Closed-form Solution to Photorealistic Image Stylization
[Li *et al.*, ECCV’18]

PhotoWCT



(a) WCT (b) PhotoWCT

PhotoWCT

(a) Style

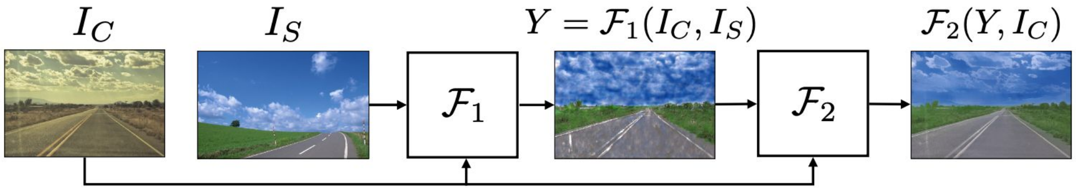
(b) Content

(c) WCT [10]

(d) PhotoWCT

PhotoWCT

Post-processing:



PhotoWCT

PhotoWCT + smoothing

PhotoWCT

Style

Content

Reinhard et al. [1]

Pitié et al. [2]

Luan et al. [9]

PhotoWCT + smoothing

PhotoWCT

Style

Content

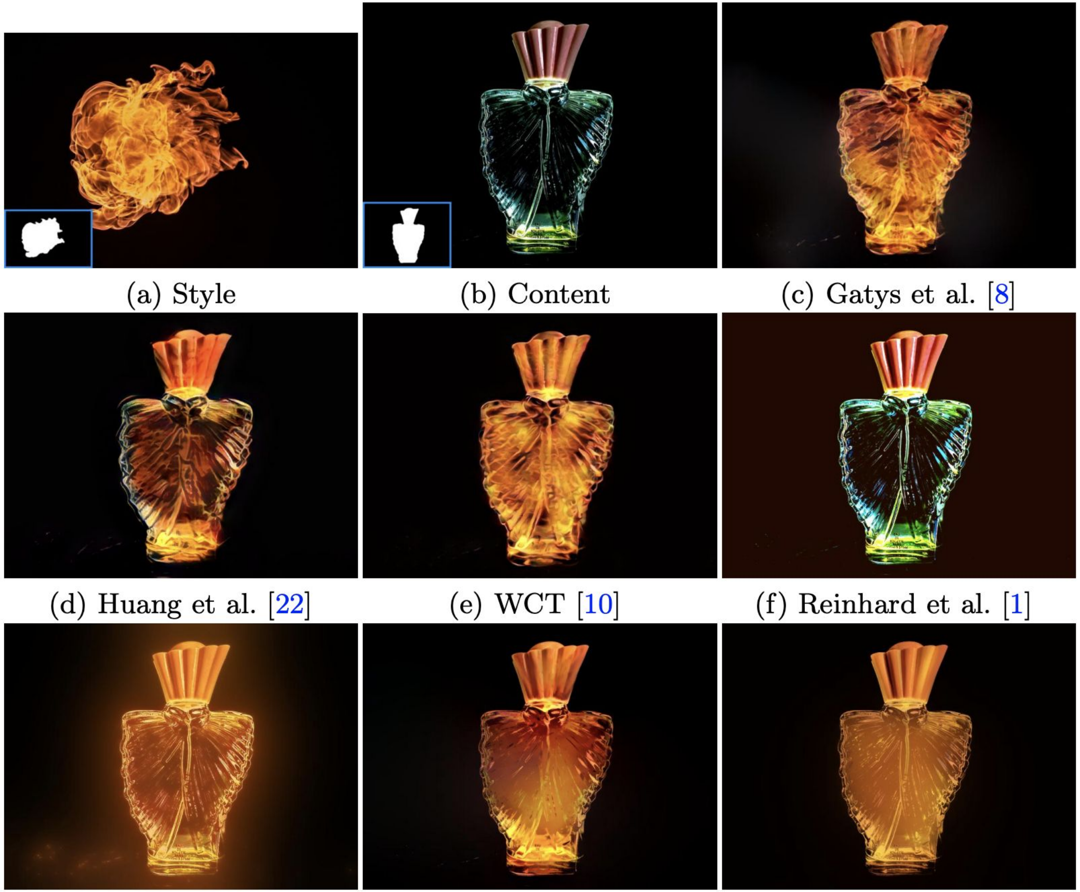
Reinhard et al. [1]

Pitié et al. [2]

Luan et al. [9]

PhotoWCT + smoothing

PhotoWCT



(a) Style (b) Content (c) Gatys et al. [8]
(d) Huang et al. [22] (e) WCT [10] (f) Reinhard et al. [1]
(g) Pitié et al. [2] (h) Luan et al. [9] (i) Ours

Comparison

	Task type	Model type	Arbitrary style?	Speed
DPST	Photorealistic	IOB	Y	Slow
PhotoWCT	Photorealistic	MOB	Y	Fast training-free

Agenda

- Motivation
- History
- Literature Review
- **Empirical Comparison**
- Future Direction
- Conclusion

Comparison

	Task type	Model type	Arbitrary style?	Speed
NST	Artistic	IOB	Y	Slow
Perceptual Loss	Artistic	MOB	N (PSPM)	Fast
WCT	Artistic	MOB	Y	Fast training-free
DPST	Photorealistic	IOB	Y	Slow
PhotoWCT	Photorealistic	MOB	Y	Fast training-free

Agenda

- Motivation
- History
- Literature Review
- Empirical Comparison
- **Future Direction**
- Conclusion

Future Direction

Future Direction

- High-resolution stylization? E.g. 12 megapixels, a typical resolution for modern smartphone cameras.

Future Direction

- High-resolution stylization? E.g. 12 megapixels, a typical resolution for modern smartphone cameras.
- Real-time application which can run on mobile camera.

Future Direction

- High-resolution stylization? E.g. 12 megapixels, a typical resolution for modern smartphone cameras.
- Real-time application which can run on mobile camera.

Thank you!

Reference

[1] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." *arXiv preprint arXiv:1508.06576* (2015).
[2] Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." *European conference on computer vision*. Springer, Cham, 2016.
[3] Huang, Xun, and Serge Belongie. "Arbitrary style transfer in real-time with adaptive instance normalization." *Proceedings of the IEEE International Conference on Computer Vision*. 2017.
[4] Luan, Fujun, et al. "Deep photo style transfer." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.
[5] Li, Yijun, et al. "Universal style transfer via feature transforms." *Advances in neural information processing systems*. 2017.
[6] Li, Yijun, et al. "A closed-form solution to photorealistic image stylization." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
[7] Yoo, Jaejun, et al. "Photorealistic style transfer via wavelet transforms." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.
[8] Li, Xueting, et al. "Learning linear transformations for fast arbitrary style transfer." *arXiv preprint arXiv:1808.04537* (2018).
[9] Mechrez, Roey, Eli Shechtman, and Lihi Zelnik-Manor. "Photorealistic style transfer with screened poisson equation." *arXiv preprint arXiv:1709.09828* (2017).
[10] Li, Yanghao, et al. "Demystifying neural style transfer." *arXiv preprint arXiv:1701.01036* (2017).

Reference

[11] Kyprianidis, Jan Eric, et al. "State of the" Art”: A Taxonomy of Artistic Stylization Techniques for Images and Video." *IEEE transactions on visualization and computer graphics* 19.5 (2012): 866-885.
[12] P. Rosin and J. Collomosse, Image and video-based artistic stylisation. Springer Science & Business Media, 2012, vol. 42.
[13] A. Hertzmann, “Painterly rendering with curved brush strokes of multiple sizes,” in Proceedings of the 25th annual conference on Computer graphics and interactive techniques. ACM, 1998, pp. 453– 460.
[14] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, “Image analogies,” in Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, 2001, pp. 327–340.
[15] M. Zhao and S.-C. Zhu, “Portrait painting using active templates,” in Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering. ACM, 2011, pp. 117–124.