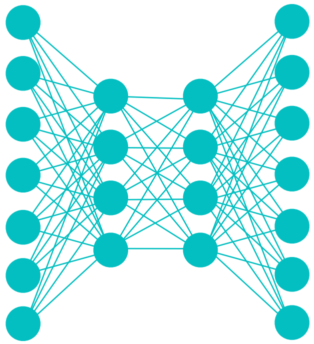
Lecture Notes for

Neural Networks and Machine Learning

One Shot Style Transfer
Photo-realistic Transfer
Non-image Styling



Logistics and Agenda

- Logistics
  - Next Assignment: Style Transfer
- Agenda
  - *A History of Style Transfer (last time)*
  - *Image Optimization Algorithms (last time)*
  - *Student Paper Presentation (last time)*
  - *Model Optimization Algorithms (last time)*
  - One Shot Algorithms, Demo (today)
  - Town Hall, Lab Style Transfer (today)
  - Evaluating Style Transfer Performance (today)
  - Extensions in Other Domains (today)

Last Time:

Whitening and Coloring with the Grammian of Activations:

content

$$A_{f,c} = flatten(A_c)$$

$$U_c, \Sigma_c, U_c^T = \text{SVD}(A_{f,c} \cdot A_{f,c}^T)$$

$$\widehat{A}_{f,c} = U_c \cdot \frac{1}{\sqrt{\Sigma_c}} \cdot U_c^T \cdot A_{f,c}$$

style

$$A_{f,s} = flatten(A_s)$$

$$U_s, \Sigma_s, U_s^T = \text{SVD}(A_{f,s} \cdot A_{f,s}^T)$$

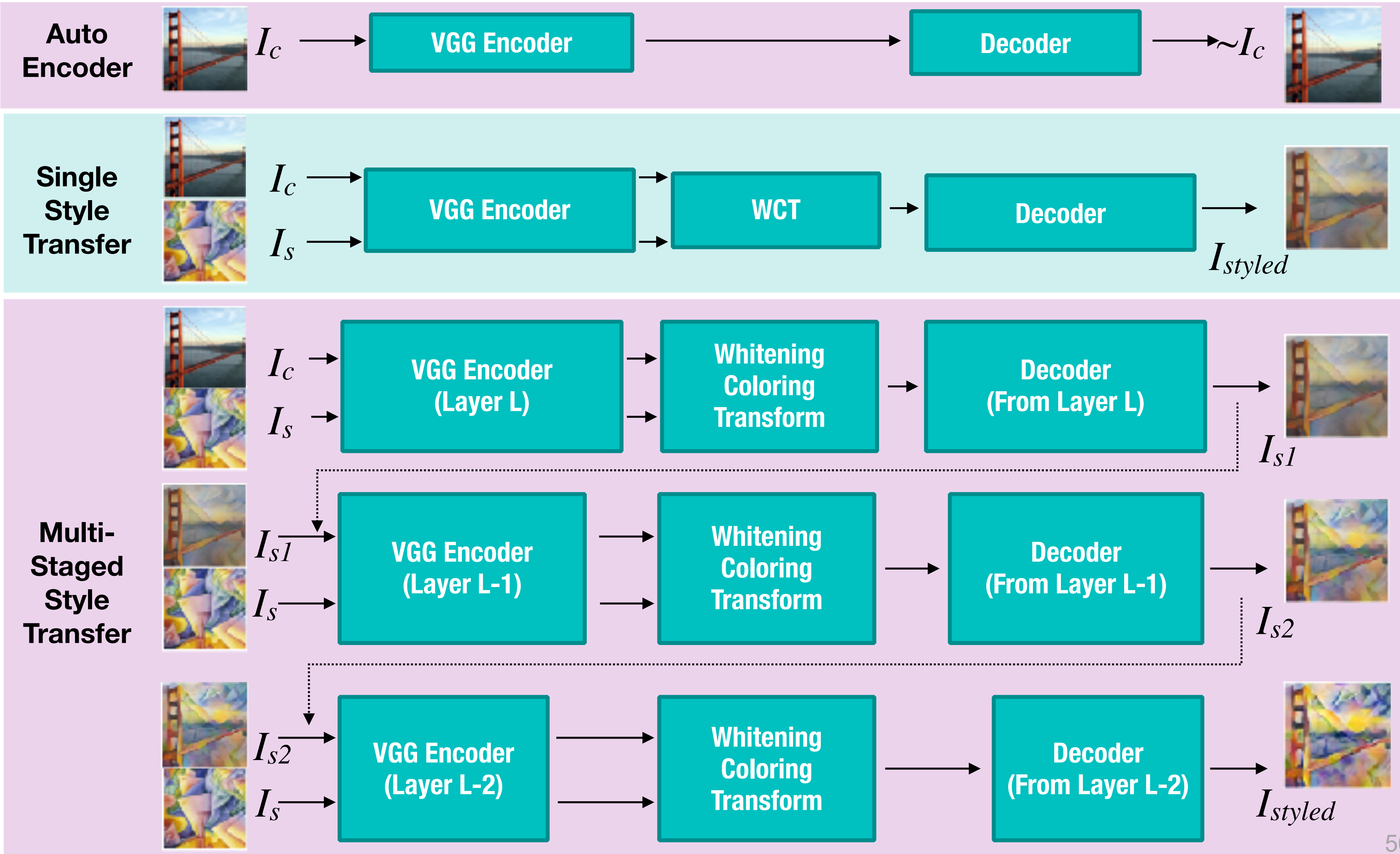
content styled

$$\widehat{A}_{f,s\leftarrow\approx c} = \underbrace{U_s \cdot \sqrt{\Sigma_s} \cdot U_s^T}_{\text{Desired Cov.}} \cdot \underbrace{\widehat{A}_{f,c}}_{\text{Whitened}}$$

$$A_{s\leftarrow\approx c} = reshape(\widehat{A}_{f,s\leftarrow\approx c})$$

WCT

Multi-Staged WCT (Last Time)



Why not go the other way?

- Start at earlier layers and apply WCT as we progress through the network
- Paper does not have good explanation, but results are subjectively poorer:

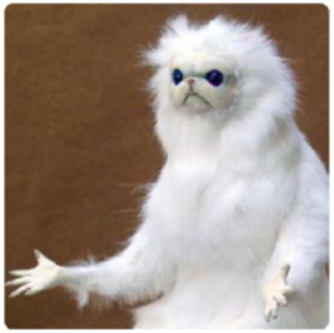
L > L-1 > L-2 > L-3

L-3 > L-2 > L-1 > L

Removing Style? Only Whitening

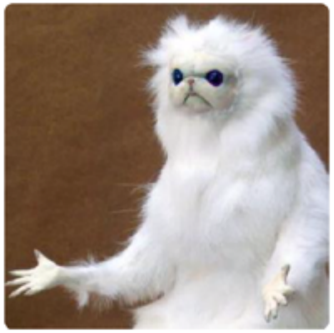
One Shot Style Transfer

Li, *et. al* Universal Style Transfer



justinledford

Justin Ledford

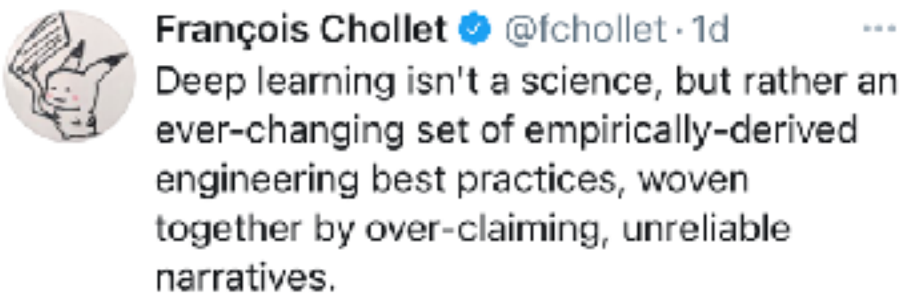


Yihao Wang

Follow Along: https://github.com/8000net/universal-style-transfer-keras

Or in the master repository:
05c UniversalStyleTransfer.ipynb

Town Hall


François Chollet @fchollet · 1d
Deep learning isn't a science, but rather an ever-changing set of empirically-derived engineering best practices, woven together by over-claiming, unreliable narratives.

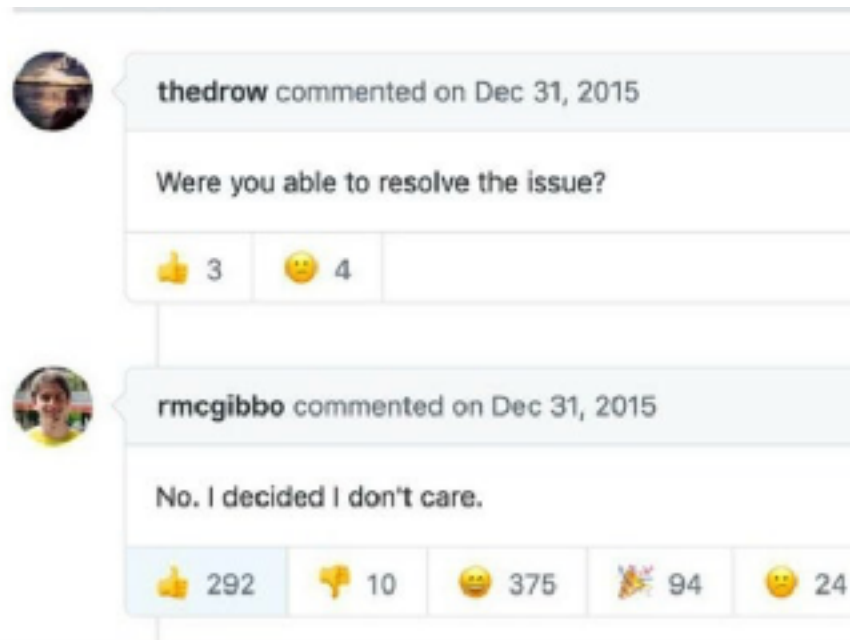

thedrow commented on Dec 31, 2015
Were you able to resolve the issue?
3
4
rmcgibbo commented on Dec 31, 2015
No. I decided I don't care.
292
10
375
94
24

Photo-Realistic Transfer

Grace Lindsay
@neurograce

C. Shannon on keeping science in order: "Authors should submit only their best efforts [...] A few first rate research papers are preferable to a large number that are poorly conceived or half-finished. The latter are no credit to their writers & a waste of time to their readers"

nature › commentary › article

MENU

nature

Commentary | Published: 30 April 1992

The growing inaccessibility of science

Donald P. Hayes

Nature **356**, 739–740(1992) | Cite this article

1015 Accesses | **44** Citations | **27** Altmetric | Metrics

Access options

Rent or Buy article	Subscribe to Journal
Get time limited or full article access on ReadCube.	Get full journal access for 1 year
from $8.99	$199.00 only $3.83 per issue
Rent or Buy	Subscribe
All prices are NET prices.	All prices are NET prices. VAT will be added later in the checkout.

Photo Style Transfer

Style | **Content** | **Results**

Photo Realistic WCT

- Use exact WCT architecture as before
  - …but use max un-pooling in upsample layers, instead of transpose convolutions (*meh*)
  - …and a smoothing constraint applied as an optimization on the result
  - Notation is borrowed from graph manifold rankings:

$$\arg\min_{R} \frac{1}{2} \sum_{i,j \in C}^{N,M} e^{-\frac{\|I_i^c - I_j^c\|^2}{\sigma_{i,j}^2}} \left\| \frac{R_i}{\sqrt{D_{ii}}} - \frac{R_j}{\sqrt{D_{jj}}} \right\|^2 + \left(\frac{1}{\alpha} - 1\right) \sum_{i}^{N} \sum_{j}^{M} \| R_{i,j} - Y_{i,j} \|^2$$

Y. Li, M.-Y. Liu, X. Li, M.-H. Yang, J. Kautz, A Closed-form Solution to Photorealistic Image Stylization, 2018

Smoothing

$\mathbf{I}^c$ is the content image, $\mathbf{Y}$ is the stylized image in graph structure

$\mathbf{R}$ is the desired result in graph structure

weighted sum of adjacent pixels i, j in $\mathbf{R}$

mixing hyperparameter

L2 norm of graphs

$$\arg\min_{R} \frac{1}{2} \boxed{\sum_{i,j\in 1\Delta}^{N,M} \boxed{e^{-\frac{\|I_i^c - I_j^c\|^2}{\sigma_{i,j}^2}}} \left\| \frac{R_i}{\sqrt{D_{ii}}} - \frac{R_j}{\sqrt{D_{jj}}} \right\|^2} + \boxed{\left(\frac{1}{\alpha} - 1\right)} \boxed{\sum_{i}^{N} \sum_{j}^{M} \| R_{i,j} - Y_{i,j} \|^2}$$

$\mathbf{W}_{i,j}$

affinity of content image as graph edges normalized by std of neighboring pixels (1Δ) known as "Matting Affinity"

$$D_{ii} = \sum_{\forall j} e^{-\frac{\|I_i^c - I_j^c\|^2}{\sigma_{i,j}^2}} = \sum_{\forall j} W_{i,j}$$

D is a diagonal matrix (degree matrix), summed from W

$$\hat{\mathbf{R}} = (1-\alpha)(\mathbf{I} - \alpha \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}})^{-1} \mathbf{Y}$$

closed form solution for smoothed result (I is identity matrix)

Y. Li, M.-Y. Liu, X. Li, M.-H. Yang, J. Kautz, A Closed-form Solution to Photorealistic Image Stylization, 2018

What is W? Connectivity of pixels as graph

$$D_{ii} = \sum_{\forall j} e^{-\frac{\|I_i^c - I_j^c\|^2}{\sigma_{i,j}^2}} = \sum_{\forall j} W_{i,j}$$

$\mathbf{I}^c$

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

$W_{i,j}$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	Δ			Δ	Δ										
2	Δ	0	Δ			Δ										
3		Δ	0	Δ		Δ	Δ	Δ								
4			Δ	0			Δ	Δ								
5	Δ				0	Δ			Δ	Δ						
6	Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ					
7			Δ	Δ		Δ	0	Δ		Δ	Δ	Δ				
8			Δ	Δ			Δ	0			Δ	Δ				
9					Δ	Δ			0	Δ			Δ	Δ		
10					Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ	
11						Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ
12							Δ	Δ			Δ	0			Δ	Δ
13									Δ	Δ			0	Δ		
14									Δ	Δ	Δ		Δ	0	Δ	
15										Δ	Δ	Δ		Δ	0	Δ
16											Δ	Δ			Δ	0

sum over each row

D_{ii}

$$W_{6,7} = e^{-\frac{(I_6 - I_7)^2}{\sigma_{1 \to 11}^2}}$$

1	2	3
5	6	7
9	10	11

sigma neighborhood

What is W, Y, and R?

$$\hat{\mathbf{R}} = (1-\alpha)\boxed{(\mathbf{I} - \alpha\mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}})}^{-1}\mathbf{Y}$$

N²xN² · N²xN² identity · N²xN² · N²xN² · N²xN² · N²xN²

Laplacian of graph

$W_{i,j}$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	Δ			Δ	Δ										
2	Δ	0	Δ			Δ										
3		Δ	0	Δ		Δ	Δ	Δ								
4			Δ	0			Δ	Δ								
5	Δ				0	Δ			Δ	Δ						
6	Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ					
7			Δ	Δ		Δ	0	Δ		Δ	Δ	Δ				
8			Δ	Δ			Δ	0			Δ	Δ				
9					Δ	Δ			0	Δ			Δ	Δ		
10					Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ	
11						Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ
12							Δ	Δ			Δ	0			Δ	Δ
13									Δ	Δ			0	Δ		
14									Δ	Δ	Δ		Δ	0	Δ	
15										Δ	Δ	Δ		Δ	0	Δ
16											Δ	Δ			Δ	0

sum over each row

D_{ii}

- **D** is diagonal and easily invertible
- **W** is sparse and efficiently inverted after multiplications
- **Y** is the stylized image pixels on a diagonal matrix
- **R** can be converted to an image by returning the diagonal

How to make this graph?

sklearn.feature_extraction.image.grid_to_graph

sklearn.feature_extraction.image.**grid_to_graph**(*n_x, n_y, n_z=1, mask=None, return_as=<class 'scipy.sparse.coo.coo_matrix'>, dtype=<class 'int'>*) [source]

Graph of the pixel-to-pixel connections

Edges exist if 2 voxels are connected.

Parameters:

n_x : ***int***
Dimension in x axis

n_y : ***int***
Dimension in y axis

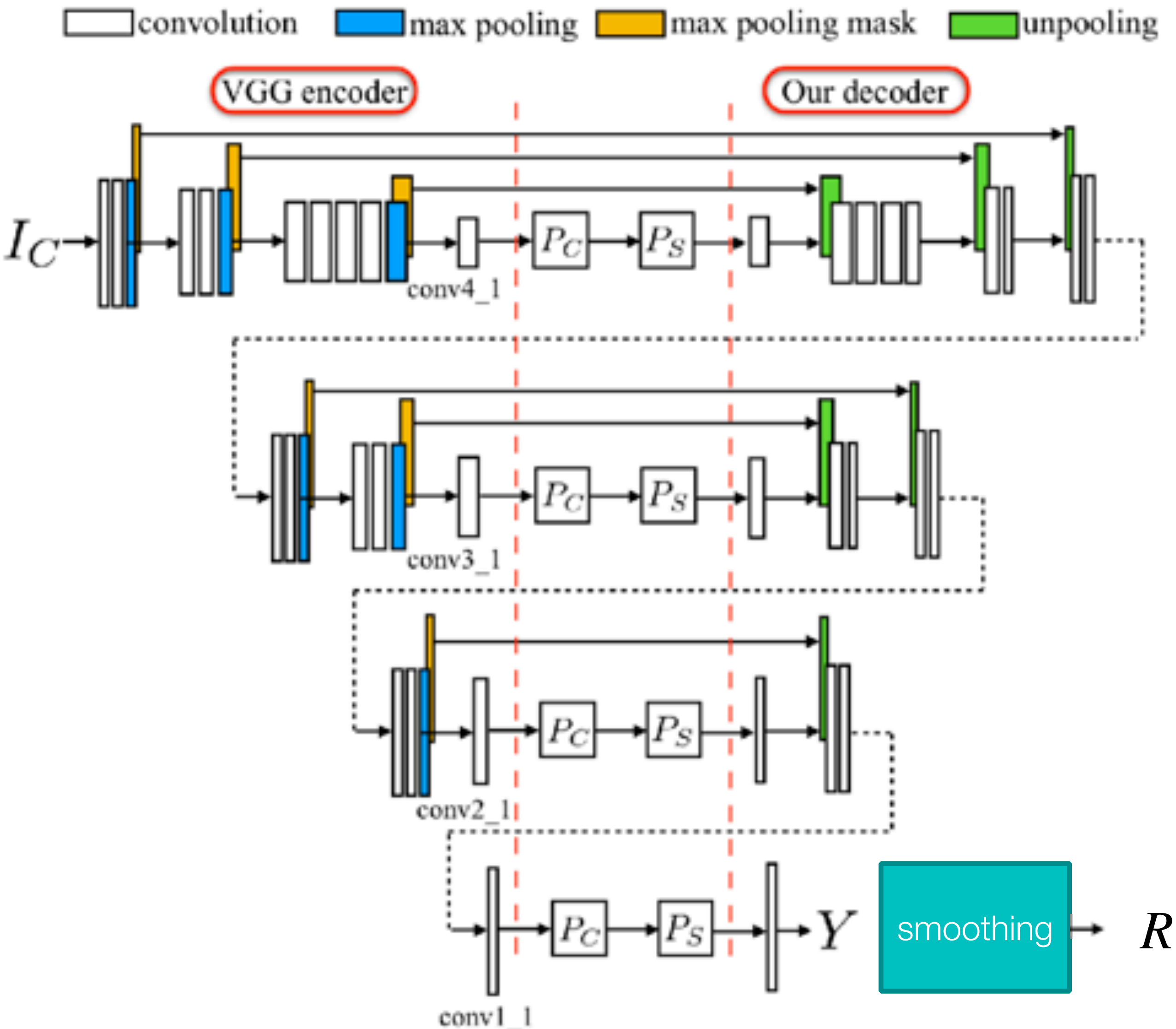
n_z : ***int, optional, default 1***
Dimension in z axis

mask : ***ndarray of booleans, optional***
An optional mask of the image, to consider only part of the pixels.

return_as : ***np.ndarray or a sparse matrix class, optional***
The class to use to build the returned adjacency matrix.

dtype : ***dtype, optional, default int***
The data of the returned sparse matrix. By default it is int

Similar Architecture as Before



Results

(a) Style
(b) Content
$Y_{\text{no unpooling}}$
(c) WCT [10]
(d) PhotoWCT
Y
$R_{\text{no unpooling}}$
(e) WCT + smoothing
(f) PhotoWCT + smoothing
R

Apply Masking to Different Segments of Image

Paper	Loss	Description
Gatys et al. [4]	*Gram Loss*	The first proposed style loss based on Gram-based style representations.
Johnson et al. [43]	*Perceptual Loss*	Widely adopted content loss based on perceptual similarity.
Berger and Memisevic [29]	*Transformed Gram Loss*	Computing *Gram Loss* over horizontally and vertically translated features. More effective at modelling style with symmetric properties, compared with *Gram Loss*.
Li et al. [51]	*Mean-substraction Gram Loss*	Subtracting the mean of feature representations before computing *Gram Loss*. Eliminating large discrepancy in scale. Effective at multi-style transfer with one single network.
Zhang and Dana [52]	*Multi-scale Gram Loss*	Computing *Gram Loss* over multi-scale features. Eliminating a few artefacts.
Li et al. [38]	*MMD Loss with Different Kernels*	*Gram Loss* is equivalent to *MMD Loss with Second Order Polynomial Kernel*. *MMD Loss with Linear Kernel* is capable of comparable quality with *Gram Loss*, but with lower computational complexity.
Li et al. [38]	*BN Loss*	Achieving comparable quality with *Gram Loss*, but conceptually clearer in theory.
Risser et al. [40]	*Histogram Loss*	Matching the entire histogram of feature representations. Eliminating instability artefacts, compared with single *Gram Loss*.
Li et al. [41]	*Laplacian Loss*	Eliminating distorted structures and irregular artefacts.
Li and Wand [42]	*MRF Loss*	More effective when the content and style are similar in shape and perspective, compared with *Gram Loss*.
Champandard [65]	*Semantic Loss*	Incorporating a segmentation mask over *MRF Loss*. Enabling a more accurate match.
Gu et al. [54]	*Reshuffle Loss*	Connecting both global and local style losses. Capable of preserving global appearance while avoiding distortions in local style patterns.
Li and Wand [48]	*Adversarial Loss*	Computed based on PatchGAN. Utilising contextual correspondence between patches. More effective at preserving coherent textures in complex images.
Jing et al. [61]	*Stroke Loss*	Achieving continuous stroke size control while preserving stroke consistency.
Wang et al. [62]	*Hierarchical Loss*	Enabling a coarse-to-fine stylisation procedure. Capable of producing large but also subtle strokes for high-resolution content images.
Liu et al. [63]	*Depth Loss*	Preserving depth maps of content images. Effective at retaining spatial layout and structure of content images, compared with single *Gram Loss*.
Ruder et al. [72]	*Temporal Consistency Loss*	Designed for video style transfer. Penalising the deviations along point trajectories based on optical flow. Capable of maintaining temporal consistency among stylised frames.
Chen et al. [70]	*Disparity Loss*	Designed for stereoscopic style transfer. Penalising bidirectional disparity. Capable of consistent strokes for different views.