
DIP 2020

Style Transfer Final Project

第12組

黃振修 王煥智 蕭皖文 莫易喆

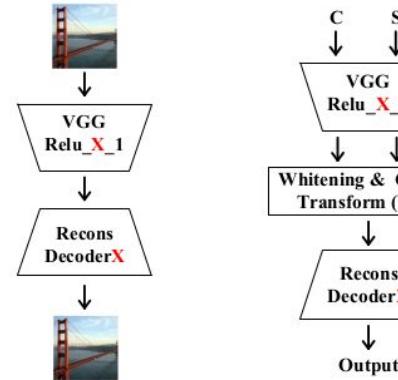
Agenda

- Method
- Artistic Effect Observation (素描、水彩、油畫、水墨)
- Our Contribution
- Result Demo
- Conclusion

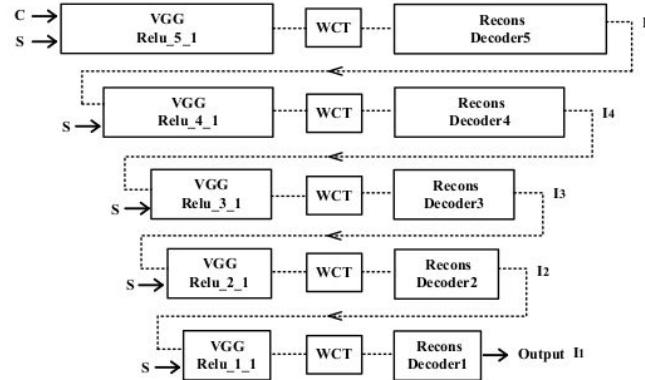
Universal Style Transfer with WCT

- WCT - whitening and coloring transforms [1]
- We use an improved WCT implementation [2] with gamma parameter

$$\hat{I}_l = d_l \left(\gamma \left(\delta C_l^s(e_l(\hat{I}_{l+1})) + (1 - \delta) C_l^s(e_l(I^c)) \right) + (1 - \gamma)e_l(I^c) \right)$$



(a) Reconstruction (b) Single-level stylization



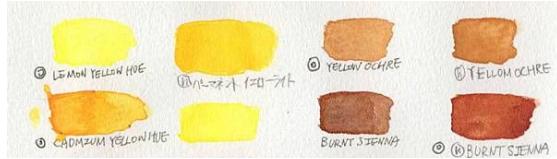
(c) Multi-level stylization

[1] Li, Yijun, et al. "Universal style transfer via feature transforms." Advances in neural information processing systems 30 (2017): 386-396.

[2] Wynen, Daan, Cordelia Schmid, and Julien Mairal. "Unsupervised learning of artistic styles with archetypal style analysis." Advances in Neural Information Processing Systems. 2018.

Artistic Effect Observation

- Artistic drawing is not simply style transfer [3]



paint by artist



style transfer by artist's painting

[3] <https://schollz.github.io/watercolor/?fbclid=IwAR0r4yPIVpCK7-7tyY1gsjtROOjWKAQL5DfCcGRK7phvQBhH1oAqS7MTLo#fn:3>

Our Contribution

- Style blending with saliency map
 - Replace constant gamma & delta value
- Content image pre-processing
 - Color, saturation adjustment, ... etc
- Style image pre-processing
 - Histogram matching to content image
- Style transfer resolution matching
 - If content image is too small, $2X \Rightarrow$ transfer $\Rightarrow 1/2X$
- Video style transfer
 - Reduce the style image weighting, smaller delta value
 - Temporal weighted average

Saliency Map

- Why use saliency? → not all regions are equal, viewer's eye is drawn more to more interesting areas, such as eyes, mouth, etc.
- Saliency map - a heatmap localizing “the most interesting” pixels of the image.
- Traditional approach: a spectral residual approach by analyzing the log-spectrum.
- Deep learning approach: CNN’s feature maps just before FC layers.

Saliency Map

- Saliency can be used in preprocessing and/or in WCT.
- In WCT we modify the formulation of parameters δ, γ (Wynen et al).
 - δ - controls the contribution of the current stylization level,
 - γ - controls the amount of stylization.

$$\hat{I}_l = d_l(\gamma(\delta C_l^s(e_l(\hat{I}_{l+1} + (1 - \delta)C_l^s(e_l(I^c)))) + (1 - \gamma)e_l(I^c))$$

$$\delta \in (0, 1) \longrightarrow \delta = \delta_0 - w I_{saliency}, \delta_0 \in (0, 1)$$

$$\gamma \in (0, 1) \longrightarrow \gamma = \gamma_0 - \frac{w}{2} I_{saliency}, \gamma_0 \in (0, 1)$$





*WCT + Spectral Residual
(pre-processing + δ , γ)*

*WCT + VGG heatmap
(pre-processing + δ , γ)*

Zero Out Low Saliency Region (寫意)



No Saliency



Spectral Saliency



VGG Activation Saliency

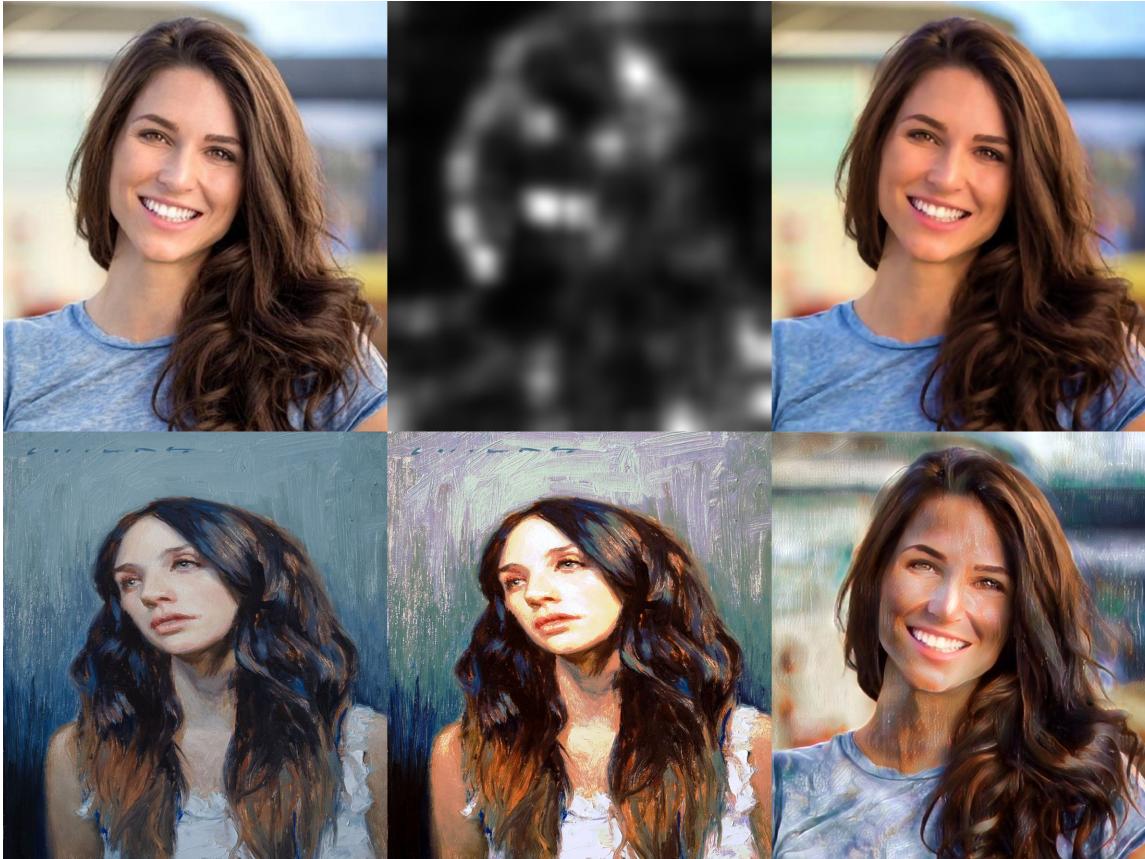


We do not use the zero out processing at last, because the uncertainty of saliency map!

Pre-Processing

- For pencil sketching, convert content image to gray image
- Style image RGB histogram matching to content image
- Oil painting, increase saturation (1.5X) and reduce luminance (0.9X) (HSV space)
- Apply bilateral filtering before oil painting, to reduce realistic granularity
- Watercolor painting, decrease saturation (0.75X) and enhance luminance (1.1X) (HSV space)
- If content image is much smaller than style image, double its size so that the brushes and strokes in style image won't be too big for the content image

Processing Example



Pre-Processing (Summary)



- gamma: $0.9 - 0.2 * (s[i])$
- delta: $0.95 - 0.1 * (s[i])$ $0 \leq s[i] \leq 1$

	Oil	Watercolor	Pencil	Ink
Grayscale			V	
Increase Saturation	V			
Reduce Saturation		V		
Bilateral filtering	V			
Style Image hist matching	V	V	V	
Upscale and Downscale	O	O	O	O
Weighted by saliency map	V	V	V	V

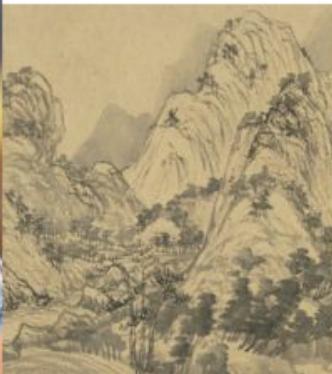
Result Demo



Result Demo



Result Demo



Result Demo



Result Demo



Result Demo



Result Demo



Result Demo



Result Demo



© David Maclellan Hodge

© David Maclellan Hodge

Result Demo



Video Processing

- Don't apply saliency map, because video is not still picture
- gamma: 0.95, delta: **0.6**
 - Delta controls if we want to accumulate the style effect from last layer
 - We preserve more single layer stylish to prevent the flickering effect
- Apply temporal frame weighted average: $1/4 * [1 \ 2 \ 1]$
- All the pre-processing is the same as image
- Encode the video with high bitrate

NTU Video

- 油畫
 - https://drive.google.com/drive/folders/1mn3T5I_EKU8NVwSkUcZOMUQov3rgf9VI?fbclid=IwAR2uyijLOAEUwXvwwVlZmM2AtikjIYApwGNgNULqTs8YGr6_o2haknwUk9WM
- 水墨
 - https://drive.google.com/drive/folders/1mn3T5I_EKU8NVwSkUcZOMUQov3rgf9VI?fbclid=IwAR2uyijLOAEUwXvwwVlZmM2AtikjIYApwGNgNULqTs8YGr6_o2haknwUk9WM
- 素描
 - https://drive.google.com/drive/folders/1mn3T5I_EKU8NVwSkUcZOMUQov3rgf9VI?fbclid=IwAR2uyijLOAEUwXvwwVlZmM2AtikjIYApwGNgNULqTs8YGr6_o2haknwUk9WM
- 水彩
 - https://drive.google.com/drive/folders/1mn3T5I_EKU8NVwSkUcZOMUQov3rgf9VI?fbclid=IwAR2uyijLOAEUwXvwwVlZmM2AtikjIYApwGNgNULqTs8YGr6_o2haknwUk9WM

Conclusion & Take Away

- Which style image is very important!
 - We spend many time to search suitable style image
 - Maybe pix2pix generation will solve this problem in the future?
- Resolution matching between content & style image
 - So that the brushes and strokes can precisely transferred
- Pre-processing is important and necessary
- Saliency detection is important, but also difficult
- For video style transfer, need to consider **temporal coherency** in the loss function [4]



[4] Chen, Xinghao, et al. "Optical flow distillation: Towards efficient and stable video style transfer." European Conference on Computer Vision. Springer, Cham, 2020.

Thank You

Q&A



+



WCT

WCT + Spectral Residual
(pre-processing + δ , γ)WCT + VGG heatmap
(pre-processing + δ , γ)