

Deep Photo Style transfer

① 작성일시	@January 21, 2022 3:15 PM
☰ Keywords	Basics 101 Style Transfer
👤 Reviewer	(L) LAIT
📎 Paper	https://arxiv.org/pdf/1703.07511.pdf
📅 발표일자	@January 14, 2022
📎 발표영상	http://gofile.me/6UrAI/1vYPMF0Ez
① 최종 편집	@April 3, 2022 2:31 PM
☑ Prof.	<input checked="" type="checkbox"/>
📎 PPT	http://gofile.me/6UrAI/mXYqrNLvf
☰ Q&A Link	
☰ 비고 (추가 노트)	
📎 추가자료	
☑ Reviewed	<input checked="" type="checkbox"/>
☰ 열	
☰ 속성	
☰ 속성 1	

v1 작성자 : 김하연
220114 발표 및 보완점 정리 : Laplacian Matrix(locally affine의 뜻)
220117 1차 초안 작성 + 최대한 Laplacian 보충 설명 정리
220121 ~ Method

Summary

- Input과 reference가 사진이여도 그림같아 보이는 문제를 처음으로 해결
(Photographic style transfer)
- Content의 구조는 그대로 유지하고 style transfer
(Structure preservation)
 - > Matting Laplacian
- Spill over(관련 없는 object 사이에도 transfer가 일어나는 문제) 해결
(Semantic accuracy and transfer faithfulness)
 - > Semantic Segmentation

Abstract

한 줄 소개 : This paper introduces a deep-learning approach to photographic style transfer that handles a large variety of image content while faithfully transferring the reference style.

기존 논문 문제 : Our approach builds upon the recent work on painterly transfer that separates style from the content of an image by considering different layers of a neural network. However, as is, this approach is not suitable for photorealistic style transfer. Even when both the input and reference images are photographs, the output still exhibits distortions reminiscent of a painting.

이 논문의 contribution : Our contribution is to constrain the transformation from the input to the output to be locally affine in colorspace, and to express this constraint as a custom fully differentiable energy term.

결과 : We show that this approach successfully suppresses distortion and yields satisfying photorealistic style transfers in a broad variety of scenarios, including transfer of the time of day, weather, season, and artistic edits.

기존의 연구들은 input과 reference image가 사진이어도 여전히 distortion이 생겨 그림같아 보이는(like-painting)문제를 가지고 있다. 이에, 이 논문은 transformation을 colorspace(RGB채널)에서 locally affine하도록 제약함으로써 사진같아 보이도록 한다.(Photorealistic style transfer)

Introduction

- Photographic style transfer : transfer the style of a reference style photo onto another input picture 예시) 다른 밝기, 낮/밤, 날씨 등
- 현재 존재하는 연구들은 diversity가 제한적이고 stylistic match에 대한 faithfulness가 부족(목표는 broad and faithful = 다양한 이미지 & reference style을 정확하게 옮기기). Gatys를 기본 모델로 사용하나, 여전히 그림 같아 보이는 문제 발생
- 정리 : transformation model that is locally affine in colorspace(Matting Laplacian), consider the difference in context between style and input(spill over방지, segmentic segmentation)
- Challenges and Contributions
 - Structure preservation : local drastic effect(극소적인 부위에 대한 급격한 변화)를 주는 것이 목표지만, 이러한 effect는 edge / regular pattern은 그대로 보존되어야 한다. 기존 Reinhard의 모델의 경우 global color transform을 이용했지만, 원하는 style에 match하기엔 한계가 존재했다. 따라서, 이 연구에서는 solution space를 photorealistic image로 제한한다.



위 사진의 경우 건물의 창문 하나하나에 불을 켜야하지만(local drastic effect), 창문은 여전히 네모 grid 안 (edge/regular pattern)에서 변형되지 않고 보존되어야 한다.

- Semantic accuracy and transfer faithfulness : transfer should respect the semantics of the scene(예를 들어, 도시의 경우 빌딩은 빌딩끼리 하늘은 하늘끼리 matching되어야 한다.) → 그래서, 우리는 style image에서 각각 input neural patch 와 가장 유사한 patch끼리 matching하고자 한다. (incorporate a semantic labeling of the input and style images into the transfer procedure)
- 기존의 CNNMRF방식은 style image의 전체 region이 무시되기 때문에 poor한 결과가 생긴다. Gram matrix만을 사용하는 경우에도 spill over 발생
- 기존 연구들과 가장 큰 차이점 : aim for **photorealistic transfer**(기존은 local change와 large-scale consistency사이 약간의 tension이 있었으나, 이 논문에서는 generic하고 더 넓고 다양한 style image를 다룰 수 있다!)

Method

1. Background(Gatys) : 위 모델은 Gatys 모델을 기반으로 만들어졌다.

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

with: $\mathcal{L}_c^\ell = \frac{1}{2N_\ell D_\ell} \sum_{ij} (F_\ell[O] - F_\ell[I])_{ij}^2 \quad (1b)$

$$\mathcal{L}_s^\ell = \frac{1}{2N_\ell^2} \sum_{ij} (G_\ell[O] - G_\ell[S])_{ij}^2 \quad (1c)$$

N_l filters with a vectorized feature map of size $D_l \cdot F_l[\cdot] \in R^{N_l \times D_l}$, Gram matrix $G_l[\cdot] = F_l[\cdot]F_l[\cdot]^T \in R^{N_l \times N_l}$, α_l, β_l : weights to configure layer preference, Γ : weight balancing tradeoff between content and style

style image S 를 input image I 에 transfer를 진행하여 output image O 를 생성한다.(위 목적함수를 minimizing하는 방향으로)

2. Method

a. Photorealism regularization

$$\mathcal{L}_m = \sum_{c=1}^3 V_c[O]^T \mathcal{M}_I V_c[O] \quad (2)$$

\mathcal{M}_I : laplacian matting matrix(depend on input image I , $N \times N$ pixels), $V_c[O]$: vectorized version ($N \times 1$) of the ouput image O in channel c

- 목적 : penalize image distortion(LS penalty : minimize) → image transform locally affine in color space
- affine function(input RGB → output counterparts)
 - edge patch의 경우 모든 채널에 동일한 위치에 놓여져 있으므로 edge 자체는 움직일 수 없다.
- Matting Laplacian 이용(grayscale matte → locally affine combination of the input RGB channels) : input O 이미 photorealistic 하므로 output을 제약하기보

다 input이 특징을 잃지 않도록 제약, 즉 이미지에 왜곡이 생겼을때, penalty를 가 함)

▼ **Closed form matting**

Matting: 이전의 background & foreground 中 F_i 를 뺀다!

$$I_i = \alpha_i \cdot F_i + (1 - \alpha_i) \cdot \frac{B_i}{\text{fore}} \quad \text{back} : \text{전경과 후경을 확장 배제 나눠!}$$

④ R, G, B

1) Color line assumption

: 전경과 후경의 색상들은 RGB공간에서 주변기하위에 존재한다고 가정

$$\beta_i \quad F_1 \quad F_2$$

$$\text{전경 } F_i = \beta_i F_1 + (1 - \beta_i) F_2$$

$$\text{후경 } B_i = \delta_i B_1 + (1 - \delta_i) B_2$$

→ 전경과 후경을 같은 두 원본으로 표현 ○

$$(즉, 한 학생은 두 경의 선생님일지도!) \Rightarrow \alpha_i = \alpha_i^T I_i + b \quad (\text{true matte } \alpha)$$

증명) α 는 I_i 의 linear fn임을 보자!

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i \leftarrow F_i \text{과 } B_i \text{가 같은 선형에 있으면 } \alpha \text{도 선형}$$

$$= \alpha_i (\beta_i F_1 + (1 - \beta_i) F_2) + (1 - \alpha_i) (\delta_i B_1 + (1 - \delta_i) B_2)$$

$$= \alpha_i \beta_i F_1 + \alpha_i (1 - \beta_i) F_2 - \alpha_i \delta_i B_1 - \alpha_i (1 - \delta_i) B_2$$

$$I_i - B_2 = \begin{bmatrix} F_2 - B_2 \\ F_1 - F_2 \\ B_1 - B_2 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \alpha_i \beta_i \\ \alpha_i - \alpha_i \delta_i \end{bmatrix}$$

$$I_i - B_2 = [F_2 - B_2, F_1 - F_2, B_1 - B_2] \begin{bmatrix} \alpha_i \\ \alpha_i \beta_i \\ \alpha_i - \alpha_i \delta_i \end{bmatrix}$$

$$\begin{bmatrix} \alpha_i \\ \alpha_i \beta_i \\ \alpha_i - \alpha_i \delta_i \end{bmatrix} = \begin{pmatrix} r_1^T \\ r_2^T \\ r_3^T \end{pmatrix} [I_i - B_2]$$

$$= \begin{pmatrix} r_1^T \\ r_2^T \\ r_3^T \end{pmatrix} [I_i - B_2]$$

$$\therefore \alpha_i = r_1^T (I_i - B_2) = r_1^T I_i - r_1^T B_2 \leftarrow \text{이제 } \alpha_i \text{를 colorize해보자!}$$

(즉, original color linear한)

o1 matting도 외울 수 있는가?

→ 이를 미구한 α 와 실제 α 차이가 얼마나 되는가? a^* , b^* 인가

→ cost function

$$J(\alpha, a, b) = \sum_{j=1}^N \left(\sum_{i \in \text{window}_j} (a_i - a_j^* I_i + b_j)^2 + \epsilon \|a_j\|^2 \right) \text{ 이 최소화되는 경우: } \alpha^*, L^*, \alpha \text{ 일 때!}$$

$$J(\alpha) = \alpha^T L \alpha$$

$$\min J(\alpha) = \frac{d}{d\alpha} J(\alpha) = 0 \quad \text{eigenvalue} \quad \text{sum of all pixel} \quad \text{make window} \quad \text{window line} \quad \text{all pixel}$$

$$\frac{d}{d\alpha} = 2L\alpha = 0 \Rightarrow L\alpha = 0 \quad \text{Null vector of } L \text{ solves Matting eqn}$$

?

• 편리한 방식은 eigen vector

→ 제작 과정 (비밀: $\alpha=0$, 허용: $\alpha=1$)

→ 이로 부터 정답이 되는 어려움

$$J(\alpha, a, b) = \sum_{j=1}^N \sum_{i \in \text{window}_j} (a_i - a_j^* I_i + b_j)^2 + \epsilon \|a_j\|^2 \quad \epsilon \text{ regularization}$$

window enforce coupling between pixels

$$\sum_{j=1}^N \left\| \begin{bmatrix} I_i^T & 1 \\ I_i^T & 1 \\ \vdots & \vdots \\ I_i^T & 1 \\ \hline I_i^T & 0 \end{bmatrix} \begin{bmatrix} a_j \\ b_j \end{bmatrix} - \begin{bmatrix} a_j^* \\ b_j^* \\ \vdots \\ a_j^* \\ b_j^* \end{bmatrix} \right\|^2$$

regularization

$$= \sum_{j=1}^N \|G_j \begin{bmatrix} a_j \\ b_j \end{bmatrix} - \begin{bmatrix} a_j^* \\ b_j^* \end{bmatrix}\|^2 + \text{정규화하는 데 사용되는 것}$$

depend on image colors

unknown (or Matic)

LS Problem

α 를 만드는 해로 때, optimal a^* , b^*

제작 과정, optimal

$$\begin{bmatrix} a_j \\ b_j \end{bmatrix}^* = (G_j^T G_j)^{-1} G_j^T d_j$$

pseudo inverse

제작 과정 a^* 와 b^* 는 α 의 fn

$$J(\alpha) = \sum_{j=1}^N \|G_j(G_j^T G_j)^{-1} G_j^T d_j - \alpha_j\|^2$$

$$= \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}^T \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}$$

all α 's in img

can be built from img colors

$$\therefore J(\alpha) = \alpha^T L \alpha \quad (\alpha: \text{matrix})$$

$$\min J(\alpha) \Rightarrow \frac{d}{d\alpha} J(\alpha) = 0$$

$$= 2L\alpha = 0 \Rightarrow L\alpha = 0$$

Eigen vector of L

⇒ null vector of L solves matting eqn

b. Augmented style loss with semantic segmentation

$$\mathcal{L}_{s+}^{\ell} = \sum_{c=1}^C \frac{1}{2N_{\ell,c}^2} \sum_{ij} (G_{\ell,c}[O] - G_{\ell,c}[S])_{ij}^2 \quad (3a)$$

$$F_{\ell,c}[O] = F_{\ell}[O]M_{\ell,c}[I] \quad F_{\ell,c}[S] = F_{\ell}[S]M_{\ell,c}[S] \quad (3b)$$

c : semantic segmentation mask 채널 수 $M_{l,c}$: layer l, channel c에 해당하는 segmentation mask, $F_{l,c}$: feature map

- 목적 : Gram matrix의 entire image에 대해 계산되어져, 방향성이 없으므로 (vectors up to an isometry) semantic context 변화를 적응할 능력이 부족하고 spill over가 발생하기 쉽다 → additional channel and augment neural style algorithm(즉, semantic segmentation 별로 스타일을 맞추겠다 = 하늘은 하늘끼리)
- channel을 추가하여 마치 mast처럼 하늘은 하늘, 땅은 땅끼리 semantic context 를 구분한다.(즉 하늘은 하늘끼리 gram matrix 결과값이 작다.)
- orphan semantic labels 방지 : input semantic label이 reference style image 의 label안에서 골라지도록 제약(lake와 sea를 똑같이 봄)

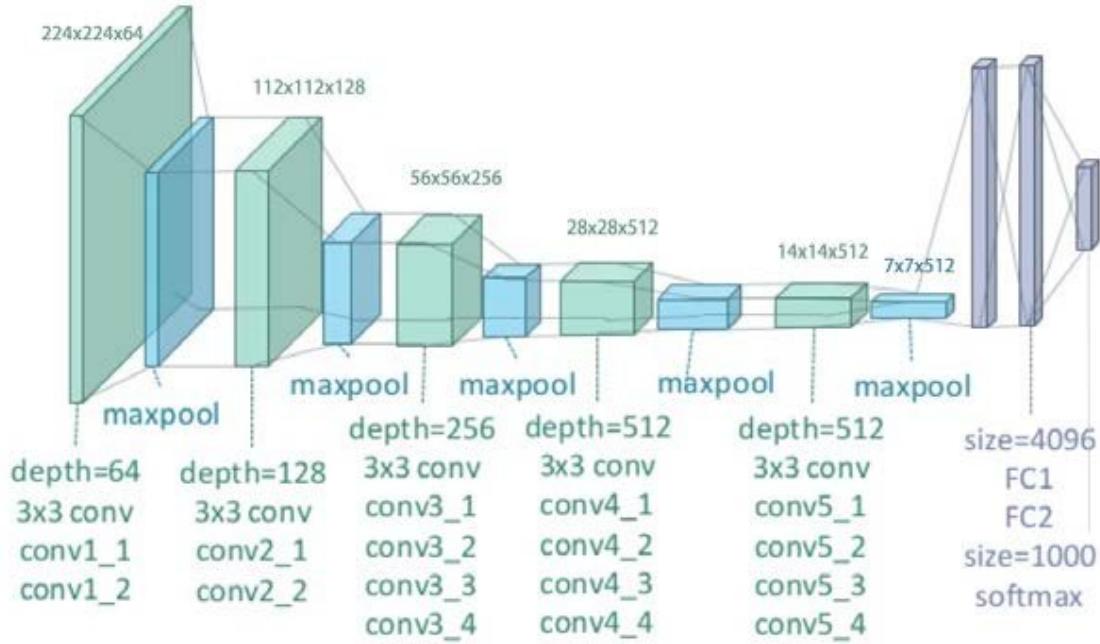
c. Our approach

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^L \alpha_l \mathcal{L}_c^l + \Gamma \sum_{\ell=1}^L \beta_{\ell} \mathcal{L}_{s+}^{\ell} + \lambda \mathcal{L}_m \quad (4)$$

\mathcal{L}_{s+}^l : augmented style loss, \mathcal{L}_m : photorealism regularization

3. Implementation

a. feature extractor : the pre-trained VGG-19



b. content representation : conv4_2($\alpha_l=1$, 나머지 layer는 0)

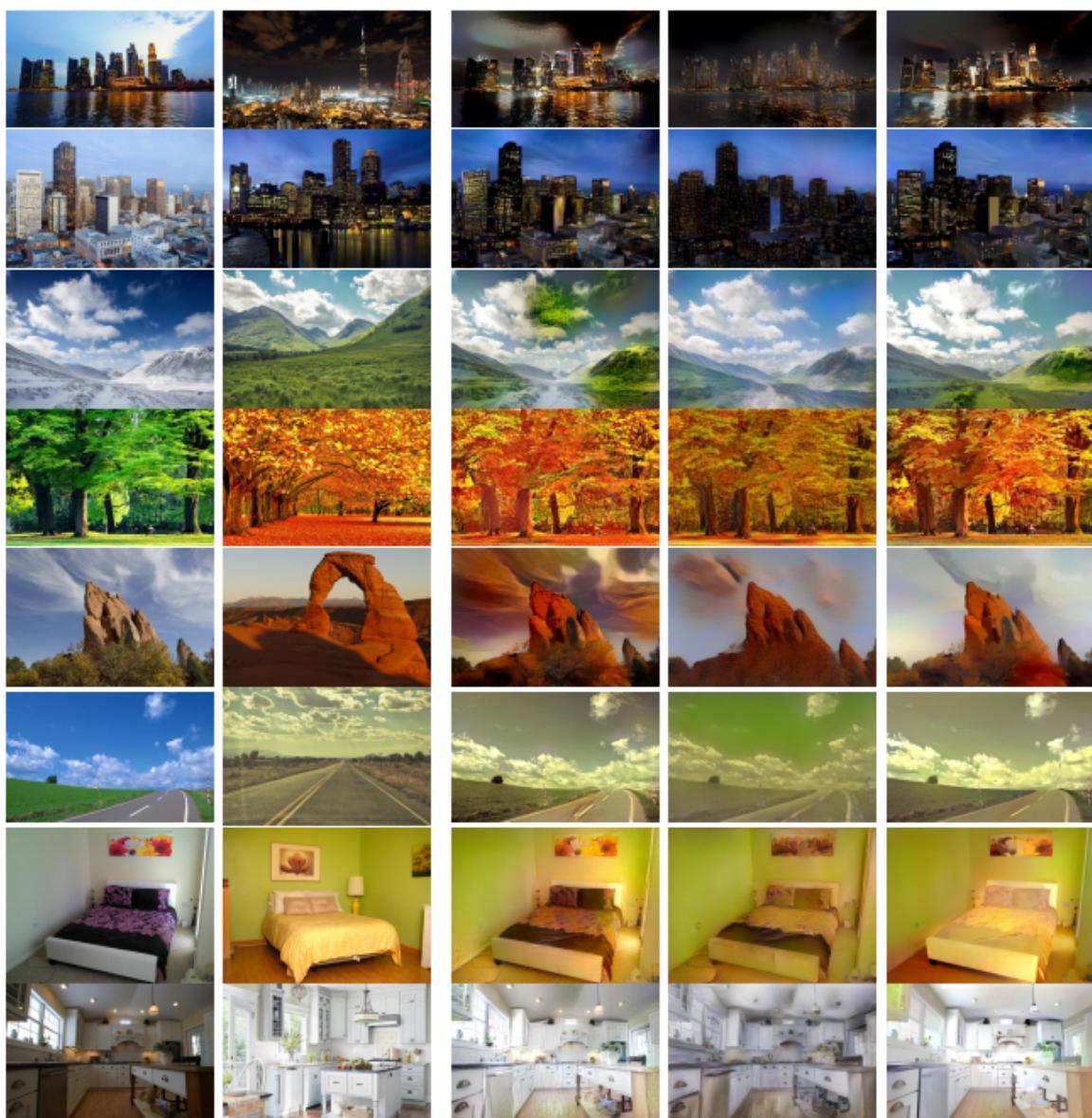
($\beta_l = 1/5$, 나머지 layer는 0)

layer preferences and parameters : $\Gamma = 10^2$, $\lambda = 10^4$

(λ 의 경우 너무 작으면 distortion 막기 힘들고, 너무 크면 style transfer되는 것을 막는다)

Result and Conclusion

▼ Fig 4(with Neural style, CNNMRP)



(a) Input image

(b) Reference style image

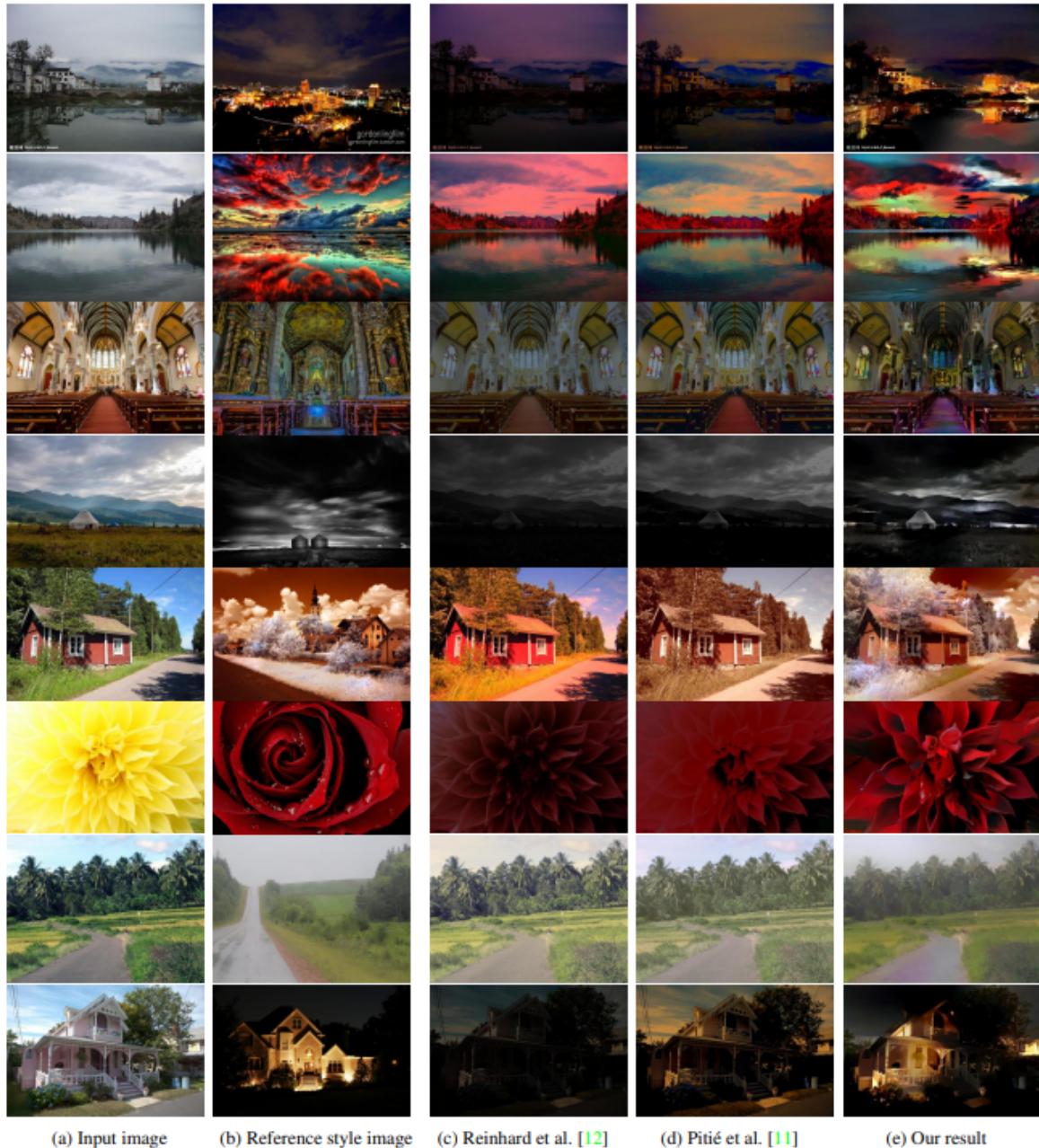
(c) Neural Style

(d) CNNMRF

(e) Our result

(c),(d)의 경우, painting-like distortion 발생

▼ Fig 5(with global style transfer methods)



(a) Input image

(b) Reference style image

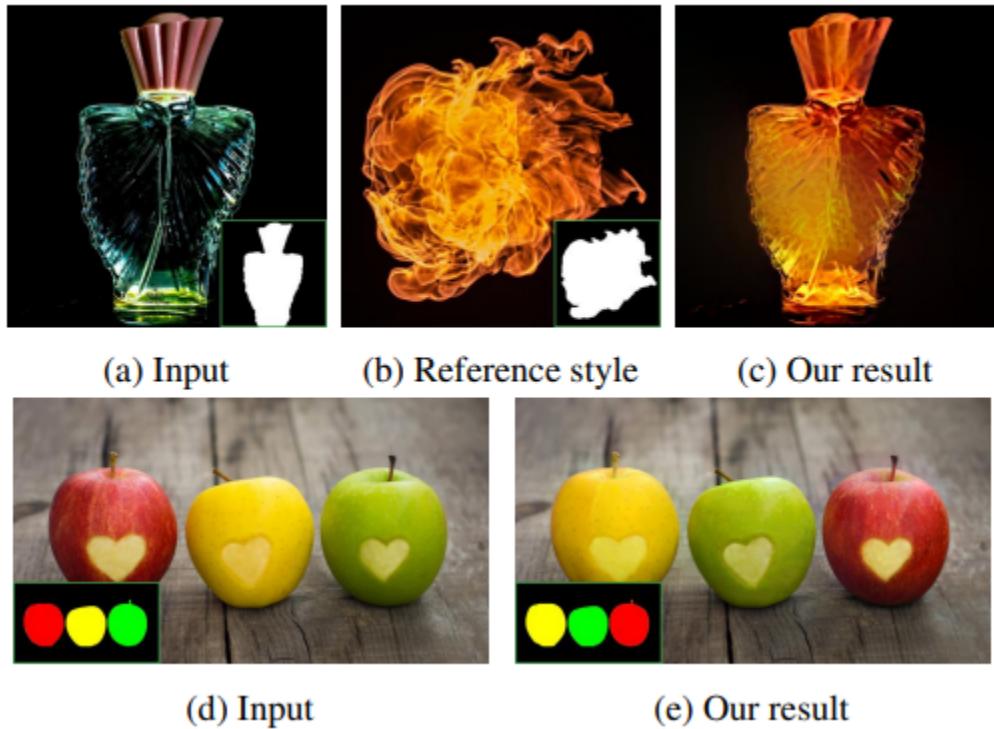
(c) Reinhard et al. [12]

(d) Pitié et al. [11]

(e) Our result

(c), (d)의 경우 distortion은 없으나, global color mapping 사용 → limit the faithfulness. 반대로 (e)의 경우 local, handle context-sensitive color change

▼ Fig 7(manual semantic mask : control 가능)



user가 semantic mask를 제공함으로써, transfer를 통제할 수 있다.

Fig 8의 경우 extreme mismatch로 semantic mask transfer의 실패 예시를 보여준다. → manual segmentation을 이용하여 fix 가능



We use the **Matting Laplacian** to constrain the transformation from the input to the output to be locally affine in colorspace. **Semantic segmentation** further drives more meaningful style transfer yielding satisfying photorealistic results in a broad variety of scenarios, including transfer of the time of day, weather, season, and artistic edits.

References

1. Regularized term : <https://dogfoottech.tistory.com/168>
 2. Laplacian : <https://www.cis.upenn.edu/~cis515/cis515-14-graphlap.pdf>
 3. A closed form solution to Natural Image Matting :
<https://ieeexplore.ieee.org/document/4359322>[Matting Laplacian 관련 논문,
Reference[9]]
- 수식관련 YouTube 설명: <https://youtu.be/Mvd93DdgqAY>