Style Transfer

논문 스터디

2019.02.19

발표자 박성현

DAVIAN
Data and Visual Analytics Lab

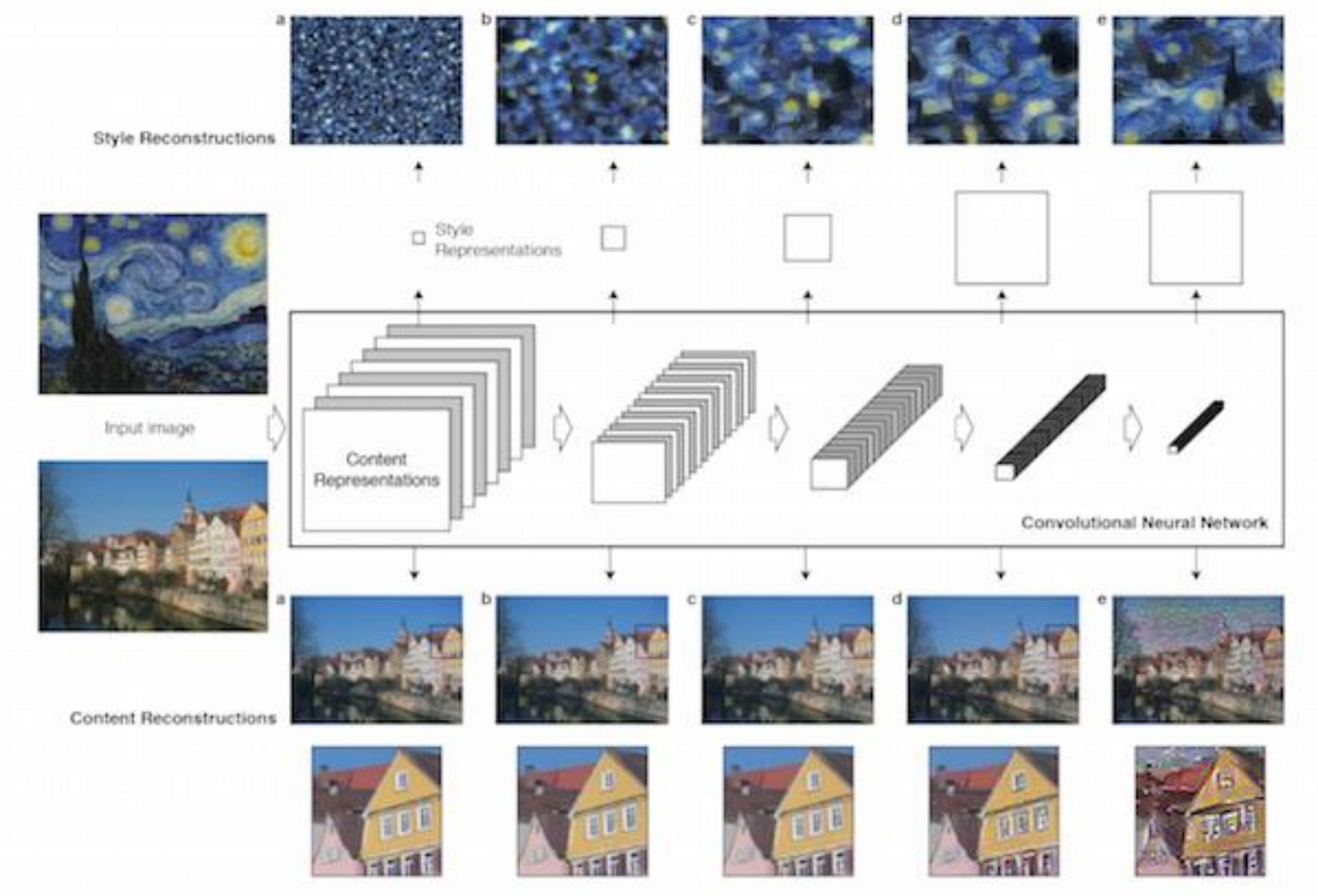
KOREA UNIVERSITY

1 Introduction

Style Transfer

1 Introduction

Style Transfer



1 Introduction

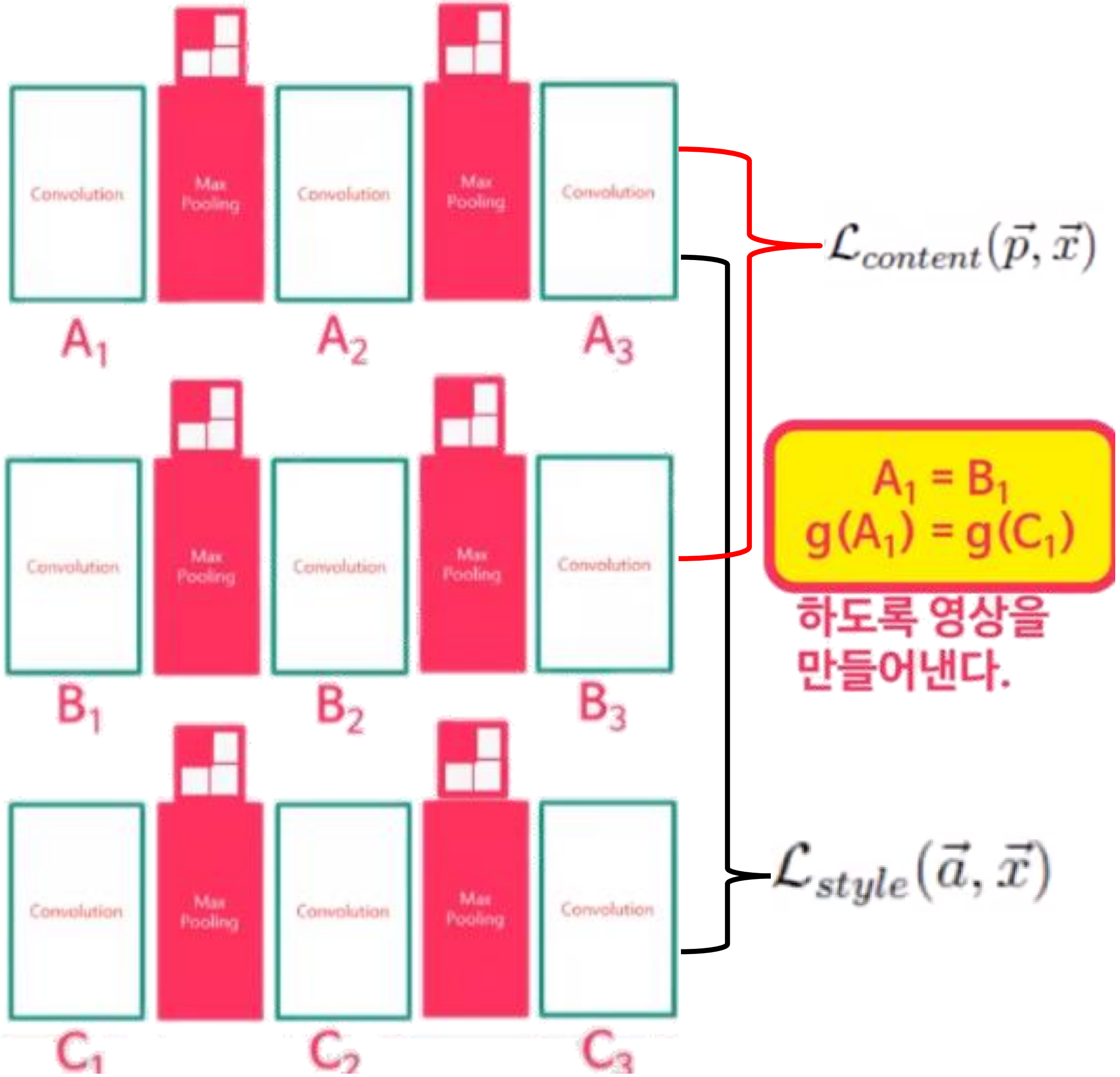
Style Transfer

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

만드는 영상

사진 영상

Art 영상



$A_1 = B_1$

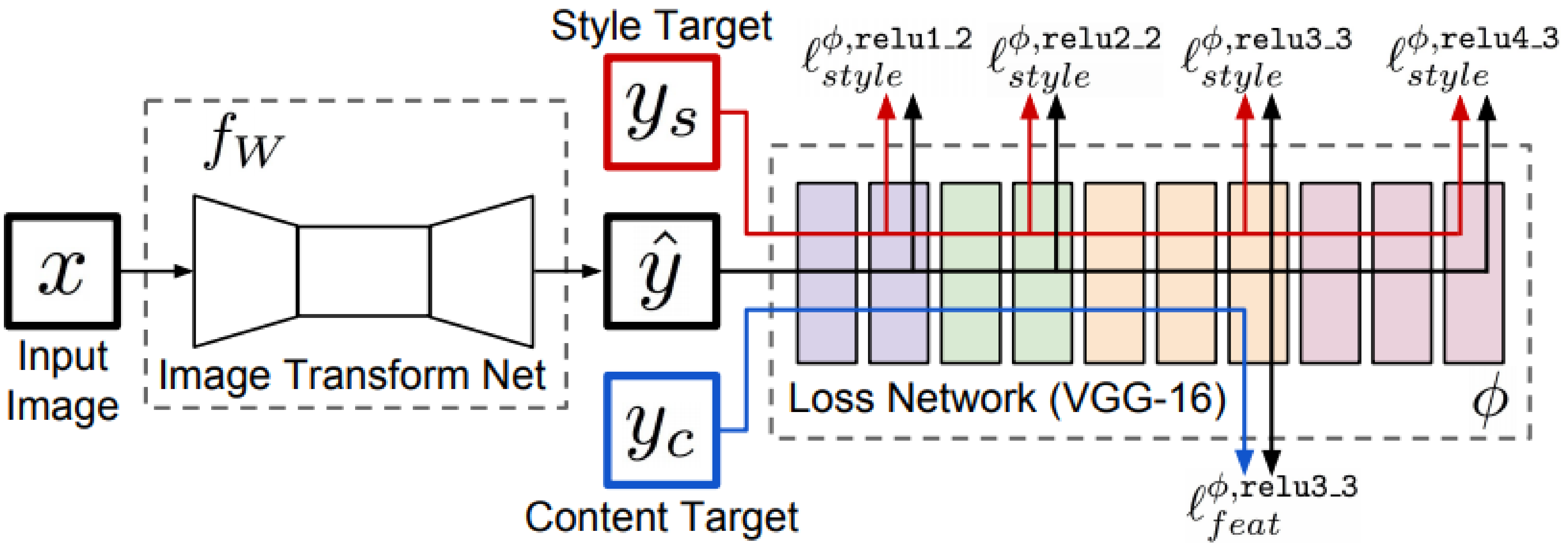
$g(A_1) = g(C_1)$

하도록 영상을 만들어낸다.

2

Fast Style Transfer

Perceptual Losses for Real-Time Style Transfer (ECCV 2016)



3 Ada-IN

Arbitrary Style Transfer with Adaptive Instance Normalization(ICCV 2017)

기존의 Style Transfer Method 문제점 존재

1. Gatys' Image style transfer using CNN(CVPR, 2016)은 임의의 style에 적용가능. 그러나 최적화가 굉장히 느림.
2. Ulyanov's Feed-forward synthesis of textures and stylized images(ICML, 2016) 등은 feed-forward 기법으로 빠른 학습이 가능하지만, single style로만 transfer 가능

→ AdaIN(**Ada**ptive **I**nstance **N**ormalization)을 이용하여 임의의 style에 대해 실시간으로 Style Transfer를 가능하게 하는 방법 제시

3 Ada-IN

Arbitrary Style Transfer with Adaptive Instance Normalization(ICCV 2017)

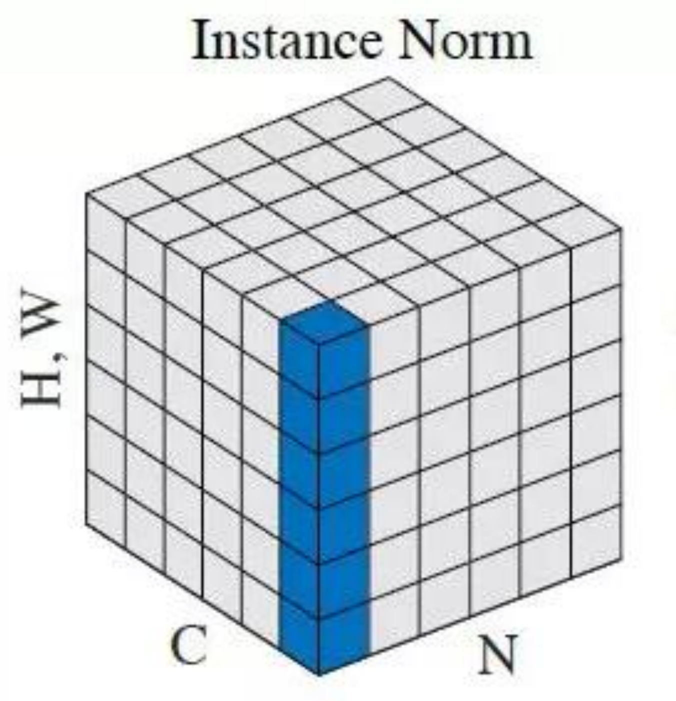
[Instance Normalization]

- Instance Norm은 각 channel과 sample별로 평균 및 표준편차 계산

$$\mathrm{IN}(x) = \gamma \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \beta$$

$$\mu_{nc}(x) = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} x_{nchw}$$

$$\sigma_{nc}(x) = \sqrt{\frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} (x_{nchw} - \mu_{nc}(x))^2 + \epsilon}$$



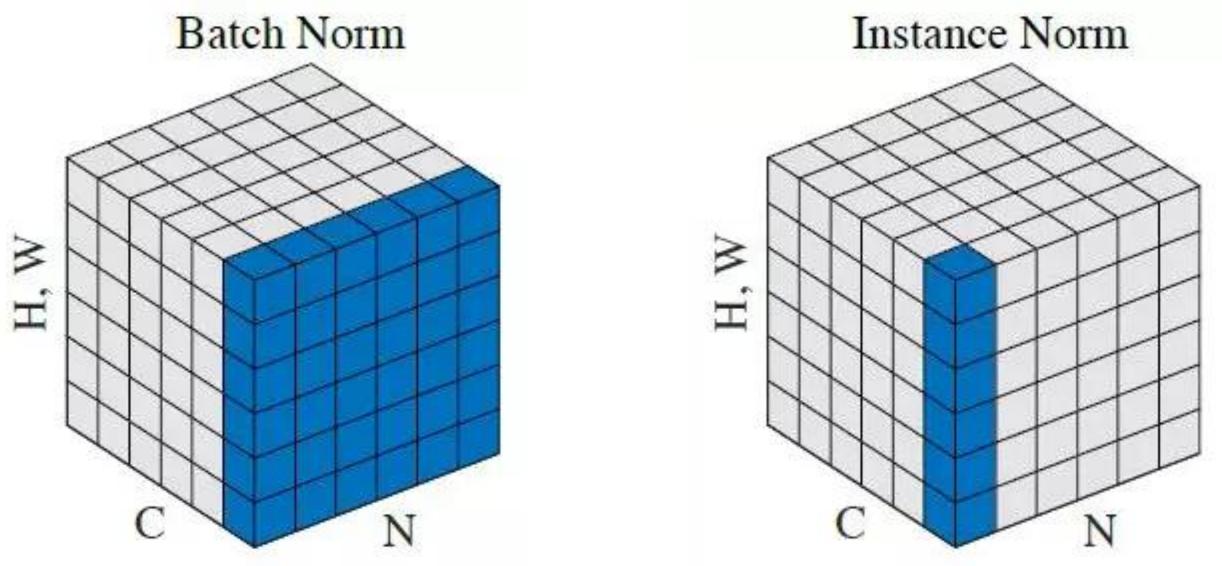
3 Ada-IN

Arbitrary Style Transfer with Adaptive Instance Normalization(ICCV 2017)

- BN : sample들이 single style 중심으로, 각각의 고유한 style을 유지한 상태로 정규화 진행

- IN : sample별로도 정규화 하기 때문에 모든 style을 하나의 target style로 정규화 할 수 있음

→ Instance Norm을 이용하면 정규화된 하나의 target style으로부터
새로운 style을 transfer할 수 있기 때문에 학습이 더욱 수월



3 Ada-IN

Arbitrary Style Transfer with Adaptive Instance Normalization(ICCV 2017)

[Conditional Instance Normalization]

- s는 Style Index로 Style Image마다 parameter γ^s, β^s를 학습하여, Style Transfer 진행
- Network는 똑같은 Convolutional parameters를 가지고 있으면서, 다른 affine parameters를 가지면, 다른 Style의 이미지를 생성하는 게 가능

$$\mathrm{CIN}(x; s) = \gamma^s \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \beta^s$$

3 Ada-IN

Arbitrary Style Transfer with Adaptive Instance Normalization(ICCV 2017)

[Adaptive Instance Normalization]

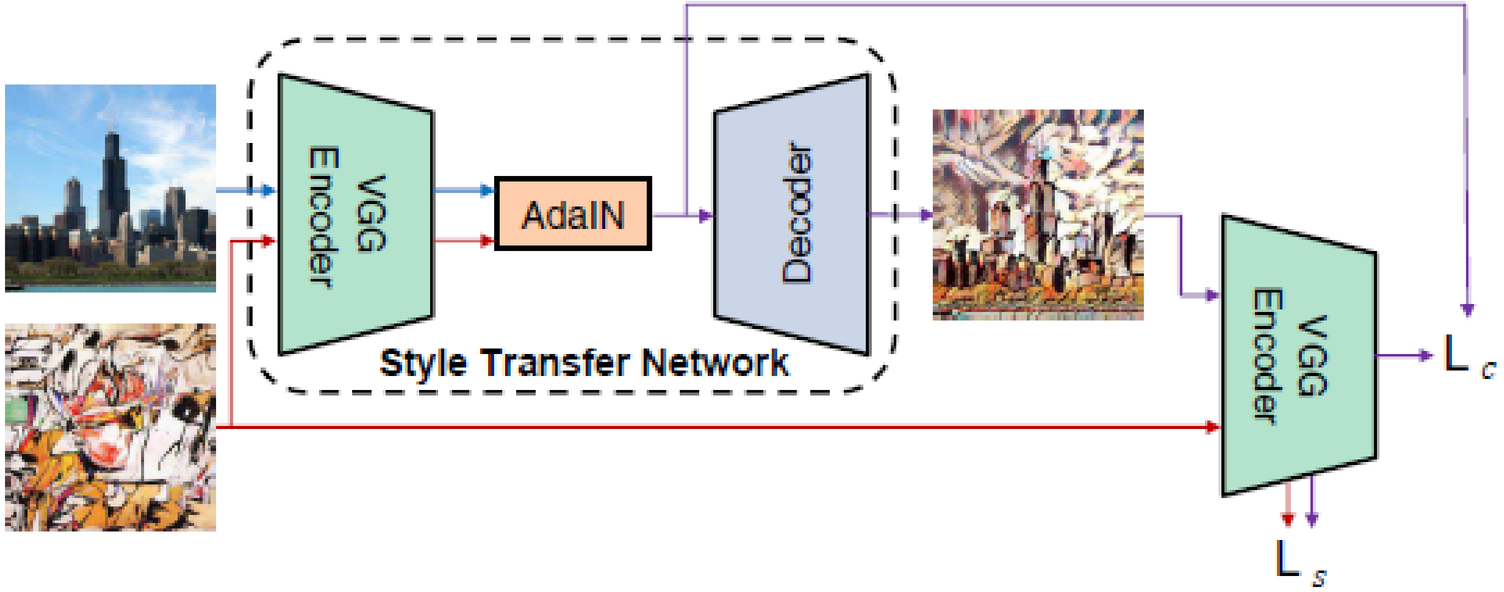
- Affine parameters를 통해 content input feature map x의 분포를 style input feature map y의 분포와 동일하게 함

→ 원하는 input y의 style에 대한 feature statistics(mean, variance)를 x에 transfer할 수 있음

$$\mathrm{AdaIN}(x, y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y)$$

3 Ada-IN

Arbitrary Style Transfer with Adaptive Instance Normalization(ICCV 2017)



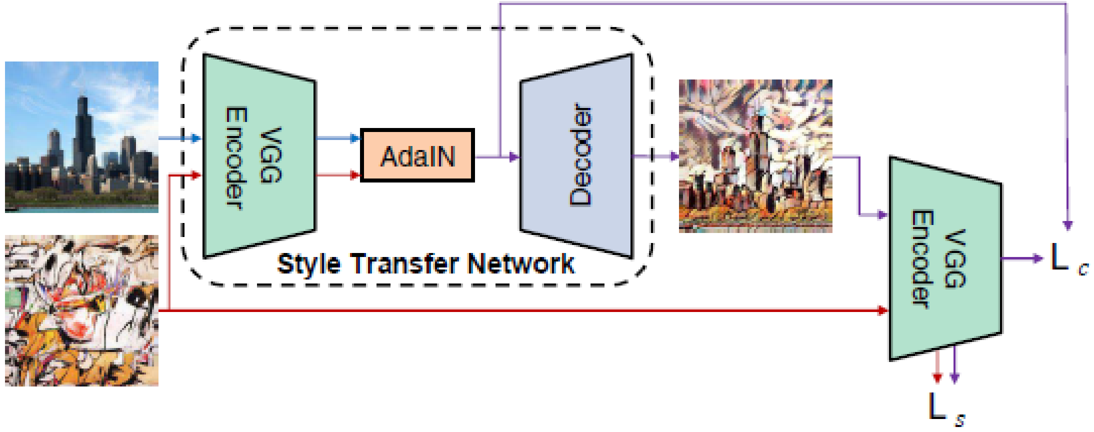
$$t = \mathrm{AdaIN}(f(c), f(s))$$

$$T(c, s) = g(t)$$

- f = Encoder
- t = Style transferred feature map
- g = Decoder
- T(c, s) = Style transferred image

3 Ada-IN

Arbitrary Style Transfer with Adaptive Instance Normalization(ICCV 2017)



$$\mathcal{L} = \mathcal{L}_c + \lambda \mathcal{L}_s$$

[전체 Loss]

$$\mathcal{L}_c = \| f(g(t)) - t \|_2$$

[Content Loss]

$$\mathcal{L}_s = \sum_{i=1}^{L} \| \mu(\phi_i(g(t))) - \mu(\phi_i(s)) \|_2 \; + \sum_{i=1}^{L} \| \sigma(\phi_i(g(t))) - \sigma(\phi_i(s)) \|_2$$

[Style Loss]

3 Ada-IN

Arbitrary Style Transfer with Adaptive Instance Normalization(ICCV 2017)

Method	Time (256px)	Time (512px)	# Styles
Gatys *et al.*	14.17 (14.19)	46.75 (46.79)	∞
Chen and Schmidt	0.171 (0.407)	3.214 (4.144)	∞
Ulyanov *et al.*	**0.011** (N/A)	**0.038** (N/A)	1
Dumoulin *et al.*	**0.011** (N/A)	**0.038** (N/A)	32
Ours	**0.018** (0.027)	**0.065** (0.098)	∞

3

Ada-IN

Arbitrary Style Transfer with Adaptive Instance Normalization(ICCV 2017)

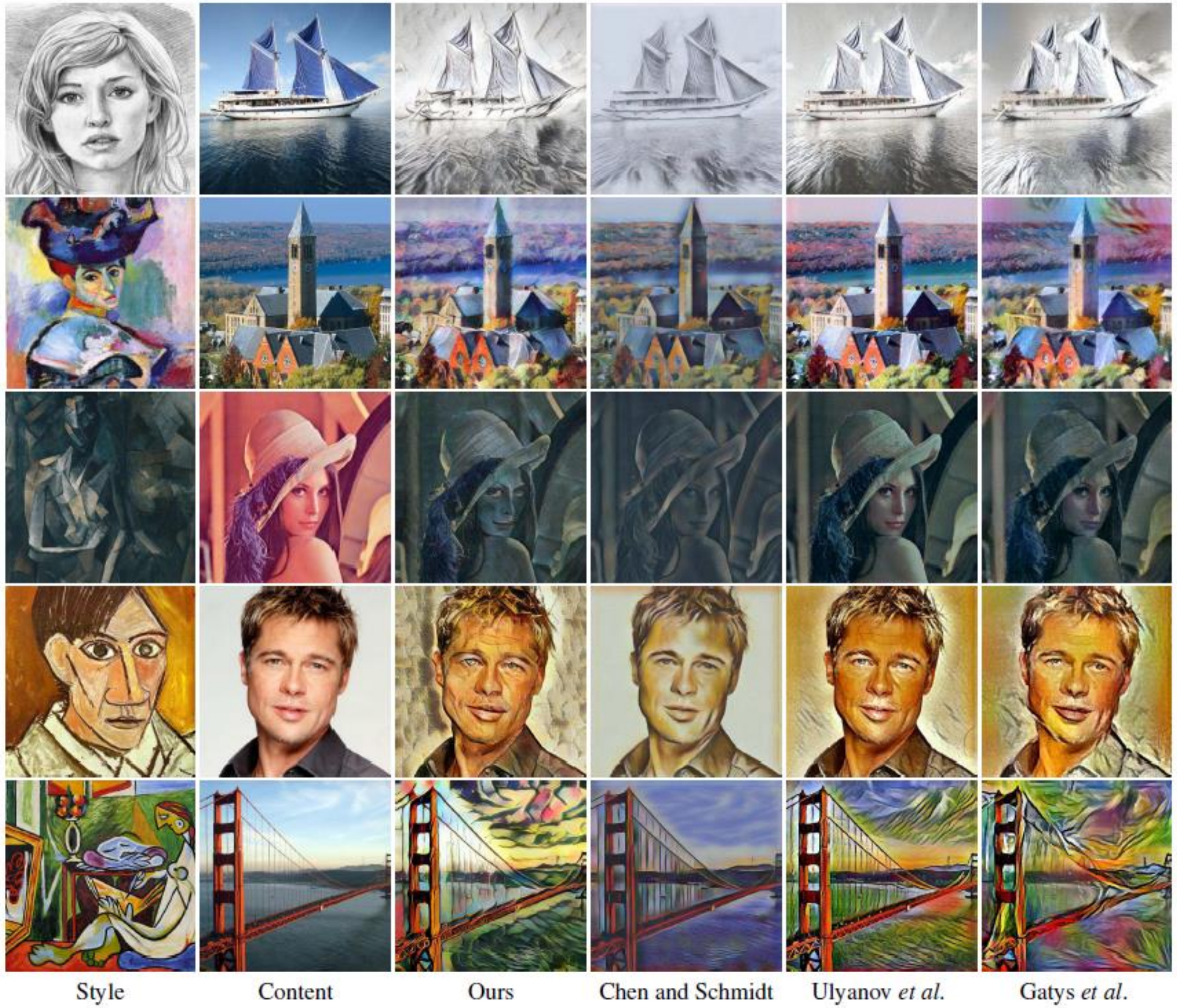


Figure 4. Example style transfer results. All the tested content and style images are never observed by our network during training.

4 Separating Style and Content

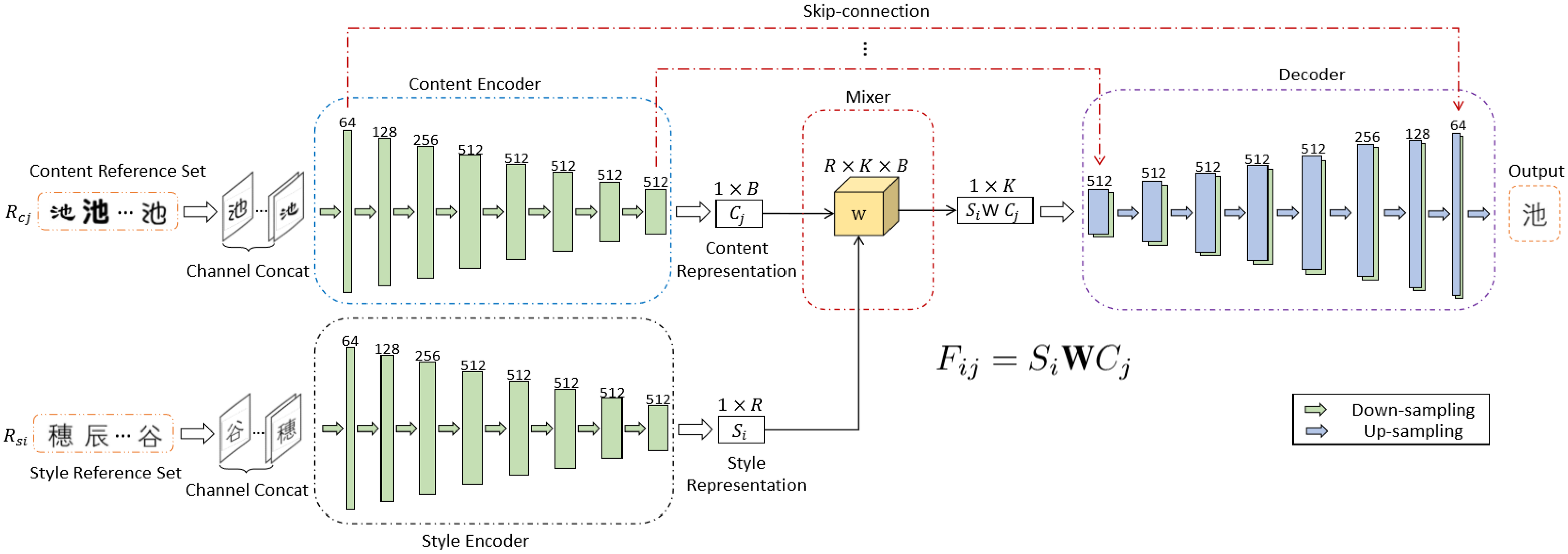
Separating Style and Content for Generalized Style Transfer (CVPR 2018)

Table 1. Comparison of *EMD* with existing methods.

Methods	Data format	Generalizable to new styles?	Requirements for new style transfer	What the model learned?
Pix2pix [10]	paired	The learned model can only transfer images to styles which appeared in the training set. For new styles, the model has to be retrained.	Retrain on a lot of training images for a source style and a target style.	The translation from a certain source style to a specific target style.
CoGAN [14]	unpaired			
CycleGAN [28]	unpaired			
Rewrite [1]	paired			
Zi-to-zi [2]	paired			
AEGN [16]	paired			
Perceptual [12]	unpaired		Retrain on many input content images and one style image.	Transformation among specific styles.
StyleBank [5]	unpaired			
Patch-based [6]	unpaired	The learned model can be generalized to new styles.	One or a small set of style/content reference images.	The swap of style/content feature maps.
AdaIn [9]	unpaired			The transferring of feature statistics.
EMD	triplet			The feature representation of style/content.

4 Separating Style and Content

Separating Style and Content for Generalized Style Transfer (CVPR 2018)



$$F_{ij} = S_i \mathbf{W} C_j$$

[Model]

4 Separating Style and Content

Separating Style and Content for Generalized Style Transfer (CVPR 2018)

$$\theta = \arg\min_{\theta} \sum_{I_{ij} \in \mathcal{D}_t} L(\hat{I}_{ij}, I_{ij} | \mathcal{R}_{S_i}, \mathcal{R}_{C_j})$$

[Training Objective]

$$L(\hat{I}_{ij}, I_{ij} | \mathcal{R}_{S_i}, \mathcal{R}_{C_j}) = W_{st}^{ij} \times W_b^{ij} \times ||\hat{I}_{ij} - I_{ij}||$$

[Loss]

$$W_{st}^{ij} = 1/N_b^{ij}$$

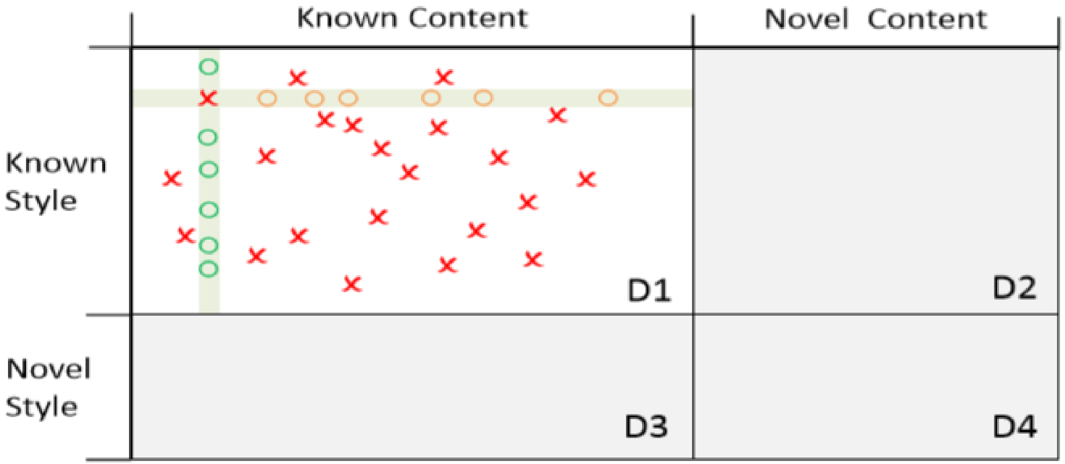
($N_b{}^{ij}$: Number of black pixels of target image I_{ij})

$$W_b^{ij} = \frac{exp(\text{mean}_{ij})}{\sum_{I_{ij} \in \mathcal{D}_t} exp(\text{mean}_{ij})}$$

(Mean value of black pixels)

4 Separating Style and Content

Separating Style and Content for Generalized Style Transfer (CVPR 2018)



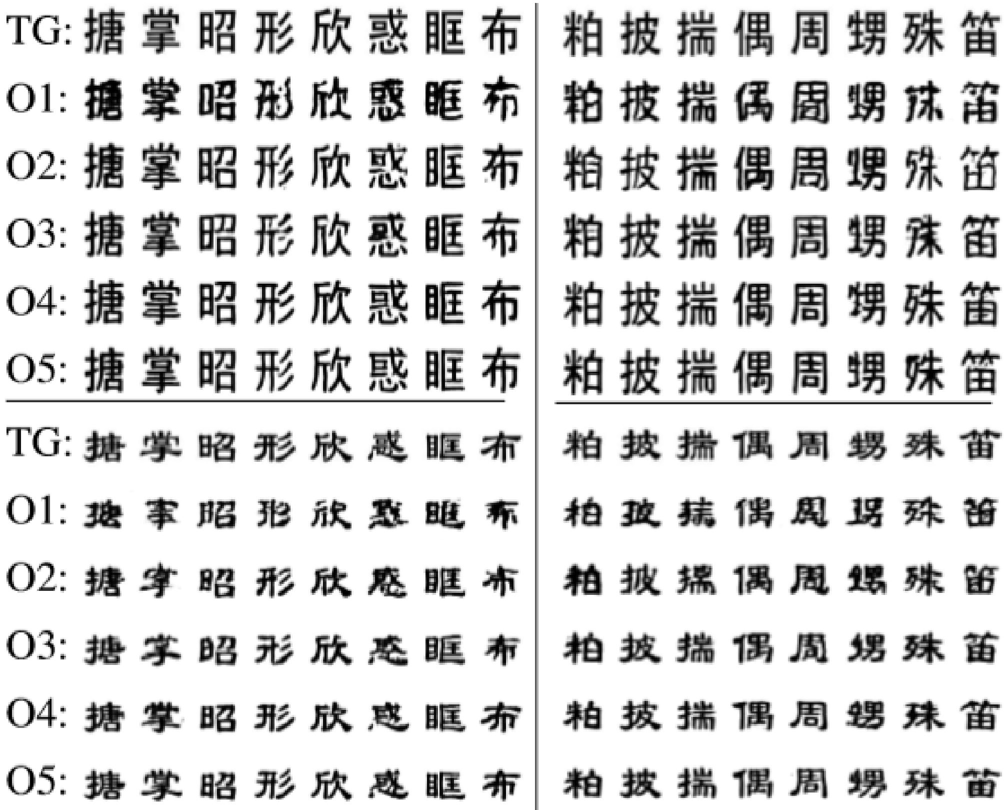


Figure 4. Generation results for D_1, D_2, D_3, D_4 (from upper left to lower right) with different training set size. TG: Target image, O1: Output for N_t=20k, O2: Output for N_t=50k, O3: Output for N_t=100k, O4: Output for N_t=300k, O5: Output for N_t=500k. In all cases, r=10.

4 Separating Style and Content

Separating Style and Content for Generalized Style Transfer (CVPR 2018)

	L1 loss	RMSE	PDAR
Pix2pix	0.0105	0.0202	0.17
AEGN	0.0112	0.0202	0.3001
Zitozi	0.0091	**0.0184**	0.1659
C-GAN	0.0112	0.02	0.3685
EMD	**0.0087**	**0.0184**	**0.1332**