

## A Multi-domain Style Transfer by Modified Generator of GAN

Geum-Boon Lee\*

\*Professor, SW Convergence Education Institute, Chosun University, Gwangju, Korea

### [Abstract]

In this paper, we propose a novel generator architecture for multi-domain style transfer method not an image to image translation, as a method of generating a styled image by transferring a style to the content image. A latent vector and Gaussian noises are added to the generator of GAN so that a high quality image is generated while considering the characteristics of various data distributions for each domain and preserving the features of the content data. With the generator architecture of the proposed GAN, networks are configured and presented so that the content image can learn the styles for each domain well, and it is applied to the domain composed of images of the four seasons to show the high resolution style transfer results.

▶ **Key words:** Multi domain, Latent vector, Generator architecture, Gaussian noise, Style transfer

### [요 약]

본 논문은 콘텐츠 이미지에 스타일 이미지를 입력 스타일이 적용된 이미지를 생성하고자 기존의 이미지 대 이미지 방법이 아닌 다중 도메인 스타일 트랜스퍼(style transfer) 방법을 적용한다. 도메인별로 데이터 분포에 대한 다양한 특성을 고려하고, 콘텐츠 데이터의 특징을 잘 보존하면서 높은 품질의 이미지가 생성되도록 잠재 벡터(latent vector)와 가우시안 노이즈를 추가하는 새로운 GAN의 생성자(generator) 아키텍처를 제안한다. 콘텐츠 이미지가 도메인별 스타일을 잘 학습할 수 있도록 네트워크를 구성하고 사계절 이미지로 구성된 도메인에 적용하여 고해상도의 스타일 트랜스퍼 결과를 보여준다.

▶ **주제어:** 멀티 도메인, 잠재 벡터, 생성자 아키텍처, 가우시안 노이즈, 스타일 트랜스퍼

- 
- First Author: Geum-Boon Lee, Corresponding Author: Geum-Boon Lee
  - \*Geum-Boon Lee (goldpalm@chosun.ac.kr), SW Convergence Education Institute, Chosun University
  - Received: 2022. 05. 12, Revised: 2022. 06. 24, Accepted: 2022. 07. 08.

## I. Introduction

콘텐츠 이미지를 다른 이미지의 스타일로 트랜스퍼(translation, transfer)하는 딥러닝 기법들이 최근 많은 발전을 이뤄왔다 [1-4]. 인간은 직관적으로 이미지의 스타일 및 이미지 내의 정보를 인식하는 것이 가능하지만 기계가 무감각으로 자동 인식을 가능케 하는 것은 매우 어려운 과제였다. 한 이미지의 콘텐츠를 보존하면서 다른 이미지의 스타일로 입히는 스타일 트랜스퍼 연구들은 주로 얼굴 이미지에 스타일(헤어, 수염, 안경 등)을 합성시키는 연구에서 뛰어난 결과를 보여주고 있다 [3-4].

이미지 간(image-to-image) 스타일 트랜스퍼는 크게 두 가지 방법이 제시되어왔다 [2, 5-8]. 먼저, 단일 순방향(feed-forward network)으로 콘텐츠 이미지와 스타일 이미지 간의 합성을 위해 훈련하는 방법이다. 이 방법은 두 이미지 간 스타일 변환으로 제한되므로 콘텐츠 이미지와 스타일 이미지 한 쌍으로 훈련 데이터를 얻는 것이 어렵고, 스타일 변환의 유연성과 다양한 특성을 다루는데 어려움이 있다. 또 다른 방법은 서로 다른 두 도메인 상의 영상이 주어졌을 때 한 도메인에서 다른 도메인으로 트랜스퍼하도록 적대적 학습 방법(adversarial learning method)을 사용하는 것으로 두 도메인 간의 매핑만을 고려한 방법이다 [1, 3, 4, 9-10]. 두 도메인 간의 매핑만 고려하므로 다른 도메인을 학습하고자 할 때는 새로운 네트워크를 형성해야 하는 확장성(scalability)의 한계를 갖게 된다. 이를 해결하고자 등장한 방법들도 다중 도메인의 매핑을 제시하지만 여전히 각 도메인에 대한 결정론적인 매핑을 학습하고 있으며, 도메인별 데이터 분포가 상이한 다중 모델 특성을 고려하지 못하고 있다. 각 도메인을 미리 정해진 라벨로 표시하고 one-hot-vector를 적용함으로써 소스 이미지와 라벨링된 도메인에 해당하는 동일한 output을 생성하므로 여러 스타일이 적용되기 어려운 스타일 다양성(style diversity)의 한계를 보여주고 있다.

본 논문은 이러한 스타일 트랜스퍼(style transfer)에 있어서 확장성과 다양성의 한계를 극복하고, 다양한 스타일의 이미지들을 생성하기 위해 이미지 데이터 분포에 대한 특성을 반영하고 데이터 특징을 잘 보존하면서 낮은 차원으로 효율적으로 압축하는 것을 학습하는 VAE(Variational Auto Encoder)의 매니폴드 공간인 latent vector를 도입하고 disentangled 되도록 [11-13], GAN 입력으로 랜덤 노이즈를 추가하는 매핑 네트워크를 적용하여 콘텐츠가 도메인별로 스타일을 학습할 수 있도록 새로운 딥러닝 모델을 구축하여 고해상도의 스타일 변환 이미지들을 생성하는 방법을 제시한다.

## II. Related Works

### 1. Style transfer methods

스타일 합성은 콘텐츠 이미지와 스타일 이미지가 주어졌을 때, 이미지의 주요 형태는 콘텐츠 이미지로 유지되고 그 위에 스타일 이미지의 스타일을 입히는 것을 말한다. 각 콘텐츠 이미지와 스타일 이미지에 대한 DNN 합성곱 레이어(convolution layer)로부터 특징 맵(feature maps)을 보존하고, 이들의 합성의 결과인 새로운 이미지의 특징 맵이 콘텐츠와 스타일에 대해 유사한 특성을 갖도록 이미지를 최적화하여 생성한다. Gatys et al. 은 콘텐츠 손실(content loss)과 스타일 손실(style loss)이 최소화하도록 이미지의 출력 픽셀들을 최적화함으로써 콘텐츠와 스타일 각각의 특성을 담고 있는 새로운 영상을 만들었다 [14].

히스토그램 손실(histogram losses)을 사용한 스타일 합성 방법들은 텍스처를 합성할 때 생기는 문제들을 보완해 줄 수 있는 오류 항(error term)인 히스토그램 손실을 추가하고, 이미지 영역을 나누어서 스타일을 합성하는 국부적 스타일 합성도 다루고 있다 [15].

사진-스타일 합성 방법인 Deep photo style transfer는 실제 사진을 스타일 이미지로 합성할 때 세부적인 영역에 왜곡이 일어나는데, 세부적인 형태(shape)와 특징(features)을 보존하도록 손실 항(loss term)을 추가하여 이를 해결하고 있다 [16].

### 2. latent vector for style transfer

이미지 생성의 핵심은 잠재 벡터(latent vector)이다. 이 잠재 벡터는 보통 랜덤한 확률분포에서 얻는다. 제일 많이 사용되는 것이 무상관의 독립항등분포(iid) 상태의 다변량 가우시안(multi variate Gaussian)을 사용한다.

#### 2.1 Latent Vector

VAE(Variational Autoencoder)에서 입력 데이터가 들어오면 인코딩 과정을 거쳐 압축된 표상(compressed representation)을 만드는데 이것이 latent vector  $z$ 가 된다. 인코더(encoder)는 입력 데이터를 추상화하여 잠재적인 특징을 추출하여  $z$ 를 만들고 디코더(decoder)는  $z$ 를 바탕으로 이미지 데이터를 복원하는 역할을 한다.  $z$ 는 평균과 분산을 모수로 하는 정규분포로 매니폴드(manifold)를 가정한다. 매니폴드(manifold)는 고차원의 공간의 데이터들을 잘 표상(representative)할 수 있는 저차원의 공간을 말한다. 원래 데이터 공간에서 얽혀 있는(entangled) 데이터들이 매니폴드에서 풀린다면(disentangled) 해석

이 쉽고, 이미지 생성 작업이 쉬워진다. 잠재 벡터  $z$ 의 평균과 분산을 조금씩 바꾸면 매니폴드의 연속 경로 (continuous path)를 따라 한 이미지에서 다른 이미지들로 자연스럽게 변하는 모습을 얻을 수 있다.

## 2.2 Latent Space

GAN(Generative Adversarial Network)은 실제 이미지의 확률 분포를 잘 근사하여 같은 확률 분포를 갖는 이미지를 생성하도록 모델을 학습시킨다. 생성자(Generator)와 판별자(Discriminator)의 두 확률 분포 간의 차이 (divergence)를 최소화하도록 하여 서로 적대적으로 학습하면서 목표 이미지(target image)를 생성한다. 이 과정에서 생성자는 저 차원의 랜덤 샘플(random sample)을 고 차원에 매핑하도록 학습한다. 만약 생성자가 훈련 데이터셋의 분포를 잘 학습했다면 그 데이터 분포로부터 추출된 샘플은 GAN의 생성자로부터 나온 결과와 매우 유사하다고 할 수 있다. 그러나 GAN은 생성자의 입력으로 들어가는 노이즈 벡터  $z$ 에 아무런 제약을 주지 않기 때문에 생성 모델이 심하게 얽혀 있는 방식으로 표상(representation)을 학습할 여지가 있다. 따라서 잠재 공간(latent space)에 표상이 없는 포인트로부터 이미지를 생성한다면 와핑(warping)되거나 왜곡(distorted) 될 수 있다.

## 3. Normalization for style transfer

### 3.1 Adaptive instance normalization

특징 공간(feature space) 내에서 콘텐츠 이미지의 통계적 특적인 평균과 분산을 스타일 이미지의 특징에 대해 정렬해 스타일을 입히는 방식이다 [17].

$$AdaIN(x, y) = \sigma(y) \left( \frac{x - \mu(x)}{\sigma(x)} \right) + \mu(y) \quad (1)$$

식(1)에서 콘텐츠 입력  $x$ 와 스타일 입력  $y$ 가 있을 때  $x$ 의 분포를 자신의 평균과 분산으로 정규화한 후 정규화된 콘텐츠 정보를 스타일 정보를 이용하여 선형 변환(affine transformation)하기 위해  $y$ 의 평균과 분산을 적용한다. 선형 변환하는 파라미터들은 학습되지 않고 스타일 정보가 들어올 때마다 계산된다.

### 3.2 Batch Normalization

은닉층(hidden layer)의 활성 함수마다 입력값들의 분포가 계속 바뀌는 내부 공변량 변화(internal covariance shift)로 인한 불안정한 훈련(training) 과정을 전체적으로 안정화하여 학습 속도를 가속화하기 위한 방법으로 연구되었다. 내부 공변량 변화 현상은 은닉층의 출력에 의한

활성 분포의 변화 때문에 그 다음 층의 입력 분포가 달라지는 현상이다. 이를 해결하기 위해 네트워크를 학습시킬 때 미니 배치 단위로 데이터를 가져와서 학습을 시킨다. 식(3), (4)와 같이 각 특징별로 평균과 표준편차를 구하여 정규화하고, 식(2)와 같이 스케일 팩터(scale factor)인 감마와 시프트 팩터(shift factor)인 베타를 학습해 스타일이 적용된 적절한 분포를 갖도록 한다 [18-19].

$$BN(x) = \gamma \left( \frac{x - \mu(x)}{\sigma(x)} \right) + \beta \quad (2)$$

$$\mu_c(x) = \frac{1}{NHW} \sum_{n=1}^N \sum_{h=1}^H \sum_{w=1}^W (x_{nchw}) \quad (3)$$

$$\sigma_c(x) = \sqrt{\frac{1}{NHW} \sum_{n=1}^N \sum_{h=1}^H \sum_{w=1}^W (x_{nchw} - \mu_c(x))^2 + \epsilon} \quad (4)$$

### 3.3 Instance Normalization

IN(instance normalization)은 BN과 같이 감마와 베타를 학습시키는 파라미터를 가진다.

$$\in(x) = \gamma \left( \frac{x - \mu(x)}{\sigma(x)} \right) + \beta \quad (5)$$

$$\mu_{nc}(x) = \frac{1}{WH} \sum_{w=1}^W \sum_{h=1}^H (x_{ncwh}) \quad (6)$$

$$\sigma_{nc}(x) = \sqrt{\frac{1}{WH} \sum_{w=1}^W \sum_{h=1}^H (x_{ncwh} - \mu_{nc}(x))^2 + \epsilon} \quad (7)$$

식(5-7)은 BN과 동일하지만 평균과 표준편차에서 배치 축(batch axis)을 나타내는  $N$ 과 채널축(channel axis)  $C$ 에 대해서도 정규화한다 [20-21]. 이렇게 함으로써 BN은 배치와 추론(inference) 시에 사용하는 통계가 다르지만, IN은 같게 된다.

## III. A Proposed Style Transfer Model

이미지는 의미적으로 중요한 특징들로 표현될 수 있다. 데이터가 고차원으로 갈수록 이런 특징들을 학습하기가 매우 어렵고 표상이 복잡하게 얽혀있지 않은 방식(disentangled)으로 학습될 가능성은 낮다. 학습 모델은 무감독 학습으로 데이터의 중요 특징들을 보여줄 수 있어야 한다. 복잡하게 얽히지 않고 풀어져 있다는 것

(disentanglement)은 이미지 생성 모델에서 잠재 공간이 선형적인 구조를 갖게 되어 하나의 팩터를 움직였을 때 하나의 특성이 바뀌도록 만들 수 있다는 것이다 [22-24]. 그런데 GAN과 같은 생성 모델은 랜덤 노이즈로부터 출발해서 우리가 원하는 스타일 이미지 분포로 변환되도록 매핑 과정을 학습하면서 노이즈에 조금씩 변화를 주었는데 스타일 이미지는 급격하게 바뀔 수 있다. 한 차원에서 약간의 변화된 값을 부여하면서 약간의 다른 이미지를 원했으나 전혀 다른 이미지로 변환되는 일이 벌어지게 되는 것이다. 그리고 GAN은 하나의 스케일에서의 하나의 이미지밖에 생성할 수 없는 제약이 있다. 따라서 본 논문은 다중 스케일에서 여러 스타일 이미지를 만들어 낼 수 있도록 GAN의 생성자의 구조를 변형하여 다중 이미지 샘플링이 가능하도록 하였다.

### 1. A Generator for Style Transfer

$X$ 를 이미지의 집합,  $T$ 를 도메인들의 집합이라 할 때, 생성자  $G$ 는 임의의 도메인  $t$ 에 대응하는 다양한 스타일들이 합성된 이미지들을 생성한다.

입력 이미지  $x$ 를 매핑 시키기 위한 네트워크  $F$ 와 스타일 인코더  $E$ 로부터 나온 스타일 벡터  $s$ 를 적용하여  $G(x, s)$ 로 변경한다.  $G$ 에  $s$ 를 적용시키기 위해 AdaIN을 적용한다. 매핑 네트워크  $F$ 는 도메인  $t$ 와 잠재 벡터  $z$ 가 입력되면 스타일 코드  $s$ 를 만들 수 있다.  $F$ 는  $z$ 와  $t$ 를 샘플링하여 다양한 스타일 코드들을 생성한다.

$$s = F_t(z) \tag{8}$$

스타일 네트워크에서는 스타일 참조 이미지  $x_{ref}$ 에 대응되는 도메인  $t$ 에 대해 스타일 코드를 추출한다. 생성자  $G$ 는 콘텐츠 이미지  $x$ 가 스타일 참조 이미지  $x_{ref}$ 의 스타일을 반영하여 스타일이 합성한 이미지를 생성한다.

이때 판별자  $D$ 는 해당 도메인  $t$ 의 실제 이미지인지,  $G(x, s)$ 로 생성된 가짜 이미지인지를 판별하게 된다. 매핑 네트워크의 잠재 벡터  $z$ 와 타겟 도메인  $\tilde{t}$ 를 랜덤하게 샘플링하여 목표로 하는 스타일 이미지  $\tilde{s}$ 를 생성한다.

$$\tilde{s} = F_{\tilde{t}}(z) \tag{9}$$

생성자  $G$ 는  $\tilde{s}$ 를 통해 타겟 도메인  $\tilde{t}$ 와 구분되지 않는 이미지를 생성하는 것을 목표로 한다. 이렇게 생성자  $G$ 와 판별자  $D$ 는 적대적으로 학습하며 스타일 이미지를 생성하는데, 이때 적대적 손실(adversarial loss)은 다음과 같다.

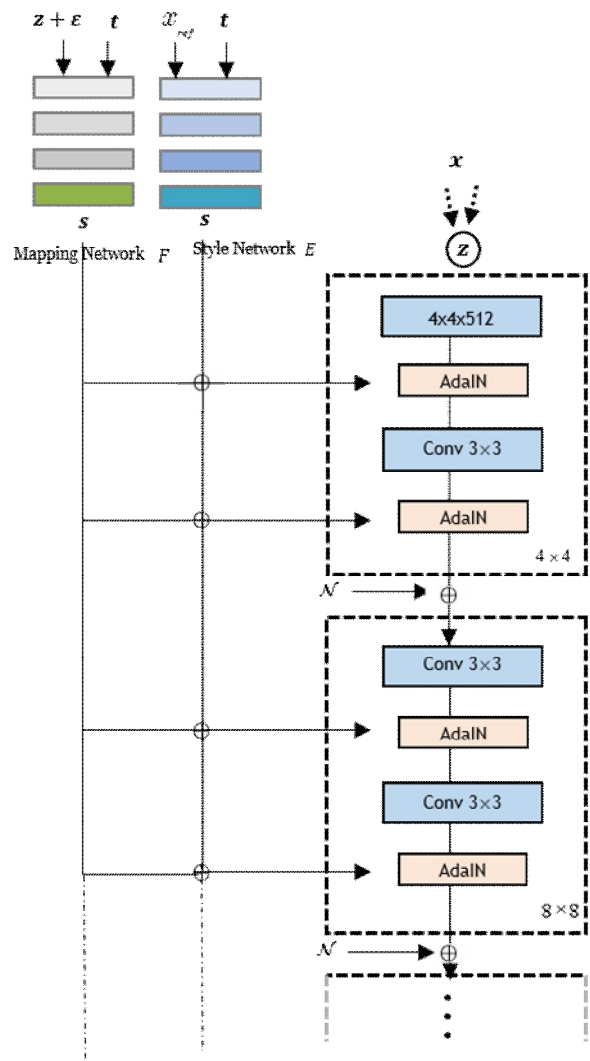


Fig. 1. A Modified Generator for style transfer

$$L_{adv} = E_{x,t}[\log D_t(x)] + E_{x,\tilde{t},z}[\log(1 - D_{\tilde{t}}(G(x, \tilde{s})))] \tag{10}$$

콘텐츠 이미지의 중요한 속성들을 보존하면서 타겟 스타일로 변환하도록 주기 일관성 손실(cycle consistency loss)을 적용할 수 있다. GAN의 생성자는 각 도메인을 해당 도메인으로 매핑하도록 훈련된다. 그러나 매핑에 있어 하나의 도메인만을 출력하는 모드 붕괴(mode collapse) 문제가 발생할 수 있다. 따라서 두 도메인간 양방향 매핑을 적용하고, 이 때 발생하는 주기 일관성 손실을 계산한다.

$$C_{loss} = E_{x,t,\tilde{t}}[\|x - G(G(x,t), t')\|_1] \tag{9}$$

실제 이미지  $x$ 를 새로운 스타일로 바꾼 이미지  $G(x,t)$ 를 원본 스타일로 다시 바꿔  $G(G(x,t), t')$ 를 적용함으로써 원본 이미지와의 차이를 비교하여 원본 이미지의 특징을 유지할 수 있도록 하고, 손실로  $L_1$  norm을 계산한다.

## 2. A Proposed Style Transfer Model

본 논문은 이미지 대 이미지 스타일 합성이 아닌 다중 도메인 스타일 변환 방식에 관한 것으로 기존의 얼굴과 헤어 중심 스타일 변환에 대한 것이 아닌 사계절 이미지에 대한 스타일 변환을 하기 위해 생성자의 아키텍처를 설계하고 이를 적용한다.

생성자는 AdaIN에 스타일  $s$ 를 임베딩시키는 네트워크와 VAE의 인코더를 도입한다. VAE는 주어진 데이터를 잘 설명할 수 있는 확률분포를 가정하고 적절한 확률분포에 대한 파라미터를 무감독으로 학습한다. 매니폴드를 가정하므로 일련의 확률분포(분산과 평균)에 따른 샘플링으로 자연스런 스타일 이미지들을 얻을 수 있다. 그러나 잠재 벡터 집합  $Z$ 를 고정 확률 분포로 가정하고 훈련 데이터의 확률 분포를 학습시키므로 왜곡된 스타일 이미지를 생성할 수 있다. 이 문제를 해결하기 위해 가우시안 노이즈를 추가한다. 가우시안 노이즈는 잠재 벡터를 유연하게 만들어 콘텐츠 이미지에 스타일 도메인이 잘 적용될 수 있도록 한다.

Fig. 1의 AdaIN에 스타일을 적용할 때 레이어별로 다른 스타일을 적용 시켜야 하는데, 이전 레이어의 결과에 의존한 상태로 다음 레이어에 스타일을 적용하게 되므로 목표로 하는 스타일 이미지를 얻을 수 없다. 그러므로 제안하는 생성자 구조는 Fig. 1과 같이 하나의 스타일을 적용한 후 다음 레이어에 노이즈를 추가하여 훈련 데이터 분포를 독립항등 분포로 만들고, 다시 스타일을 적용하는 구조로 되어있다.

## IV. Experiment Result

본 논문에서 실험에서 사용한 콘텐츠 및 스타일 트랜스퍼 이미지의 크기는  $512 \times 512$ , 훈련에 사용된 학습률은 0.0001, 배치 크기는 4, Adam optimizer  $\beta_1 = 0.500$ ,  $\beta_2 = 0.999$ 를 적용했다. Fig. 2와 같이 사계절의 이미지 데이터셋 3690개를 사용하였다. 그리고 모든 실험은 Nvidia GeForce RTX 2080 Ti GPUs and Intel® Core™ i9-9900KF CPU에서 사용되었다.

다중 도메인 스타일 변환에 사용한 기존의 방법[1]은 성별에 따른 도메인별 생성 모델로 헤어 색상, 안경 등 해당 스타일에 대해 각 도메인에 대해 학습하여 새로운 이미지를 합성한다. 도메인에 국한된 학습을 하므로 도메인의 확률 통계에 민감하다. 고차원의 데이터 분포 상에서 밀집(dense)하지 않고 흩어져 있는(sparse)한 데이터들은 훈련이 잘되지 않는다. 이것은 도메인의 평균과 분산으로부터 벗어난 데이터들이 전체 성능에 영향을 미치게 되는 것이

다. 수집한 사계절의 데이터셋은 봄, 여름, 가을, 겨울의 도메인으로 분류되어 도메인별 훈련을 하였으나 얼굴 데이터셋이 성공한 것만큼 학습이 되지 않았다. 학습이 전혀되지 않았으며 모드 붕괴도 일어났다. 따라서 본 논문은 GAN의 생성자를 수정하여 도메인 데이터 분포를 따르면서도 좀 더 유연한 학습을 할 수 있는 생성자를 만들고자 하였다.

스타일 트랜스퍼 방법 [2]은 스케일별 생성 모델이다. 학습된 분포에서 각 스타일에 대한 샘플을 그려나가는 방법으로 스케일별로 새로운 스타일 이미지를 생성한다. 네트워크 스케일을  $4 \times 4$ ,  $8 \times 8$ , ...,  $256 \times 256$ ,  $512 \times 512$ 로 확장해 가면서 고해상도의 이미지를 생성한다. 채널 지역화로 2-3개 채널들이 특정 스타일에 관여한다. 그러므로 스케일별 이미지의 특정 스타일에만 영향을 미칠 수 있다.

본 논문은 생성기 구조에 잠재 벡터와 노이즈를 함께 고려하여 사계절 데이터셋에 적용하였다. Fig 3는 콘텐츠 이미지에 스타일이 적용되어가는 과정을 보여준다. 참조되는 타겟 도메인(사계절) 이미지에 따라 콘텐츠 이미지의 특성이 바뀌게 되고, 훈련을 증가할수록 타겟 도메인의 스타일을 생성하게 된다.



(a) Spring



(b) Summer



(c) Autumn



(d) Winter

Fig. 2. Four Seasons Datasets



Fig. 2는 각 계절이 보여주는 고유한 의미있는 특징들이 있으므로 이러한 동일한 속성(형태, 컬러 등)을 공유하는 이미지 집합을 도메인으로 만든다.

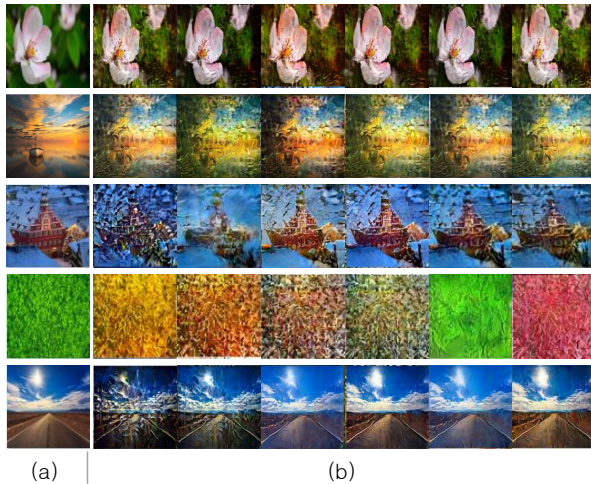


Fig. 3. (a) contents image, (b) Styled images

Fig. 3은 콘텐츠 이미지들에 스타일이 적용되는 과정의 이미지들을 보여준다. 사계절의 이미지들이 스타일 이미지들의 형태와 컬러 등이 적용되는 과정을 보여준다. 콘텐츠에 해당하는 이미지들이 주어지고, 참조 이미지의 도메인이 주어지면 도메인 내의 이미지들을 가져와 스타일 잠재 벡터를 만들어 스타일 이미지를 생성한다.

## V. Conclusions

본 논문은 도메인을 갖는 이미지들의 스타일 트랜스퍼에 관한 것으로 사계절의 콘텐츠 이미지들로 멀티 도메인 스타일 트랜스퍼를 하였다. 도메인에 특화된(specific) 방법들이 얼굴 이미지와 같은 통계량의 변화가 심하지 않은 데이터에는 잘 적용되었으나 사계절 이미지와 같은 변화량이 큰 이미지들을 학습시키는 데 어려움이 많았다. 본 논문은 이를 해결하고자 VAE의 잠재벡터와 가우시안 노이즈를 사용하여 샘플들의 특징 공간을 유지하면서 다른 분포를 학습할 수 있는 매핑 네트워크를 만들어 여러 스케일에 적용하였다. 하나의 스케일 처리 블록이 다른 스케일 블록에 의존성을 주지 않도록 블록마다 가우시안 노이즈를 함으로써 도메인의 새로운 이미지들로 학습할 수 있었으며, 스타일이 적용된 이미지들을 결과로 얻을 수 있었다.

향후 제안한 생성자로 평균과 분산의 차이가 다양한 도메인에 적용하고, 스타일 트랜스퍼의 다른 방법들과 비교 분석하는 연구가 필요하다.

## ACKNOWLEDGEMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R111A1A01075149)

## REFERENCES

- [1] A. Almahairi, S. Rajeshwar, A. Sordoni, P. Bachman, and A. Courville, "Augmented cyclegan: Learning many-to-many mappings from unpaired data," In *Proceeding of the 35th International Conference on Machine*. PMLR. Stockholmsmässan, pp. 195-204, 2018.
- [2] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," 2018. *arXiv preprint arXiv: https://arxiv.org/abs/1812.04948*,
- [3] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "StarGAN v2: Diverse Image Synthesis for Multiple Domains," 2020. *arXiv preprint arXiv: https://arxiv.org/abs/1912.01865*
- [4] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator architecture for Generative Adversarial Networks," 2019. <https://arxiv.org/abs/1812.04948>
- [5] S. Na, S. Yoo, and J. Choo, "MISO: Mutual Information Loss with Stochastic Style Representations for Multimodal Image-to-Image Translation," 2019. *arXiv preprint arXiv: https://arxiv.org/abs/1902.03938*
- [6] J.-Y. Zhu, R. Z. D. Pathak, T. Darrell, A. A. Efros, O. Wang and E. Shechtman, "Toward Multimodal Image-to-Image Translation," 2017. *arXiv preprint arXiv: https://arxiv.org/abs/1711.11586*
- [7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017. *arXiv preprint arXiv: https://arxiv.org/abs/1611.07004*
- [8] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," 2018. *arXiv preprint arXiv: https://arxiv.org/abs/1703.10593*
- [9] Q. Mao, H.-Y. Lee, H.-Y. Tseng, S. Ma, and M.-H. Yang, "Mode Seeking Generative Adversarial Networks for Diverse Image Synthesis," 2019. *arXiv preprint arXiv: https://arxiv.org/abs/1903.05628*
- [10] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," 2018. *arXiv preprint arXiv: https://arxiv.org/abs/1711.09020*
- [11] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," 2014. *arXiv preprint arXiv: https://arxiv.org/abs/1312.6114*

- [12] A. Odena, C. Olah, and J. Shlens, "Conditional Image Synthesis With Auxiliary Classifier GANs," 2017. *arXiv preprint arXiv*: <https://arxiv.org/abs/1610.09585>
- [13] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," 2014. *arXiv preprint arXiv*: <https://arxiv.org/abs/1406.2661>
- [14] L. A. Gatys, A. S. Ecker, and Matthias Bethge, "Image Style Transfer Using Convolutional Neural Networks," In *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*. CVPR. Las Vegas, NV, pp. 2414-2423, 2016.
- [15] E. Risser, P. Wilmot, and C. Barnes, "Stable and Controllable Neural Texture Synthesis and Style Transfer Using Histogram Losses," 2017. *arXiv preprint arXiv*: <https://arxiv.org/abs/1701.08893>
- [16] F. Luan, S. Paris, E. Shechtman, and K. Bala. "Deep Photo Style Transfer," 2017. *arXiv preprint arXiv*: <https://arxiv.org/abs/1703.07511>
- [17] X. Huang and S. Belongie, "Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization," 2017. *arXiv preprint arXiv*: <https://arxiv.org/abs/1703.06868>
- [18] S. Loffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," 2015. *arXiv preprint arXiv*: <https://arxiv.org/abs/1502.03167>
- [19] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial network," 2016. *arXiv preprint arXiv*: <https://arxiv.org/abs/1511.06434>
- [20] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images," 2016. *arXiv preprint arXiv*: <https://arxiv.org/pdf/2005.09635>
- [21] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," 2017. *arXiv preprint arXiv*: <https://arxiv.org/pdf/1701.02096>
- [22] Y. Shen, C. Yang, X. Tang, and B. Zhou, "InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs," 2020. *arXiv preprint arXiv*: <https://arxiv.org/pdf/2005.09635>
- [23] J. Chang, Q. Mao, Z. Zhao, S. Wang, S. Wang, H. Zhu and S. Ma, "Layered Conceptual Image Compression Via Deep Semantic Synthesis," 2019. In *Proceeding of the IEEE Conference on Image Processing*. ICIP. Taipei, Taiwan, 694-698. DOI: <https://doi.org/10.1109/ICIP.2019.8803805>.
- [24] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," 2018. *arXiv preprint arXiv*: <https://arxiv.org/abs/1706.08500>

## Authors



Geum-Boon Lee received the M.S. and Ph.D. degrees in Computer Engineering from Daejeon University and Chosun University, Korea, in 2002 and 2010, respectively. Dr. Lee joined the faculty of Chosun University,

Gwangju, Korea, in 2018. She is currently a professor, National program of Excellence in Software center of Chosun University. She is interested in artificial intelligence, pattern recognition, bigdata analysis, and especially deep learning.