

# 물체 변형 성능을 향상하기 위한 U-net 및 Residual 기반의 Cycle-GAN

## U-net and Residual-based Cycle-GAN for Improving Object Transfiguration Performance

김 세 운<sup>1</sup> · 박 광 현<sup>†</sup>  
Sewoon Kim<sup>1</sup>, Kwang-Hyun Park<sup>†</sup>

**Abstract:** The image-to-image translation is one of the deep learning applications using image data. In this paper, we aim at improving the performance of object transfiguration which transforms a specific object in an image into another specific object. For object transfiguration, it is required to transform only the target object and maintain background images. In the existing results, however, it is observed that other parts in the image are also transformed. In this paper, we have focused on the structure of artificial neural networks that are frequently used in the existing methods and have improved the performance by adding constraints to the exiting structure. We also propose the advanced structure that combines the existing structures to maintain their advantages and complement their drawbacks. The effectiveness of the proposed methods are shown in experimental results.

**Keywords:** Deep learning, AI, Image-to-image translation, CNN, GAN

### 1. 서 론

이미지-이미지 번역(Image-to-Image Translation)<sup>[1]</sup>은 이미지를 활용하는 딥러닝의 응용 분야 중 하나로서, 인공신경망의 입력과 출력 데이터가 모두 이미지 형태이다. 입력 이미지의 도메인을 원하는 다른 도메인으로 변환하는 일을 수행하며, 도메인을 어떻게 설정하느냐에 따라 다양한 응용이 가능하다. 흑백 사진을 컬러 사진으로<sup>[2]</sup>, 지도를 위성사진으로<sup>[3]</sup>, 선화를 채색된 이미지로 변환<sup>[3]</sup>하는 등의 예시가 있다. 다른 데이터의 형태가 이미지이기 때문에 합성곱 인공신경망(Convolutional Neural Networks)을 기반으로 하여, 최근에는 딥러닝 분야에서 가장 활발하게 연구되고 있는 생성모델 중 하나인 GAN (Generative Adversarial Networks)<sup>[4]</sup>을 많이 사용하고 있다. GAN은 판별(Discriminator)과 생성(Generator)을 담당하는 두 개의 인공신경망으로 구성되며, 학습 과정에서

두 개의 인공신경망이 서로 경쟁을 한다. 본 논문에서는 이미지-이미지 번역의 일종인 물체 변형(Object Transfiguration)의 성능을 개선하고자 하였다. 물체 변형에서는 이미지 상의 특정 물체 자체를 도메인으로 보고 다른 물체로 변형하는 일을 수행한다. 말을 얼룩말로 변형하거나 사과를 오렌지로 변형하는 등의 예시를 들 수 있다. 본 논문에서는 이를 위해 GAN에 Cycle-Consistency 개념을 추가한 Cycle-GAN<sup>[5]</sup>의 형태를 사용하였다. 일반적으로 특정 물체가 포함된 이미지에 대해 해당 물체만 다른 물체로 변경된 이미지를 구하여 학습 데이터로 사용하기는 어려운데, Cycle-GAN은 쌍이 맞지 않는 데이터로도 학습이 가능하다. 즉, 변형할 두 도메인만 결정되면 각 데이터에 대응하는 상대 도메인의 데이터가 없어도 되기 때문에 물체 변형에 적합한 형태이다. 이 때, 이미지 상에서 목표가 되는 물체만 정확하게 인식하고 해당 부분만 적절하게 변형하는 것이 필요한데, 기존 결과에서는 목표가 되는 물체 이외의 부분에서도 변형이 생기는 현상을 관찰할 수 있다. 본 논문에서는 GAN을 구성하는 인공신경망들 중 직접적으로 도메인 간의 변환을 학습하는 생성망의 구조에 초점을 맞추어 기존 방법의 한계점을 개선하고자 하였다. 즉, 기존 방법과 같이 생성망이 상대 도메인의 전체 이미지를 생성하는 것이 아니라

Received : Jan. 16, 2018; Revised : Feb. 19, 2018; Accepted : Feb. 20, 2018

1. School of Robotics, Kwangwoon University, Seoul, Korea (sewoon7301@naver.com)

† Corresponding author: School of Robotics, Kwangwoon University, Seoul, Korea (akaii@kw.ac.kr)

Copyright © KROS

특정 영역에 대한 마스크를 생성하여 물체 변형을 수행하도록 제약을 주었으며, 이를 통해 목표가 되는 물체에 집중하는 효과를 얻을 수 있었다. 또한 기존에 사용된 생성망 구조와 이미지-이미지 번역에서 자주 사용되는 구조를 분석하고 결합하여 각각의 장점은 살리고 단점을 보완하는 새로운 형태의 구조를 제안하였다. 실험에는 기존 Cycle-GAN에서 사용된 horse2zebra 데이터<sup>[6]</sup>를 사용하였으며, 본 논문의 목적에 따른 지표를 제시하여 정량적으로 성능을 평가하고자 하였다.

## 2. 관련 연구

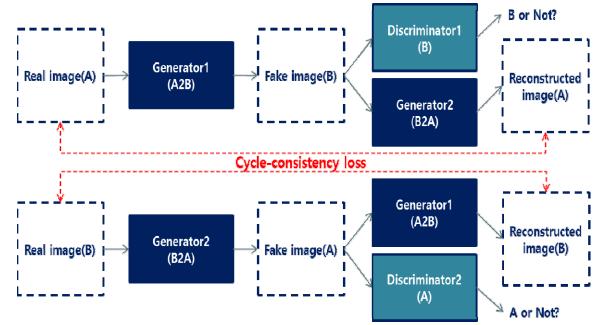
### 2.1 GAN & Cycle-GAN

GAN (Generative Adversarial Networks)은 현재 딥러닝 분야에서 가장 활발하게 연구되고 있는 생성모델 중 하나이며, 판별(Discriminator)과 생성(Generator)을 담당하는 두 개의 인공신경망으로 구성되어 있다. 판별망은 입력된 데이터가 실제 데이터인지 생성망이 생성한 가짜 데이터인지를 판별하도록 학습하고, 생성망은 판별망을 속일 수 있는 데이터를 생성하도록 학습한다. 이러한 경쟁 과정을 통해 이미지 생성을 학습할 수 있으며, 이미지-이미지 번역에서도 좋은 성과를 보여주고 있다. GAN은 수많은 변형 및 응용들이 존재하는데, 그 중 Cycle-GAN은 Cycle-Consistency의 개념이 추가된 모델이다. 데이터를 다른 도메인으로 변형하였다가 다시 본래의 도메인으로 변형하였을 때 원본 데이터와 동일하게 유지되어야 한다는 개념으로, 이 덕분에 쌍이 맞지 않는 데이터로도 학습이 가능하다. 동일한 개념을 사용하는 연구로 Disco-GAN<sup>[7]</sup>, Dual-GAN<sup>[8]</sup>이 있다. 본 논문에서는 Cycle-GAN을 물체 변형을 위한 큰 틀로 사용하고 내부에서 사용되는 인공신경망의 세부적인 구조에 초점을 맞추어, 구조적인 제약을 추가하고 기존 이미지-이미지 번역에서 빈번하게 사용되는 구조들을 결합함으로써 성능 향상을 얻고자 하였다.

### 2.2 Adversarial Loss

Cycle-GAN은 [Fig. 1]과 같이 두 쌍의 GAN으로 구성된다. 변환을 수행할 두 도메인이 결정되면, 각각의 GAN을 통해 두 도메인 간의 양방향 변환이 동시에 학습된다. 판별망과 생성망의 적대적인 구도를 위한 기본적인 손실함수(Loss)는 식 (1)과 같고, 서로의 입장에서 이를 최적화하기 위해 경쟁한다.

$$L_{x2y\_adv} = E_y [\log D_y(y)] + E_x [\log(1 - D_y(G_{x2y}(x)))] \quad (1)$$



[Fig. 1] Cycle-GAN: There are two domains (A, B) to perform object transfiguration, and a pair of generator and discriminator for learning transfiguration from each domain to a relative domain. Cycle-consistency loss is caused by the difference between the real image and the reconstructed image transformed back into original domain

여기서 판별망  $D_y$ 는  $y$ 도메인의 실제 이미지 또는 가짜 이미지  $G_{x2y}(x)$ 를 입력받아  $y$ 도메인의 실제 데이터가 맞는지를 판별하고, 생성망  $G_{x2y}$ 는  $x$ 도메인의 이미지를 입력으로 받아 판별망  $D_y$ 를 속일 수 있는  $y$ 도메인의 이미지를 생성한다. 두 도메인 사이에서 이러한 매핑을 학습하므로 각각  $L_{A2B\_adv}$  와  $L_{B2A\_adv}$ 가 존재한다.

### 2.3 Cycle-consistency Loss

Cycle-consistency Loss는 상대 도메인으로 변형한 이미지를 다시 원래의 도메인으로 변형한 이미지와 원본 이미지 간의 차이에서 발생하며, 식 (2)와 같이 계산된다.

$$L_{cyc} = E_A [\| G_{B2A}(G_{A2B}(A)) - A \|_1] + E_B [\| G_{A2B}(G_{B2A}(B)) - B \|_1] \quad (2)$$

여기서  $A, B$ 는 각 도메인에서의 실제 이미지 데이터이다. 손실함수에 이를 추가함으로써 Adversarial loss만 사용했을 때보다 입력과 출력 간의 매핑에 제약이 생긴다.

### 2.4 Full Objective

최종적으로 Cycle-GAN을 통해 최적화할 목적함수는 식 (3)과 같다.

$$L = L_{A2B\_adv} + L_{B2A\_adv} + \lambda L_{cyc} \quad (3)$$

여기서  $\lambda$ 는 각 loss들 간의 중요도를 조절하기 위한 가중치이다. 이 목적함수를 판별망들은 최대화, 생성망들은 최소화하기 위해 서로 경쟁하며 학습하고, 이 과정을 통해 두 도메인 간의 변환을 생성망에 학습시킬 수 있다.

### 3. 제안하는 방법

본 논문에서는 물체 변형 성능을 향상시키기 위해 생성망의 세부 구조에 초점을 맞추었다. 첫 번째로, 간단한 구조적 제약(Input-residual)을 추가하고, 두 번째로는 기존에 사용되던 구조들의 특성을 분석하고 이들을 결합하여 장점은 유지하고 단점은 보완할 수 있는 구조( $\omega$ -net)를 제안한다. 최종적으로는 이 두 가지를 조합하여( $\omega$ -net과 Input-residual의 조합) 본 논문의 목표를 달성하고자 한다.

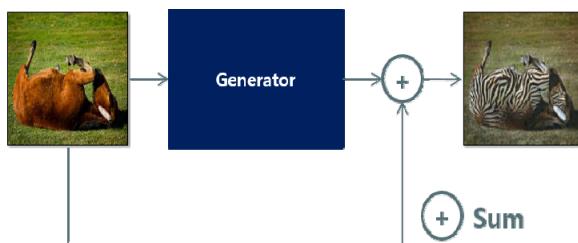
#### 3.1 구조적 제약을 추가한 Input-residual

이미지 상의 물체가 변형되기 위해서는 목표가 되는 물체에 해당하는 특정 영역의 픽셀 값들이 조정되어야 한다. 본 논문에서는 이러한 아이디어를 생성망에 제약으로 주고자 하였다. [Fig. 2]와 같이 생성망의 입력과 출력을 더하여 결과를 내도록 덧셈 연결을 추가함으로써 이러한 제약을 주었다.

이 연결을 통해 생성망은 변형된 전체 이미지를 생성하는 것이 아니라 입력 이미지에 더해질 마스크를 생성하여 물체 변형을 수행하도록 제약된다. 이미지의 생성 과정을 식(4)와 같이 변경한 것으로 볼 수 있다.

$$b = a + G(a) \quad (4)$$

여기서  $a$ 는 생성망으로 들어가는 입력 이미지,  $b$ 는 생성된 이미지이다. 이 때,  $G(a)$ 를 입력에 더할 마스크로 볼 수 있다. 인공신경망의 전체 입력과 출력에 대한 덧셈연산을 추가했을



[Fig. 2] Input-residual: We added the connection to make a result by adding the input and output of the generator. This constrains the object transfiguration to be performed through mask generation

뿐이며, 형태는 Residual Block과 동일하다. 이를 통해 물체 변형을 수행할 때 목표가 되는 물체에 집중하는 효과를 얻을 수 있었다.

#### 3.2 기준의 Residual-net과 U-net 분석

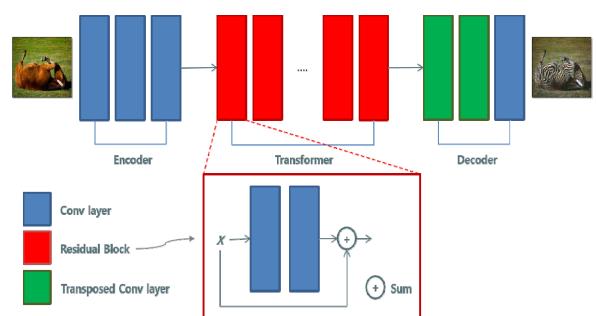
##### 3.2.1 Residual-net

딥러닝에서 Residual<sup>[9,10]</sup>은 인공신경망이 깊어질수록 학습이 잘 되지 않는 점을 개선하기 위한 방법으로 연구되었다. [Fig. 3]에서처럼 Residual Block은 층을 건너뛰는 덧셈 연결을 통해 입력과 출력의 차이를 학습하게 되며, 기울기의 흐름이 원활해져 깊은 인공신경망도 학습이 가능해진다. 이러한 구조는 입력에 어떤 작은 변화를 생성하여 더하는 형태로도 볼 수 있는데, 이러한 특성으로 인해 이미지-이미지 변역에서는 변환기(Transformer)의 역할로 자주 사용되고 있다. Cycle-GAN의 생성망 역시 Residual Block을 기반으로 구성되어 있다<sup>[11]</sup>.

[Fig. 3]의 구조는 크게 인코더(Encoder)와 변환기(Transformer), 디코더(Decoder) 부분으로 나눌 수 있다. Residual Block이 담당하고 있는 변환기 부분을 제외하면 인코더와 디코더는 비교적 얇은 층으로 구성되어 있다. 합성곱 인공신경망은 층이 깊어질수록 좀 더 상위 레벨의 추상화된 특징을 추출할 수 있으며, 이러한 과정에서 동반되는 다운샘플링을 통해 크기, 위치 등에 대해 어느 정도 불변의 특성을 가지게 된다. 하지만 인코더-디코더 단이 알다는 것은 물체를 정확하게 인식해야 하는 경우에는 단점이 될 수 있다. 이에 따라 실제 변형이 잘 된 예시들에서도 물체 이외의 부분 또한 변형이 발생하는 현상을 관찰할 수 있다.

##### 3.2.2 U-net

U-net<sup>[12]</sup>도 이미지-이미지 변역에서 자주 사용되는 구조 중 하나이며 의료영상 분할을 위해 제안된 구조이다. 앞서 기술



[Fig. 3] Residual-net: The composition of the residual-net can be roughly divided into an encoder, a decoder, and a transformer. Among them, the residual block is used as a transformer

한 것처럼 다운샘플링은 합성곱 인공신경망이 어느 정도 불변한 특성을 가지는데 기여하는 반면, 정확한 공간적인 정보의 손실을 야기하기도 한다. 이미지 분할(Image Segmentation)에서는 정확한 공간적인 정보가 필요하기 때문에 이를 보완하고자 인코더와 디코더 사이에 연결을 추가한 것이 U-net의 핵심이다. [Fig. 4]에서와 같이 디코딩 때 인코더 단에서 나온 특정 정보도 같이 활용할 수 있도록 연결된다. 즉, 공간적인 정보를 유지하면서도 층을 더 깊게 쌓을 수가 있고, 이러한 구조가 인코딩-디코딩 측면에서는 Residual-net보다 유리하다고 볼 수 있다. 하지만 반대로 Residual Block을 사용하지 않기 때문에 변환의 개념이 빈약하다.

### 3.3 Residual Block을 추가한 $\omega$ -net

본 논문에서는 [Fig. 5]와 같이 Residual-net과 U-net을 결합하여 서로의 장점은 살리고 단점은 보완할 수 있는 형태의 구조를 제안한다. 인코딩-디코딩 측면에서 좀 더 유리한 U-net의 구조를 기본 골격으로 하였으며, 인코더와 디코더 사이에 Residual Block을 거치도록 추가하여 변환의 개념을 보완하였다. Residual-net에서처럼 특정 해상도에서만 Residual Block을 통한 변환을 거치는 것이 아니라 여러 해상도에서 이러한 과정을 거치게 되는데, 이는 컴퓨터 비전 분야에서 자주 쓰이

는 기법과도 유사하다. 또한 같은 크기의 Residual Block이라 하더라도 더 낮은 해상도에서 연산을 수행하기 때문에 구성에 따라 연산량 절감 효과도 얻을 수 있다. Residual-net에 비해 더 큰 해상도에서의 연산도 추가되지만 큰 해상도에서는 비교적 적은 채널 수를 사용하기 때문에 연산량을 줄일 수 있다.

## 4. 실험

### 4.1 실험 방법

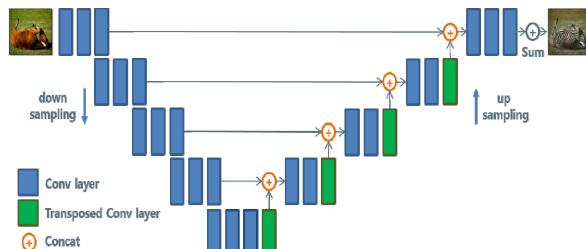
실제 구현에서는 좀 더 안정적인 것으로 알려져 있는 LSGAN<sup>[13]</sup>의 Adversarial Loss를 사용하였으며, Cycle-GAN의 최적화를 위해 Adam을 사용하였다. 이 때, 학습률은 0.0002로, beta는 각각 0.5와 0.999로 하였으며, 배치 크기는 4로 하였다. 총 150 세대(epoch) 동안 판별망과 생성망을 각각 한 번씩 갱신하였고, 100세대 이후부터는 0까지 선형적으로 학습률을 감쇄시켰다. 이미지의 크기는 256×256으로 고정하였고, 30×30 크기의 패치 형태 판별망(Patch Discriminator)을 사용하였다. 이때, 50장 크기의 이미지 풀<sup>[14]</sup>을 두었다. 위와 같은 요소들은 모두 동일하게 적용하고 생성망의 구조를 변경해가며 실험을 진행하였다.

### 4.2 데이터 세트

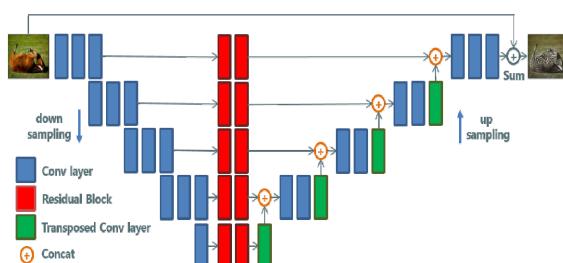
실험에는 기존 Cycle-GAN에서 사용되었던 horse2zebra 데이터<sup>[6]</sup>를 사용하였다. 본 논문의 실험과 관련성이 낮은 데이터들은 일부 삭제하여 총 말 816장, 얼룩말 948장으로 학습 데이터를 구성하였다. 테스트 데이터 세트로는 Weizmann Horses 데이터 세트<sup>[15]</sup>의 말 이미지 328장을 사용하였다. 이는 정량적인 지표를 측정하기 위한 것으로서 다음 절에서 설명한다.

### 4.3 배경 유지 성능을 평가하기 위한 정량적 지표

생성모델이 생성한 데이터를 정량적으로 평가하는 것은 어려운 일이다. 예를 들어 말이 얼룩말로 얼마나 잘 변형되었는지는 주관적인 판단으로 평가되기 때문이다. 따라서 이미지-이미지 번역에서는 주로 Amazon Mechanical Turk (AMT)를 이용한 정성적인 설문조사가 사용된다. 하지만 배경 유지 성능은 색차 계산을 통해 정량적인 평가가 가능하다. Weizmann Horses 데이터 세트에는 328장의 말 이미지와 함께 말 부분만 분할되어 있는 Segmentation map이 존재한다. 이를 통해 배경 부분만 분할할 수 있으며 분할된 배경 부분에 대해 변형 전 후의 색차를 계산함으로써 정량적인 평가를 할 수 있다. 색차 계



[Fig. 4] U-net: U-net compensates for the loss of spatial information by concatenating the information of the encoder.



[Fig. 5]  $\omega$ -net: U-net, which has advantage in terms of encoding-decoding, is used as a basic skeleton, and we supplemented the concept of transformation by adding a residual block between the encoder and the decoder.

산에는 인간의 눈에 의한 색차 인식을 가능한 가깝게 수학적으로 재현하기 위해 만들어진 Lab 색 공간을 사용하였다. 이 때, 색차는 식(5)과 같이 계산된다. (CIELAB Delta E\* - CIE76)

$$\Delta E_{ab} = \sqrt{(\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2} \quad (5)$$

$L, a, b$ 는 Lab 색공간의 각 채널로서,  $L$ 은 밝기를,  $a$ 는 빨간색과 초록색 중 어느 쪽으로 더 치우쳐져 있는지를,  $b$ 는 노란색과 파란색 중 어느 쪽으로 더 치우쳐져 있는지를 나타낸다. 각 픽셀에 대해 색차를 계산한 후, 식(6)과 같이 전체 이미지 중 색차가 일정한 문턱 값 이하로 유지되는 픽셀의 비율을 표로 사용하였다.

$$\frac{\sum x}{h \times w} \times 100, \quad x = \begin{cases} 1, & \text{if } \Delta E_{ab} < \text{threshold} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

여기서  $h, w$ 는 각각 이미지의 높이와 너비를 나타낸다. 전체 테스트 데이터들에 대해 식(6)을 적용하고, 평균값을 구하여 정량적인 평가를 하였다. 문턱 값은 CIE76 색차 값에 따른 사람의 인지에 대한 조사를 참고하여 0~10으로 변경해가며 평가를 진행하였다. [Table 1]은 이러한 조사의 예시로서, 조사 및 규격에 따라 조금씩 차이가 있다.

#### 4.4 정성적 평가

Input-residual을 추가한 후 생성망의 동작이 마스크를 생성하도록 변경되는 것을 [Fig. 6]을 통해 확인할 수 있다. 여기서 가장 왼쪽의 이미지는 입력 이미지이고, Residual-net을 통해 물체 변형한 결과가 두 번째 이미지이다. 세 번째 이미지는 Residual-net에 Input-residual 연결을 추가한 후 생성망이 생성한 마스크를 나타내며, 입력 이미지와 더해져 네 번째의 최종 결과를 만들어 낸다. 네 번째 이미지가 두 번째 이미지에 비해 배경이 더 잘 유지되는 것을 확인할 수 있다.

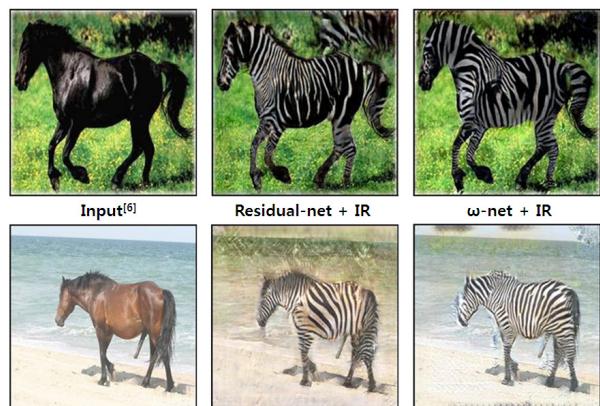
[Fig. 7]은 Residual-net과  $\omega$ -net에 Input-residual을 적용하고 결과를 비교한 예시이다. 두 번째 열의 결과들과 같이

[Table 1] CIE76 survey<sup>[16]</sup>

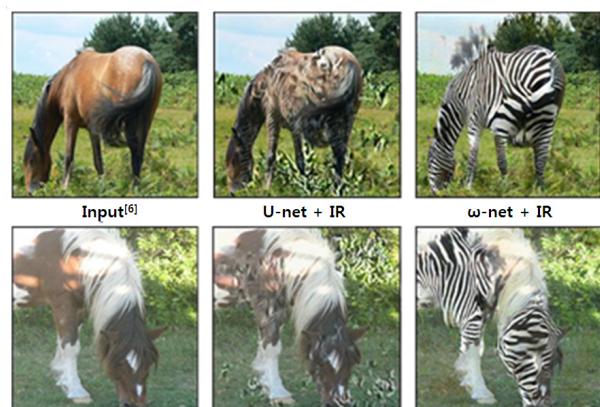
$0 < \Delta E_{ab} < 1$	Observer does not notice the difference.
$1 < \Delta E_{ab} < 2$	Only experienced observer can notice the difference.
$2 < \Delta E_{ab} < 3.5$	Unexperienced observer also notice the difference.
$3.5 < \Delta E_{ab} < 5$	Clear difference in color is noticed.
$5 < \Delta E_{ab}$	Observer notice two different colors.



[Fig. 6] Input-residual: The leftmost image<sup>[6]</sup> is the input image, and the second image is the result of object transfiguration through residual-net. The third image shows the mask generated by the generator after connecting input-residual to residual-net. This mask is added to the input image and the fourth image is the final result. We can see the background is better maintained with input-residual



[Fig. 7] Residual-net vs.  $\omega$ -net: These are some examples of comparing the results of residual-net and  $\omega$ -net. Input-residual is applied to both neural networks. In residual-net, we can see through the second column images that background noise is still generated even though input-residual is applied. In third column images, we can see this tendency is relatively improved by  $\omega$ -net



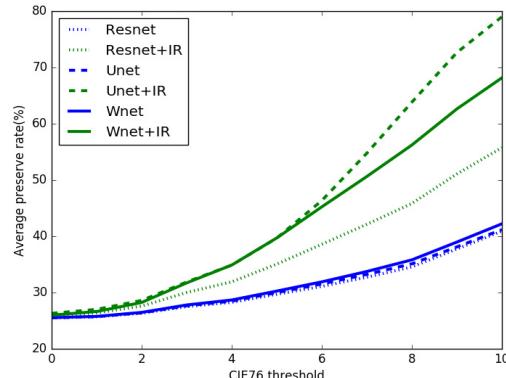
[Fig. 8] U-net vs.  $\omega$ -net: These are some examples of comparing the results of u-net and  $\omega$ -net. Input-residual is applied to both neural networks. In u-net, there are some examples that do not transfigurate well like the second column images, while we can see these examples can be transfigurated well by  $\omega$ -net (third column images)

Residual-net에 Input-residual을 추가하였음에도 여전히 배경에 노이즈가 생기는 것을 살펴볼 수 있다. 세 번째 열의 결과들은 이러한 현상이  $\omega$ -net을 통해 개선됨을 보여준다.

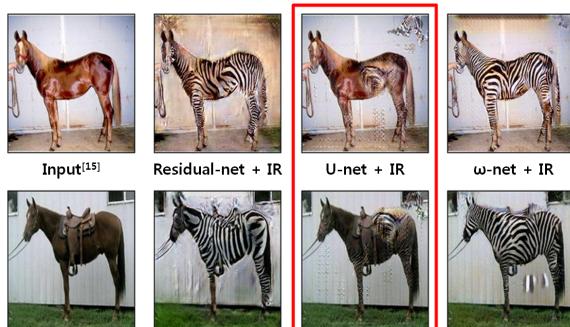
[Fig. 8]은 U-net과  $\omega$ -net에 Input-residual을 적용하고 결과를 비교한 예시이다. 두 번째 열의 결과들과 같이 U-net에서는 대상 물체가 잘 변형되지 않지만 세 번째 열의 결과들과 같이  $\omega$ -net에서는 잘 변형됨을 확인할 수 있다.

#### 4.5 배경 유지에 대한 정량적 평가

[Fig. 9]는 328장의 테스트 데이터 세트에 대하여 4.3절에서 제시한 지표를 통해 각 구조들을 배경 유지 관점에서 평가한 결과를 나타낸다. 즉, 목표가 되는 물체가 얼마나 잘 변형되었는지를 평가하는 것이 아니라 배경이 얼마나 잘 유지되는지만 평가한다. Input-residual을 추가하기 전의 각 구조들은 배경 유지의 측면에서 모두 낮은 성능의 결과를 보이지만, Input-residual을 추가한 후에는 각 구조에서 모두 배경 유지 성능이 개선됨을 볼 수 있다. 이 때, U-net과 Input-residual의 조합이 배경 유지 측면에서 수치적으로 가장 높은 성능을 보이고 있지만 [Fig. 10]과 같이 변형 자체가 상대적으로 크게 미흡하다는 문제점을 확인할 수 있다. 즉, U-net과 Input-residual의 조합이 배경을 잘 유지하는 것처럼 보이지만 이는 이미지 변형 자체가 잘 수행되지 않기 때문에 배경 또한 변형이 거의 발생하지 않은 것이다. 따라서 본 논문의 목표에 부합한다고 할 수 없다. 이에 반해 제안하는 방법인  $\omega$ -net과 Input-residual의 조합은 물체를 잘 변형하면서도 Residual-net과 Input-residual의 조합에 비해 배경이 더 잘 유지된다는 것을 확인할 수 있다.



[Fig. 9] This graph shows the result of evaluating background maintenance for each structure through the metric presented in this paper. In all three structures, we can see that the background maintenance are improved through input-residual. Here, the combination of u-net and input-residual shows the highest performance, but there is a problem that the transfiguration is relatively insufficient ([Fig. 10]).



[Fig. 10] U-net's failure case: The combination of u-net and input-residual shows the best performance in the background maintenance, but there is a problem that the transfiguration is relatively insufficient as in third column images.

는지를 평가하는 것이 아니라 배경이 얼마나 잘 유지되는지만 평가한다. Input-residual을 추가하기 전의 각 구조들은 배경 유지의 측면에서 모두 낮은 성능의 결과를 보이지만, Input-residual을 추가한 후에는 각 구조에서 모두 배경 유지 성능이 개선됨을 볼 수 있다. 이 때, U-net과 Input-residual의 조합이 배경 유지 측면에서 수치적으로 가장 높은 성능을 보이고 있지만 [Fig. 10]과 같이 변형 자체가 상대적으로 크게 미흡하다는 문제점을 확인할 수 있다. 즉, U-net과 Input-residual의 조합이 배경을 잘 유지하는 것처럼 보이지만 이는 이미지 변형 자체가 잘 수행되지 않기 때문에 배경 또한 변형이 거의 발생하지 않은 것이다. 따라서 본 논문의 목표에 부합한다고 할 수 없다. 이에 반해 제안하는 방법인  $\omega$ -net과 Input-residual의 조합은 물체를 잘 변형하면서도 Residual-net과 Input-residual의 조합에 비해 배경이 더 잘 유지된다는 것을 확인할 수 있다.

## 5. 결 론

본 논문에서는 물체 변형을 수행할 때 물체 이외의 부분도 같이 변형되는 기존 방법의 문제를 개선한 새로운 생성망 구조를 제안하였다. 직접적으로 변형을 담당하는 인공신경망에 제약을 추가하고, 기존 구조들을 결합하여 세부적인 구조를 변경함으로써 배경을 유지하면서 물체만 변형될 수 있도록 하였다. 또한 정량적인 평가를 위한 지표를 제시하여 기존 방법과 비교하고 성능을 검증하였다. 하지만 본 논문에서 제시한 지표는 배경 유지라는 목표를 평가하기 위한 것으로서 물체의 변형 자체가 잘 되었는지를 판단하기 위한 것은 아니다. 향후 이를 평가하고 검증할 수 있는 방법에 대한 연구가 필요하며, 지속적인 실험을 통해 더욱 최적화하는 것이 필요하다.

## References

- P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, pp. 5967-5976, 2017.
- R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," *European Conference on Computer Vision*, Amsterdam, Netherlands, pp. 649-666, 2016.
- Preferred Networks, *PaintsChainer*, [Online], <https://paintschainer.preferred.tech>, Accessed: January 16, 2018.
- I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *28th Annual Conference on Neural Information Processing Systems*, Montreal, Canada, pp. 2672-2680, 2014.
- J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired

- image-to-image translation using cycle-consistent adversarial networks,” *2017 IEEE International Conference on Computer Vision*, Venice, Italy, pp. 2242-2251, 2017.
- [6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large scale visual recognition challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211-252, April, 2015.
- [7] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, “Learning to discover cross-domain relations with generative adversarial networks,” *International Conference on Machine Learning*, Sydney, Australia, pp. 1857-1865, 2017.
- [8] Z. Yi, H. Zhang, P. Tan, and M. Gong, “Dualgan: Unsupervised dual learning for image-to-image translation,” in *IEEE International Conference on Computer Vision*, Venice, Italy, pp. 2868-2876, 2017.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *2017 IEEE International Conference on Computer Vision*, Las Vegas, USA, pp. 770-778, 2016.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” *European Conference on Computer Vision*, Amsterdam, Netherlands, pp. 630-645, 2016.
- [11] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” *European Conference on Computer Vision*, Amsterdam, Netherlands, pp. 694-711, 2016.
- [12] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, pp.234-241, 2015.
- [13] X. Mao, Q. Li, H. Xie, R. Y.K. Lau, Z. Wang, and S. P. Smolley, “Least squares generative adversarial networks,” *2017 IEEE International Conference on Computer Vision*, Venice, Italy, pp. 2813-2821, 2017.
- [14] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, “Learning from simulated and unsupervised images through adversarial training,” *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, pp. 2242-2251, 2017.
- [15] E. Borenstein and S. Ullman, “Combined top-down/bottom-up segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2109-2125, December, 2008.
- [16] W. Mokrzycki and M. Tatol, “Color difference  $\Delta E$  - A survey,” *Machine Graphic and Vision*, vol. 20, no. 4, pp. 383-411, April, 2011.



김 세운

2016 광운대학교 로봇학부 학사  
2016 ~ 현재 광운대학교 대학원 로봇학과  
석사과정

관심분야: 인공지능, 딥러닝, 머신러닝, 영상처리



박 광 현

1994 KAIST 전기및전자공학과 학사  
1997 KAIST 전기및전자공학과 석사  
2001 KAIST 전기및전자공학과 박사  
2008 ~ 현재 광운대학교 로봇학부 교수

관심분야: 로봇 소프트웨어, 기계학습, 교육용 로봇