

# R2와 어텐션을 적용한 유넷 기반의 영상 간 변환에 관한 연구<sup>☆</sup>

## Image-to-Image Translation Based on U-Net with R2 and Attention

임 소 현<sup>1</sup>                      전 준 철<sup>1\*</sup>  
So-hyun Lim                      Jun-chul Chun

### 요 약

영상 처리 및 컴퓨터 비전 분야에서 하나의 영상을 통해 다른 영상으로 재구성하거나 새로운 영상을 생성하는 문제는 하드웨어의 발전에 따라 꾸준히 주목받고 있다. 그러나 컴퓨터를 통해 생성한 이미지를 사람의 눈으로 바라봤을 때 자연스럽게 않다는 문제 또한 계속해서 대두되고 있다. 최근 딥러닝 분야에 대한 연구가 활발히 진행됨에 따라 이를 활용한 영상 생성 및 개선 문제 또한 활발히 연구되고 있으며 그 중에서도 적대적 생성 신경망(Generative Adversarial Network)이라는 네트워크가 영상 생성 분야에 있어 좋은 결과를 보이고 있다. 적대적 생성 신경망이 제안된 이후 이를 기반으로 하는 다양한 네트워크가 제시됨에 따라 영상 생성 분야에서 더 자연스러운 영상을 생성하는 것이 가능해졌다. 그 중 pix2pix는 조건 적대적 생성 신경망 모델로 다양한 데이터셋에서도 좋은 성능을 보이는 범용적인 네트워크이다. pix2pix는 U-Net을 기반으로 두고 있으나 U-Net을 기반으로 하는 네트워크 중에서는 더 좋은 성능을 보이는 네트워크가 다수 존재한다. 때문에 본 연구에서는 pix2pix의 U-Net에 다양한 네트워크를 적용해 영상을 생성하고 그 결과를 상호 비교 평가한다. 각 네트워크를 통해 생성된 영상을 통해 기존의 U-Net을 사용한 pix2pix 모델보다 어텐션, R2, 어텐션-R2 네트워크를 적용한 pix2pix 모델이 더 좋은 성능을 보이는 것을 확인하고 그 중 가장 성능이 뛰어난 네트워크의 한계점을 향후 연구로 제시한다.

☞ 주제어 : 영상 간 변환, 조건 적대적 생성 신경망, 유넷

### ABSTRACT

In the Image processing and computer vision, the problem of reconstructing from one image to another or generating a new image has been steadily drawing attention as hardware advances. However, the problem of computer-generated images also continues to emerge when viewed with human eyes because it is not natural. Due to the recent active research in deep learning, image generating and improvement problem using it are also actively being studied, and among them, the network called Generative Adversarial Network(GAN) is doing well in the image generating. Various models of GAN have been presented since the proposed GAN, allowing for the generation of more natural images compared to the results of research in the image generating. Among them, pix2pix is a conditional GAN model, which is a general-purpose network that shows good performance in various datasets. pix2pix is based on U-Net, but there are many networks that show better performance among U-Net based networks. Therefore, in this study, images are generated by applying various networks to U-Net of pix2pix, and the results are compared and evaluated. The images generated through each network confirm that the pix2pix model with Attention, R2, and Attention-R2 networks shows better performance than the existing pix2pix model using U-Net, and check the limitations of the most powerful network. It is suggested as a future study.

☞ keyword : Image-to-Image Translation, conditional GAN, U-Net

## 1. 서 론

최근 하드웨어가 발전함에 따라 이전에는 기술적인 한계로 인해 연구되지 못했던 영상 처리 및 컴퓨터 비전 분야에 대한 연구가 활발히 진행되고 있다. 특히, 하나의 영상을 다른 영상으로 변환, 재구성하거나 혹은 새로운 영상을 생성하는 문제는 컴퓨터의 처리 방식과 사람의 처리 방식이 달라 컴퓨터로 생성되거나 변환되는 영상이 자연스럽게 않기 때문에 자연스러운 영상을 생성하는 것에 대해 계속 연구되어야 할 주제로 인식되고 있다. 일반적으로 영상 간 변환은 입력 영상과 출력 영상 간의 매핑

<sup>1</sup> Department of Computer Science, Kyonggi University, Gyeonggi-do, 16227, Korea

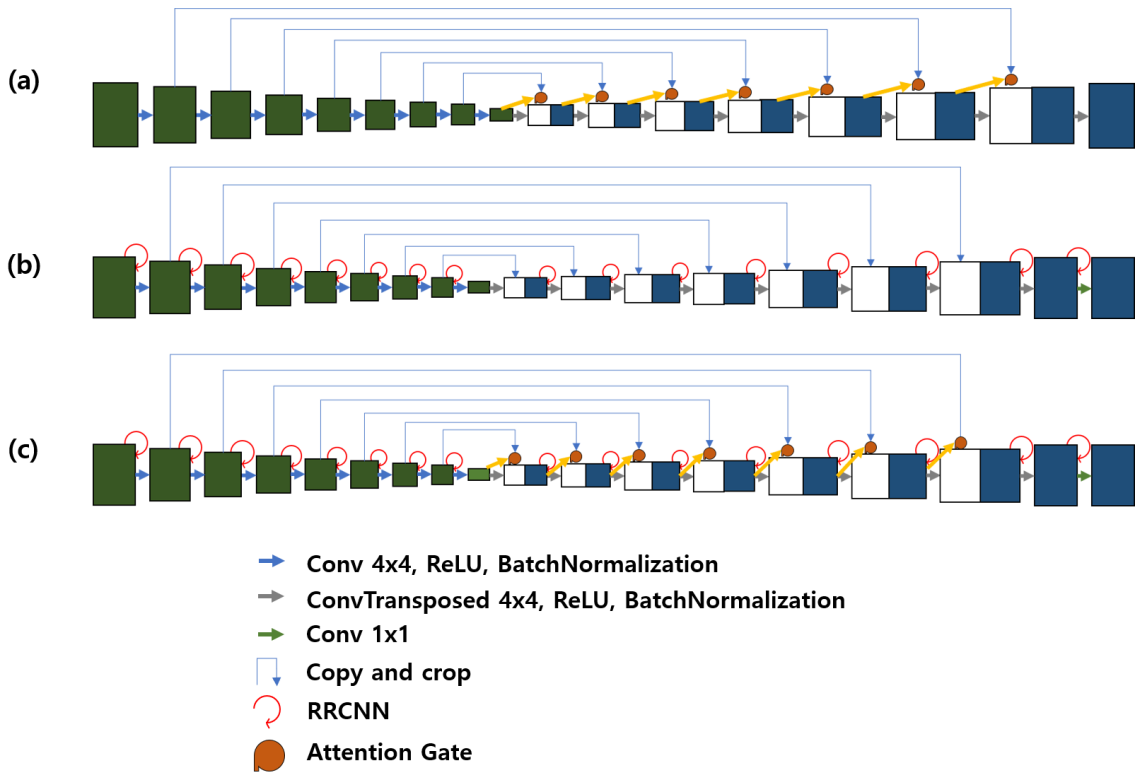
[Received 10 March 2020, Reviewed 23 March 2020(R2 19 April 2020), Accepted 6 May 2020]

\* Corresponding author (jochun@kyonggi.ac.kr)

☆ 본 연구는 경기도의 경기도 지역협력연구센터 사업의 일환으로 수행하였음. [GRRG KGU 2017-B04. 영상 및 네트워크 기반 지능정보 제조 서비스 연구]

☆ 본 연구는 2020년 경기대학교 대학원 연구원장학생 장학금 지원에 의하여 수행되었음

☆ 본 연구는 2019 한국인터넷정보학회 추계학술발표대회 논문을 수정 보완하였음.



(그림 1) 네트워크 구조. (a) 어텐션 유넷, (b) R2 유넷, (c) 어텐션-R2 유넷

(Figure 1) Network structure. (a) Attention U-Net, (b) R2 U-Net, (c) Attention-R2 U-Net

을 통해 영상을 생성하는 방식이다. 영상 간 변환은 매핑을 위해서 기존에는 인코더-디코더[1]를 사용했으나 이는 낮은 수준의 영상 정보를 공유하게 되어 병목현상이 발생하여 출력 영상이 훼손되는 단점을 가지고 있다. 또한, 입력 영상을 분석하는데 있어 분류나 회귀[2][3]와 같은 단순한 방식을 사용했기 때문에 영상 전체에 대한 처리가 이뤄지지 않고 픽셀이 독립적으로 처리되어 부자연스러운 영상을 생성한다[4].

최근 딥 러닝(Deep Learning)이 크게 발전함에 따라 영상 처리는 물론 컴퓨터 비전 분야에도 접목되기 시작했다. 딥러닝은 기존의 연구들에 비해 영상 생성 또는 재구성 측면에서 뛰어난 성능을 보여줄 뿐만 아니라 최근 초기에 대두된 딥러닝에 더 많은 개념들이 적용됨에 따라 다양한 네트워크 또한 제안되고 있기 때문이다. 특히 적대적 생성 신경망(GAN, Generative Adversarial Network)은 다양한 딥러닝 네트워크 가운데 영상 생성에 있어 탁월한 성능을 보인다. 적대적 생성 신경망은 성능이 뛰어난

은 물론, 다양한 원리나 구조를 추가해 새로운 모델을 생성하고 그에 따른 그 결과 혹은 과정에서 활용 범위가 굉장히 넓다. 본 논문에서는 U-Net 구조를 가지는 조건적 생성 신경망(conditional GAN)을 기반으로 한 pix2pix에 edges2shoes 데이터 셋을 사용해 영상을 변환하고, U-Net[5]에 다양한 연산 혹은 방식을 적용해 그 결과를 확인하고 비교함으로써 연산을 추가하는 것이 네트워크에 있어 어떤 영향을 미치는 지에 대하여 분석하였다. 위 (그림 1)은 본 논문에서 적용한 다양한 네트워크 구조를 제시하고 있다.

## 2. 관련 연구

### 2.1 유넷

영상 간 변환 문제는 입력 영상과 출력 영상을 매핑하는 것이 특징이다. 입력 영상과 출력 영상의 특징을 분석

해 매핑하고 변환 시에 매핑한 특징들을 사용해 변환한다. 기존 연구에서는 인코더-디코더 방식을 사용해 변환 문제를 해결했다[1]. 인코더-디코더 방식은 입력과 출력에서의 영상은 같지만 중간의 신경망을 축소함으로써 병목 현상을 일으켜 입력 영상의 특징들을 압축하는 방식이다. 이는 높은 수준의 영상 정보들이 오가는 것이 아닌 압축된 특징들과 같은 낮은 수준의 정보들을 공유하게 만들기 때문에 출력되는 영상이 훼손되거나 자연스럽게 않다는 단점을 가진다. 이를 해결하기 위해 사용된 방법이 U-Net이다[5]. 인코더-디코더와 U-Net의 가장 큰 차이는 skip connection에 있다. 병목 현상을 일으켜 특징을 압축한다는 것은 U-Net에도 똑같이 적용되는 방식이나 U-Net은 skip connection을 사용해 크기가 같은 단계의 입력과 출력 영상을 연결해 낮은 수준 즉, 압축된 특징들은 물론 축소되기 이전의 높은 수준의 영상 정보를 출력 영상에 영향을 끼친다. 또한, 압축된 영상의 특징만이 출력에 영향을 끼치는 것이 아닌 압축되지 않은 입력 영상의 특징 또한 출력에 영향을 끼치기 때문에 인코더-디코더 방식보다 자연스러운 영상을 생성할 수 있다. 아울러 유넷 모델에 다른 구조를 추가해 구현하는 것이 가능해 다양한 방식으로 활용할 수 있다[6][7].

## 2.2 적대적 생성 신경망

적대적 생성 신경망[8]은 영상을 생성하는 생성자와 생성자에서 생성한 영상과 정답 영상을 비교해 참 거짓을 구분하는 구분자가 적대적으로 상호작용하며 구분자가 구분할 수 없는 영상을 만드는 것을 목표로 하는 네트워크이다[9]. 적대적 생성 신경망이 제안되기 이전에는 영상 간 변환 문제를 픽셀 단위를 분류와 회귀를 통해 해결했다[2][3]. 그러나 이 방법은 공식을 통해 입력 영상과 출력 영상을 처리하는 방식으로 입력 영상과 출력 영상 간의 특징들이 서로 매핑되지 않고 독립적으로 처리되어 출력 영상이 자연스럽게 생성되지 않는 경우가 많다. 하지만 적대적 생성 신경망의 경우에는 영상을 압축하면서 생기는 손실에 대해서 구조적인 학습이 가능해 자연스러운 출력 영상을 생성한다. 적대적 생성 신경망에서는 입력 영상의 특징을 가지는 영상을 생성하지만 동시에  $z$ 라는 새로운 입력 값 혹은 잡음을 주어 입력 영상과 같은 특징을 갖지만 세부 사항이 다른 이미지를 생성한다. 또한, 잡음 외에도 다른 조건을 추가하거나 혹은 컨볼루션(convolution) 구조를 추가하는 등의 다양한 모델을 생성해 사용할 수 있으며[10][11] 특정 객체를 제거한

영상을 생성하는 방식도 존재한다[12][13]. 또한, 영상 압축 시스템[14], 자연 언어 처리[15]과 같이 단순한 영상 생성 분야 이외의 다른 컴퓨터 비전 및 이미지 프로세싱 분야에도 적용 가능하다.

## 3. 영상 변환 네트워크

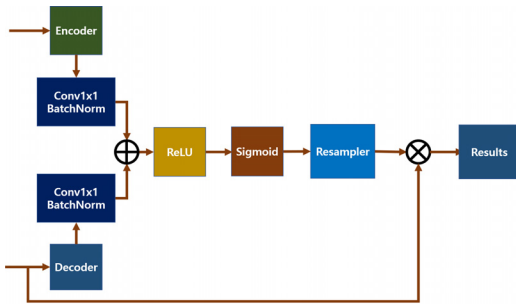
### 3.1 pix2pix 유넷

pix2pix에 적용된 네트워크는 유넷으로 컨볼루션 연산을 사용해 특징을 추출하고 확장 경로에서 인코더에서 추출한 특징을 skip connection을 통해 사용하는 구조를 가진다. pix2pix 네트워크는 본 논문에서 진행한 실험의 가장 기본이 되는 구조로 해당 구조에 다양한 모델을 추가해 다음의 3.2~3.4절에서 제시된 네트워크를 통해 실험을 진행하였다. pix2pix는 영상 간 변환 문제에서 입력 영상과 출력 영상 사이에 존재하는 연관성에 사용한 대표적인 영상 변환 기법으로 Philip Isola et al.[4]가 제안하였다. pix2pix는 엣지-영상, 라벨-영상과 같이 짝을 이루는 데이터 셋을 사용하는 것이 특징이며 학습 과정에서 두 영상 사이의 연관성을 매핑하여 효율적인 학습을 진행하는 구조를 가진다. 영상을 생성할 때 단순한 적대적 생성 신경망을 사용하는 것이 아닌 조건(condition)을 추가하는 조건 적대적 생성 신경망[11]을 사용해 두 영상 간의 연관성을 한 번 더 제시함으로써 영상 변환의 성능을 높이고자 하였다. 또한, 일정 크기의 패치(patch)를 사용해 전체 영상의 부분적 요소에 대해 자세히 학습하는 patchGAN을 구분자에 적용시켜 생성 영상에 있어 고품질의 결과를 생성한다. 하지만, 부분적으로 학습하는 만큼 전체적인 문맥 정보를 해석하는데 있어 어려움이 있다[16].

### 3.2 어텐션 유넷

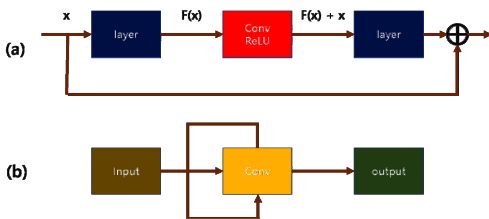
어텐션 유넷[6]은 유넷에 어텐션(Attention)을 추가로 적용한 네트워크로 본 실험에 적용된 네트워크의 구조는 (그림 1)의 (a)의 구조와 같다. 어텐션은 이전에 사용된 특징을 다시 한 번 주목하는 것으로 영상을 확대하는 디코더 부분에서 같은 단계의 인코더 영상을 그대로 가져와 사용할 뿐만 아니라 유넷과 달리 인코더 영상의 특징과 이전 단계의 디코더 영상의 특징에 어텐션 게이트(Gate) 연산을 적용해 나온 결과 값을 특징 디코더에서 다시 사용하는 방식으로 어텐션을 추가했다. 이때 어텐션 게이트의 구조는 (그림 2)와 같다. (그림 2)를 보면 인

코더 영상의 특징과 디코더 영상의 특징에 컨볼루션과 배치 정규화 연산을 적용한 값을 합한 뒤 특징을 추출하는 ReLU, 시그모이드(Sigmoid) 연산을 차례로 적용해 추출된 특징들에 전 단계의 디코더 특징을 결합해 나온 결과 값을 다음 디코더 연산에 사용하는 구조를 가진다. 이를 통해 이전의 인코더 특징과 디코더 특징을 어텐션하는 구조를 구현하였다.



(그림 2) 어텐션 게이트 구조  
(Figure 2) Attention Gate Structure

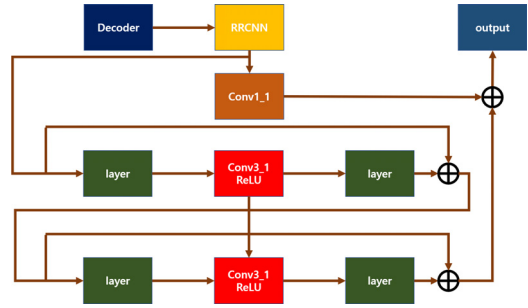
### 3.3 R2 유넷



(그림 3) R2 구조. (a)Residual, (b) Recurrent  
(Figure 3) R2 structure. (a) Residual, (b) Recurrent

R2는 Recurrent와 Residual를 뜻하는 용어로 (그림 3)의 (a)는 Residual의 구조를 (그림 3)의 (b)는 Recurrent의 구조를 가진다[7]. (그림 1)의 (b)는 R2 유넷으로 (그림 3)의 (a)와 (b) 두 가지를 동시에 유넷에 적용한 네트워크이다. R2 연산은 들어온 입력 값에 두 번의 Residual를 적용하고 그 사이에 Recurrent를 적용해 결과 값을 출력하는 구조로 구현했다. (그림 4)는 R2 연산을 구조화한 것으로 (그림 1)에서 사용되는 RRCNN 블록의 구조와 같다. 디코더에서만 어텐션 연산을 적용했던 어텐션 유넷과 달리 R2 유넷은 인코더와 디코

더의 모든 단계에서 R2 연산이 적용되어 최종적으로 생성되는 출력 값에 보다 많은 단계의 특징과 연산들이 사용되는 것이 특징이다.



(그림 4) 유넷에 적용된 R2 구조  
(Figure 4) R2 structure applied U-Net

### 3.4 어텐션-R2 유넷

어텐션-R2 유넷은 3.2절에서 설명한 어텐션과 3.3절에서 설명한 R2 연산을 동시에 적용한 네트워크로 (그림 1)의 (c)는 어텐션-R2 유넷의 전체적인 구조를 나타낸 그림이다. (그림 1)의 (c)는 (b)의 구조에 디코더 단계에서 어텐션 게이트 연산을 적용해 구현했다. 디코더 단계에서 특징을 추출할 때 RRCNN을 먼저 적용한 뒤, 그 결과 값과 같은 단계의 인코더에서 가져온 결과 값에 어텐션 게이트 연산을 적용해 디코더 단계에서의 최종 결과 값을 출력한다.

## 4. 실험

### 4.1 실험 환경

본 논문의 실험은 우분투 16.04 LTS 운영체제에서 Pytorch 프레임워크를 기반으로 진행하였다. 가중치 최적화 모델로는 Adams를 사용하였으며, 학습률은  $2e-4$ 로 설정했다. 학습 데이터의 크기는  $256 \times 256$ , 배치의 크기는 4, 에폭은 15로 설정해 학습을 진행하였다. 사용된 데이터 셋은 edges2shoes로 엣지 영상과 원본 영상이 한 쌍으로 이루어진 데이터 셋으로 학습 데이터에는 49825 개, 검증 데이터로는 200개를 사용했다.



(그림 5) 적용한 네트워크를 통해 생성된 영상. (a) 엣지 영상, (b) 원본 영상, (c) pix2pix, (d) 어텐션, (e) R2, (f) 어텐션-R2  
 (Figure 5) Generated Image applied by Network. (a) edges, (b) Ground Truth, (c) pix2pix, (d) Attention, (e) R2, (f) Attention-R2

#### 4.2 실험 결과

(그림 5)는 입력 값으로 들어간 엣지 영상과 정답 값으로 들어간 원본 영상, 3절에서 설명한 네트워크를 적용해 생성된 영상들을 나타낸 그림으로 각 네트워크가 얼마나 자연스러운 영상을 생성했는가를 확인할 수 있는 그림이다. (그림 5)를 보면 전반적으로 신발 영상이 깔끔하게 생성되는 것을 볼 수 있다. 특히 (5), (6)과 같이 신발 객체의 크기가 크고 엣지 영상의 엣지가 뚜렷할수록 pix2pix, 어텐션, R2, 어텐션-R2 네트워크 순으로 영상이 명료하고 자연스럽게 생성되는 것을 확인할 수 있다. 그러나 (1)과 같이 광택이 나는 소재의 신발 영상은 광택에 엣지가 생겨 R2, 어텐션-R2 네트워크를 적용해 생성된 영상은 정답 영상과 비교했을 때 새로운 색상이 생성되는 것을 확인할 수 있다.

또한, (2), (7), (8)과 같은 얇은 구두와 부츠, (3), (4)와 같은 샌들, 슬리퍼의 영상과 같이 신발 객체가 작은 영상의 경우에 신발 객체가 아닌 부분에 색상이 생성되는 것을 볼 수 있다. 이는 불필요한 데이터로 잡음(noise)에 속한다. 이러한 잡음이 많으면 영상을 처리하는데 있어 방해가 되기 때문에 잡음이 적은 영상을 생성하는 것은 영상 생성 분야에서 중요한 문제이다. (그림 6)은 네트워크를 적용해 생성된 영상들의 잡음의 정도를 확인할 수 있는 그림으로 (그림 6)의 홀수 번호는 (그림 5)의 (1), (4), (6), (8)과 같은 영상들이고 짝수 번호는 (그림 5)의 RGB 값을 얻어 R=255, G=255, B=255인 픽셀은 그대로 값을 유지하고 그 외 다른 값이 검출되면 검은 픽셀로 변환한 영상을 나타낸 것이다. (그림 6)을 보면 pix2pix와 어텐션 유넷의 경우에는 생성된 영상에 사람의 눈에 보이는 잡



(그림 6) 생성된 이미지들의 잡음. (a) pix2pix, (b) 어텐션, (c) R2, (d) 어텐션-R2  
 (Figure 6) Noise of Generated images (a) pix2pix, (b) Attention, (c) R2, (d) Attention-R2

음뿐만 아니라 눈에 보이지는 않지만 이미지를 처리할 때 방해되는 잡음이 존재하며 이와 비교해 R2 유넷과 어텐션-R2는 이러한 잡음이 훨씬 적게 생성되는 것을 확인할 수 있다. 특히 (그림 5)에서 눈에 보이는 잡음이 생긴 영상은 (그림 6)에서도 그 정도가 심한 것을 확인할 수 있다.

$$FID^2 = \|m_f - m_r\|_2^2 + Tr(C_f + C_r - 2(C_f C_r)^{1/2}) \quad (\text{수식 1})$$

실험 결과의 성능 평가에서는 GAN를 기반으로 하는 네트워크의 성능을 평가할 때 쓰이는 지표[17]인 FID(Frechet Inception Distance)[18]를 사용해 영상 생성 성능을 비교하였다. FID는 생성된 영상과 정답 영상의 특징을 비교해 그 거리를 측정하는 방식으로 수식 (1)과 같은 연산을 적용하는 방식이다. 표 1은 수식 (1)을 통해 구한 FID의 값을 나타낸 것으로 소수점 셋째자리까지 표기했으며 FID의 값이 작을수록 네트워크의 성능이 좋다고 판단할 수 있다. 표 1을 보면 pix2pix, 어텐션, R2, 어텐션-R2 순으로 값이 작아지는 것을 확인할 수 있으며 이를 통해 해당 순서로 네트워크의 성능이 개선된다고 판단할 수 있다.

(표 1) 생성된 영상들의 FID  
 (Table 1) FID of generated images

U-Net	pix2pix	어텐션	R2	어텐션-R2
FID	10.097	9.870	9.228	9.214

## 5. 결 론

본 논문에서는 다양한 네트워크 모델을 적용한 유넷을 pix2pix에 적용하고 그 차이를 확인 및 비교하였다. 기본적인 유넷에 비해 추가적인 연산을 가한 것이 더 좋은 결과를 보이지만 어텐션보다 R2 모델을 추가했을 때 더 좋은 성능을 보이는 것을 확인할 수 있으며, 어텐션-R2 유넷의 경우엔 R2 유넷에 비해 그 성능이 살짝 떨어지는 것을 확인할 수 있다. 이를 통해 어텐션과 같이 단계 차이가 많이 나는 인코더 혹은 디코더의 특징을 사용하기 보단 R2 개념과 같이 직전의 결과를 사용하는 것이 자연스러운 영상을 생성하는 것에 좋은 영향을 미치는 것으로 본다. 그러나 (그림 6)을 보면 성능이 좋은 R2 유넷과 어텐션-R2 유넷 또한 잡음이 생성되는 경우가 있어 후처리를 진행하지 않는 한 온전한 이미지로 사용되기에는 어렵다는 한계를 보여주고 있다. 향후 연구에서는 생성된 영상들을 바로 다른 영상 처리에 사용 가능하도록 R2

유넷을 기반으로 잡음을 최소화하는 것에 대한 연구를 수행하고자 한다.

## 참고문헌(Reference)

- [1] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. "Context encoders: Feature learning by inpainting.", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016. <https://doi.org/10.1109/cvpr.2016.278>
- [2] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation." In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015. <https://doi.org/10.1109/cvpr.2015.7298965>
- [3] G. Larsson, M. Maire, and G. Shakhnarovich. "Learning representations for automatic colorization.", In Computer Vision - ECCV ,pp. 577 - 593, 2016. [https://doi.org/10.1007/978-3-319-46493-0\\_35](https://doi.org/10.1007/978-3-319-46493-0_35)
- [4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE., 2017.
- [5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", In Lecture Notes in Computer Science, pp.234-241, 2015. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [6] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, Daniel Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas", In MIDL, arXiv preprint arXiv:1804.03999, 2018. <https://arxiv.org/abs/1804.03999>
- [7] Md Zahangir Alom, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari, "Nuclei Segmentation with Recurrent Residual Convolutional Neural Networks based U-Net (R2U-Net)", In NAECON - IEEE National Aerospace and Electronics Conference. IEEE, 2018. <https://doi.org/10.1109/naecon.2018.8556686>
- [8] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, "Generative Adversarial Networks", In NIPS, arXiv preprint arXiv:1406.2661, 2014. <https://arxiv.org/pdf/1406.2661.pdf>
- [9] R. Mu, X. Zeng, "A Review of Deep Learning Research", In TIIS, Vol. 13, No.4, 2019. <https://doi.org/10.3837/tiis.2019.04.001>
- [10] Alec Radford & Luke Metz, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", In CoRR, arXiv preprint arXiv:1511.06434, 2016. <https://arxiv.org/pdf/1511.06434.pdf>
- [11] Mehdi Mirza, Simon Osindero, "Conditional Generative Adversarial Nets", In CoRR, arXiv preprint arXiv:1411.1784, 2014. <https://arxiv.org/pdf/1411.1784.pdf>
- [12] Hee-jin Yoon, Kang-jik Kim, Jun-chul Chun, "GAN-based shadow removal using context information", Journal of Internet Computing and Services, Vol. 20, No. 6, pp. 29-36, 2019. <https://doi.org/10.7472/jksii.2019.20.6.29>
- [13] Yao Wang, Woojin Jeong, Young Shik Moon, "Single Image Dehazing Based on Depth Map Estimation via Generative Adversarial Networks", Journal of Internet Computing and Services, Vol. 19, No. 5, pp. 43-54, 2018. <https://doi.org/10.7472/jksii.2018.19.5.43>
- [14] Eirikur Agustsson, Michael Tschannen, Fabian Mentzer, Radu Timofte, Luc Van Gool, "Generative Adversarial Networks for Extreme Learned Image Compression", The IEEE International Conference on Computer Vision (ICCV), pp. 221-231, 2019. <https://arxiv.org/pdf/1804.02958.pdf>
- [15] Liqun Chen, Yizhe Zhang, Ruiyi Zhang, Chenyang Tao, Zhe Gan, Haichao Zhang, Bai Li, Dinghan Shen, Changyou Chen, Lawrence Carin, "Improving Sequence-to-Sequence Learning via Optimal Transport", In ICLR, 2019.

- <https://arxiv.org/pdf/1901.06283.pdf>
- [16] Zhang, H., Sindagi, V., & Patel, V. M. "Image De-raining Using a Conditional Generative Adversarial Network". IEEE Transactions on Circuits and Systems for Video Technology, 1 - 1. 2019.  
<https://doi.org/10.1109/tcsvt.2019.2920407>
- [17] Han Zhang, Ian Goodfellow, Dimitris Metaxas, Augustus Odena, "Self-Attention Generative Adversarial Networks", In CoRR, 2018.  
<https://arxiv.org/pdf/1805.08318.pdf>
- [18] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S, "Gans trained by a two time-scale update rule converge to a local nash equilibrium", In Advances in Neural Information Processing Systems, pp. 6626 - 6637, 2017.  
<https://arxiv.org/pdf/1706.08500.pdf>

## ◎ 저 자 소 개 ◎



**임 소 현(So-hyun Lim)**

2020 B.S. in Computer Engineering, Kyonggi University, Suwon, Korea  
2020~Present : M.S. Student in Computer Science, Kyonggi University, Suwon, Korea  
Research Interests : Image Processing, Computer Vision  
E-mail : gip.shlim1017@gmail.com



**전 준 철(Junchul Chun)**

1984 B.S. in Computer Science, Chung-Ang University, Seoul, Korea  
1986 M.S. in Computer Science(Software Engineering), Chung-Ang University, Seoul, Korea  
1992 M.S. in Computer Science and Engineering (Computer Graphics), The Univ. of Connecticut, USA  
1995 Ph.D. in Computer Science and Engineering (Computer Graphics), The Univ. of Connecticut, USA  
2001.02~2002.02 Visiting Scholar, Michigan State Univ. Pattern Recognition and Image Processing Lab.  
2009.02~2010.02 Visiting Scholar, Univ. of Colorado, Wellness Innovation and Interaction Lab.  
1995~Current Professor at Dept of Computer science, Kyonggi University.  
Research Interests : Augmented Reality, Computer Vision, Human Computer Interaction  
E-mail : jchun@kgu.ac.kr