

# 생성적 적대 신경망을 활용한 부분 위변조 이미지 생성에 관한 연구

송성헌\* · 최봉준\*\* · 문미경\*\*\*

## A Study on Image Creation and Modification Techniques Using Generative Adversarial Neural Networks

Seong-Heon Song\* · Bong-Jun Choi\*\* · M-Ikyeong Moon\*\*\*

### 요 약

생성적 적대 신경망(Generative Adversarial Networks, GAN)은 내부의 두 신경망(생성망, 판별망)이 상호 경쟁하면서 학습하는 네트워크이다. 생성자는 현실과 가까운 이미지를 만들고, 구분자는 생성자의 이미지를 더 잘 감별하도록 프로그래밍 되어있다. 이 기술은 전체 이미지 X를 다른 이미지 Y로 생성, 변형 및 복원하기 위해 다양하게 활용되고 있다. 본 논문에서는 원본 이미지에서 부분 이미지만 추출한 후, 이를 자연스럽게 다른 객체로 위변조할 수 있는 방법에 대해 기술한다. 먼저 원본 이미지에서 부분 이미지만 추출한 후, 기존에 학습시켜놓은 DCGAN 모델을 통해 새로운 이미지를 생성하고, 이를 전체적 스타일 전이(overall style transfer) 기술을 사용하여 원본 이미지의 질감과 크기에 어울리도록 리스타일링(re-styling) 한 후, 원본 이미지에 자연스럽게 결합하는 과정을 거친다. 본 연구를 통해 원본 이미지의 특정 부분에 사용자가 원하는 객체 이미지를 자연스럽게 추가/변형할 수 있음으로써 가짜 이미지 생성의 또 다른 활용 분야로 사용될 수 있을 것이다.

### ABSTRACT

A generative adversarial network (GAN) is a network in which two internal neural networks (generative network and discriminant network) learn while competing with each other. The generator creates an image close to reality, and the discriminator is programmed to better discriminate the image of the constructor. This technology is being used in various ways to create, transform, and restore the entire image X into another image Y. This paper describes a method that can be forged into another object naturally, after extracting only a partial image from the original image. First, a new image is created through the previously trained DCGAN model, after extracting only a partial image from the original image. The original image goes through a process of naturally combining with, after re-styling it to match the texture and size of the original image using the overall style transfer technique. Through this study, the user can naturally add/transform the desired object image to a specific part of the original image, so it can be used as another field of application for creating fake images.

### 키워드

Generative Adversarial Networks, DCGAN, Machine Learning, Unsupervised Learning, Image, Computer Vision, OpenCV  
생성적 적대 신경망, DCGAN, 머신 러닝, 비지도 학습, 이미지, 컴퓨터 비전, OpenCV

\* 동서대학교 소프트웨어학과 학부연구원 (ssh97540@gmail.com)

\*\* 동서대학교 소프트웨어융합대학 교수 (bjchoi@dongseo.ac.kr)

\*\*\* 교신저자 : 동서대학교 소프트웨어학과

· 접수일 : 2022. 03. 07

· 수정완료일 : 2022. 03. 27

· 게재확정일 : 2022. 04. 17

· Received : Mar. 07, 2022, Revised : Mar. 27, 2022, Accepted : Apr. 17, 2022

· Corresponding Author : Mi-Kyeong Moon

College of Software Convergence, Dongseo University,

Email : mkmoon@dongseo.ac.kr

## 1. 서 론

생성적 적대 신경망(Generative Adversarial Networks, GAN)은 딥러닝 모델 중 이미지 생성에 널리 쓰이는 모델이다[1]. 기본적인 딥러닝 모델인 CNN(: Convolutional Neural Network)은 미리 학습된 모델을 기반으로 이미지를 분류하고 인식하는 문제에 사용되었다[2-3]. 이와 다르게 GAN은 두 개의 딥러닝 (생성자, 판별자)이 서로 경쟁하며 학습하면서 새로운 유사한 이미지를 만들도록 한다. 생성자는 자신이 만든 이미지를 판별자가 진짜라고 판별할 수 있을 정도의 진짜 같은 가짜 이미지를 만들도록 학습된다. 판별자는 생성자가 만든 이미지를 가짜라고 판별하고 실제 이미지를 진짜라고 판별할 수 있도록 학습된다. GAN은 노이즈 입력을 받아 원하는 범주의 새로운 이미지를 생성해내거나 입력 이미지를 다른 형태의 이미지 또는 동영상으로 변환하는 기술로 계속 발전하고 있다[4-6]. 최근에는 이미지를 손쉽게 편집하거나 그럴 수 있는 프로그램으로 서비스되어 실제 디자인 분야에서도 활용되고 있다.

노이즈를 이미지로 변환하는 대표적인 기술인 DCGAN(: Deep Convolutional GAN)은 지도학습 알고리즘인 CNN을 GAN에 적용한 것이다[7]. 이는 입력 노이즈의 잠재공간(latent space)을 변화시켜서 원하는 이미지를 다양하게 생성할 수 있었으며, GAN보다 좋은 성능을 보여주었다. 이미지를 새로운 이미지로 생성하는 기술에는 Pix2PixHD[8]와 SPADE[9], LostGANs[10] 등이 있다. 이들은 이미지의 테두리(edge) 정보만을 이용하여 이미지를 복원하거나 배치(layout) 정보만으로 이미지를 생성할 수 있는 기술이다. CycleGAN[11]은 하나의 이미지를 고희, 모네 등의 특정 화풍으로 변형시키기 위해 활용되는 알고리즘이다. 이미지의 부분적인 변형을 할 수 있는 기술인 StarGAN[12]은 사람의 머리색, 피부색과 같이 한 개만 변화시키고 싶을 때 사용할 수 있는 기술이다. 이와 유사한 StyleGAN[13]은 사람의 머리색, 나이, 성별 등의 style 정보를 적용하여 이미지를 변환할 수 있는 기술이다. 이러한 기술들은 모두 전체 이미지를 다른 이미지로 생성, 변환 및 복원하기 위해 활용되고 있다.

본 논문에서는 원본 이미지에서 부분 이미지 객체

만 추출하여 다른 이미지 객체로 자연스럽게 위변조할 수 있는 방법에 대해 기술한다. 먼저 원본 이미지에서 수정하고자 하는 부분 이미지를 도려내어 학습된 인공신경망을 활용해 새로운 이미지를 생성한다. 생성된 이미지를 전체적 스타일 전이(overall style transfer) 기술[14]로 원본 이미지의 질감과 크기에 맞게 리스타일링(re-styling) 과정을 거친다. 이러한 과정을 거쳐 만들어진 이미지를 컴퓨터 비전의 한 오픈소스 라이브러리인 OpenCV를 활용하여 원본 이미지에서 추출한 부분에 자연스럽게 결합시킨다. 본 연구를 통해서 기존 연구들과 다르게 원본 이미지의 특정 객체 이미지가 다른 객체 이미지로 자연스럽게 대체될 수 있도록 함으로써 가짜 이미지 생성의 또 다른 활용 분야로 사용될 수 있을 것이다.

## II. 관련연구

### 2.1 GAN Paint

현재 GAN을 활용하여 이미지의 일부분을 변형하는 연구들이 다양하게 진행되고 있다. 그중 2019년 발표된 연구 [15]에서는 GAN Dissection을 활용하여 단위(Unit), 물체(Object) 및 장면(Scene) 수준에서 GAN을 시각화하고 이해하기 위한 분석 프레임워크를 제시하였다. 먼저 분할기반 네트워크(segmentation-based Network)를 사용하여 물체 개념(object concepts)에 관련된 해석 가능한 단위들을 식별한다. 그리고 출력된 물체, 사라진 물체를 컨트롤할 수 있는 단위 셋(Set of units)을 식별하여 인과적 효과를 정량화한다. 마지막으로 임시적 물체 단위(casual object units)와 생성된 물체 사이의 맥락적 관계를 조사한다. 위 연구에서는 이러한 프레임워크를 사용하여 GAN Paint 프로그램을 제작하였다. 특정 물체를 선택하여 사진의 일부분을 드래그하면 해당 위치에 어울리게 물체가 생성된다. 본 연구는 관련연구와 같이 부분 이미지를 생성하는 것뿐만 아니라 원본 이미지의 일부분을 없앤 후, 새로운 부분 이미지를 원본 이미지 질감과 크기에 어울리도록 결합하는 과정을 다루고 있는 점에서 차이점이 있다.

## 2.2 이미지 복원 (Image Inpainting)

이미지 복원(Image Inpainting)은 컴퓨터 비전 분야 중 하나로 원본 이미지로부터 가려진(Masking) 부분들을 주위 이미지로 생성하여 복원하는 기술이다. NVIDIA에서 발표된 연구 [16]에 따르면 기존 이미지 복원 기술은 원본 이미지에 가려진 부분이 있는 경우 비어있는 픽셀의 값을 주변의 값으로 바탕을 채워나간다. 위 연구에서는 기존에 사용하지 않던 테두리 연결(edge connection)을 추가하여 연구의 성능을 높였다. 빠른 테두리 탐지(canny edge detection)를 활용하여 가려진 부분의 테두리를 찾아내 결합하여 좀 더 세부적으로 복원한 이미지를 만들어낸다. 공개된 데이터 셋인 CelebA, Places2, ParisStreetView를 사용하여 연구를 진행하였으며 정량적인 성능 평가를 통해 기존 논문보다 좋은 성능을 보였다.

또한 오픈소스 라이브러리인 OpenCV에서는 두 가지의 이미지 복원 알고리즘을 제공한다. 첫 번째 알고리즘은 Alexandru Telea의 논문 [17]을 기반으로 한 것이다. 이는 영역의 경계로부터 시작하여 내부를 외곽선부터 점진적으로 채우는 기술로서, 복원 지점부터 정규화를 통해 구해진 가중치 합으로 픽셀이 교체된다. 두 번째 알고리즘은 Marcelo Bertalmío의 논문 [18]을 기반으로 한다. 이는 부분 편차 방정식과 유체 역학에 기반을 두어 휴리스틱 방식을 사용하여 이미지 복원을 구현하였다. 본 논문에서는 이미지의 일부분의 물체를 지우는 것뿐만 아니라 새로운 부분 이미지를 자연스럽게 추가 및 변형함으로써 가짜 이미지를 생성한다는 점에서 차이점이 있다.

## III. 생성적 적대 신경망을 활용한 부분 위변조 이미지 생성

본 논문에서 제안하는 생성적 적대 신경망을 활용한 부분 위변조 이미지 생성 기법의 목표는 원본 이미지의 일부분을 생성된 새로운 이미지로 수정하는데 있어 이미지를 자연스럽게 정확하게 병합하는 것이다. 이를 위해 이미지 전처리 기법을 활용하여 이미지의 원활한 학습 및 스타일 재건을 효과적으로 할 수 있게 한다. 스타일 재건에 있어 스타일은 생성된 이미지의 배경(전체적인 스타일) 뿐만 아니라 질감(구

체적인 스타일)까지를 뜻한다.

그림 1은 생성적 적대 신경망을 활용한 부분 위변조 이미지 생성과정을 보여주는 그림이다. 원본 이미지(Original Image)에서 부분 이미지 A를 추출한다. 추출한 부분 이미지 A를 기반으로 잠재공간 Z로 인코딩한다. 학습된 DCGAN을 기반으로 새로운 이미지 B를 생성한다. 원본 이미지를  $N \times M$ 으로 나누어 랜덤으로 이미지 C를 추출한다. 생성된 이미지 B와 추출된 이미지 C를 기반으로 스타일 전이(style transfer)를 거쳐 이미지의 질감을 변형시킨 이미지 D를 생성한다. 배경과 질감이 변형된 이미지 D는 원본 이미지의 일부분으로 대체하여 들어가게 된다.

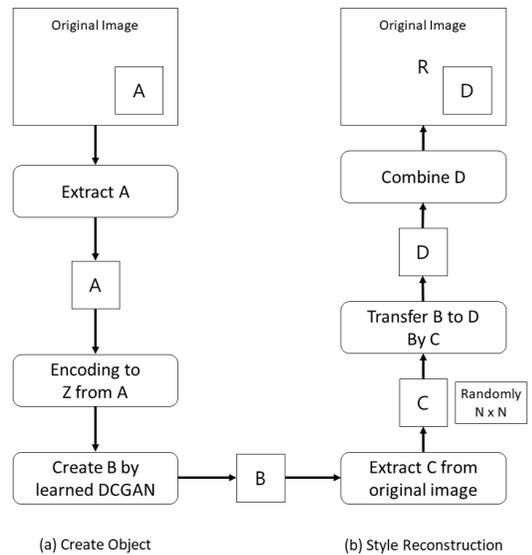


그림 1 부분 위변조 이미지 생성기 구조  
Fig. 1 Partially forged image generator structure

### 3.1 물체 이미지 생성

본 논문에서는 새로운 부분 이미지를 생성하기 위하여 생성적 적대 신경망을 확장시킨 DCGAN을 활용한다. 물체 이미지 생성을 위해 토마토 이미지 5,103개, 파프리카 이미지 2,478개, 사과 이미지 1,440개 등 과일 이미지들을 학습에 사용하였다. 배경이 없는 과일 이미지를 생성하기 위해 OpenCV의 한계점(Thresholding)을 활용하여 배경 제거를 한 후 과일 이미지들을 데이터로 사용하였다. 새로운 물체 이미지

를 생성하기 위해 GAN의 생성자 모델과 판별자 모델을 구축한다. 구축된 모델을 통해 하나의 과일 이미지에 대해 64개의 서로 다른 이미지가 생성되며, 이 중 무작위로 하나의 이미지를 사용한다. 생성된 이미지는 스타일을 재건(Reconstruction) 시키기 전 병합될 위치에 존재하는 배경을 생성된 이미지에 덮어주기 위한 이미지 전처리를 진행한다. 본 논문의 연구에서 이미지 전처리 기법으로 OpenCV의 Thresholding, Bitwise Operation을 이용한다.

그림 2는 본 논문의 연구에서 사용한 새로운 부분 이미지 배경 추가에 대한 과정을 보여준다. 먼저 그림 2의 A, B는 한계점(Thresholding)을 활용하여 새로운 물체의 배경을 모두 제거한 뒤 검은색으로 마스크(mask) 처리된 것을 보여준다. 그림 2의 C, D는 Bitwise Operation으로 마스크 된 물체 외의 부분을 마스크 처리한 후 원본 이미지의 배경과 결합한 것을 보여준다. 그림 2의 E는 A과 D를 합성한 것이다. 이 방법을 사용하기 위해서는 새로운 물체 이미지와 원본 이미지의 부분 객체 이미지 크기를 bitwise 연산을 처리하여 크기를 동일하게 한다.

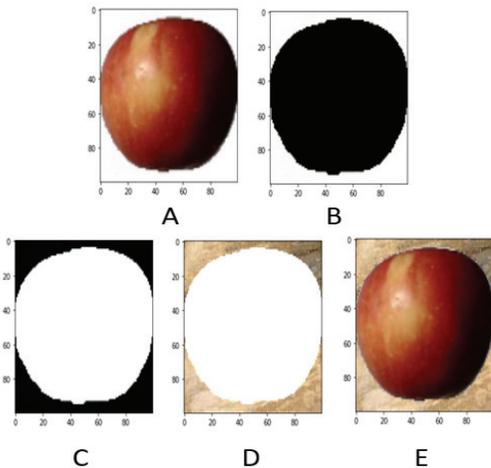


그림 2 본 연구에 사용된 이미지 전처리를 활용한 배경 추가 과정  
 Fig. 2 The process of adding a background using image preprocessing used in this study

### 3.2 스타일 재건

스타일 재건은 새로 생성된 부분 객체 이미지의 배경 및 질감을 원본 이미지와 유사하게 하여 전혀 어색함 없이 자연스럽게 원본 이미지에 융화될 수 있도록 하는 것이다. 그림 3의 A는 전처리 기법을 활용하여 배경이 추가된 생성 이미지(콘텐츠 이미지)이고, 그림 3의 B는 스타일 재건을 위해 사용될 원본 이미지의 일부분(스타일 이미지)이다. 이 두 개의 이미지를 결합하기 위하여 전체적 이미지 전이 학습을 진행한다. 이 두 이미지는 pretrained vgg19를 기반으로 만들어진 생성자 모델을 통해 스타일이 변경된다. 그림 3의 C는 생성자를 통해 스타일이 변경된 이미지이다.

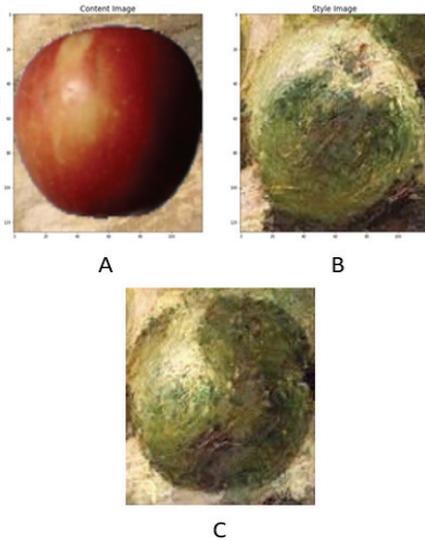


그림 3 스타일 재건 결과  
 Fig. 3 Result of the style reconstruction

수식 (1)은 스타일 이미지를 추출하기 위해 사용된 수식이다.

$$I(wh) = \left\{ I(w_i, h_j) : 0 \leq i < \frac{w_r}{w_c}, 0 \leq j < \frac{h_r}{h_c} \right\} \quad (1)$$

스타일 재건에 사용될 스타일 이미지는 원본 이미지의 너비  $w_r$  \* 높이  $h_c$ 을 생성된 부분 이미지 너비  $w_c$  \* 높이  $h_r$ 로 등분한 후, 이 중 무작위로 이미지를 추출한다. 효과적인 스타일 재건 이미지를 생성하기 위해 여러 개의 스타일 이미지를 추출하여 n 번 반복하여 생성한다. 콘텐츠 이미지는 스타일 이미지와 크기가 동일하게 하기 위해 OpenCV의 이미지 크기조

절(Image Resizing) 함수를 활용하여 이미지 사이즈를 조정한다. 스타일이 변형된 이미지를 생성한 후 원본 이미지의 추출한 부분에 이미지를 결합한다.

#### IV. 결과 및 실험 평가

본 논문의 연구를 구현하기 위해 프로세서는 i7-9700, 메모리는 32G, 그래픽 카드는 Tesla T4가 장착된 컴퓨터에 Ubuntu 18.04.5 LTS를 설치하여 사용하였다. 또한, 빠른 학습을 위해 GPU 사용을 위한 CUDA 11.0을 설치하였으며 Python 3.7에서 Pytorch를 사용하여 연구를 진행하였다.

##### 4.1 스타일 이미지에 따른 생성 이미지 차이

그림 4는 스타일 이미지에 따라 스타일 재전 성능을 비교한 것이다. 그림 4의 A는 스타일 이미지를 원본 이미지의 일부분으로 지정한 것이다. 그림 4의 B는 스타일 이미지를 원본 이미지 전체로 지정한 것이다. 그림 4의 C는 수식 (1)을 활용하여 임의의 이미지를 반복 지정한 것이다. 이는 A, B에 비해 C에서 이미지의 질감 변화가 효과적임을 보여준다. 스타일 재전을 위하여 원본 이미지 전체 중 일부분을 랜덤하게 반복하여 스타일 이미지를 지정함으로써 자연스럽게 결합 될 수 있음을 확인하였다.

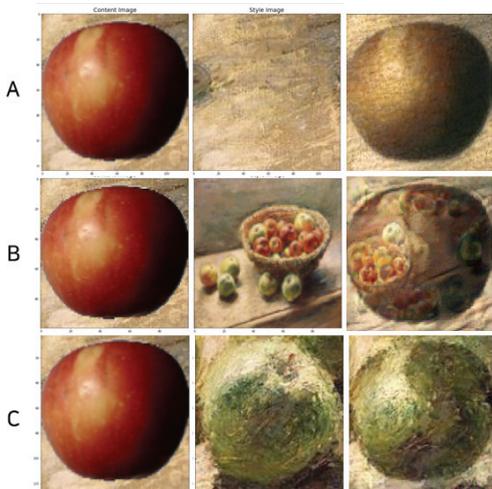


그림 4 Style 담당 이미지에 따른 결과 변화  
Fig. 4 Changes in results according to the image in charge of the style

##### 4.2 부분 위변조 이미지 생성

본 논문의 연구에서는 실험 이미지로 클로드 모네의 사과 한 그릇 그림을 사용하였다. 그림 5의 위는 원본 이미지이며, 바구니 안의 사과 하나(네모 박스)를 위변조 하려 한다. 그림 5의 아래는 사과 이미지가 토마토 이미지로 변형된 결과를 보여준다. 직관적으로 보았을 경우 토마토 이미지가 주변에 존재하는 사과들의 색감과 질감, 배경 등과 유사하게 생성해 내는 것을 볼 수 있다.

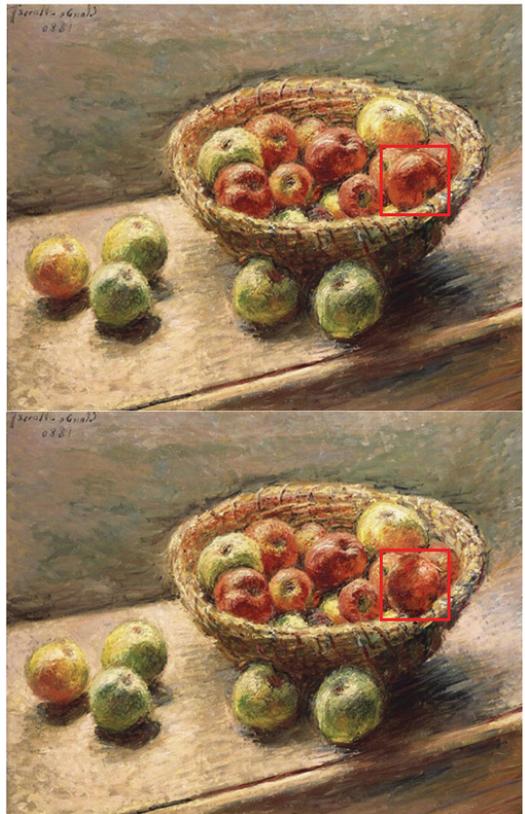


그림 5 부분 위변조 이미지 생성 결과  
Fig. 5. Results of partial forgery and alteration image generation

#### V. 결론

본 논문에서는 이미지 전체를 다른 이미지로 생성하는 방식과 다르게 이미지 일부분을 다른 부분 객체 이미지로 위변조하는 방법에 대해 기술하였다. 위변조

할 부분 이미지는 DCGAN 모델을 통해 학습을 시킨 후 생성하였다. 생성된 이미지를 전체적 이미지 전이 기술로 원본 이미지의 질감과 크기에 맞게 리스타일링 과정을 거쳤으며, 이러한 과정을 거쳐 만들어진 이미지를 컴퓨터 비전의 오픈소스 라이브러리인 OpenCV를 활용하여 원본 이미지의 추출한 부분에 자연스럽게 결합시켰다. 실험을 통해 기존의 사과 이미지가 원본 이미지에 어색함 없이 토마토 이미지로 변형됨을 볼 수 있었다. 본 연구를 통해 원본 이미지의 특정 부분에 사용자가 원하는 객체 이미지를 자연스럽게 추가/변형할 수 있음으로써 가짜 이미지 생성의 또 다른 활용 분야로 사용될 수 있을 것이다. 향후 연구에서는 정적인 이미지뿐만 아닌 동영상에서도 부분 객체가 변형되어 대체되는 방법에 대해 연구가 진행될 것이다.

#### 감사의 글

이 논문은 2021년도 동서대학교 "Dongseo Cluster Project" 지원에 의하여 이루어진 것임 (DSU-20210007)

#### References

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, A. Ozair, S. Courville, and Y. Bengio, "Generative adversarial nets," *Report*, 2014.
- [2] Y. Jeong and G. Choi, "Efficient iris recognition using deep-learning convolution neural network(CNN)," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 15, no. 3, 2020, pp. 521-526.
- [3] J. Yoo, "A Design of Small Scale Deep CNN Model for Facial Expression Recognition using the Low Resolution Image Datasets," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 16, no. 1, 2021, pp. 75-80.
- [4] C. Donahue, J. Macauley, and M. Puckette. "Adversarial audio synthesis," *Report*, 2018.
- [5] S. Huang, Q. Li, C. Anil, X. Bao, S. Oore, and R. Grosse, "Timbretron: A wavenet (cyclegan (cqt (audio))) pipeline for musical timbre transfer," *Report*, 2018.
- [6] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," *Information Fusion*, vol. 64, 2020, pp. 131-148.
- [7] A. Radford, L. Metz, and S. Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks," *Report*, 2015.
- [8] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, June 2017, pp. 5967-5976.
- [9] T. Park, M. Liu, T. Wang, and J. Zhu, "Semantic image synthesis with spatially-adaptive normalization," In *Proc. of the IEEE/CVF conference on computer vision and pattern recognition*, Long Beach USA. June 2019, pp. 2337-2346.
- [10] W. Sun and T. Wu. "Image synthesis from reconfigurable layout and style," In *Proc. of the IEEE/CVF International Conference on Computer Vision*, Seoul, Korea, Oct. 2019.
- [11] J. Y. Zhu, T. Park, P. Isola, and A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," In *Proc. of the IEEE international conference on computer vision*, Venice Italy, Oct. 2017, pp. 2223-2232.
- [12] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," In *Proc. of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, USA, June 2018, pp. 8789-8797.
- [13] T. Karras, S. Laine, and T. Aila. "A style-based generator architecture for generative adversarial networks," In *Proc. of the IEEE/CVF conference on computer vision and pattern recognition*, Long Beach USA. June 2019, pp. 4401-4410.
- [14] L. A. Gatys, A. S. Ecker, and M. Bethge. "Image

style transfer using convolutional neural networks," In *Proc. of the IEEE conference on computer vision and pattern recognition*, Las Vegas, USA, June 2016, pp.2414-2423.

[15] Bau, D., J. Zhu, H. Strobel, B. Zhou, J. Tenenbaum, W. T. Freeman, and A. Torralba, "Gan dissection: Visualizing and understanding generative adversarial networks," *Report*, 2018.

[16] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, "Edgeconnect: Generative image inpainting with adversarial edge learning," *Report*, 2019.

[17] A. Telea. "An image inpainting technique based on the fast marching method," *Journal of graphics tools*, vol. 9, no. 1, 2004, pp. 23-34.

[18] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting." In *Proc. of the 27th annual conference on Computer graphics and interactive techniques*, New Orleans, USA, July 2000, pp.417-424.



**문미경(Mi-Kyeong Moon)**

1990년 이화여자대학교 전자계산학과 졸업(이학사)  
 1992년 이화여자대학교 대학원 전자계산학과 졸업(이학석사)  
 2005년 부산대학교 대학원 컴퓨터공학과 졸업(공학박사)  
 2008년 동서대학교 컴퓨터공학부 교수  
 ※ 관심분야 : 소프트웨어공학, AI융합기술 응용

저자 소개



**송성현(Seong-Heon Song)**

2016~ 동서대학교 컴퓨터공학부  
 ※ 관심분야 : 생성모델, 컴퓨터 비전, 자연어처리



**최봉준(Bong-Jun Choi)**

2013년 동서대학교 컴퓨터정보공학부 졸업(공학사)  
 2015년 연세대학교 대학원 컴퓨터과학과 졸업(공학석사)

2014년~2019년 LG전자 CTO부문 근무  
 2019년~2021년 Aalborg University OR Lab 근무  
 2021년 동서대학교 소프트웨어융합대학 교수  
 ※ 관심분야 : 자연어처리, 데이터생성, 딥러닝

