한국정보전자통신기술학회논문지(jkiiect)'22-10, Vol.15 No.5
논문 22-15-05-360 http://dx.doi.org/10.17661/jkiiect.2022.15.5.360

Pix2Pix의 수용 영역 조절을 통한 전통 고궁 이미지 복원 연구

황원용* 김효관**

A Study on the Restoration of Korean Traditional Palace Image by Adjusting the Receptive Field of Pix2Pix

Won-Yong Hwang* Hyo-Kwan Kim**

요 약 본 논문은 흑백 사진으로만 남아 있는 한국의 전통 고궁 사진을 적대적 생성 신경망 기법의 하나인 Pix2Pix를 활용하여 컬러 사진으로 복원하기 위한 학습 모델 구조를 제시한다. Pix2Pix는 합성 이미지를 생성기와 합성 여부를 판정하는 판별기의 학습 모델 조합으로 구성된다. 본 논문은 판별기의 수용 영역을 조절하여 인공지능 모델을 학습하고 그 결과를 고궁 사진이 가지는 특성을 고려하여 분석하는 내용을 다룬다. 기존에 흑백 사진 복원에 사용하는 Pix2Pix의 수용 영역은 주로 고정된 크기로 사용하였으나 이미지의 변화가 다양한 고궁 사진을 복원함에 있어서는 고정된 수용 영역을 일률적으로 적용하기에 적합하지 않다. 본 논문에서는 고궁의 특성을 반영할 수 있는 판별기의 수용 영역을 확인하기 위해 기존의 고정된 수용 영역의 크기를 변화시켜 나타나는 결과를 관찰하였다. 실험은 사전에 준비한 고궁 사진을 기반으로 판별기의 수용 영역을 조정하고 모델의 학습을 진행하였다. 판별기의 수용 영역 변화에 따른 모델의 손실을 측정하고 최종 학습한 학습 모델을 복원 대상 흑백 사진에 대입하여 복원 결과를 확인한다.

Abstract This paper presents a AI model structure for restoring Korean traditional palace photographs, which remain only black-and-white photographs, to color photographs using Pix2Pix, one of the adversarial generative neural network techniques. Pix2Pix consists of a combination of a synthetic image generator model and a discriminator model that determines whether a synthetic image is real or fake. This paper deals with an artificial intelligence model by adjusting a receptive field of the discriminator, and analyzes the results by considering the characteristics of the ancient palace photograph. The receptive field of Pix2Pix, which is used to restore black-and-white photographs, was commonly used in a fixed size, but a fixed size of receptive field is not suitable for a photograph which consisting with various change in an image. This paper observed the result of changing the size of the existing fixed a receptive field to identify the proper size of the discriminator that could reflect the characteristics of ancient palaces. In this experiment, the receptive field of the discriminator was adjusted based on the prepared ancient palace photos. This paper measure a loss of the model according to the change in a receptive field of the discriminator and check the results of restored photos using a well trained AI model from experiments.

Key Words : Discriminator, GAN, Pix2Pix, Receptive field, Restoring photo

1. 서론

적대적 생성 신경망은 주어진 이미지 또는 영상을 기반으로 두 개의 딥러닝 모델이 서로 경쟁적으로 학습하여 실존과 유사한 데이터를 합성 및 생성하는 강화 학습 방법 중에 하나로 딥 러닝의 한 분야이다[1]. 기계 학습은 수치적 분석과 예측에 많이 적용되는 반면에 딥 러닝은 음성과 영상 등 방대한 데이터를 기반

* Corresponding Author: Department of Fintech Korea Polytechnics (wyhwang@kopo.ac.kr)
** Department of Fintech, Korea Polytechnics
Received September 23, 2022 Revised October 08, 2022 Accepted October 19, 2022

으로 하는 예측에 주로 적용하고 있다[2]. 적대적 생성 신경망(Generative Adversarial Networks)은 현재 사회적으로 문제가 되는 유명인사의 가짜 합성 이미지 및 영상 제작에 악용되는 문제가 있지만, 이미지 화질 개선, 워터마크 제거, 흑백 사진의 복원, 이미지 채색 등 기술의 효용성을 확인한 분야도 많다. 적대적 생성 신경망은 두 개의 학습 모델을 사용하는데, 하나는 거짓의 합성 데이터를 생성하는 모델로 생성기(Generator)라 부르며 최대한 진짜 같은 데이터를 생성하도록 학습을 진행한다. 다른 하나의 모델은 판별기(Discriminator)라고 부르며 생성기에서 생성한 합성된 데이터의 조작 여부를 판단하도록 학습을 진행한다. 이 두개의 모델이 서로 경쟁하듯 학습을 진행하게 되면 이상적으로는 생성자에서 생성한 가짜 이미지를 판별기에서 진짜 이미지로 판독을 하게 되고, 이때 생성되는 데이터는 육안으로 확인하였을 때 그럴듯한 이미지나 영상으로 표출되는 것이다. 본 논문에서는 적대적 생성 신경망의 여러 기술 중 Pix2Pix를 기반으로 흑백으로 남아있는 한국의 고궁 사진을 컬러사진으로 복원할 때 판별기에 적용되는 수용영역(receptive field)의 조절을 통해 변화되는 결과에 대해 살펴보고자 한다.

2. Pix2Pix 개요

Pix2Pix는 적대적 생성 신경망의 기술 중 하나로 이미지를 Patch라는 단위로 훈련하는 개념을 적용하였다. 본 논문에서는 Pix2Pix 아키텍처 중 판별기의 수용 영역을 조절하는 관점에서 연구를 진행하였다.

2.1 Pix2Pix 적용 방안

Pix2Pix는 입력 이미지와 출력 이미지가 하나의 짝으로 구성되는 데이터가 모델 학습의 전제 조건이다. 가짜의 합성 이미지를 생성하는 생성기(Generator)는 U-Net 아키텍처를 활용하며, 판별기(Discriminator)는 생성기에서 합성한 이미지의 진위여부를 판별하게 된다[3]. 본 논문에서는 흑백 고궁 사진과 컬러 고궁 사진을 하나의 짝으로 구성하여 생성기와 판별기를 학습시킨다. 이후, 학습한 모델을 통하여 현재는 소실되어 흑백 사진으로만 존재하는 고궁 사진을 채색함에 있어, 판별기의 파라미터 조절을 통해서 도출된 결과를 다뤄볼 것이다. Pix2Pix의 생성기와 판별기가 학습하는 구조는 아래 그림 1과 같다. 판별기는 2개의 파라미터를 받아서 진위 여부를 판단하는데, 첫 번째 파라미터는 입력(흑백) 이미지이며, 두 번째 파라미터는 생성기에서 합성한 이미지 또는 입력(흑백)과 짝을 이루는 정답지인 출력(컬러) 이미지이다. 그림 1의 좌측 하단에서는 생성기에서 합성한 가짜 이미지를 판별기에 전달하고 이에 대한 정답지를 '가짜' 라고 알려주어 이에 대한 손실을 최소화하는 방향으로 학습하고, 그림 1의 우측 하단에서는 실제 출력(컬러) 이미지를 판별기에 전달하고 이에 대한 정답지를 '진짜' 라고 알려주어 이에 대한 손실을 최소화하는 방향으로 판별기를 학습한다. 생성기는 이 과정에서 판별기를 속이는 방향으로 학습하게 된다.

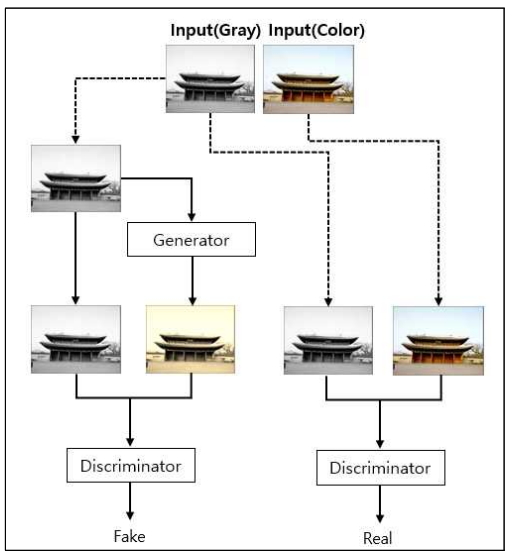

그림 1. Pix2Pix 적용 흐름도
Fig. 1. Pix2Pix Application Flowchart

2.2 판별기 모델의 수용 영역

Pix2Pix 판별기는 PatchGan으로 구현된다[3]. PatchGan은 판별 대상 입력 이미지를 특정 크기 단위로 나누어 판별하게 되는데 이때의 특정 크기 단위를 수용 영역(receptive field)이라고 한다. 기본적으로 PatchGan은 70*70 크기의 수용 영역을 사용하는데, 이를 판별기에서 몇 번의 컨볼루션 연산을 통해 최초의

70*70에 해당하는 이미지의 각 영역에 대한 진위 여부를 결과 값으로 도출하게 된다. 아래 그림 2와 같이 판별 대상이 되는 입력 이미지의 크기를 256*256이라고 가정할 때, 70*70 수용 영역을 설정한 판별기 모델은 최종적으로 16*16 크기의 판별 결과를 도출하게 된다. 16*16의 판별결과에서 각 하나의 픽셀은 입력 이미지에 적용한 70*70 영역에 대한 판결 결과를 의미하게 되고 1에 가까울수록 진짜 이미지라고 판단하게 되고 0에 가까울수록 가짜 이미지로 판단하게 된다.

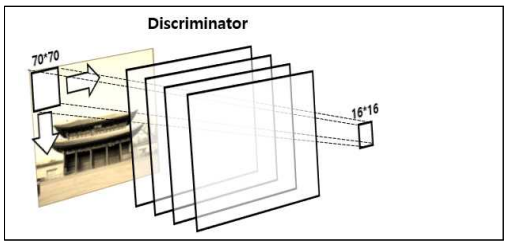

그림 2. Pix2Pix 수용 영역
Fig. 2. Pix2Pix Receptive Field

임의의 연산 층(Layer)의 수용 영역의 크기(N)는 연산을 적용하는 필터(Filter)의 크기, 순회하는 스텝을 의미하는 스트라이드(Stride)의 크기, 컨볼루션의 결과로 나오는 다음 레이어의 크기(Output)에 의해 결정되고 이를 요약하면 다음 수식과 같다[5].

$$N = (Output - 1) * Stride + Filter \quad (1)$$

판별기에서 사용하는 각 레이어별 수용 영역이 위 수식에 의해 어떻게 도출되는지를 살펴보고 이를 기반으로 스트라이드와 필터 크기의 변경을 통해 수용 영역을 변화시켜 성능 변화를 도출하는 것이 본 논문의 주요 요지이다.

일단, PatchGAN은 C64-C128-C256-C512 아키텍처를 사용한다[4]. 각 레이어의 'C' 문자는 Convolution, BatchNormalization, ReLu를 순차 적용한다는 의미이고, 각 레이어의 숫자는 필터의 개수를 의미한다[4].

모든 레이어는 기본적으로 필터 크기를 4*4, 스트라이드 크기를 2*2로 적용한다. (단, 마지막 C512 레이어의 스트라이드 크기는 1*1로 적용) 상기 수식을 적용하여 수용 영역을 도식화하면 아래 그림 3과 같다.

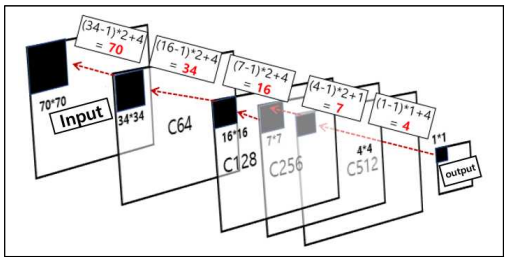

그림 3. 수용 영역의 계산
Fig. 3. Calculation of Receptive Field

3. 본론

3.1 이론적 고찰

수용 영역이 70*70인 PatchGAN이 Pix2Pix에서 일반적인 이미지 복원시 가장 좋은 성능을 나타낸다는 실험 결과가 있으나[4], 본 논문에서는 한국의 고궁 이미지를 채색 복원함에 있어서 최적의 수용 영역 설정 값을 찾기 위해 필터 및 스트라이드의 크기를 조정하여 실험을 진행한다. 성능의 평가는 판별기의 손실값을 기준으로 한다. 판별기의 손실은 아래 그림 4, 그림 5와 같이 2가지를 측정한다.

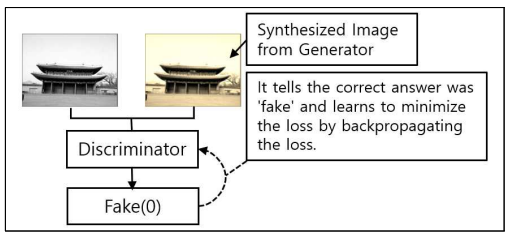

그림 4. 가짜 이미지의 판별기 학습 흐름
Fig. 4. Discriminator Learning Flow of Fake Image

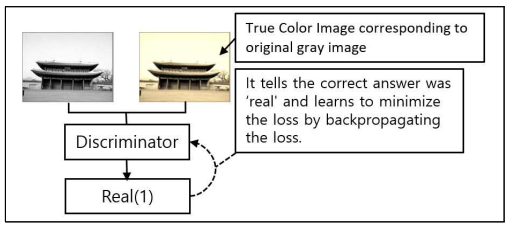

그림 5. 진짜 이미지의 판별기 학습 흐름
Fig. 5. Discriminator Learning Flow of Real Image

판별기에서 손실은 2가지로 측정할 수 있는데, 하나는 그림 4와 같이 가짜 이미지를 가짜라고 판단하는데 발생하는 손실과 다른 하나는 그림 5와 같이 진짜 이미지를 진짜라고 판단하는데 발생하는 손실이다.

판별기 훈련을 위한 고궁 이미지는 흑백 382장(그림 6)과 이와 짝이 되는 컬러 382장(그림 7)을 준비한다.

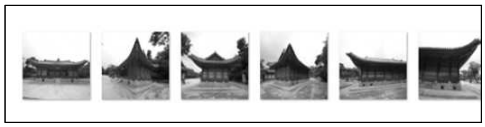

그림 6. 훈련용 흑백 이미지
Fig. 6. Gray Scale Image for Model Training

그림 7. 훈련용 칼라 이미지
Fig. 7. Color Image for Model Training

실험은 382장의 흑백과 컬러 이미지를 총 10번의 에포크(Epoch)를 거쳐 판별기를 훈련시킨다. 각 에포크에서는 임의의 1장의 이미지를 선택하여 판별기와 생성기의 손실을 계산하고 모델 훈련을 실시하므로 총 3820회의 훈련을 수행하게 된다. 생성기의 아키텍처는 2017년에 발표된 논문 Image-to-Image Translation with Conditional Adversarial Networks에 적용된 인코더와 디코더를 U-Net으로 연결한 구조를 동일 적용하였다[4].

3.2 실험 결과

그림 4에서 발생하는 손실 값을 d_loss_fake, 그림 5에서 발생하는 손실 값을 d_loss_real로 명명하고 수용 영역의 변화에 따르는 판별기의 손실 값 변화를 측정한 결과는 아래 표 1과 같다. 아래 표의 각 행의 학습 모델은 타 행의 학습 모델과 독립이며, 각 손실은 전체 에포크에서 발생한 손실에 대한 평균과 표준편차를 계산하였다.

표 1. 수용 영역에 따른 학습 모델의 손실 값
Table 1. Loss Value of Learning Model according to Receptive Field

Receptive Field	Filter	Stride		Loss Value	
				d_loss_fake	d_loss_real
47*47	3*3	2*2	avg.	0.164848405	0.17976252
			std.	0.186950959	0.16444898
70*70	4*4	2*2	avg.	0.328153781	0.35145922
			std.	0.095073231	0.07776048
93*93	5*5	2*2	avg.	0.231567161	0.24459372
			std.	0.194432537	0.18423063
135*135	3*3	3*3	avg.	0.218216024	0.22635299
			std.	0.196121886	0.17419041
269*269 All	4*4	3*3	avg.	0.034326785	0.03748992
			std.	0.114282713	0.11420653

위 실험 결과를 그래프로 표현하면 아래와 같다. 그래프의 가로축은 에포크이며 세로축은 손실 값이다. 그래프의 선 표현은 가시성을 고려하여 지수 이동 평균을 적용하였다.

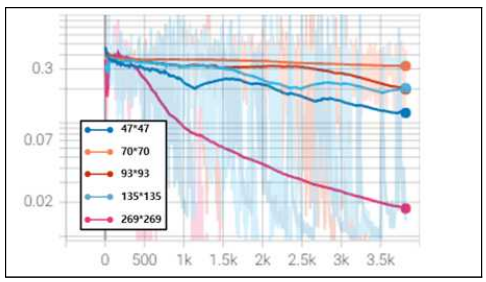

그림 8. 가짜 이미지에 대한 손실 그래프
Fig. 8. Loss Graph of Fake Image

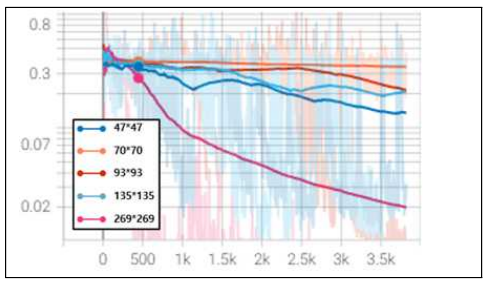

그림 9. 진짜 이미지에 대한 손실 그래프
Fig. 9. Loss Graph of Real Image

참고로, 동일 학습 에포크에서 측정된 생성기의 손실 값은 아래 그림 10과 같다.

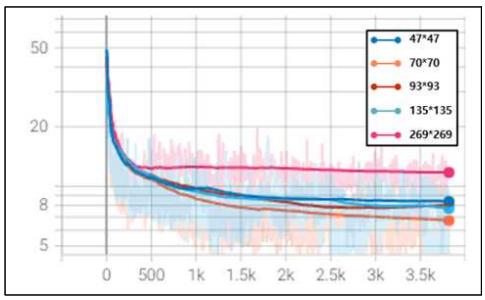

그림 10. 생성기 모델의 손실 그래프

Fig. 10. Loss Graph of Generator Model

3.3 결과 해석

수용 영역 296*296인 판별기 성능이 가장 좋게 도출되었으나, 생성기의 성능은 오히려 더 떨어진다. 이는 수용 영역의 크기가 타 실험 조건 대비 상대적으로 크기 때문에 이미지 변화가 많은 부분에서 세밀한 판단이 힘든 것으로 사료된다. 이로 인해, 손실 값은 작게 도출되었지만 생성기 학습에는 도움이 되지 못하였다. 생성기의 학습은 아래 그림 11에서 ①에 해당하는 손실과 ②에 해당하는 판별기로 부터 도출된 손실의 합이 최소화되는 방향으로 학습하게 된다. 이상적인 흐름이라면 생성기에서 생성한 질 낮은 이미지에 대해서는 판별기 측에서 ②의 손실 값이 크게 발생해야 ①과 ②에서 합산된 손실 값이 역전파 되어 원하는 의도대로 생성기의 모델이 개선이 되는데, 실제로는 ②의 손실 값이 크지 않아서 생성기를 개선할 여지가 줄어든 결과로 해석이 된다.

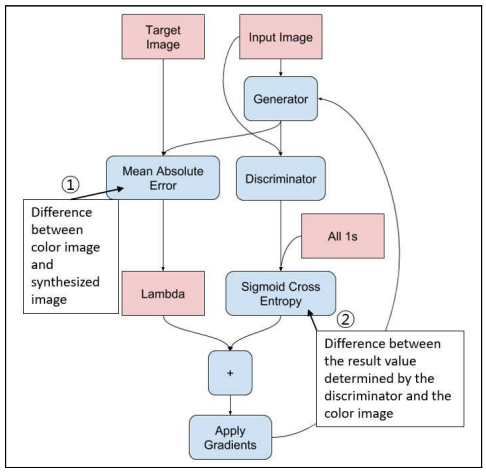

그림 11. 생성기의 학습 구조 [7]

Fig. 11. Learning Mechanism of Generator Model

적대적 생성 신경망의 평가는 단순 수치로 평가하기에는 어려워서, 블라인드 테스트 등 정성적인평가를 통해서도 품질을 측정할 수 있다[6].

나머지 수용 영역에 대한 수치적 해석은 무의미할 정도로 근사하게 분포하고 있어서 정성적인 해석으로 결과를 살펴보겠다. 아래 그림 12의 결과에서 수용 영역이 70*70인 경우, 복원한 사진에서 초록색 번짐(그림12에서 적색 표기)이 보이고 있다. 이는 다른 복원 사진에서도 비슷한 결과를 보인다. 잔디, 나뭇잎 등 세밀한 변화가 보이는 부분에서 주로 색의 번짐 현상이 자주 일어난다. 이에 비하여, 수용 영역이 47*47인 경우에는 색 번짐 현상은 덜하나, 배경 채색이 약하다. 수용 영역이 작으면 이미지의 작은 영역 복원에는 효과가 있지만, 전체적인 채색 효과가 떨어진다고 사료된다. 수용 영역이 93*93 에서는 70*70 대비 복원 이미지의 해상도가 떨어지며, 그 이상에서는 색 번짐이 거의 나타나지 않으나 해상도가 수용 영역 크기에 비례하여 급격히 떨어지는 것을 확인할 수 있었다.

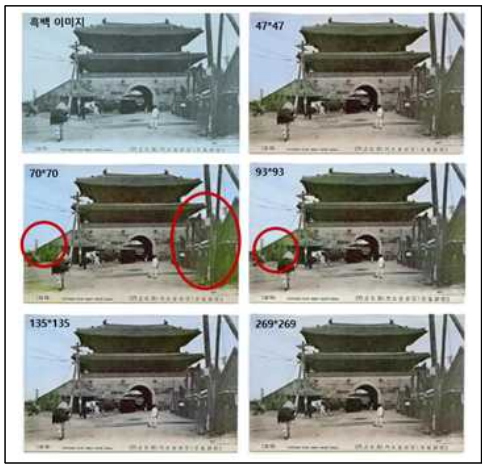

그림 12. 복원 결과
Fig. 12. Restore Results

즉, 수용 영역 설정은 복원하려는 이미지의 특성에 따른 부분 최적화와 전체 최적화 사이의 트레이드오프(trade-off) 관계로 보인다. 고궁 이미지의 특성은 기와와 단청 등의 부분 영역에서 최적화된 복구를 고려함과 동시에 고궁 이미지 주위의 풍경에 대해서 전체적인 최적화를 고려해야 한다.

아래 그림 13에서 우측 복원 이미지는 수용 영역이 70*70으로 수용 영역이 47*47(좌측 복원 이미지)의 복원 결과 대비 하늘의 배경색과 바닥의 잔디 색이 잘 표현되어 있음을 볼 수 있다.

그림 13. 수용 영역에 따른 배경 복원
Fig. 13. Restoring Background according to Receptive Field

4. 결론

본 논문에서는 흑백으로 보존된 고궁 이미지를 적대적 신경망 기법 중 Pix2Pix를 이용하여 컬러로 채색하기 위해 판별기의 수용 영역을 조정하여 결과를 도출하였다. 고궁 이미지의 특성상 건물의 문양과 색상의 변화가 보통의 단조로운 건물 이미지와 다르다. 그에 비해 고궁 주위의 객체는 주로 나무, 하늘, 잔디 등 일반적인 자연 환경의 출현 빈도가 높은 특징이 있다. 단청, 현판 등 세밀히 채색을 고려해야 하는 영역은 수용 영역을 작게 잡는 것이 복원 품질에 유리하고, 부분적인 최적화 보다 전체 최적화가 중요한 풍경과 배경이 중요한 영역은 수용 영역을 좀 더 크게 잡는 것이 유리하다고 사료된다. 이를 바탕으로 고궁 이미지를 객체 단위로 분리하여 학습하는 모델을 구현하는 것이 향후의 과제이다. 기와 무늬와 나뭇잎, 배경 하늘, 흙바닥, 돌담 등 고궁 이미지의 주로 구성되는 오브젝트를 분리하거나 고궁과 고궁 이외의 객체를 분리하여 학습하는 모델을 구현하여 복원 품질을 향상 시키는 방법을 고안하여 복원 품질을 향상시킬 필요가 있다. 또한, 복원 대상 사진이 보존 상태가 좋지 못한 사진의 경우 모델 예측이 잘 되지 않는 경향도 있으므로 향후 이에 대한 전처리를 고려해야 한다[8][9].

REFERENCES

[1] Goodfellow, Ian, et al., “Generative Adversarial Nets”, Advances in Neural Information Processing Systems, Vol. 2, pp. 2672-2680, 2014.

[2] Park, Hong-Jin, “Trend Analysis of Korea Papers in the Fields of ’Artificial Intelligence’, ‘Machine Learning’ and ‘Deep Learning,’” The Journal of Korea Institute of Information, Electronics, and Communication Technology, vol. 13, no. 4, pp. 283-292, Aug. 2020.

[3] Schonfeld, E., Schiele, B., & Khoreva, A., “A u-net based discriminator for generative adversarial networks”, In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8207-8216, 2020

[4] P. Isola et al., “Image-to-Image Translation with Conditional Adversarial Nets”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1127-1128, 2017

[5] Phillip Isola, https://github.com/phillipi/pix2pix/blob/master/scripts/receptive_field_sizes.m, 2014
[6] Ugur Demir and Gozde Unal, "Patch-based image inpainting with generative adversarial networks", arXiv preprint arXiv:1803.07422, pp 6, 2018
[7] Google Tensorflow, https://www.tensorflow.org/tutorials/generative/pix2pix?hl=ko, 2022
[8] Yanyun Qu, Yizi Chen, Jingying Huang, Yuan Xie, "Enhanced Pix2pix Dehazing Network", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8160-8168, 2019
[9] Sehwan Ki, Hyeonjun Sim, Jae-Seok Choi, Saehun Kim, Munchurl Kim, "Fully End-to-End Learning Based Conditional Boundary Equilibrium GAN With Receptive Field Sizes Enlarged for Single Ultra-High Resolution Image Dehazing", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 817-824, 2018

저자약력

황 원 용 (Won-Yong Hwang)

- 1997년 3월 ~ 2004년 2월: 고려대학교 전자 및 정보공학 학사
- 2009년 2월 ~ 2011년 2월: KAIST 소프트웨어대학원 공학석사 졸업
- 2004년 1월 ~ 2007년 2월: LG전자 MC연구소 휴대폰 솔루션 개발
- 2007년 3월 ~ 2008년 12월: LIG넥스원 지대공유도무기 개발
- 2011년 3월 ~ 2017년 5월: 삼성SDS 솔루션 개발팀(EFSS 팀)
- 2017년 6월 ~ 현재: 한국폴리텍대학 스마트금융과 교수

관심분야 인공지능, 블록체인, 핀테크

김 효 관 (Hyo-Kwan Kim)

- 2001년 3월 ~ 2007년 8월: 성균관대학교 정보통신공학부 학사
- 2011년 9월 ~ 2017년 2월: 한국교통대학교 컴퓨터공학과 석/박사
- 2007년 3월 ~ 2017년 11월: 도담시스템스 소프트웨어개발 & 삼성 SDS 데이터분석그룹
- 2017년 12월 ~ 현재: 한국폴리텍대학 스마트금융과 교수

관심분야 인공지능, 금융데이터 분석, 핀테크