

# Pix2Pix의 활용성을 위한 학습이미지 전처리 모델연계방안 연구

김호관\* 황원용\*\*

## A Study on the Image Preprocessing model linkage method for usability of Pix2Pix

Hyo-Kwan Kim\* Won-Yong Hwang\*\*

**요약** 본 논문은 적대적 생성 신경망 기법의 하나인 Pix2Pix를 활용하여 컬러색상을 입히는 경우 학습된 이미지의 빛 반사 정도에 따라 예측결과가 손상되어 나오는 부분에 집중하여 Pix2Pix 모델 적용 전 이미지 전처리 프로세스 및 모델 최적화를 위한 파라미터 테이블을 구성한다. 기존 논문에 나온 Pix2Pix 모델을 활용하여 실생활에 적용하기 위해서는 해상도에 대한 확장성을 고려해야한다. 학습 및 예측결과 이미지 해상도를 키우기 위해서는 동시에 모델의 커널 사이즈 등을 같이 맞춰주는 부분을 수정해줘야 하는데 이부분은 파라미터로 튜닝 가능하도록 설계했다. 또한 본 논문에서는 예측 결과가 빛 반사에 의해 손상된 부분만 별도 처리하는 로직을 같이 구성하여 예측결과를 왜곡시키지 않는 전처리 로직을 구성하였다. 따라서 활용성을 개선하기 위하여 Pix2Pix 모델의 학습이미지에 공통적인 빛반사 튜닝 필터를 적용하는 부분과 파라미터 구성부분을 추가하여 모델 정확도를 개선하였다.

**Abstract** This paper proposes a method for structuring the preprocessing process of a training image when color is applied using Pix2Pix, one of the adversarial generative neural network techniques. This paper concentrate on the prediction result can be damaged according to the degree of light reflection of the training image. Therefore, image preprocessing and parameters for model optimization were configured before model application. In order to increase the image resolution of training and prediction results, it is necessary to modify the of the model so this part is designed to be tuned with parameters. In addition, in this paper, the logic that processes only the part where the prediction result is damaged by light reflection is configured together, and the pre-processing logic that does not distort the prediction result is also configured. Therefore, in order to improve the usability, the accuracy was improved through experiments on the part that applies the light reflection tuning filter to the training image of the Pix2Pix model and the parameter configuration.

**Key Words** : Discriminator, Generator, GAN, Pix2Pix, Homomorphic Filter,

### 1. 서론

4차 산업혁명 시대에 들어서서 빅데이터 및 인공지능 산업이 크게 확장해 나가고 있다. 특히 시각인지 영역에 있어서 GAN(Generative Adversarial Networks), 적대적 생성 신경망이 등장함에 따라 이미지 생성 및 변환 기술이 빠르게 성장하고 있다.

GAN 기술은 인공 신경망이 다양한 노이즈(Noise) 입력을 받아 원하는 이미지를 생성하거나 입력 이미지를 다른 형태나 정보를 지닌 이미지로 변환하는 기술로 최근 급속도로 성장하는 기술이다.

GAN은 기존에 인공신경망 하나를 통해 학습을 시켰던것과 다르게 2개의 인공신경망을 구성하여 입력된 이미지를 원하는 정답이미지로 변화 시킬 수 있다. 본

\* Corresponding Author: Department of Fintech Korea Polytechnics (haiteam@kopo.ac.kr)

\*\* Department of Fintech, Korea Polytechnics

Received September 24, 2022

Revised October 23, 2022

Accepted October 24, 2022

연구에서는 우리나라의 고궁사진을 기준으로 색상복원 측면에서 회색 이미지를 컬러 이미지로 변경하는 GAN모델 중 Pix2Pix 방법을 활용하여 구현한다. 또한 Pix2Pix 모델을 최대한 활용할 수 있도록 해상도 크기조절 및 모델 파라미터등을 파라미터로 관리하여 실험한다. 또한 단순 Pix2Pix 모델만 적용하는게 아닌 분석 시 가장 시간이 많이 소요되는 전처리 부분을 모델 구동 전 적용하게 하여 활용성을 증대시키기 위한 연구를 진행한다.

## 2. 본론

### 2.1 이론적 고찰

#### 2.1.1 GAN (Pix2Pix)

GAN은 '생성적 적대 신경망'의 약자로 생성자와 식별자가 서로 경쟁(Adversarial)하며 데이터를 생성(Generative)하는 모델(Network)를 뜻한다.

이중 Pix2Pix 모델은 원 이미지  $x$ 와 변환하고자 하는 목표 이미지  $y$ 의 쌍  $(x,y)$ 가 필요하다. 생성자(G)는 이미지를 입력으로 받아서 이미지를 생성한다. 판별자(D)는 입력으로 두가지를 받는다. 생성자에 넣었던 입력 이미지와 생성자에서 출력된 이미지를 받는 경우와, 생성자에 넣었던 입력 이미지와 진짜 이미지를 넣어 학습한다.[1]



그림 1. Pix2Pix 모델  
Fig. 1. Pix2Pix Model

#### 2.1.2 학습모델 구성

본 연구에서 사용하는 학습모델 프로세스는 다음 그림과 같다. 생성자는 가짜 이미지를 만들어내고 판별자를 잘 속이도록 설계되었다. 판별자는 가짜 이미지는 가짜라고 잘판단해야하며, 진짜 이미지는 진짜 이미지라고 판단할 수 있도록 설계되었다. 주의할 점은 생성자가 학습할때는 판별자는 학습되지 않고, 반대로 판별자가 학습할때는 생성자는 학습되지 않는다. 학습 완료된 결과는 예측 시 바로 활용할 수 있도록 h5 형태로 저장한다.



그림 1. 생성자 학습모델 프로세스  
Fig. 1. Generator Training Model Process

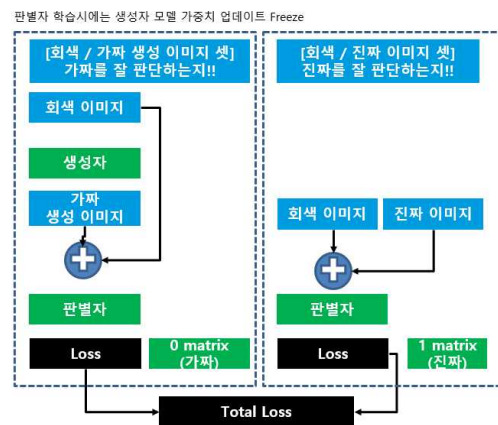


그림 2. 판별자 학습모델 프로세스  
Fig. 2. Discriminator Training Model Process

2.1.3 예측모델 구성

본 연구에서 사용하는 예측 모델의 프로세스는 다음 그림과 같다. 테스트 이미지를 불러오고 훈련모델에서 생성한 훈련결과를 불러와서 예측한다.



그림 3. 예측모델 프로세스  
Fig. 3. Predict Model Process

2.1.4 Pix2Pix 생성자 모델

Pix2Pix 모델은 생성자가 추측하는 결과의 정확도를 높이고, 판별자는 생성된 결과를 잘 비교할 수 있도록 하는 특수한 구조를 가지고 있다. Pix2Pix의 생성자는 U-Net 구조로 구성되어있다.[2] 즉, 전체적으로 영상의 크기를 다운샘플링 한 후 업샘플링 하는 구조를 갖는다.[3] 영상을 줄일때는 특성을 찾는 합성곱신경망의 기본 레이어구조(Conv2d, batch normalization, activation)패턴을 따르고 있다. 이때 인코딩 (conv2d) 시에는 영상의 크기가 줄어드는 stride가 2인컨볼루션을 이용한다.[4] 다시 커질 때는 전이 컨볼루션(transposed convolution)을 이용한다. 영상이 줄어들면서 생기는 손실은 영상이 커지는 레이어에서 같은 크기일 때 residual network에서 이전 레이어를 붙이는 것처럼 더해주기 때문에 (concatenate) 원본 이미지의 특징을 다시 살려낼 수 있는 구조로 구성되어 있다.[5]

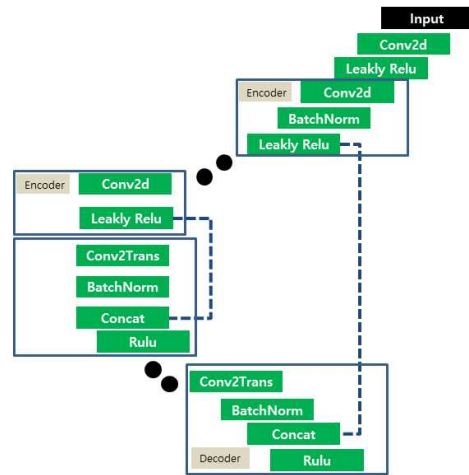


그림 4. 생성자 모델  
Fig. 4. Generator Model

2.1.5 Pix2Pix 판별자 모델

Pix2Pix 모델 판별자는 PatchGan 70사이즈 구조를 갖는다.

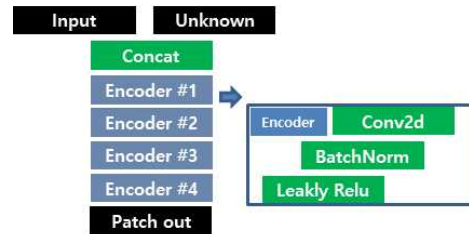


그림 5. 판별자 모델  
Fig. 5. Discriminator Model

일반적으로 전체 생성된 이미지에 대해서 전체를 가지고 sigmode 형태로 진짜/가짜를 판별하는게 아니라 이미지의 특정 크기의 패치단위로 (Patch 사이즈 70) 진짜/가짜를 판별하는 모델이다.[6][7]

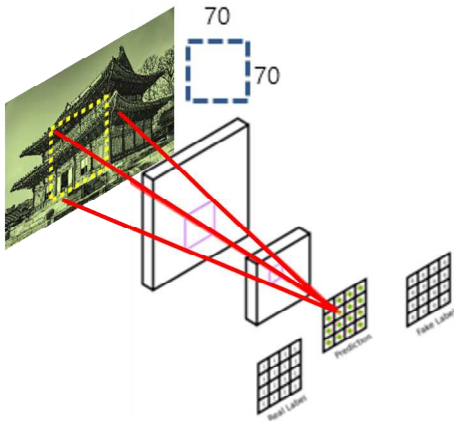


그림 6. 패치간 70  
Fig. 6. PatchGan 70

### 2.1.6 Homomorphic 필터

디지털이미지는 illumination element와 reflectance element의 곱으로 표현할 수 있다. 이 중 illumination element는 이미지에 노출된 조명을 나타내고(low frequency), reflectance element는 이미지에 존재하는 object들의 edge를 나타낸다 (high frequency). 이미지를 주파수영역으로 분리한 후 log를 취함으로써 곱셈을 덧셈연산으로 나눈후 여기에 HPF(high pass filter)를 씌우고 다시 exp 연산을 통해 log를 지워주면 조명이 제거된 이미지를 얻을 수 있다.[8]

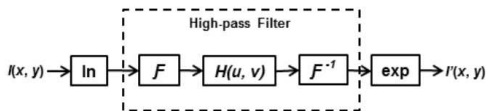


그림 7. Homomorphic 필터  
Fig. 7. Homomorphic Filter

## 2.2 연구 방법

딥러닝 모델을 다음 표와 같이 그레이/컬러 이미지 각각 431장을 준비한 후 Pix2Pix 모델에 적용한다. 실험을 위한 실측 이미지와 구매한 이미지 이다. Pix2Pix모델은 적은 수의 이미지로도 훈련 및 예측이 가능하다. 훈련모델과 예측모델은 분리하여 구현

하며, 훈련모델의 학습결과는 (.h5)로 저장한다.[9] 검증을 위한 정답지가 없는 테스트 이미지 155장을 준비한 후 모델을 검증한다.

이때 학습이미지에 전처리를 하지 않고 모델을 구동한 결과와 빛노출 및 반사효과를 전처리 과정을 통하여 제어한 후 구동한 모델결과를 비교하여 Pix2Pix 모델의 전처리 과정을 연계하는 모델 구성 방안을 제시한다.

표 1. 실험데이터  
Table 1. Test Data

구분	내용	비고	
훈련 이미지	gray	431	카테고리: 고궁
	color	431	
테스트 이미지	gray	155	

테스트 환경은. 클라우드 기반 파이썬 개발환경인 colab을 활용하여 훈련 및 예측을 진행한다.

표 2. 실험환경  
Table 2. Test Environment

Div	Contents	Comments
Train	colab (Pro Version)	
Predict	GPU 4GB PC	M2000

테스트를 원활하게 구동하기 위하여 모델은 파라미터 테이블을 참조하여 동작하도록 구성한다.

표 3. 파라미터 테이블  
Table 3. Parameter Table

No	Seg	Contents	Comments
1	CATEGORY	RANGE	COMMON: ALL SEG: GOGUNG
2	PARAM_NAME	RESOL KSIZE	RESOLUTION KSIZE SIZE
3	PARAM_VALUE	VALUE	1024
4	USE_YN	USABILITY	Y/N

## 2.3 실험

### 2.3.1 일반 Pix2Pix모델 적용

일반 Pix2Pix모델만 적용 시 수집한 이미지에 빛반

사 이미지가 존재하는 경우에는 다음 그림과 같이 반사된 빛이 결과에 반영됨을 확인할 수 있다.



그림 8. 예측결과 (Pix2Pix적용)  
Fig. 8. Result of Forecast (Pix2Pix)

### 2.3.2 이미지내 필터 적용

학습 이미지에 Homomorphic을 적용 시 다음 그림과 같이 빛반사 정도가 제어됨을 확인할 수 있다. 제어된 상태의 이미지 데이터를 활용하여 모델을 학습한다.



그림 9. 학습이미지 (Homomorphic 적용)  
Fig. 9. Training Image (Homomorphic Applied)

### 2.3.3 전처리 이미지 + Pix2Pix모델 적용

Homomorphic을 적용한 학습데이터를 사용하여 예측한 결과는 다음과 같다. 예측 결과에 반사된 빛이 반영되지 않고 정상적으로 예측됨을 확인하였다.



그림 10. 예측결과 (전처리+Pix2Pix적용)  
Fig. 10. Result of Forecast (Homomorphic+Pix2Pix)

### 2.3.4 전처리 작업 전후 비교

다음 그림은 제일 위에 위치한 사진은 흑백 테스트 이미지이며 두 번에 위치한 이미지는 Pix2Pix 모델만 적용한 예측 결과이다. 마지막에 위치한 이미지는 Homomorphic을 학습데이터에 전처리로 적용한 후 학습 및 예측한 결과 이미지이다.



그림 11 예측 결과  
Fig. 11 Predict Result

예측결과도 표 4와 같이 개선되었음을 확인할 수 있다.

표 4. 모델 손실값 비교  
Table 4. Comparison of Model Loss Values

No	proposed Logic	discriminator loss1	discriminator loss2	generator loss
1	Applied	0.28004	0.38701	3.04149
2	Not Applied	0.32232	0.40546	4.03307



### 3. 결론

Pix2Pix 모델은 이미지 복원 또는 구글맵 활용 등 다양한 산업 분야에서 활용이 되고있다. 하지만 막상 활용하려면 기본 제시된 모델이 256\*256 해상도로 너무 낮고 학습데이터를 구하기고 힘들다. 학습데이터를 힘들게 구해도 이미지의 빛 반사 등의 데이터 품질이 슈 때문에 만족할 만한 예측 결과를 얻기 힘들다. 따라서 본 논문에서는 다양한 산업현장에서 바로 적용할 수 있도록 파라미터 테이블을 구성하여 모델에 반영될 수 있는 모델 구성안을 제안한다. 해상도나 모델의 커널사이즈 등은 파라미터 설정을 통해 제어 가능하다. 또한 전처리 이미지에 필터 및 파라미터를 구성하여 모델 정확도를 개선하였다. 향후 더 나은 해상도를 적용하기 위하여 Super resolution 모델을 예측 결과에 추가로 적용하여 활용성을 개선할 수 있는 부분이 있는지 여부는 추가 연구할 과제이다.



그림 11. 제안 및 향후 발전모델 (Pix2Pix)  
 Fig. 11. Proposed and Future Model (Pix2Pix)

### REFERENCES

[1] 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

[2] Ronneberger O, U-net: Convolutional net-works for biomedical image segmentation, Int'l Conf. on Medi-cal Image Computing and Computer-assisted Intervention, pp. 235-237, 2015

[3] Milletari F, V-net: Fully convolutional neural networks for volumetric medical image segmentation, IEEE Int'l Conf. on 3D vision (3D V), pp. 565-565, 2016

[4] Ronneberger O, U-net: Convolutional net-works for biomedical image segmentation, Int'l Conf. on Medi-cal Image Computing and Computer-assisted Intervention, 2015

[5] 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

[6] Phillip Isola, [https://github.com/phillipi/pix2pix/blob/master/scripts/receptive\\_field\\_sizes.m](https://github.com/phillipi/pix2pix/blob/master/scripts/receptive_field_sizes.m), 2014

[7] Ugur Demir,Gozde Unal, Patch-Based Image Inpainting with Generative Adversarial Networks,arXiv:1803.07422v1, pp1-3, 2018

[8] MARAGOS, Petros, Morphological filtering. In: The essential guide to image processing, Academic Press, pp293-321, 2009

[9] Google Tensorflow, <https://www.tensorflow.org/tutorials/generative/pix2pix?hl=ko>, 2022

### 저자약력

#### 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

**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