



ReadMe_Pix2Pix

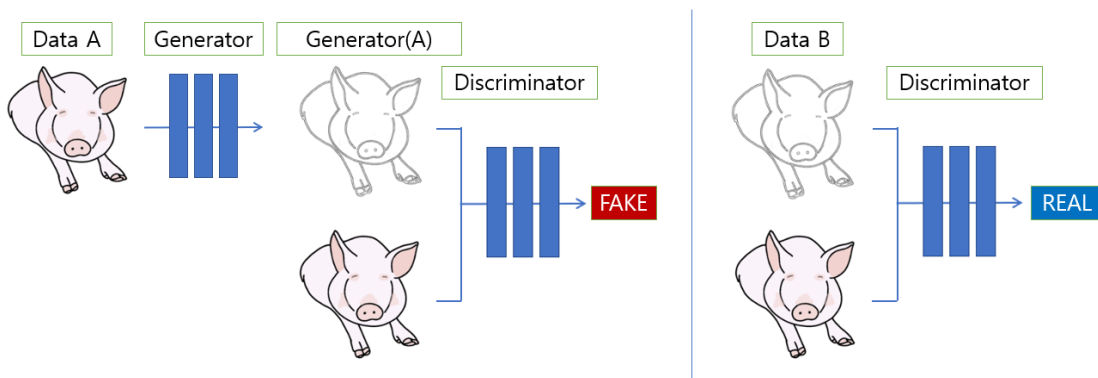
Pix2Pix 선정 배경, 모델의 구조, 학습 방법, 결과 비교

Pix2Pix란

- <https://arxiv.org/abs/1611.07004> (논문), <https://github.com/phillipi/pix2pix> (깃허브)를 기반으로 공부하고 코드를 작성하였음.

모델 선정 배경

- Pix2Pix는 CGAN(Conditional GAN)의 한 종류로 이미지로부터 이미지를 생성하는 모델. CGAN은 조건 벡터와 이미지의 짝을 데이터로 학습하는데, Pix2Pix의 경우 조건으로 이미지를 받는다.



모델의
목적함수

현실적인 이미지 만들도록 함

실제 정답과 유사하도록 함

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G)$$

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y} [\log D(x, y)] + \mathbb{E}_{x,z} [\log(1 - D(x, G(x, z)))]$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z} [\|y - G(x, z)\|_1]$$

- Pix2Pix의 장점

1. L1 loss를 최소화하는 CNN 등의 방식에서는 시각적인 결과와 무관하게 Euclidean 거리를 줄이도록 학습이 진행돼 영상이 화질이 안 좋아지는 결과가 발생함. (Pix2Pix뿐 아니라 모든 GAN의 장점)
2. CNN, 다른GAN과 달리 적은 수의 data로 결과가 나오는 편임. GAN의 경우 일반적인 모델은 고화질 이미지를 생성하기 위해서는 5만 개에서 10만 개의 데이터를 필요로 하는 반면, Pix2Pix는 천 개 내외의 데이터로도 가능함. (그러나 과적합 문제를 해결하기 위해서는 데이터를 조금 더 늘릴 필요성이 있음)
3. 변환의 형태와 무관하게 모델 구조가 범용적임. 논문 역시 같은 구조로 신발/가방 스케치>이미지, 위성 지도>약도, 낮>밤, 레이블>건물 외벽 등 여러 변환 모델을 만듦. 위성 지도에서 약도로 변환하는 경우와 같이 컬러 이미지에서 컬러링 도안 변환이 가능할 것으로 판단.

- Pix2Pix의 단점

1. 원하는 모델의 학습에 적합한 paired data를 구하기 어려움
2. 정량적 평가의 어려움. 논문의 경우에도 1)사람이 직접 결과물을 봤을 시간에 식별하는 AMT(Amazon Mechanical Turk)를 진행하고 2)사전 훈련된 classifier(분류기)를 통해 생성한 이미지가 구별되는지 FCN-score를 냄. 두 경우 모두 비용과 시간의 문제가 존재함.

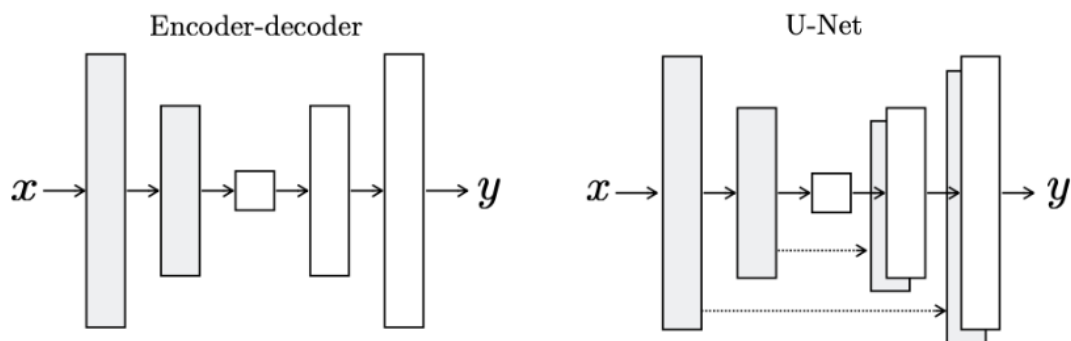
- 단점의 해결방안

1. 일러스트레이션 - 스케치 데이터를 구할 수 있었음. 도안은 아니지만 온전한 스케치가 나온다면 후처리를 통해 도안화가 가능함. 또한 1차 학습을 시킨 후 1차 모델을 바탕으로 추가 데이터를 구상할 수 있음.
2. 보고서와 발표에 다양한 케이스의 결과를 제시해 정성적 평가가 가능하도록 할 것임. (CLIP을 통해 결과물과 'coloring page'와의 정확도를 확인?)

본 모델의 구조

- 이미지를 입력 받아 도안을 출력하는 generator(생성자)와 생성한 가짜 이미지를 식별하는 discriminator(판별자)로 구성.
- generator(생성자)의 목표는 discriminator(판별자)를 속이는 것.

- 손실 함수 중 L1 loss(G pixel loss)는 generator(생성자)가 저해상도 정답과 유사한 이미지를 생성하도록 함. GAN loss는 고해상으로 진짜와 같은 이미지를 만드는 것에 집중할 수 있음.
- 모델구조는 논문과 같이 generator(생성자)는 UNet을 사용. UNet구조는 저차원 뿐만 아니라 고차원 정보도 이용하여 이미지의 특징을 추출함. 일반적인 encoder-decoder 구조는 encoder 단계에서 채널의 수를 늘리면서 차원을 축소하고 decoder 단계에서 저차원으로 인코딩된 정보만 이용하여 고차원의 이미지를 복원함. UNet 구조는 encoding 단계의 각 layer에서의 특징을 decoding 단계의 각 layer에 skip connection을 통해 직접 연결해 합침.



Encoder-decoder VS U-Net

1차 학습

[1_pix2pix_v1\(100epoch\).ipynb](#) 참고

학습에 사용한 데이터셋

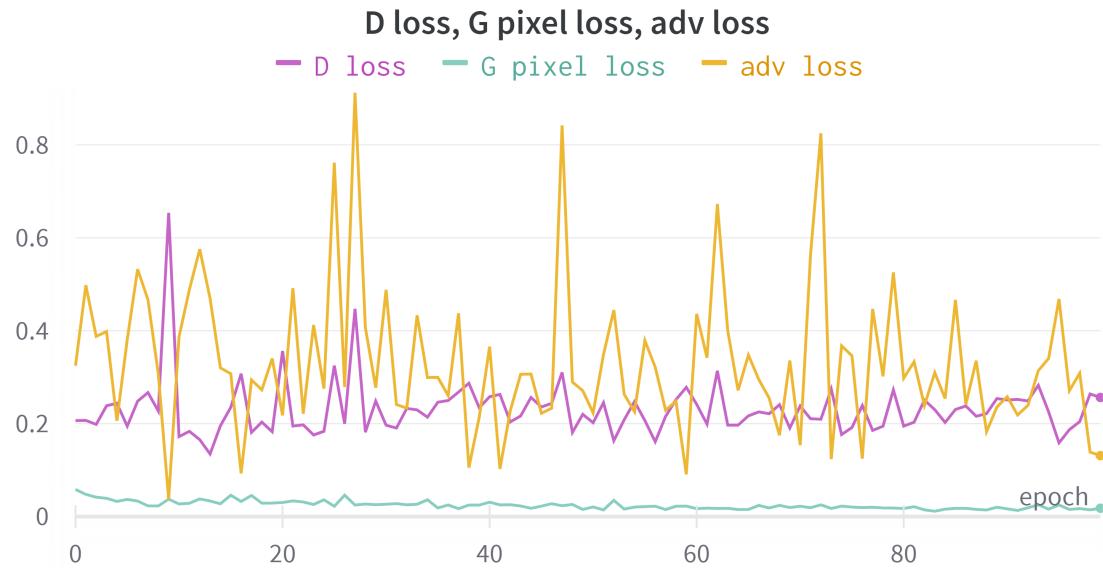
- kaggle의 'Anime Sketch Colorization Pair'



- 위와 같이 {애니 일러스트레이션 : 스케치}로 구성돼 있는 paired data 3545쌍을 활용

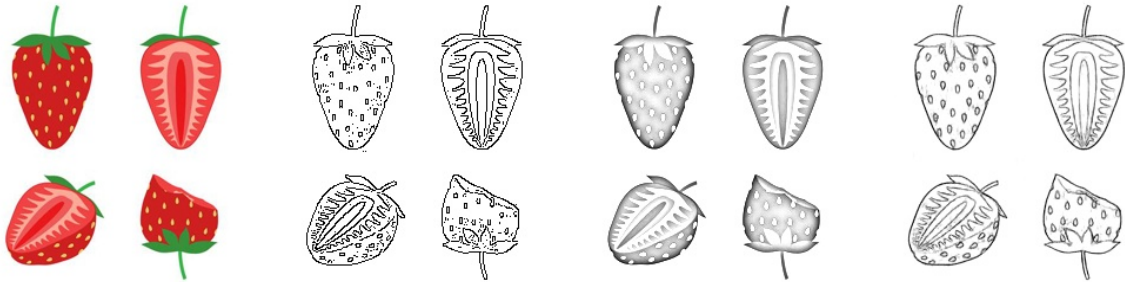
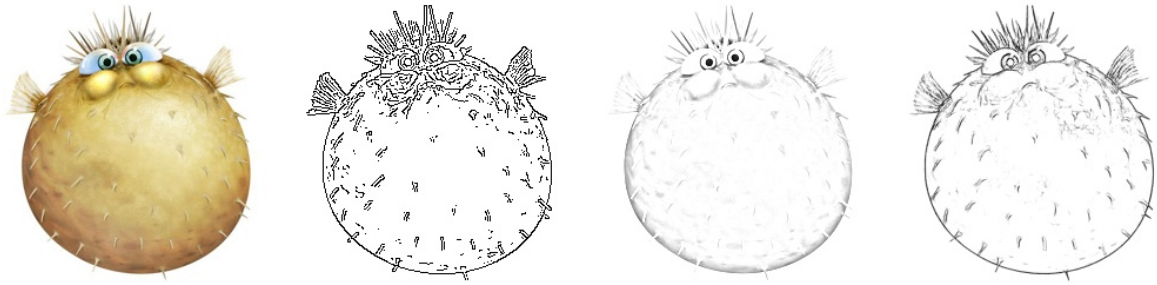
학습 시간, loss

- 100epoch, 4시간 30분



결과 분석





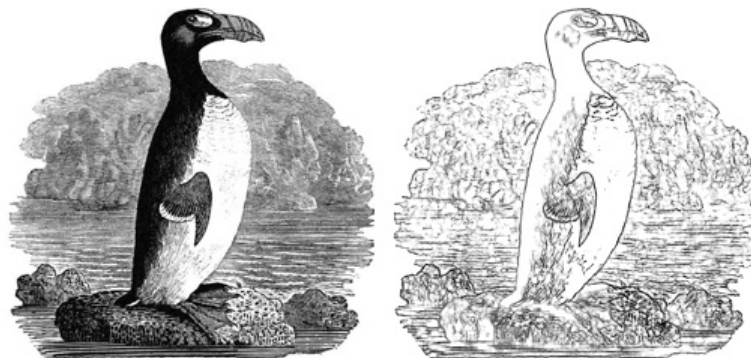
[input image]

[edge detection]

[several computer vision skills]

[our model]

- sketch와 유사한 결과물을 가져오는 기성의 모델보다 컬러링 도안에 보다 적합한 결과가 나옴을 확인할 수 있음.
- edge detection(1)의 경우 부드럽지 못한 선이 나오고 추가적인 얼룩 내지 점이 있지만, 학습 시킨 모델(4)은 깔끔한 선으로 결과물이 나옴.
- gray+invert+blur+invert(3)의 경우 검정 혹은 색이 진한 경우 색을 충분히 지우지 못하며, 연한 경우 선이 만들어지지 않은 결과물을 냄. 반면에 학습 시킨 모델(4)은 색의 명도와 무관하게 일정한 결과물이 나옴.
- 1차 학습 시킨 모델은 선이 있거나 색의 대비가 명료한 일러스트레이션의 경우, 추가 후 처리를 하면 충분히 컬러링 도안으로 사용할 수 있는 결과물이 나옴. 그러나 사진 혹은 사실적인 일러스트레이션의 경우, 뿌연 경우 결과가 잘 나오지 않음.





- 본 문제를 해결하기 위해 custom paired data를 통한 2차 학습을 진행함.

2차 학습

학습에 사용한 custom 데이터셋

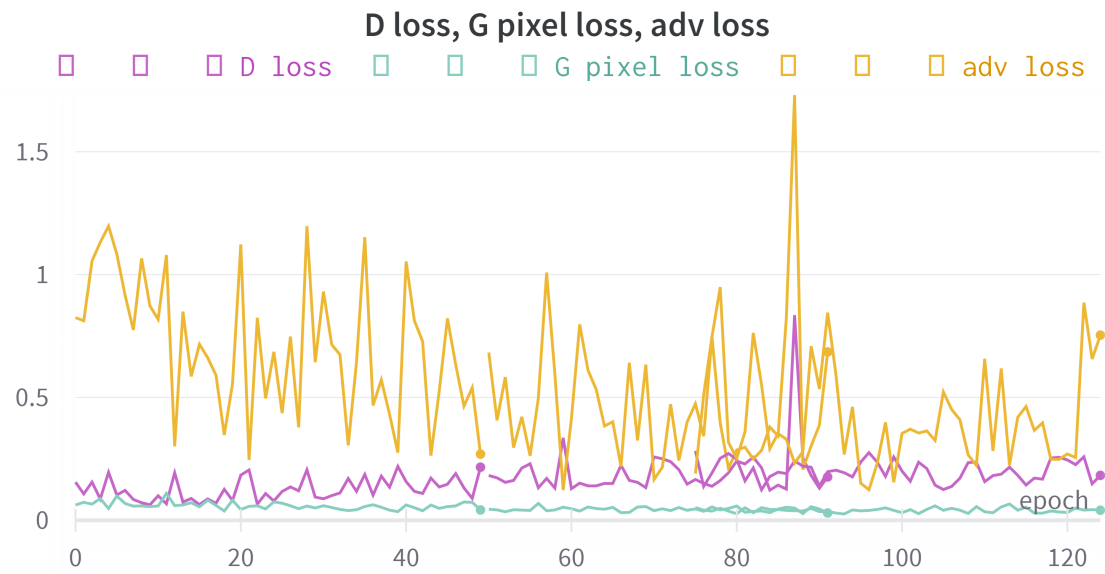
[2_crawling_images.ipynb](#) 와 [3_create_custom_paired_dataset.ipynb](#) 참고

- 원하는 형식의 데이터셋이 없어 직접 구축해 사용함.
- duckduckgo의 이미지 검색을 통한 6200장의 일러스트레이션 이미지 크롤링(a).
- 크롤링한 이미지(a)를 1차 학습한 모델을 활용해 스케치 이미지(b)를 제작하고, input 이미지에 노이즈를 추가해 blur처리 혹은 dilate처리를 함(c). 처리를 한 크롤링 이미지(c)와 스케치 데이터(b)를 붙여 8500장의 paired data로 만듦.
- 예시 paired data (왼쪽: 노이즈+blur, 오른쪽: 노이즈+dilate)

학습 시간, loss

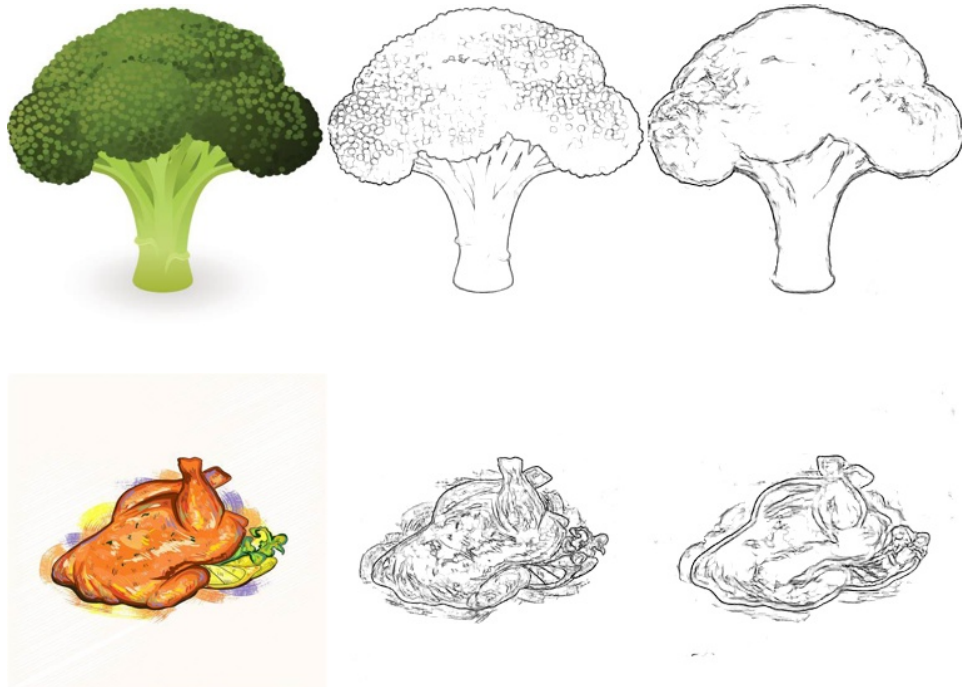
[4\(1\)_pix2pix_v2\(50epoch\).ipynb](#), [4\(2\)_pix2pix_v2\(75epoch\).ipynb](#),
[4\(3\)_pix2pix_v2\(125epoch\).ipynb](#) 참고

- 125epoch, 12시간 30분



결과 분석

- 왼쪽부터 순서대로 input image - 1차 학습 시킨 모델의 결과 - 2차 학습 시킨 모델의 결과

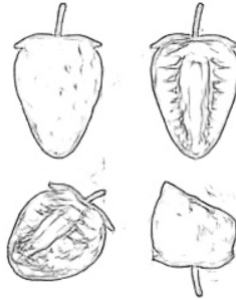




[input image]

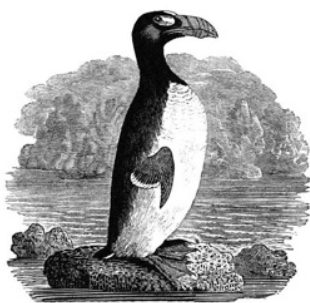


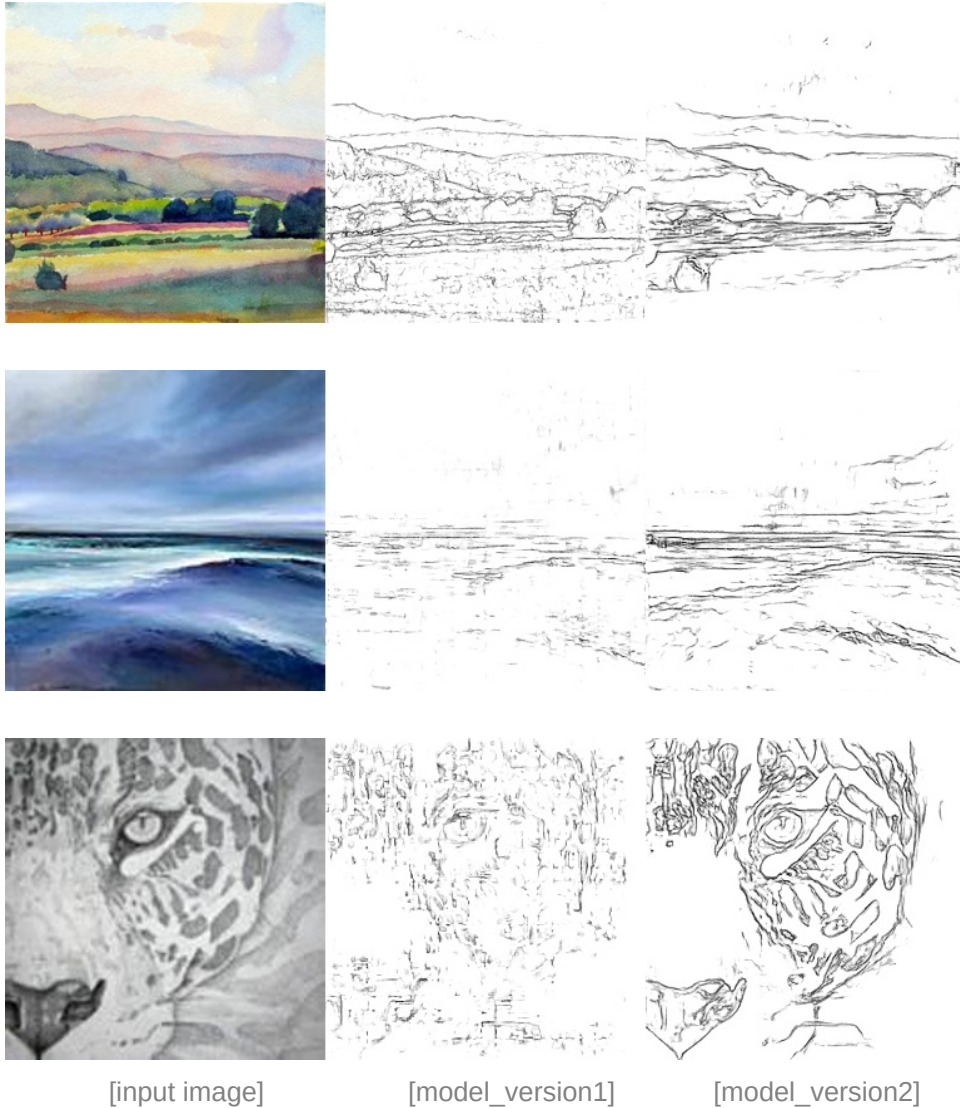
[model_version1]



[model_version2]

- 아동용 컬러링 도안에 적합하도록 자잘한 무늬를 줄이거나 단순화함. 2차 학습 데이터로 노이즈를 추가한 input 이미지로 활용함으로써, 노이즈와 유사한 작은 무늬들을 무시하도록 학습이 된 것을 확인할 수 있음. 그러나 자잘한 무늬의 중요성을 알아서 알 수 없다는 점에서 일괄적으로 무늬를 줄이는 본 모델은 단순한 일러스트레이션의 경우에서 적용은 어려움.





- 선이나 색이 뚜렷하게 구분되지 않은 경우에도 어느 정도의 스케치를 제공함. 2차 학습 데이터의 input이미지로 노이즈를 추가하고 blur 및 dilate 처리한 이미지를 활용함으로써, 뿌연 경우에도 선을 잡을 수 있도록 학습이 된 것을 확인할 수 있음. 실제 사진 및 사진과 유사한 사실적인 일러스트레이션의 경우 더 나아진 결과를 보여줌.
- 본 프로젝트는 상담에서 사용하며, 대상 내담자를 아동임. 따라서 1차 학습한 모델을 바탕으로 단순 일러스트레이션을 컬러링 도안으로 변환하는 것을 디폴트로 하고, unsplash+clip을 통한 이미지를 활용하는 경우(풍경 등을 키워드로 하는 경우에 한정) 2차 학습한 모델을 사용함.