

AI 건축 디자이너

어경권 김민욱 박성균 이도훈 임수연 진승욱

1. Introduction

우리나라 성장기 때 지어진 건물들이 지어진 지 2~30년 된 것들이 많고, 이를 아예 재건축 하거나, 아니면 외관만 리모델링 하려는 니즈도 꾸준하다. 실제로 한국외대의 경우 어문학관 외관 리모델링으로 인해 '제6회 한국 리모델링 건축대전'에서 대상을 수상하기도 하였다. 하지만 클라이언트는 전문적인 도메인지식이 부족하고 정작 자신이 뭘 원하는지 자신도 잘 모르는 경우가 많고, 그래서 설계사와 클라이언트 간의 원하는 집 디자인의 괴리가 생긴다. 그래서 설계사와 클라이언트 간의 니즈를 맞추는 과정에서 소요되는 비용이 상당하다. 따라서 클라이언트가 본인이 원하는 니즈를 정확하게 전달할 수 있도록 자신이 원하는 건축물의 형태와 건축 자재/건축가를 선택하면 새로운 집의 이미지를 생성하는 GAN모델을 만들고자 했다.

GAN은 정답을 알려주고 ai를 학습시키는 지도학습과는 달리 정답을 알려주지 않고 ai를 학습시키는 대표적인 비지도 학습이며 generator와 discriminator를 동시에 학습시키는 기술이다.

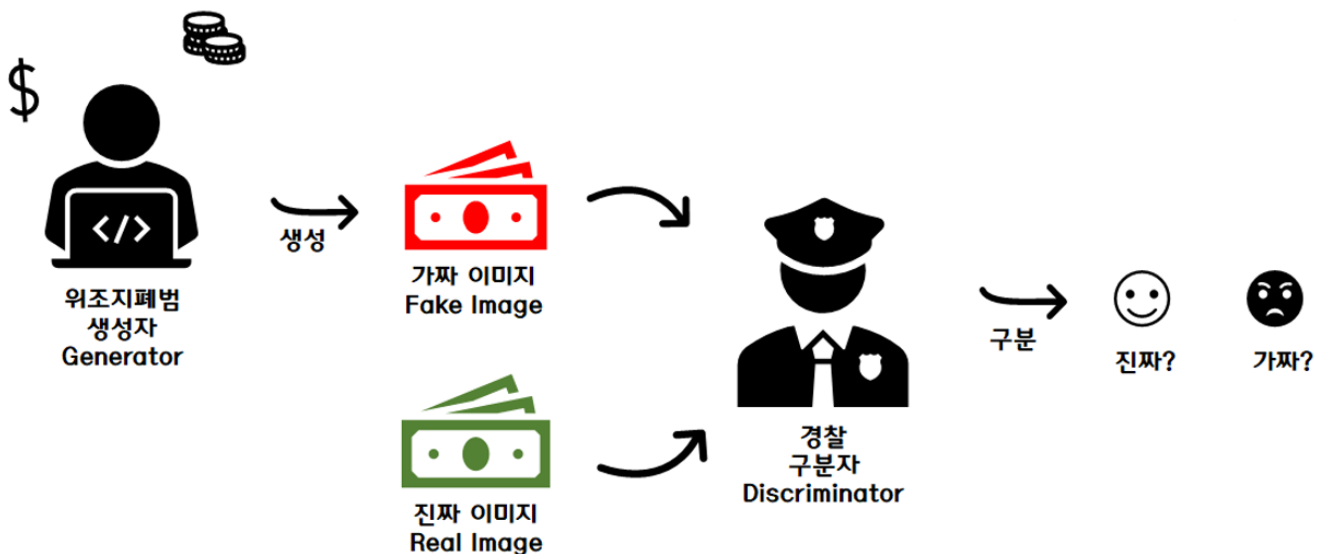


그림 1. GAN의 기본 구조

다음 그림을 보면 generator를 위조 지폐범에 비유할 수 있고 Discriminator를 위조지폐를 감별하는 경찰에 비유할 수 있다. Generator는 discriminator를 속이도록 학습하며, discriminator는 generator의 속임수에 넘어가지 않고 진짜와 가짜를 구분하도록 학습한다. 이 과정에서 generator는 진짜 같은 가짜를 만들어 discriminator를 속이도록 학습하며, discriminator는 generator의 속임수에 넘어가지 않고 가짜를 구분하도록 학습한다.

2. Development process

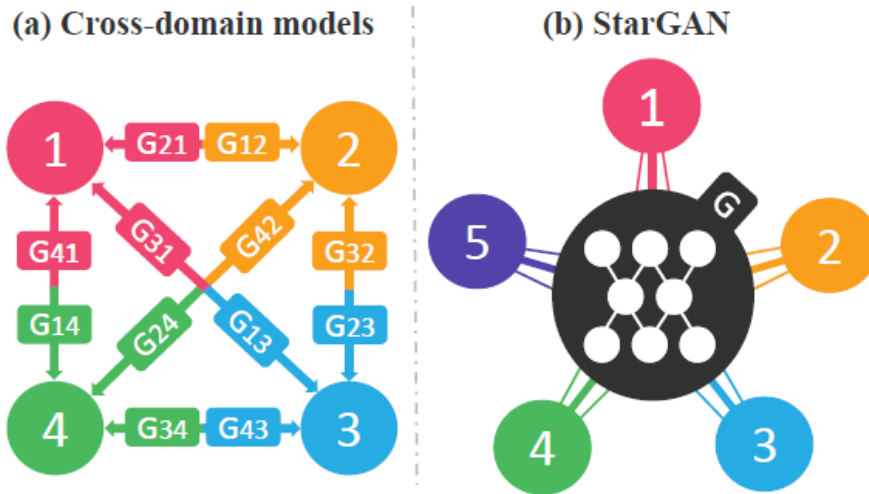


그림 2. StarGAN 구조

처음 생각했던 StarGAN 모델의 경우, 하나의 Generator로 여러 domain의 이미지를 효율적으로 생성할 수 있다는 특징이 있다. 기존의 Cross-domain GAN 모델들은 n 개의 도메인간 image translation task를 수행하기 위해 $n(n-1)$ 개의 generator가 필요하지만 StarGAN 모델의 경우 하나의 generator로 여러 domain의 이미지를 각각 따로 학습시키지 않아도 생성이 가능하다. 따라서 건물 이미지를 생성하는 모델에 적합하다고 판단하였다.

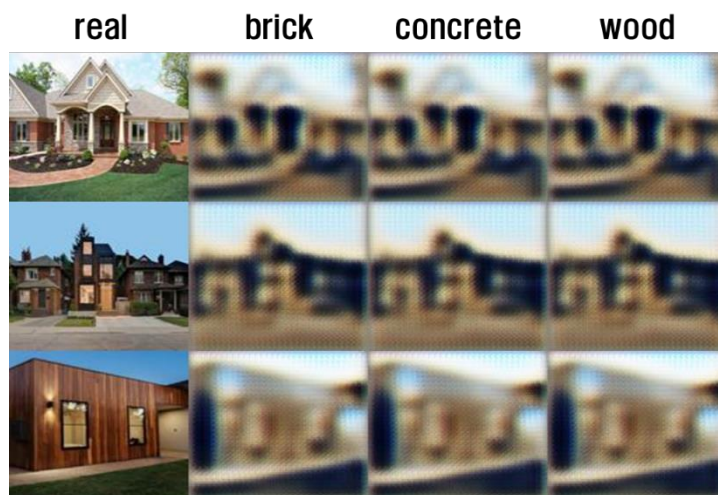


그림 3. StarGAN Output

하지만 StarGAN의 경우 하나의 Shared Generator를 사용한다는 그 특징 때문에 input 사진에 대해 큰 형태의 변화를 기대하기 힘들었다. StarGAN 논문에서 실제로 사용한 domain들도 머리색이나 피부색과 같은 얼굴 내에서 미묘한 차이들을 만들어내는 generator를 학습시킨 것에 불과하여, 건축 양식의 변화라는 큰 단위의 image translation을 하기에는 적

합하지 않았다. 실제로 저희가 dataset을 만들어 학습시켜본 결과, target domain 사이의 이미지들 간 variation이 너무 커, image generation이 제대로 되지 않는 것을 확인할 수 있었다.

결국 한 종류 domain으로의 translation이라도 제대로 학습할 수 있는 direct translation model이 필요하다고 판단하여 single-domain 내에서 unpaired image-to-image translation을 하는 cycleGAN 모델로 방향을 수정하였다.

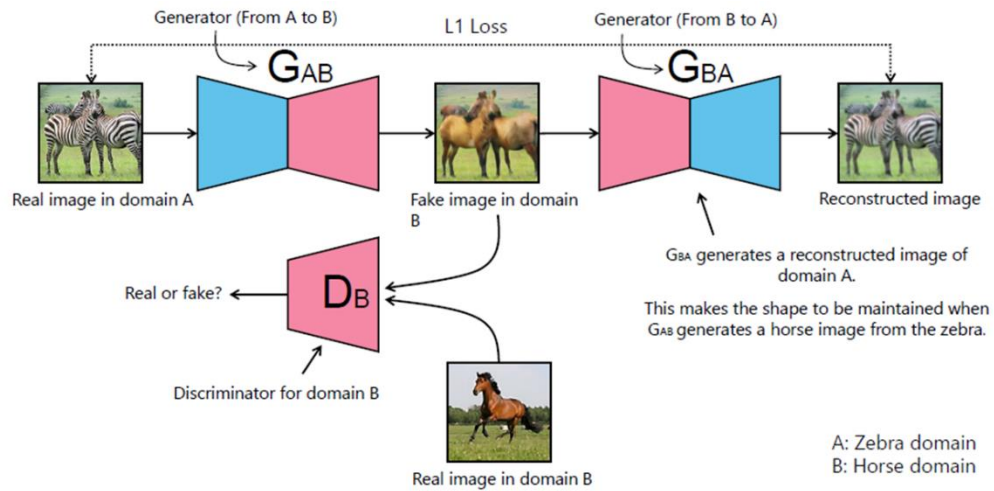


그림 4. CycleGAN 구조

CycleGAN의 경우 unpaired된 image dataset을 사용하여 Image translation을 하는데, real image를 통해 생성된 fake image를 다시 real image로 reconstruction하는 과정을 통해 이러한 unpaired domain간 학습을 가능하게 한다.

이러한 CycleGAN 모델의 경우 StarGAN보다 single domain-to-domain translation에서 향상된 성능을 기대할 수 있었다. 하지만 건축물 dataset으로는 학습을 시켰을 때 해당 domain에 대한 색상의 변화 정도만이 가능하고, 건축양식의 질감을 구현하기가 매우 어려웠다. 지향하는 정도의 형태 변화를 학습시키기 위해선 해당 domain 내에 수없이 다양한 질감과 형태들을 모두 학습시키기 위한 상당히 많은 정제된 Dataset이 필요하다고 생각되는데, 제한된 프로젝트 기간 내 저희에게 필요한 만큼의 dataset 구축은 불가능에 가깝다고 판단되었다.

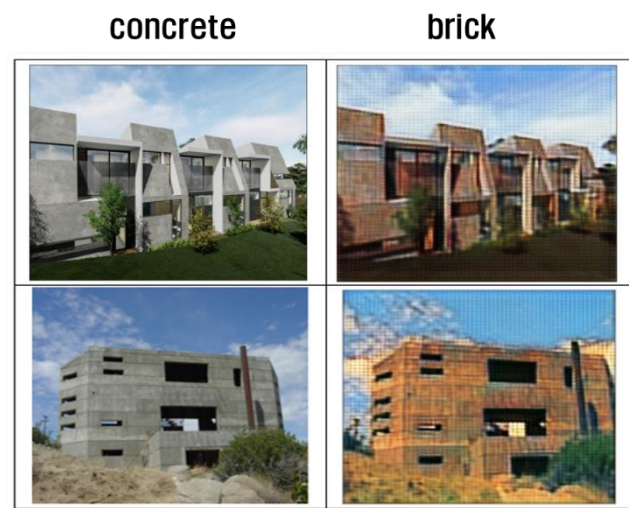


그림 5. CycleGAN Output

결국 한정된 dataset을 활용해 효율적으로 건축 양식을 학습시킬 수 있는 방법이 필요하게 되었고, 따라서 paired image-to-image translation인 방법이 최선이라고 판단하여, pix2pix 모델을 활용하기로 하였다.

Pix2pix 모델은 conditional GAN이라는 모델을 확장하여 general image-to-image translation을 해결한 모델로서 훈련하기가 쉽고, 적은 dataset만으로도 효율적으로 Generator를 학습시킬 수 있다는 특징이 있다. 실제로 pix2pix에서는 400개의 façade dataset만으로도 실제 건물 사진과 비슷한 image generation을 수행 하여 프로젝트의 목적에 적합하다고 판단이 되었다.

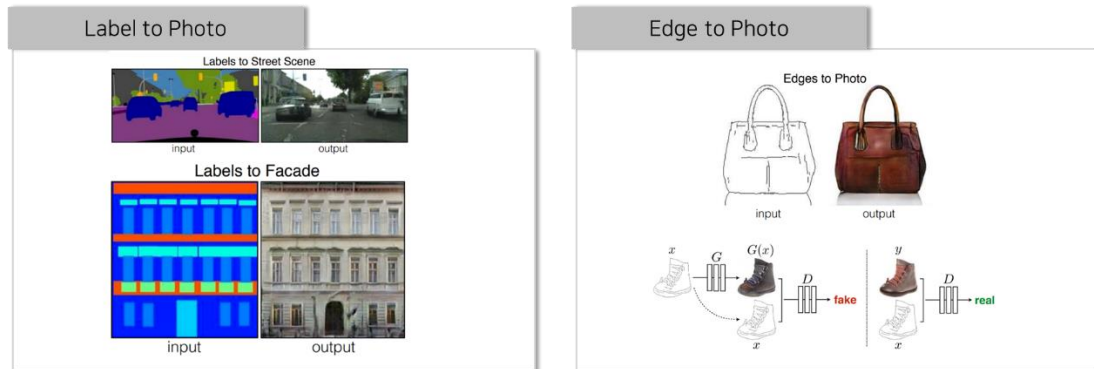


그림 6. Image Generation 방법

Pix2pix의 여러 가지 image generation 방법들 중, 원하는 image를 생성하기 위해 적합한 pix2pix 방법은 Label to Photo, Edge to Photo 두 가지 방법이 있었다. 그 중 Label to Photo의 경우 input image를 segmentation해서 label을 단 후 image generation 학습을 시킨다. 이는 segmentation과 labeling이 효과적으로 되었을 때 학습 효율이 높지만, pix2pix에서 사용한 dataset만큼 상세하게 segmentation을 하는 것은 전처리에 많은 시간이 소요된다. 반면 Edge to Photo 방법의 경우 건물과 같은 복잡한 형태는 학습 효율이 낮을 것으로 예상되지만, edge generation과정이 쉽고 빨라 다양한 형태로 학습을 진행할 수 있다는 장점이 있다. 따라서 특징이 확실한 Brick Domain에 대해서 Edge-to-Photo 학습을 통해 일반적인 건물 façade를 brick domain façade로 translation하는 Generator를 만들기로 하였고, dataset preparation을 하였다.

3. Model

pix2pix 모델을 학습시키기 위해서 구글 이미지 크롤링 과 free stock sites에서 총 405개의 brick building facade data를 수집하였고 edge detection 알고리즘을 사용해 그 이미지에 맞는 sketch data를 생성하였다. 그리고 학습시킨 generator가 벽돌의 texture를 생성하는지 알아보기 위해 CMP Façade data 100개를 사용하였다.

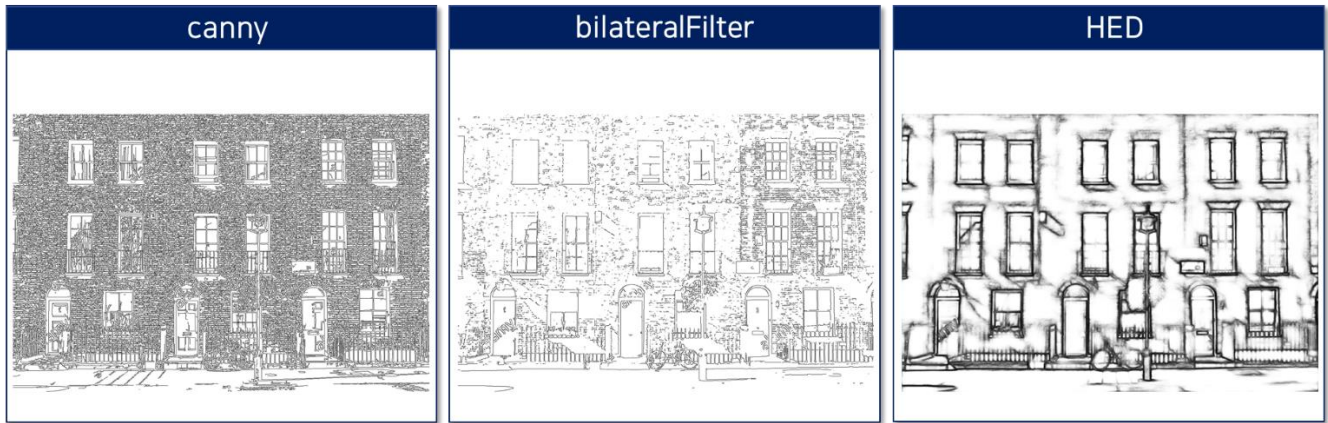


그림 7. Edge detection

데이터 전처리 단계 중 3가지 방법의 edge detection으로 sketch data를 생성하였다. 첫 번째 방법으로는 단순히 canny edge detection 알고리즘을 사용하여 sketch data를 생성하였습니다. 이 방법은 속도는 빠르지만 threshold를 작게 주면 noise가 너무 많고, threshold가 높으면 중요한 outline을 지워버리는 단점이 있다. 두 번째 방법은 bilateral filter를 사용해서 이미지를 블러 처리한 후에 첫 번째 방법과 동일한 canny edge detection 알고리즘을 사용했다. 이를 이용하면 canny보다는 noise를 줄여주지만 중요한 outline을 지워버리는 문제점이 있었다. 세 번째 방법은 HED 방법입니다. 이는 딥러닝을 활용해 중요한 edge를 판단하여 detection하지만 속도가 느리다는 단점이 있었다. 이 세가지 전처리 방법의 비교를 위해 이미지 간의 거리를 측정하는 PSNR을 사용하였다.



그림 8. Pix2pix output

위의 그림을 보면 HED를 사용해서 스케치를 만든 모델이 average PSNR이 제일 높은 것을 확인할 수 있다. 따라서 HED로 전처리를 사용한 pix2pix 모델을 사용하기로 하였다.

4. Conclusion

HED를 사용하여 만든 스케치 이미지를 벽돌 건물 이미지로 변환시키는 pix2pix 모델을 생성하는데 성공하였다. 벽돌 건물이 아닌 image를 벽돌 건물로 변환하는 데 성공하였고 건물의 전체적인 느낌도 바꾸어 주는 모습을 확인할 수 있었다.

하지만 건축 디자인을 하는 generator 모델을 만드는데 있어 향후 과제가 존재한다. 이번 프로젝트에서 생성한 pix2pix의 경우 texture를 translation하고 형태의 변화를 주진 못하였다. HED sketch 자체를 학습시켜 새로운 sketch를 생성할 수 있다면 Input Sketch를 바탕으로 Gaudi와 같은 다른 형태의 이미지를 생성한다면 형태의 변화가 가능할 것이다. 또한 Self attention 매커니즘을 기존 GAN에 도입한 SAGAN을 사용한다면 건물 이미지의 형태를 변환시키는 것이 가능할 것 이라고 기대한다.

5. Reference

- 1). (pix2pix) Isola, Phillip and Zhu, Jun-Yan and Zhou, Tinghui and Efros, Alexei A. Image-to-Image Translation with Conditional Adversarial Networks. In CVPR, 2017.
- 2). (StarGAN) Choi, Yunjey and Choi, Minje and Kim, Munyoung and Ha, Jung-Woo and Kim, Sunghun and Choo, Jaegul. StarGAN. Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. In CVPR, 2018.
- 3). (cGAN) M. Mirza and S. Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014. 3.
- 4). (CycleGAN) J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired imageto-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.
- 5). (SAGAN) Han Zhang, Ian Goodfellow, Dimitris Metaxas and Augustus Odena. Self-Attention Generative Adversarial Networks. arXiv preprint arXiv:1805.08318, 2018.