CycleGAN: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Paper's goal :

- Unpaired data를 이용한 도메인 X → 도메인 Y의 이미지 생성 모델 구성
- Paired data가 필요 없기 때문에 좀 더 자유로운 훈련 가능

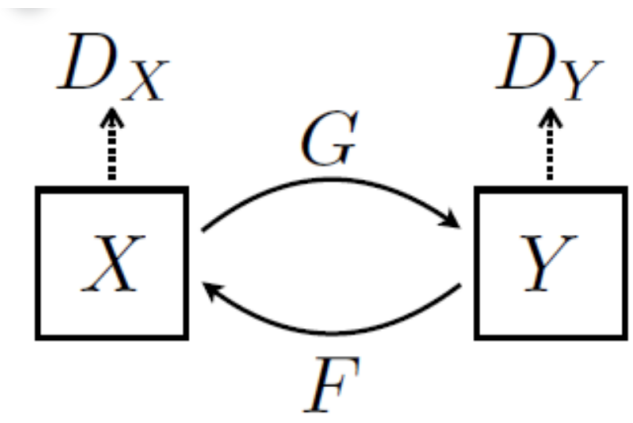
Image to Image translation 이란?

- pair 형태의 훈련 이미지 데이터셋을 사용하여 input 이미지 와 output 이미지를 매핑하는 것을 목표로 하는 생성모델의 한 분야
- 짝이 있는(paired) 훈련 데이터를 얻는 것이 어렵고 비용이 많이 듬

Unpaired 데이터가 활용되기 어려운 이유

- 특정한 이미지 x가 주어졌을 때 데이터 쌍이 존재하지 않는다면 이것을 어떤 정보로 변환을 해야할지 알지 못함
- 일반적인 GAN Loss로는 mode-collapse 발생
  - 입력 데이터로 무엇이 들어왔든 무시하고 오직 D를 속이기 위해 가이드 없이 학습을 진행하다 보니 전혀 다른 이미지를 생성
  - 어떤 input data가 들어와도 똑같은 ouput만 내놓음
- 이러한 문제를 해결하기 위해서 CycleGAN에서는 Generator가 Discriminator를 속이는 것 뿐만 아니라 이미지A → 이미지B 로 바꿀 때 다시 이미지A ← 이미지B 로 복구 가능하도록 하는 **Cycle Consistency Loss** 를 추가해서 학습을 함

CycleGAN Model



- 기본적인 구조는 다음과 같이 Generator 2개, Discriminator 2개로 구성
- X와 Y는 각각의 데이터셋을 의미
- G와 F는 역함수 관계

1. **Generator *G*** : X → Y 매핑을 수행
2. **Generator *F*** : Y → X 매핑을 수행
3. **Discriminator *Dy*** : 실제 도메인 Y의 이미지 y와 *G*가 생성한 *y^=G(x)*을 구분
4. **Discriminator *Dx*** : 실제 도메인 X의 이미지 x와 *F*가 생성한 *x^=F(y)*을 구분

Cycle Consistency Loss

- **Adversarial Loss**

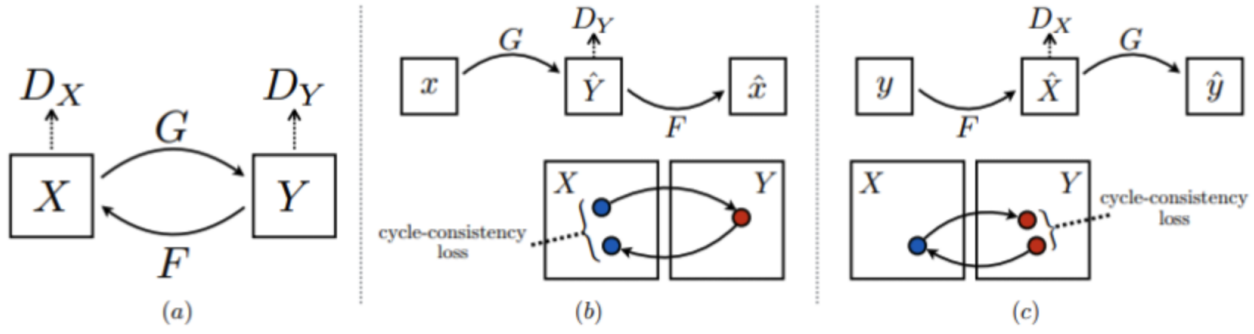
$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))], \quad (1)$$

  - 일반적인 GAN의 loss
    - G는 위의 함수를 최소화시킬려고 하고, D는 위 함수를 최대화 시키고자 함
    - $y \sim p_{data}(y)$는 Y의 데이터 분포를 따르는 원소 y를 말함
    - D_Y는 Y의 데이터 분포에서 왔는지 (도메인 Y인지 아닌지) 판단하는 Discriminator로 0~1 사이의 확률 값을 반환
    - Y 분포에서 왔다고 판단하면 1에 가까운 값을 아니면 0에 가까운 값이 나옴
  - 주의할 점은 해당 loss뿐 아니라 역에 대한 adversarial loss인 $L_{GAN}(G, D_x, Y, X)$도 고려 필요
    - 정방향, 역방향 학습 모두 GAN loss 적용
  - 현실적인 이미지를 만들도록 학습
    - 앞서 말한 것처럼 이 loss 만으로는 mode collapse 문제가 생길 수 있기 때문에 가능한 매핑 함수의 공간을 줄이기 위해, 새로운 **loss** 추가 → **Cycle Consistency Loss**

- **Cycle Consistency Loss**

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1]. \quad (2)$$

  - 입력이나 출력을 복원했을 때 실제값과 복원값의 차이를 계산
  - 복원된 이미지가 실제 이미지와 같으면 Loss가 0이 되도록 구성



  - F(G(x)), G(F(y))를 각 x, y에 가깝게 만들어 Cycle consistent 하게 만들어 줌
  - 위 이미지의 (b)는 forward cycle consistency, (c)는 backward cycle consistency
    - (b) : x → G(x) → F(G(x)) ≈ x
    - (c) : y → F(y) → G(F(y)) ≈ y

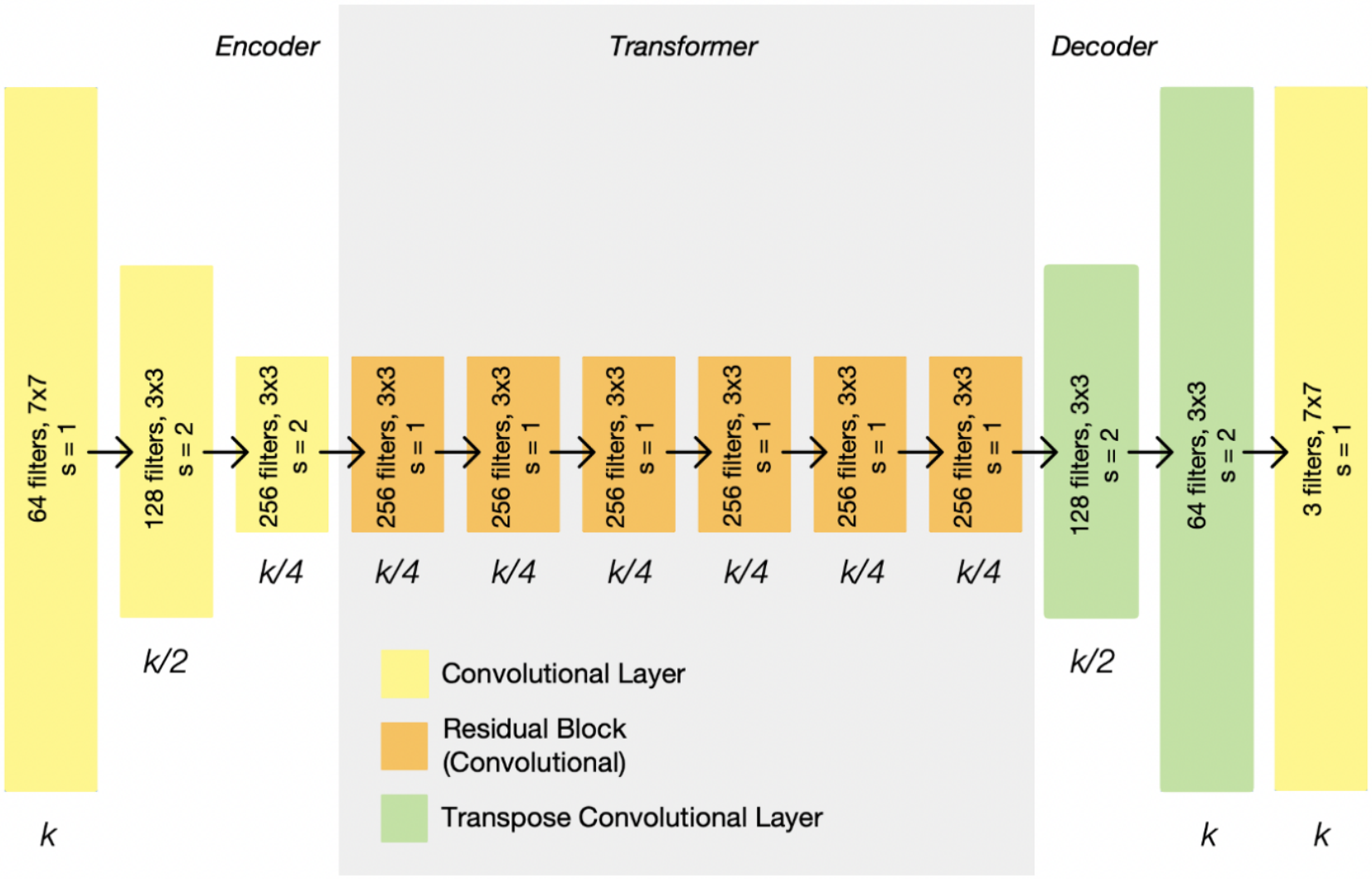
Full Objective

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ &+ \lambda \mathcal{L}_{\text{cyc}}(G, F),\end{aligned}$$

- Adversarial Loss + Cycle Consistency Loss
- λ는 cycle consistency loss 와 adversarial loss 간의 상대적인 중요도를 통제하는 하이퍼 파라미터

Implementation

- Architecture
  - Generator network: **Perceptual lossesfor real-time style transfer and super-resolution(Johnson et al. 2016)** 의 아키텍처를 기반으로 함



https://towardsdatascience.com/cyclegan-learning-to-translate-images-without-paired-training-data-5b4e93862c8d

  - Discriminator network : pix2pix와 동일하게 PatchGAN 70x70을 사용

- Training Details
  - 본 논문에서는 학습의 안정화와 더 나은 성능을 위해 **least-squared loss** 로 변경 (original : negative log likelihood objective)
  - 그동안 생성한 **image history buffer를 활용**하여 최근 생성한 50개의 이미지를 지속적으로 저장하고 그것들을 이용해 학습을 진행 (모델 진동 최소화)
  - λ = 10
  - batch size = 1 , Adam solver 이용

- 처음 100 epoch : learning rate = 0.0002
  이후 100 epoch : 선형적으로 0에 가까워지게 lr를 줄여가며 학습

Result

- **Evaluation Metrics**
  - pix2pix와 같은 데이터셋과 평가지표 사용
  - 모델 성능 평가를 위한 지표
    - AMT(Amazon Mechanical Turk) perceptual studies : 사람을 대상으로 실험 (정성적)
    - FCN score : 사람을 대상으로 한 실험이 필요하지 않은 automatic 한 양적 기준 (정량적)
    - per-pixel accuracy, IoU : Semantic segmentation metrics, 사진을 라벨링 하는 성능을 평가하기 위한 기본적인 평가 지표

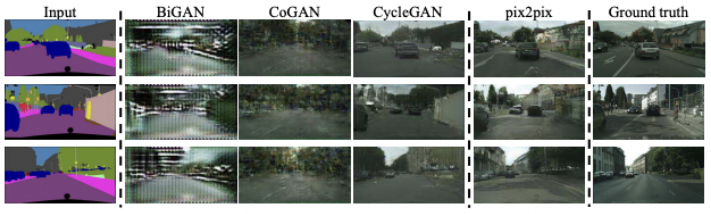


Figure 4: Different methods for mapping labels→photos trained on cityscapes. From left to right: input, BiGAN/ALI [6, 7], CoGAN [28], CycleGAN (ours), pix2pix [20] trained on paired data, and ground truth.

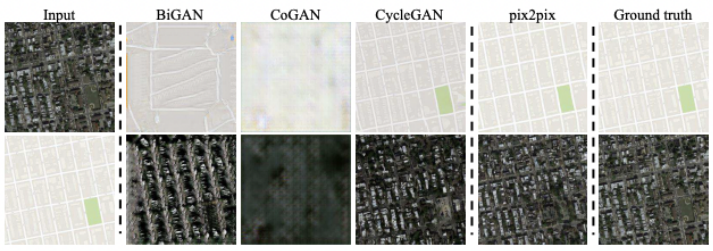


Figure 5: Different methods for mapping aerial photos↔maps on Google Maps. From left to right: input, BiGAN/ALI [6, 7], CoGAN [28], CycleGAN (ours), pix2pix [20] trained on paired data, and ground truth.

Qualitative evaluation

Loss	Map → Photo % Turkers labeled *real*	Photo → Map % Turkers labeled *real*
CoGAN [28]	0.6% ± 0.5%	0.9% ± 0.5%
BiGAN/ALI [7, 6]	2.1% ± 1.0%	1.9% ± 0.9%
Pixel loss + GAN [42]	0.7% ± 0.5%	2.6% ± 1.1%
Feature loss + GAN	1.2% ± 0.6%	0.3% ± 0.2%
CycleGAN (ours)	**26.8% ± 2.8%**	**23.2% ± 3.4%**

Table 1: AMT "real vs fake" test on maps↔aerial photos.

Loss	Per-pixel acc.	Per-class acc.	Class IOU
CoGAN [28]	0.40	0.10	0.06
BiGAN/ALI [7, 6]	0.19	0.06	0.02
Pixel loss + GAN [42]	0.20	0.10	0.04
Feature loss + GAN	0.06	0.04	0.01
CycleGAN (ours)	**0.52**	**0.17**	**0.11**
pix2pix [20]	0.71	0.25	0.18

Table 2: FCN-scores for different methods, evaluated on Cityscapes labels→photos.

Loss	Per-pixel acc.	Per-class acc.	Class IOU
CoGAN [28]	0.45	0.11	0.08
BiGAN/ALI [7, 6]	0.41	0.13	0.07
Pixel loss + GAN [42]	0.47	0.11	0.07
Feature loss + GAN	0.50	0.10	0.06
CycleGAN (ours)	**0.58**	**0.22**	**0.16**
pix2pix [20]	0.85	0.40	0.32

Table 3: Classification performance of photo→labels for different methods on cityscapes.

Quantitative evaluation

- 실험1 : Baselines 비교

        - pix2pix와 cycleGAN이 상당히 우수함을 확인
    - 실험2 : *ablation study (loss function)*
        - loss function을 넣었다 뺐다 하는 ablation study 진행.
        - 두 loss function 모두 온전히 사용하는 것이 중요
    - 실험3 : Image reconstruction quality
    - 실험4 : Application(cycleGAN for unpaired data)
        - 그림을 사진으로 변환하는 task 한정으로 새로운 loss 추가
            - 인풋과 아웃풋의 색구성을 보존하기 위해 추가한 loss

$$L_{identity}(G, F) = E_{y \sim p_{data}(y)}[\| G(y) - y \|_1] + E_{x \sim p_{data}(x)}[\| F(x)) - x \|_1]$$

Limitations and Discussion

- 다른 baseline 모델 들에 비해 여러 많은 실험 및 응용에서 좋은 성능을 보여주고 image translation의 퀄리티가 우수함을 입증
- 그러나 항상 좋은 결과만 낸 것은 아니면 몇몇 training 데이터셋에서는 실패하기도 함