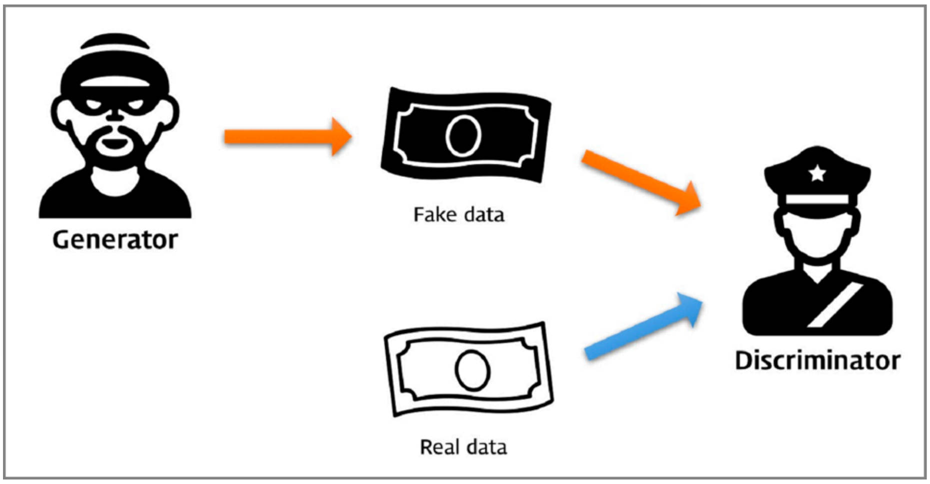
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

(CycleGAN)

1. GAN



Analogy of police and counterfeiters

- Generative model G: captures the distribution of training data
  - Maximize the probability of D making a mistake
- Discriminative model D: estimates the probability that a sample came from the training data rather than G

2. CycleGAN

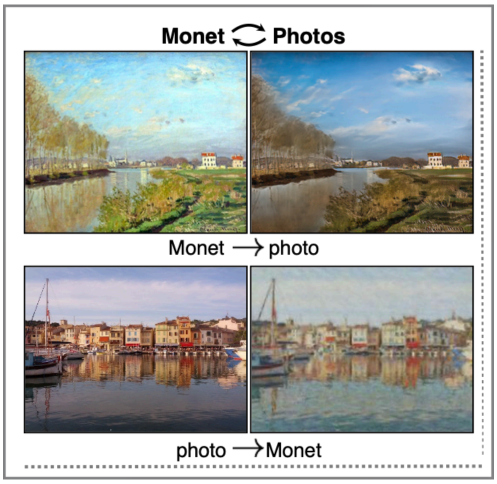


Image-to-Image Translation

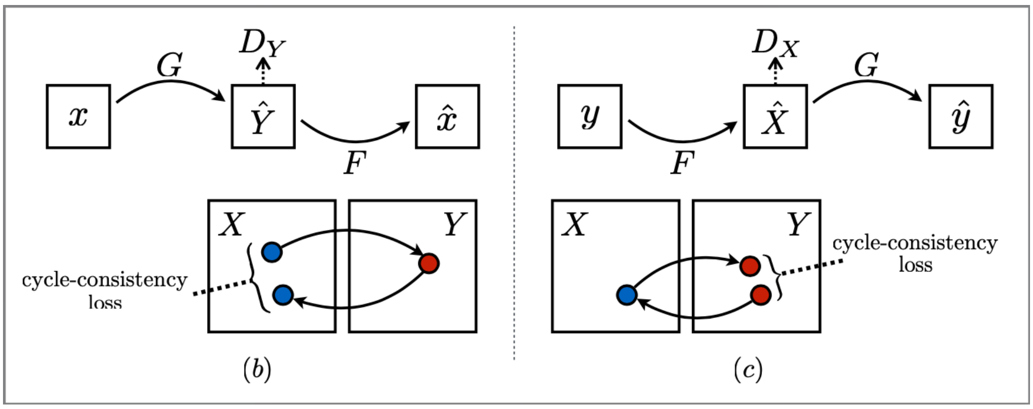
- Capture special characteristics of an image collection —> Figure out how these can be translated into the other image collection
- Approach builds on the "pix2pix" framework (supervised)
- <u>Unsupervised learning</u>: Unpaired training data (source set X and target set Y)

- Full objective:

  $l(G, F, D_X, D_Y) = l_{GAN}(G, D_Y, X, Y) + l_{GAN}(F, D_X, Y, X) + \lambda l_{CYC}(G, F)$ $(\lambda = 10)$

  $G*, F* = arg(min_{G,F} max_{D_X,D_Y} l(G, F, D_X, D_Y)$

- <u>Adversarial loss</u>: Learn a mapping $G : X \rightarrow Y$ s.t. $G(X)$ is indistinguishable from Y
  - $l_{GAN}(G, D_Y, X, Y) = E[logD_Y(y)] + E[log(1 - D_Y(G(x))]$ (cross-entropy loss)
  - $min_G max_{D_Y} l_{GAN}(G, D_Y, X, Y)$ and $min_F max_{D_x} l_{GAN}(F, D_X, Y, X)$
  - Mode collapse: all input images map to the same output image
- <u>Inverse mapping & Cycle-consistency loss</u>: train both the mapping G and F simultaneously and encourage $F(G(x)) = x$ and $G(F(y)) = y$
  - $l_{CYC}(G, F) = E[||F(G(x)) - x||] + E[||G(F(y)) - y||]$



Cycle-consistency Loss

3. Further Studies

- Evaluation Metrics
  1. AMT perceptual studies: human experiments; shown a sequence of pairs of images, click on the image they thought was a real one
  2. FCN score: apply pre-trained semantic classifier, FCN-8 to the generated image —> compare the FCN-predicted label map and the input ground truth label (How?)
     - Fully Convolutional Networks for Semantic Segmentation 참고
  3. Semantic segmentation metrics: analyze per-pixel accuracy, per-class accuracy, mean IOU

- Evaluation Baselines
  1. 여러 종류의 GAN과의 성능 비교 —> unsupervised GAN 중 가장 성능이 좋음
  2. "fully supervised" pix2pix와의 성능 비교 —> supervised model과 비슷한 성능을 보임
  3. CycleGAN 내에서의 Cycle (forward, backward), GAN loss에 대한 ablation study —> both terms are critical to the performance

- Application: GTA에 CycleGAN을 도입 —> 게임이미지를 실사화 —> 자율주행 학습

- Limitation: not uniformly positive results
  - Paired training data의 부재: 논문의 저자는 이러한 ambiguity의 해소를 위해서는 어느 정도의 weak semantic supervision이 필요할 수도 있으며, 위 논문의 핵심인 unsupervised learning에 "semi-supervised data"를 합치면 substantially more powerful translators가 탄생할 수 있다고 언급하였다.