CVPR 2017

Image-to-Image Translation with Conditional Adversarial Networks

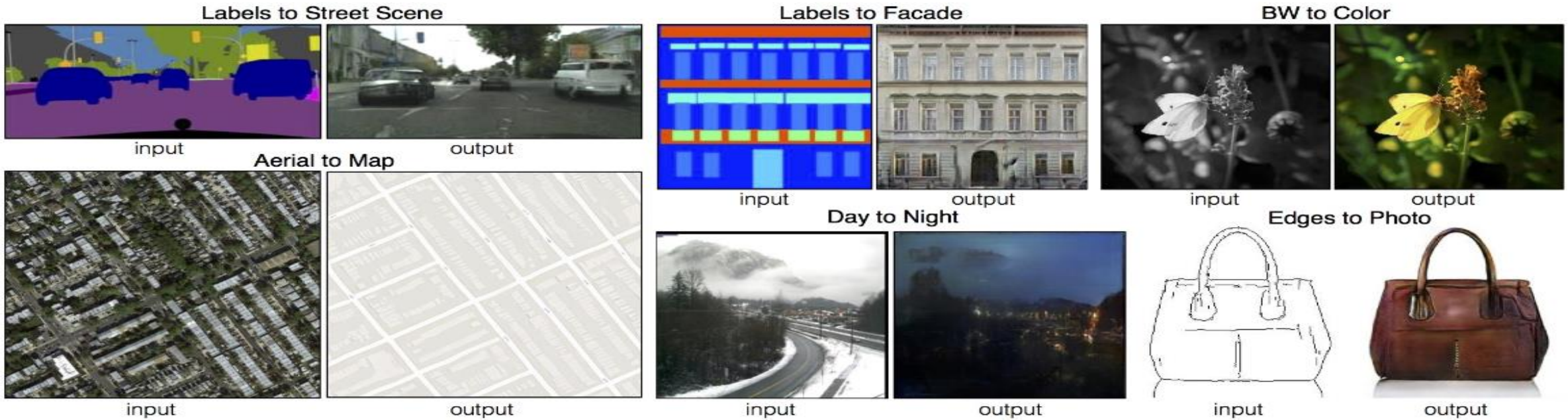
김대현

2022.11.18

목차

01 Overview

- Contributions
  - Conditional GAN을 활용하여 image-to-image translation을 제안
  - 다양한 task에 공통적으로 적용가능함을 제시
  - Loss function & Hyper Parameter의 fine tuning이 요구되지 않음



01 Inference images by 정연자수

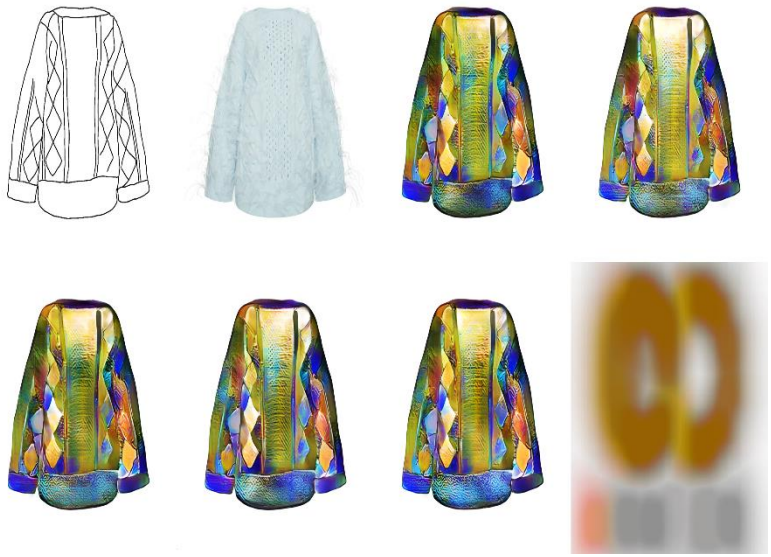


fig1. M1:Pre-trained Model(edges2handbags)

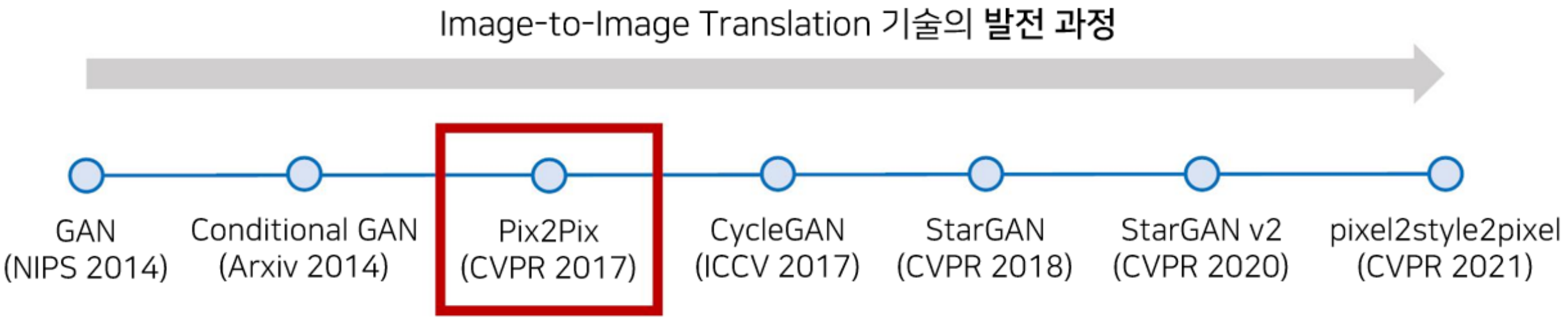
fig2. M2:Pre-trained Model(edges2shoes)

fig3. M3 : Trained Model(jy_256)

GitHub - junyanz/pytorch-CycleGAN-and-pix2pix: Image-to-Image Translation in PyTorch

목차

02 Flow of Research



- 이미지 데이터의 분포를 근사하는 모델 G를 만드는 것이 위 연구들의 목표
- 모델 G가 잘 동작한다는 의미는 원래 이미지들의 분포를 잘 모델링 할 수 있다는 것을 의미
- GAN의 후속 연구들

https://www.youtube.com/watch?v=ImiD4npRj7k

02 GAN

Overview

- GAN은 다양한 데이터를 생성할 수 있는 뉴럴 네트워크의 한 유형

Fig4. Visualization of samples from the generator model

- Not cherry-picked images
- Not memorized the training set
- Images represent sharp

Fig5. Digits obtained by linearly interpolating between coordinates in z space of the model

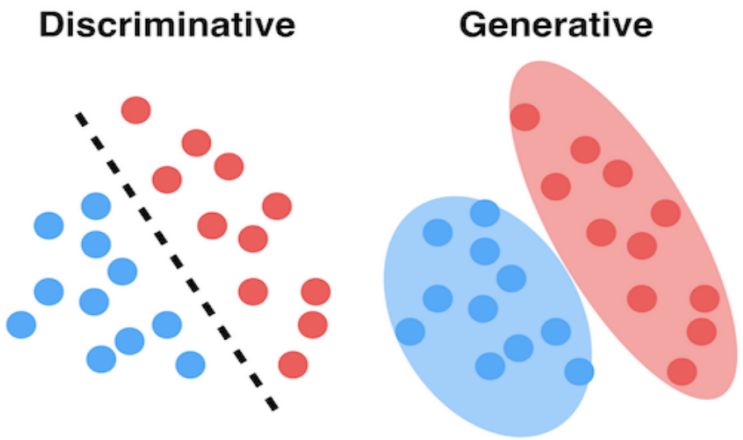
02 GAN

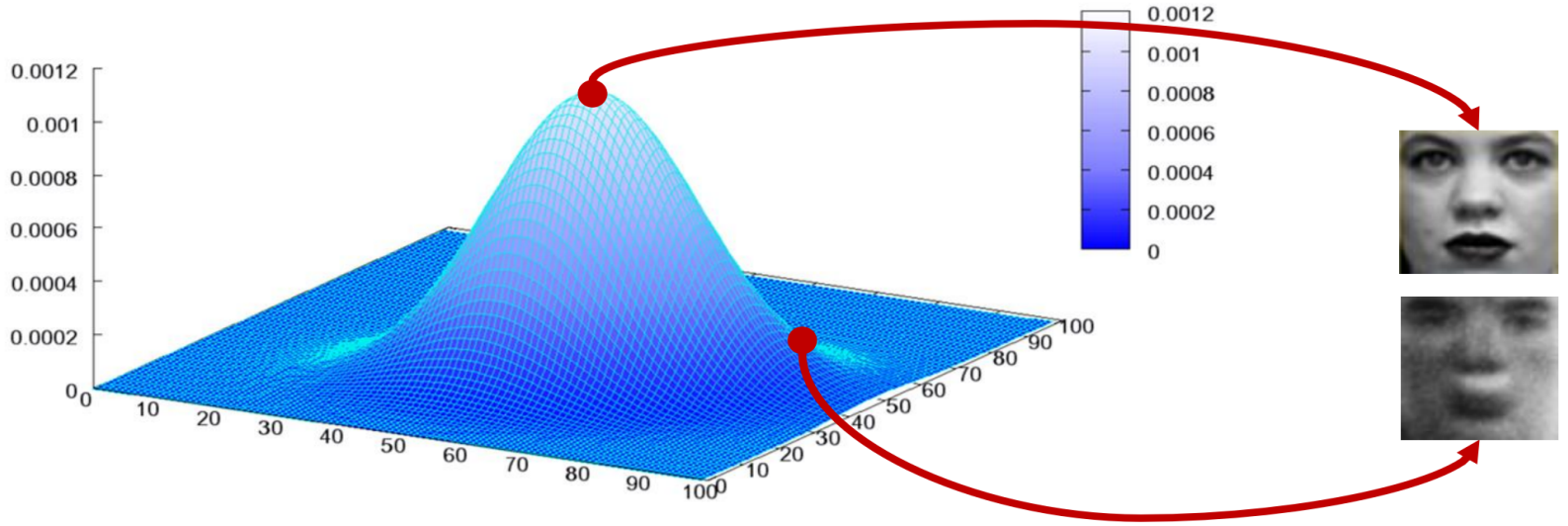
Generative Models

- 본 연구의 목표는 실존하지는 않지만 있을 법한(= semantic) 이미지를 생성할 수 있는 모델을 만드는 것

Generative Model (produce) → An image that does not exist but is likely to exist

- A statistical model of the joint probability distribution
- An architecture to generate new data instances





02 GAN

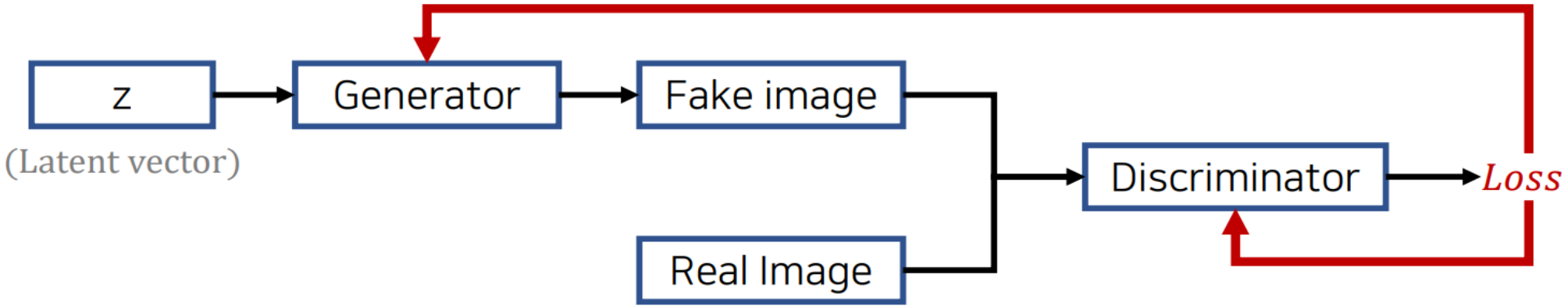
GAN 구조 이해

- **생성자(generator)**와 **판별자(discriminator)** 두 개의 네트워크를 활용한 생성 모델
- 생성자는 아래 loss function을 통해 이미지 분포를 학습

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))]$$

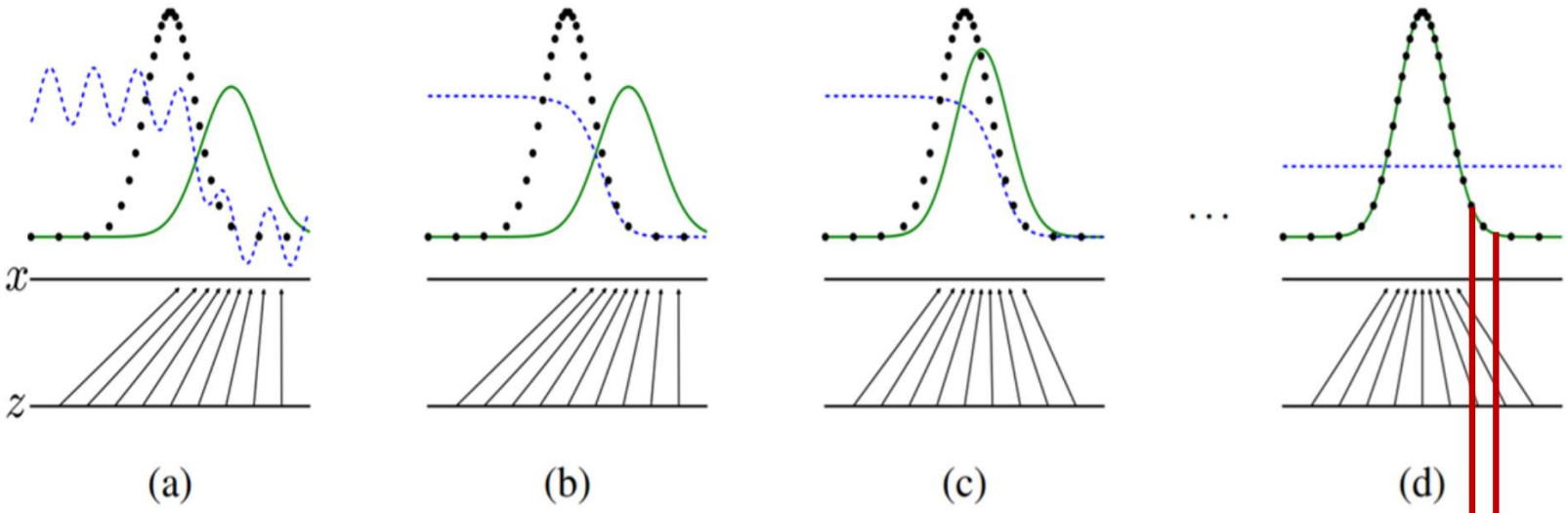
Generator $G(z)$: *new data instance*

Discriminator $D(x) = Probability$: *a sample came from the real distribution* $(Real: 1 \sim Fake: 0)$



02 GAN

GAN 수렴 과정



원본 데이터의 분포
생성 모델의 분포
판별 모델의 분포

시간이 지나면서 생성 모델 G가 **원본 데이터의 분포를 학습**

original data instance

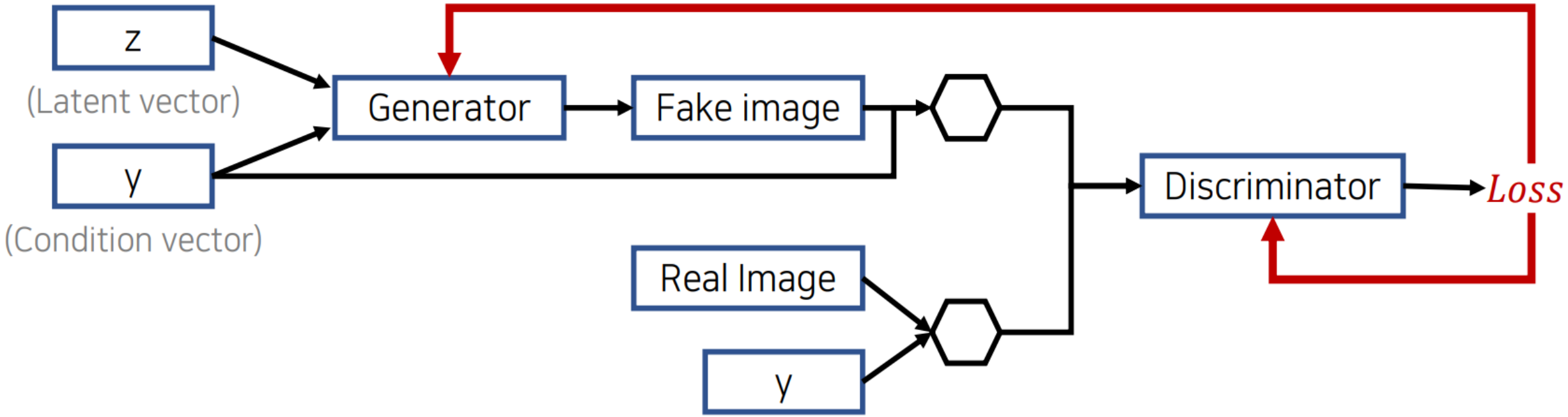
new data instance

02 Conditional GAN(cGAN)

Overview

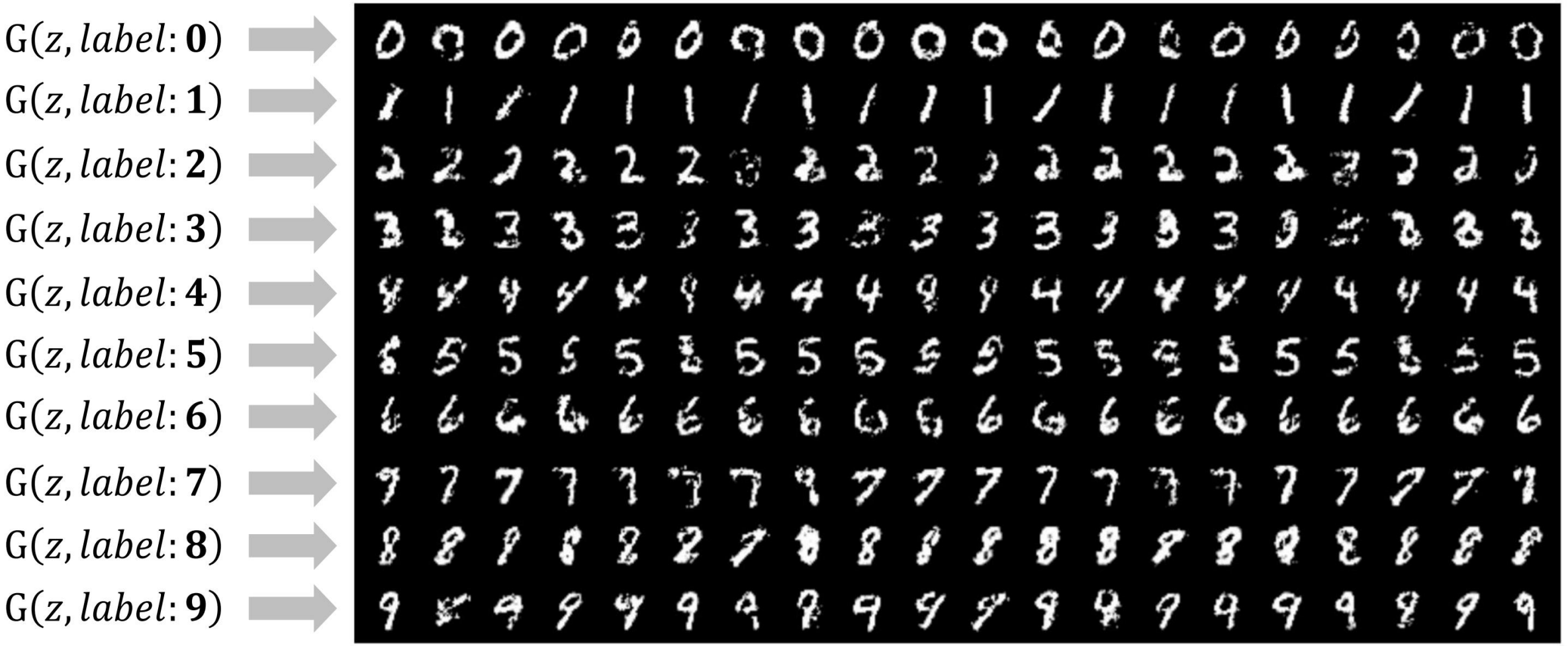
- 본 연구의 목표는 데이터의 모드를 제어할 수 있도록 조건(condition) 정보를 함께 입력하는 모델을 만드는 것

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[logD(x|y)] + E_{z\sim p_z(z)}\left[\log\left(1 - D\left(G(z|y)\right)\right)\right]$$



02 Conditional GAN(cGAN)

Experimental procedure



목차

03 Proposed Solution

Overview

- 본 연구의 목표는 이미지의 특정 양상을 다른 양상으로 바꾸는 것
- Pix2Pix는 대표적인 image-to-image translation Architecture
  - 학습 과정에서 이미지 x 자체를 조건(condition)으로 입력받는 cGAN의 한 유형
  - Pix2Pix은 픽셀들을 입력 받아, 픽셀들을 예측한다는 의미
  - 네트워크에 Dropout를 사용함으로써 random한 이미지 생성에 어느정도 영향을 미칠 수 있음

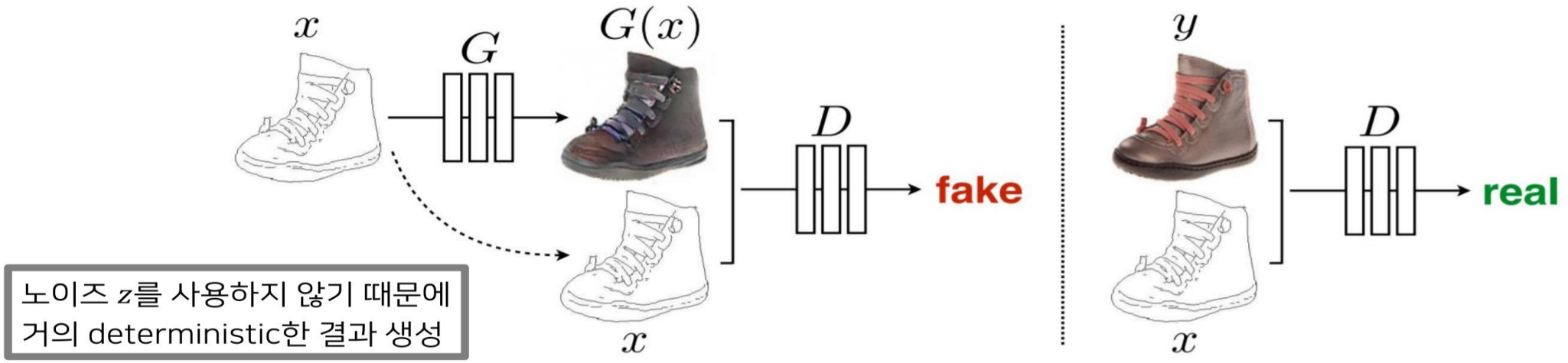
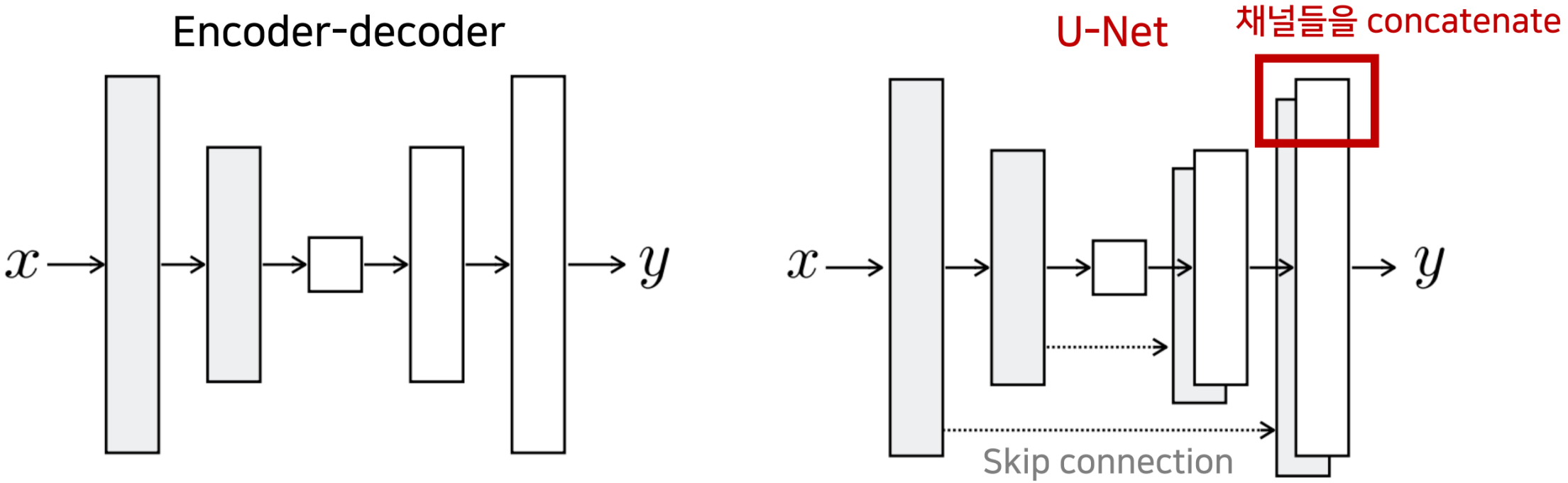


Image-to-Image Translation with Conditional Adversarial Networks (CVPR 2017)

03 Proposed Solution

Pix2pix : Architecture

- Pix2Pix는 이미지를 조건(condition)으로 입력받아, 이미지를 출력으로 내보냄
- 이를 효과적으로 처리하기 위해 U-Net 기반의 네트워크 아키텍처를 사용함 (Encoder 파트의 정보들을 활용)
- Encoder-decoder와 같이 입/출력의 차원이 같도록 하는 구조는 다양함



*U-Net: Convolutional Networks for Biomedical Image Segmentation (MICCAI 2015)

03 Proposed Solution

Pix2pix : Loss function

- GAN은 기본적으로 다른 생성 모델에 비해 blurry한 결과가 적게 나오는 편
- GAN의 성능을 더 향상시키기 위해 L1 loss function를 함께 사용 (ground – truth와 유사한 결과)
  - L2손실(Euclidean distance) 보다 L1 손실을 이용했을 때 blurry 현상이 덜 발생

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

현실적인 이미지를 만들도록 / 실제 정답과 유사하도록

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

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

*U-Net: Convolutional Networks for Biomedical Image Segmentation (MICCAI 2015)

03 Proposed Solution

Pix2pix : Discriminator

- Pix2Pix의 discriminator는 convolutional PatchGAN 분류모델을 사용
  - 이미지 전체를 판별하지 않고, 이미지 내 패치 단위(local image patch)로 real or fake 여부를 판별
  - 이는 고 해상도의 모델을 생성하기 위한 것

- 기대효과
  - 적은 파라미터 개수 필요
  - 빠른 동작 가능
  - Can be applied to arbitrarily large images

목차

04 Evaluation

Advantage 1 : U-Net Architecture (w/ image)

- U – Net 구조와 함께 본 논문이 제안한 loss(L1 + CGAN)를 사용할 때 가장 우수(real) 한 결과를 보임

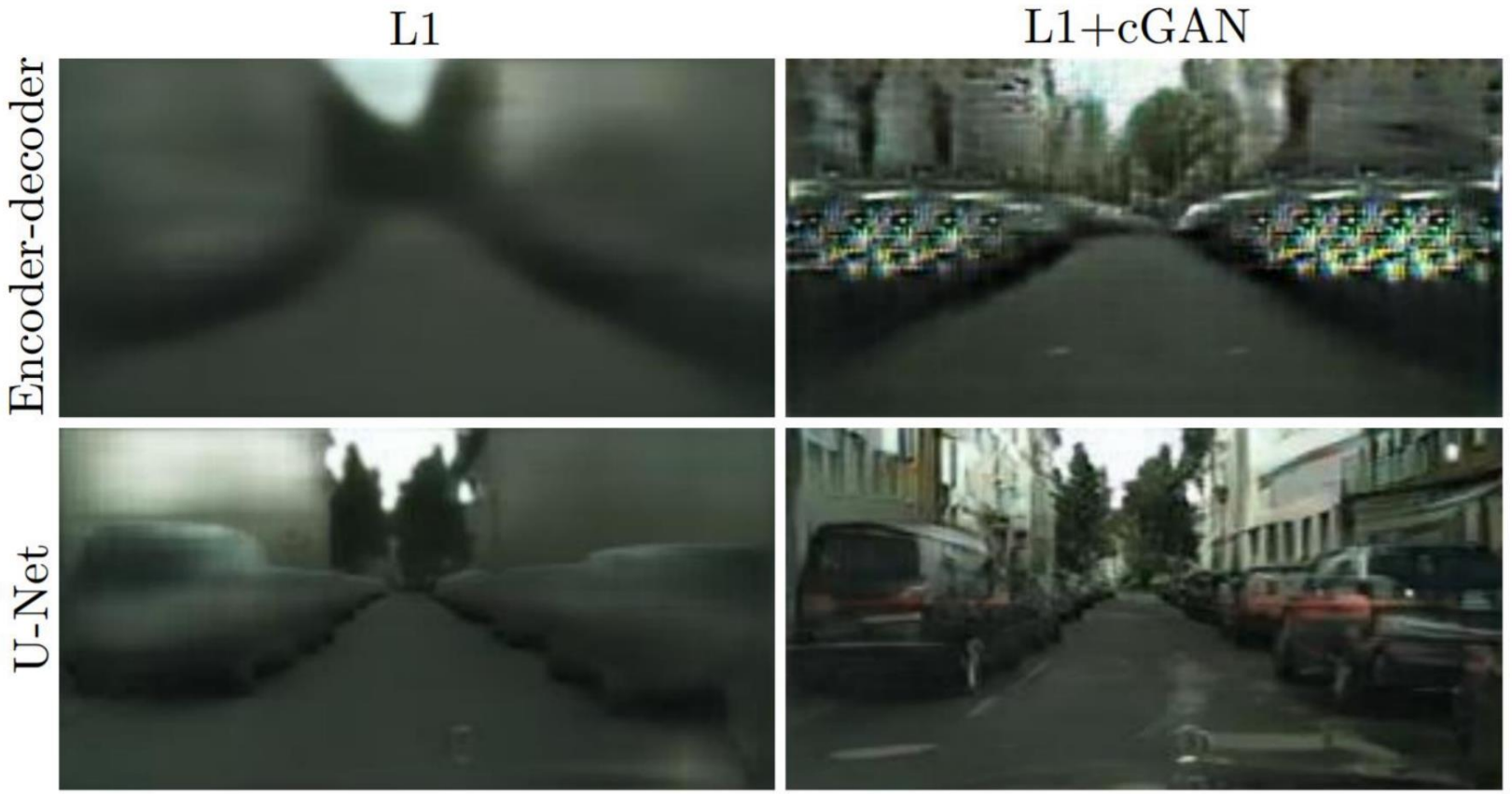


Image-to-Image Translation with Conditional Adversarial Networks (CVPR 2017)

04 Evaluation

Advantage 1 : U-Net Architecture (w/ statistics)

- U – Net 구조와 함께 본 논문이 제안한 loss(L1 + CGAN)를 사용할 때 가장 우수한 결과를 보임

Loss	Per-pixel acc.	Per-class acc.	Class IOU
Encoder-decoder (L1)	0.35	0.12	0.08
Encoder-decoder (L1+cGAN)	0.29	0.09	0.05
U-net (L1)	0.48	0.18	0.13
U-net (L1+cGAN)	**0.55**	**0.20**	**0.14**

IoU*(Intersection over Union) : Detection 이나 Semantic Segmentation의 결과를 평가하기 위한 지표 중 하나

04 Evaluation

Advantage 2 : Proposed loss function (w/ image)

1) L1 loss만 사용하는 경우 : blurry한 결과를 보임

2) cGAN loss만 사용하는 경우 : sharp 하지만 visual artifacts가 있음

3) 두 개의 loss를 적절히 혼용하는 경우 : cGAN의 결과보다 artifact가 적으면서 우수한(그럴싸한) 결과가 나옴

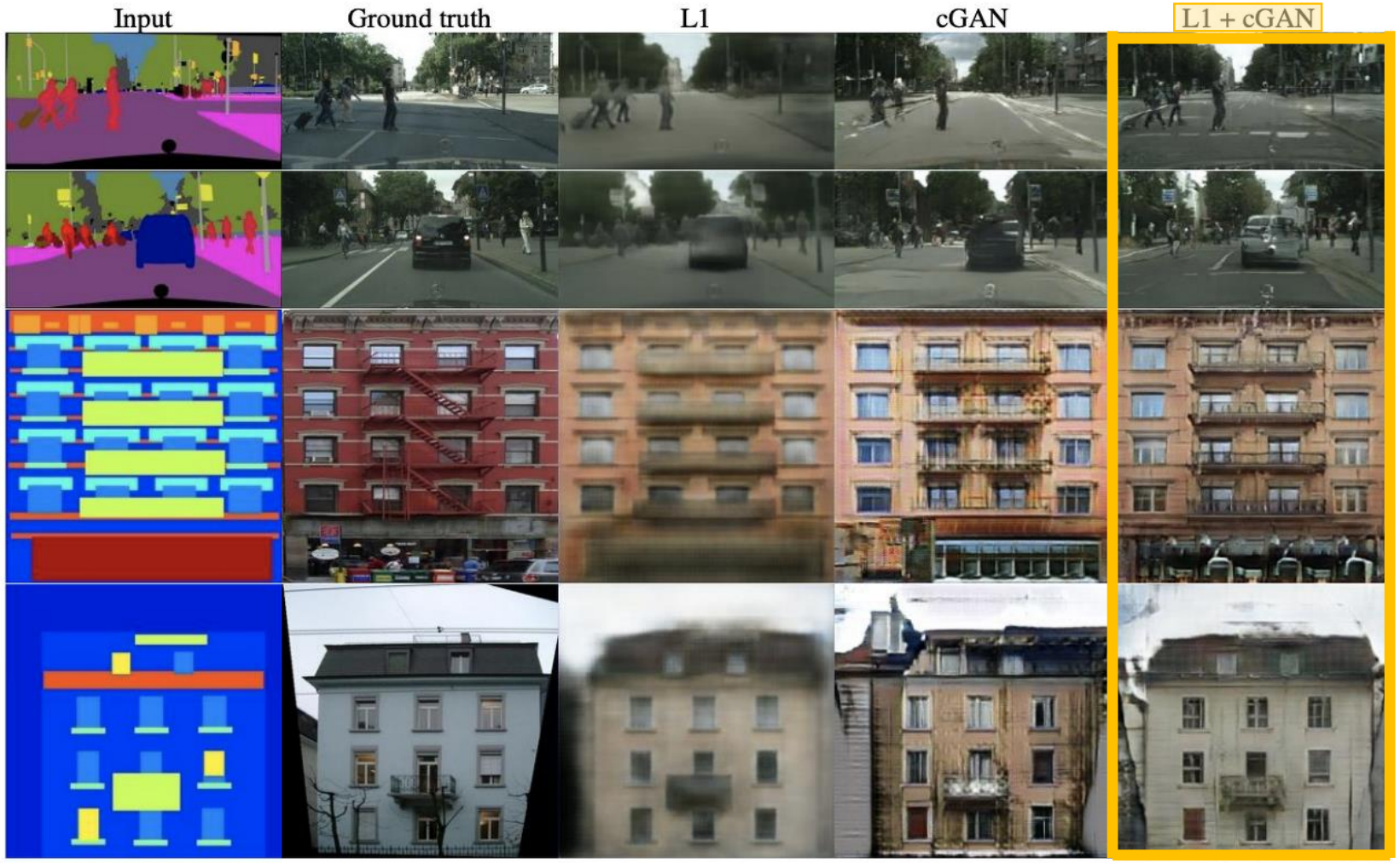


Image-to-Image Translation with Conditional Adversarial Networks (CVPR 2017)

04 Evaluation

Advantage 2 : Proposed loss function (w/ statistics)

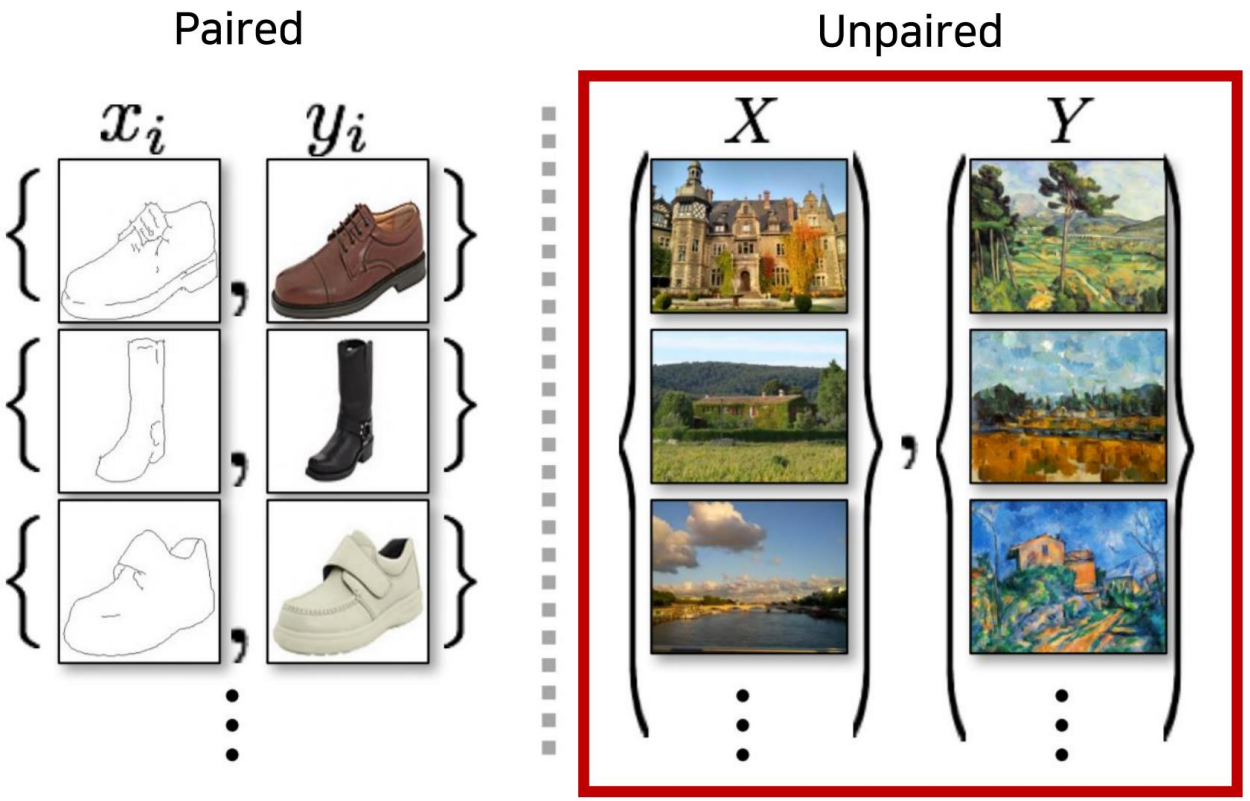
Loss	Per-pixel acc.	Per-class acc.	Class IOU
L1	0.42	0.15	0.11
GAN	0.22	0.05	0.01
cGAN	0.57	0.22	0.16
L1+GAN	0.64	0.20	0.15
L1+cGAN	**0.66**	**0.23**	**0.17**
Ground truth	0.80	0.26	0.21

Fig6. FCN-scores for different losses, evaluated on between Cityscapes labels and photos

04 Evaluation

Limitation of pix2pix

- Pix2Pix는 서로 다른 두 도메인 x,y의 데이터를 한 쌍으로 묶어 학습 진행
  - 다만 colorization과 같은 task에서는 데이터 셋을 준비하는 것에 큰 cost가 소모될 여지가 있음



한 쌍으로 묶이지 않은(unpaired)

데이터 셋에 대해서도 적용이 가능할까요?

CycleGAN을 이용해 해결 가능

Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks(ICCV 2017)

Thank you!