

생성 모델과 시각 지능

Generative Model and Visual Intelligence

09 주차 |

GAN 발전 4

상명대학교 컴퓨터과학과
민 경 하

학습목차

1. pix2pix의 개념
2. pix2pix의 구조
3. pix2pix의 구성 요소



1. pix2pix의 개념

1. pix2pix의 개념

Pix2Pix

➡ 한 영상으로부터 새로운 스타일의 영상을 변환하는 기법

- ✓ 영상의 스타일을 변환하는 기존의 기술을 발전시켜서 영상에 대한 새로운 정보를 추가하는 기능을 수행

- ✓ **Pix2pix trilogy**

Isola, P., Zhu, J. Y., Zhou, T., and Efros, A.,
Image-to-image translation with conditional adversarial networks,
Proc. of CVPR 2017, pp. 1125-1134, 2017. (11,083회 인용)

Zhu, J. Y., Park, T., Isola, P., and Efros, A.,
Unpaired image-to-image translation using cycle-consistent adversarial networks,
Proc. of CVPR 2017, pp. 2223-2232, 2017. (10,532회 인용)

Park, T., Liu, M.-Y., Wang, T.-C., Zhu, J. Y.,
Semantic Image Synthesis with Spatially-Adaptive Normalization,
Proc. of CVPR 2019, pp. 2337-2346, 2019. (1,045회 인용)

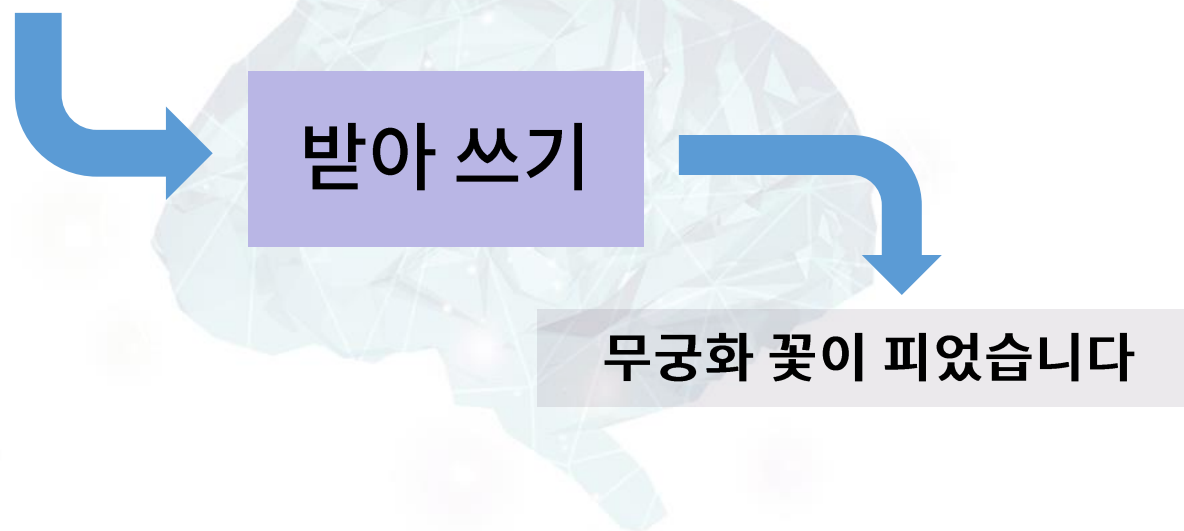
1. pix2pix의 개념

Pix2pix의 배경 (1)

㉠ 받아쓰기

- ☑ 음성 신호를 받아서 text로 바꾸는 과정
- ☑ 음성 신호를 처리해서 음성에 맞는 단어를 찾고, 그 단어를 쓰는 과정

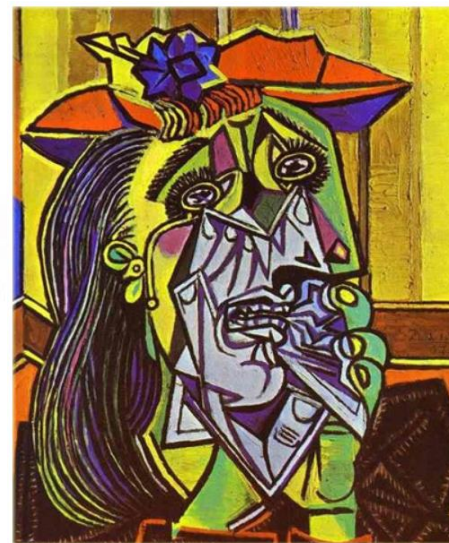
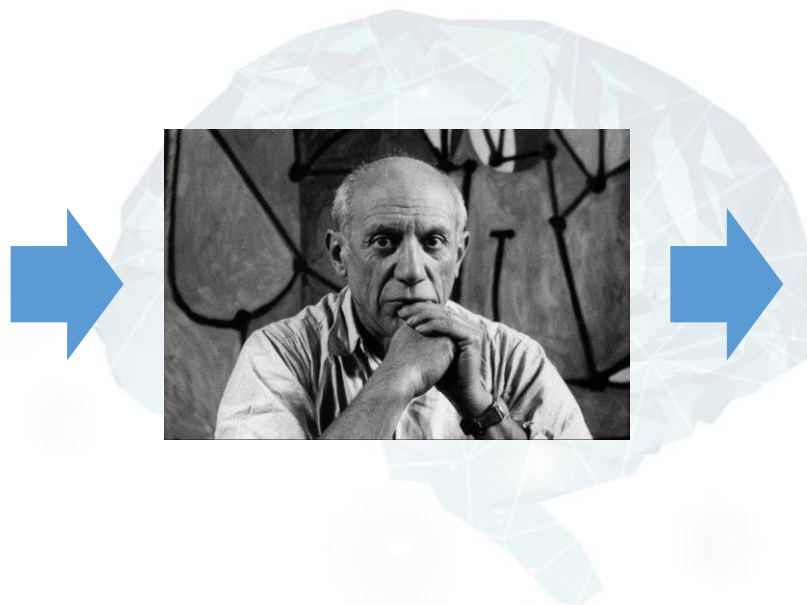
“무궁화 꽃이 피었습니다.”



Pix2pix의 배경 (2)

초상화 그리기

- ✓ 실물로부터 추출한 특징을 예술가가 가공해서 본인의 스타일로 표현하는 과정
- ✓ 영상으로부터 특징 추출 → 가공 → 스타일화된 특징 표현

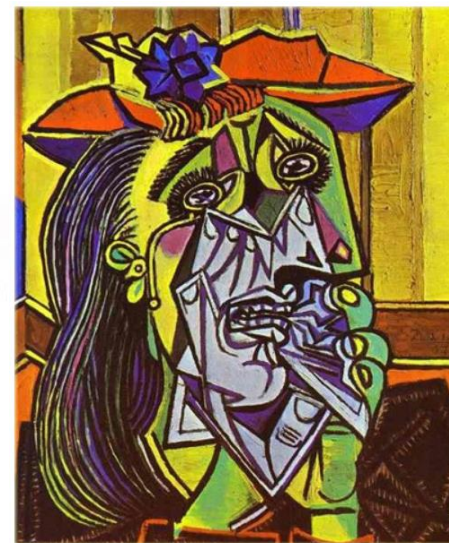
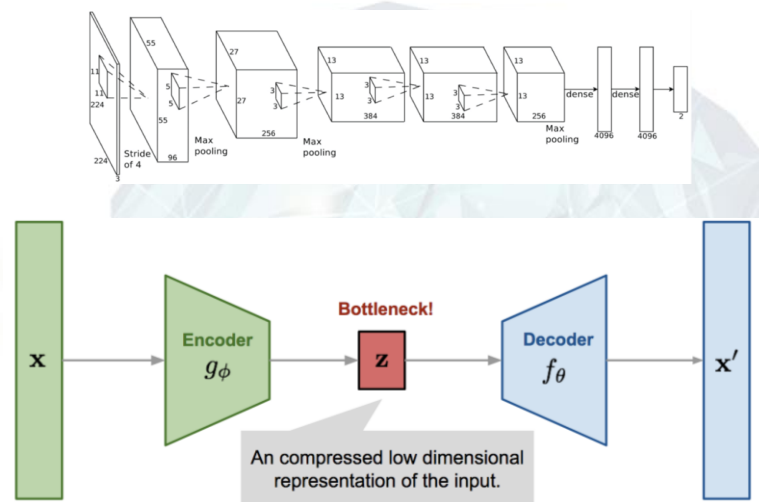


1. pix2pix의 개념

Pix2pix의 배경 (2)

초상화 그리기

- ✓ 실물로부터 추출한 특징을 예술가가 가공해서 본인의 스타일로 표현하는 과정
- ✓ 영상으로부터 특징 추출 → 가공 → 스타일화된 특징 표현
- ✓ 특징 추출, 특징 가공 및 표현을 CNN이 맡게 된다면?

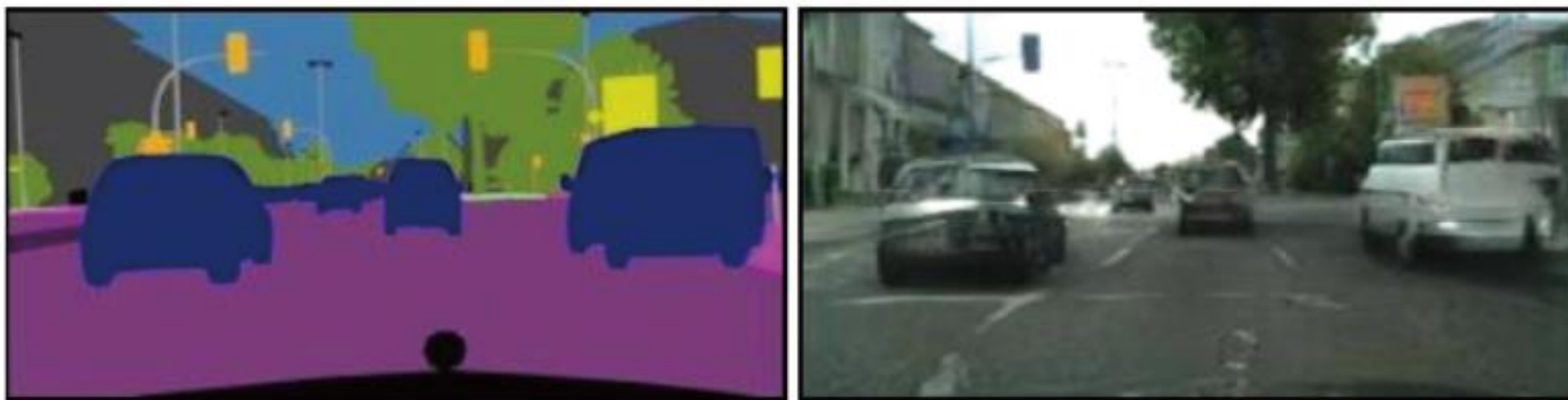


1. pix2pix의 개념

Pix2Pix

- 한 영상으로부터 새로운 스타일의 영상을 변환하는 기법

Labels to Street Scene

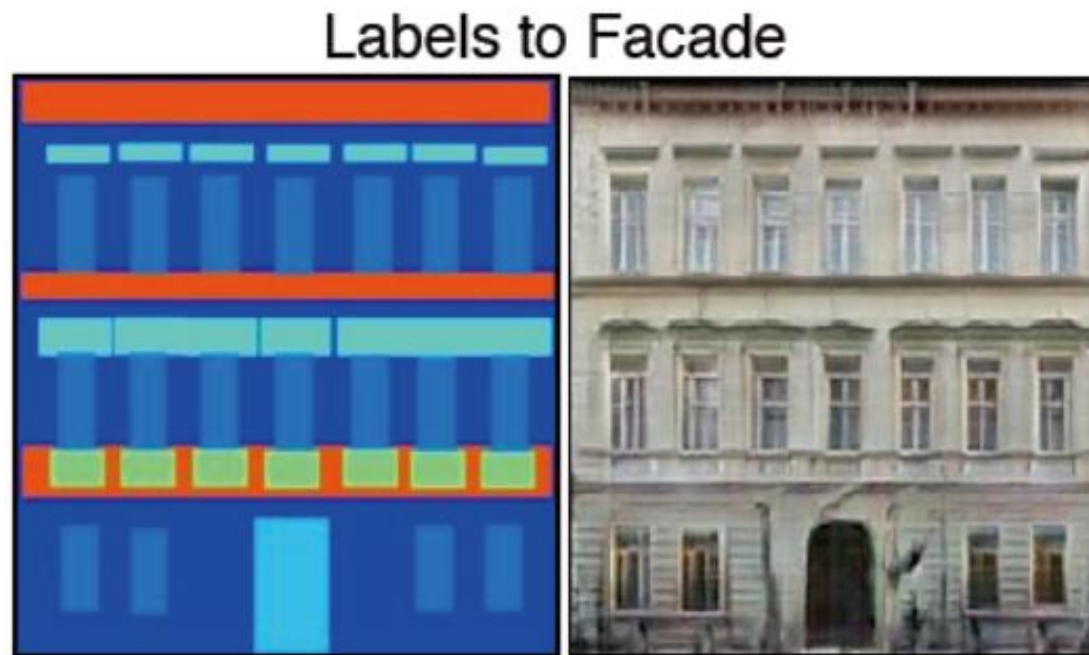


영역별로 분할된 도시 영상으로부터 원본 영상을 복원

1. pix2pix의 개념

Pix2Pix

- 한 영상으로부터 새로운 스타일의 영상을 변환하는 기법



영역별로 분할된 영상으로부터 원본 영상 (Façade)을 복원

1. pix2pix의 개념

Pix2Pix

- 한 영상으로부터 새로운 스타일의 영상을 변환하는 기법

BW to Color



흑백 영상을 컬러 영상으로 변환

1. pix2pix의 개념

Pix2Pix

- 한 영상으로부터 새로운 스타일의 영상을 변환하는 기법

Aerial to Map



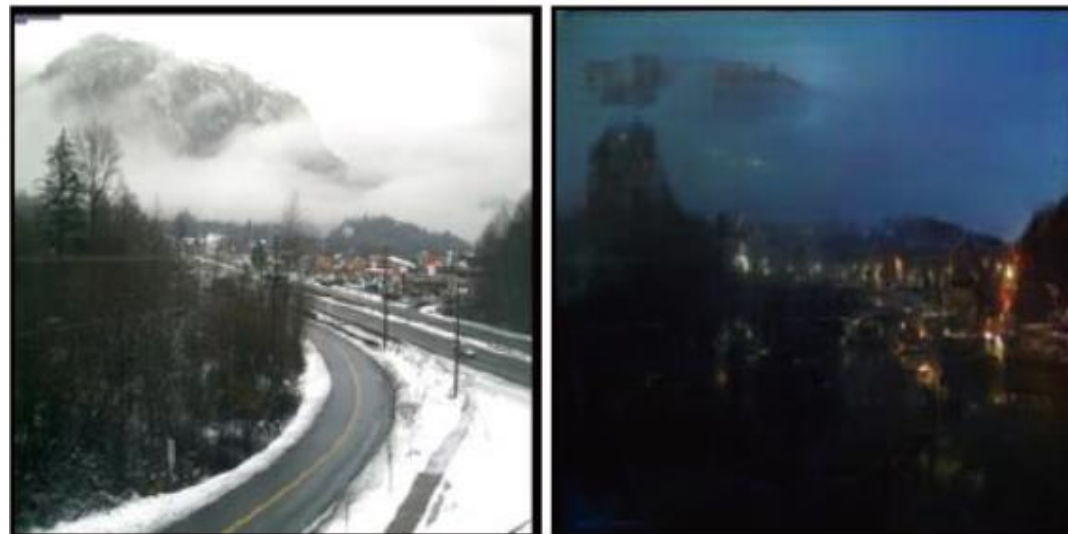
항공 사진 ↔ 지도

1. pix2pix의 개념

Pix2Pix

- 한 영상으로부터 새로운 스타일의 영상을 변환하는 기법

Day to Night



낮 영상 ↔ 밤 영상

1. pix2pix의 개념

Pix2Pix

- 한 영상으로부터 새로운 스타일의 영상을 변환하는 기법

Edges to Photo



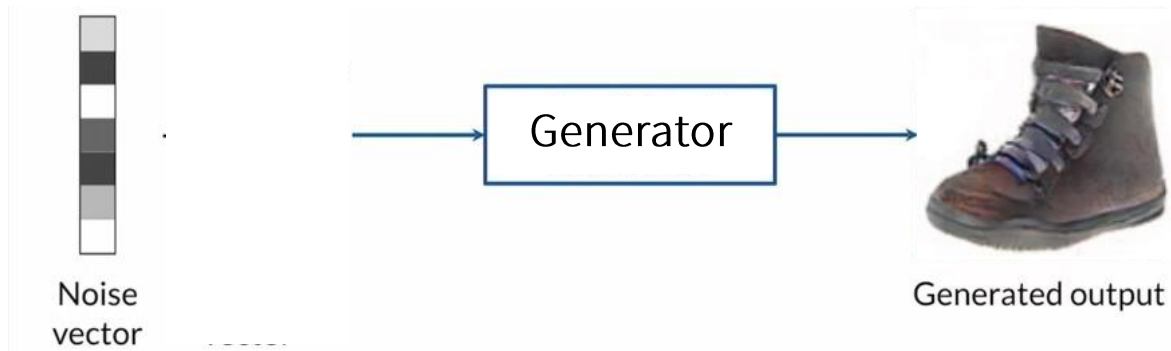
스케치에서 컬러 영상 복원



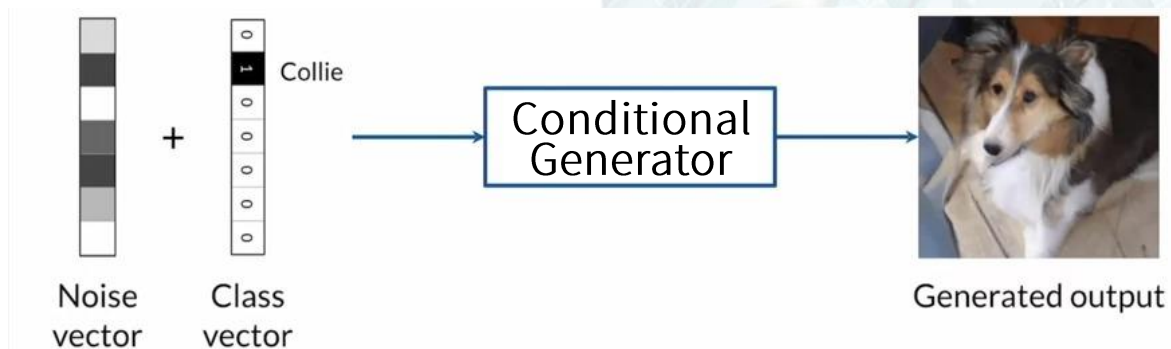
2. pix2pix의 구조

CGAN의 개념을 발전시킴

Vanilla GAN (or other GANs)의 Generator



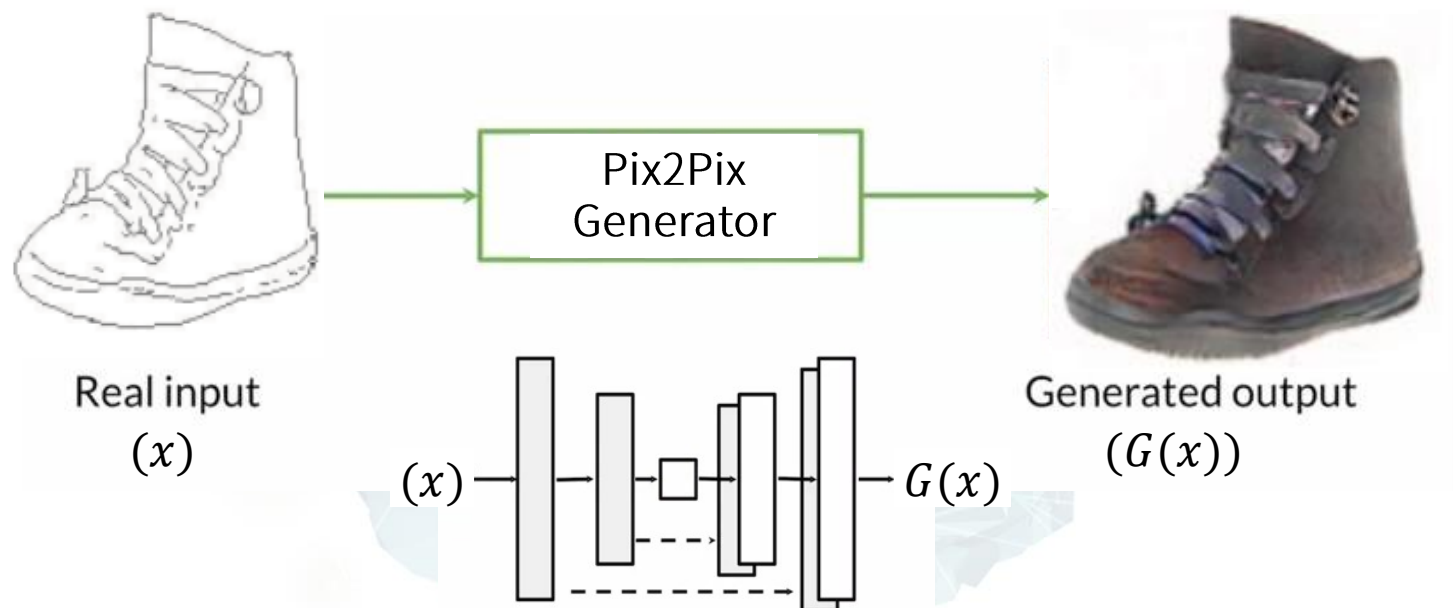
CGAN의 Generator



CGAN의 개념을 발전시킴

● Pix2pix의 Generator

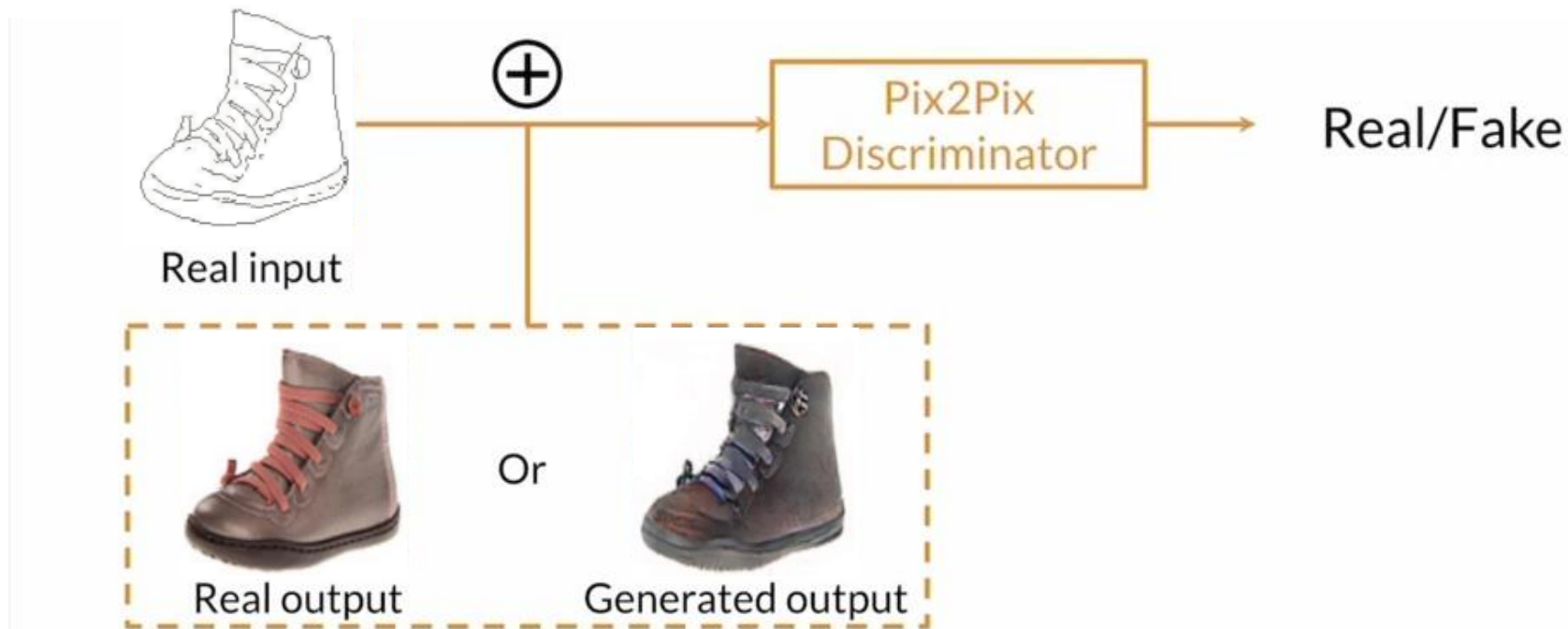
- ☑ Encoder-decoder 구조의 Generator를 사용함으로써 Latent vector를 요구하지 않음



CGAN의 개념을 발전시킴

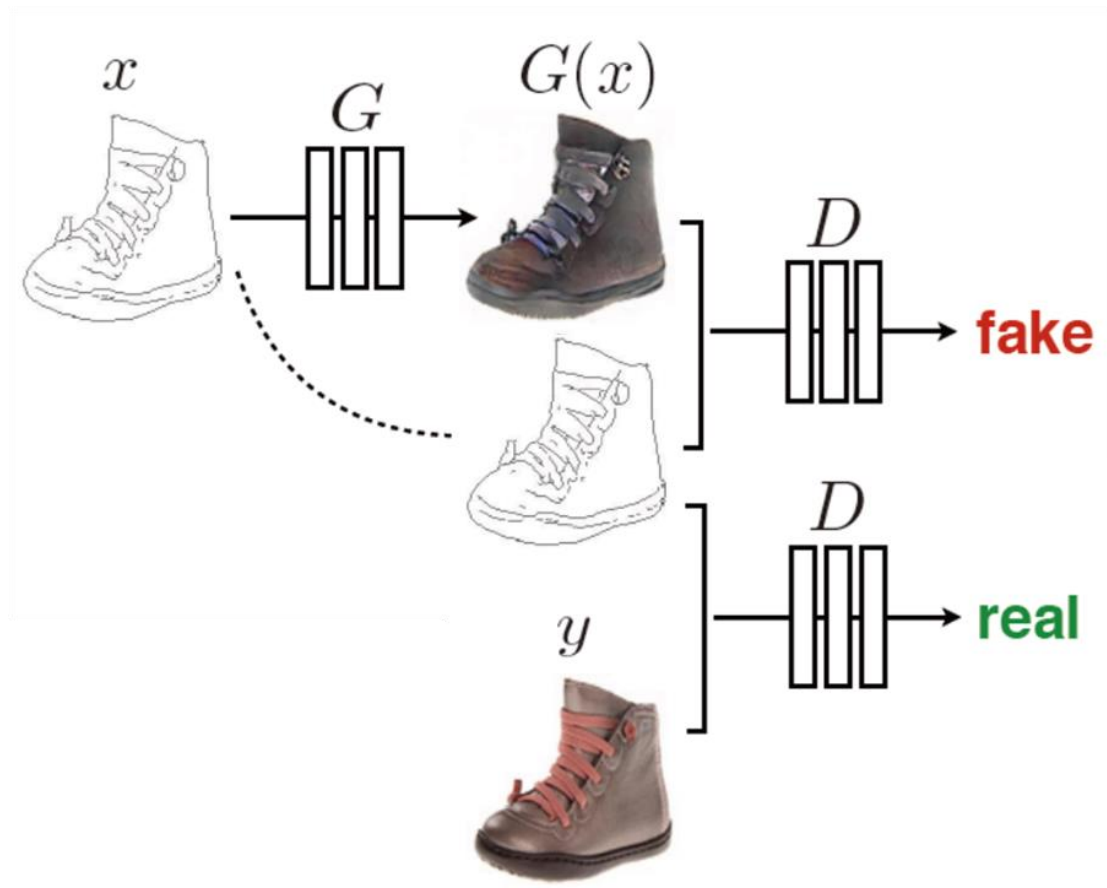
• Pix2pix의 Discriminator

- ✓ x 와 y 를 비교해서 Real을 Return하고 $G(x)$ 와 y 를 비교해서 Fake를 Return



2. pix2pix의 구조

전체적인 구조



- Generator

- ✓ 입력 영상 (x)로부터 Style 영상 (y)를 참고해서 새로운 영상 ($G(x)$)를 생성

- Discriminator

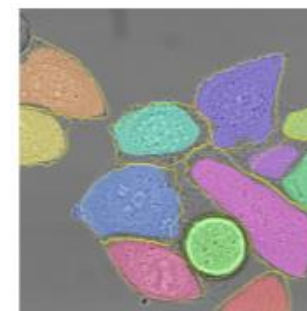
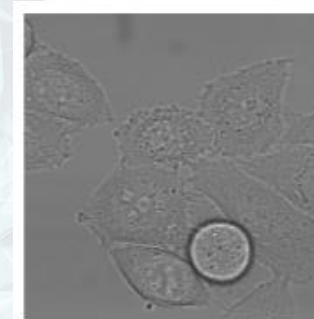
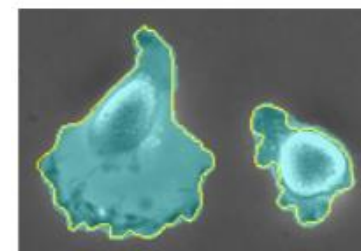
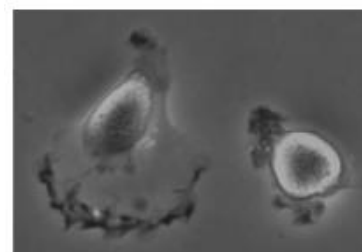
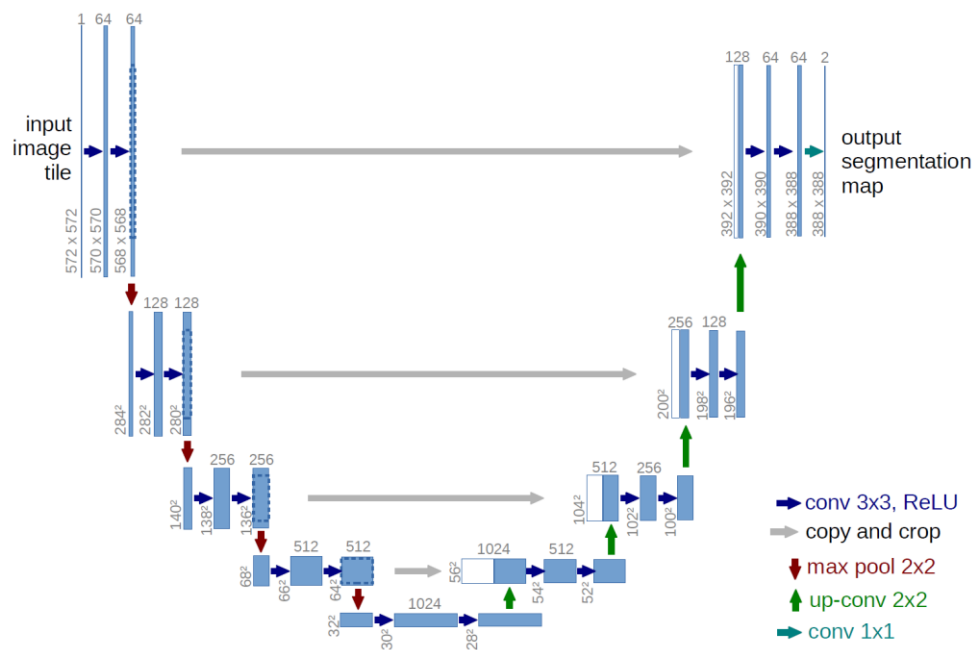
- ✓ x 와 y 를 비교해서 Real을 Return하고 $G(x)$ 와 y 를 비교해서 Fake를 Return하도록 훈련



3. pix2pix의 구성 요소

Generator

U-net 구조를 이용한 Generator



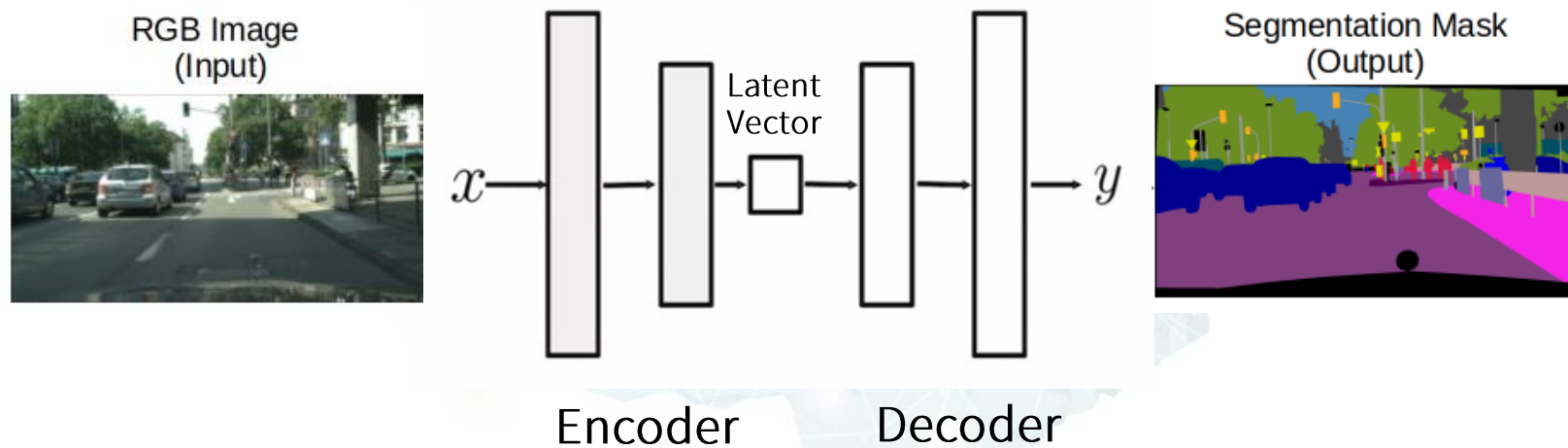
3. pix2pix의 구성 요소

Generator

U-net 구조를 이용한 Generator

Encoder-decoder 구조

- » Deconv-net이라고도 불리움
- » Convolution VS Transposed convolution



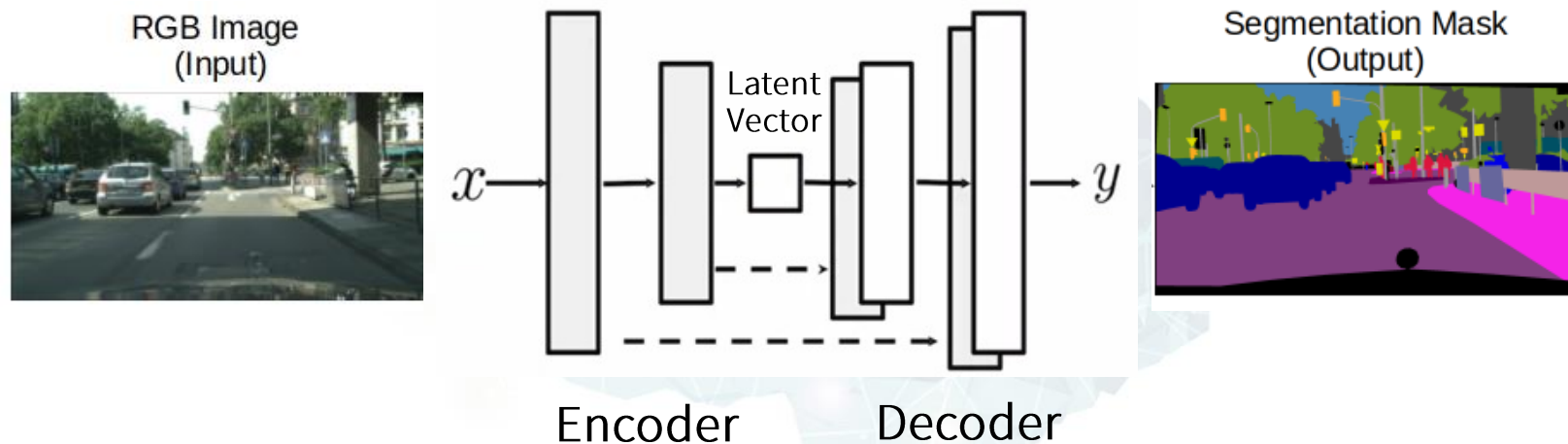
3. pix2pix의 구성 요소

Generator

U-net 구조를 이용한 Generator

✓ Skip connection

- » Forward pass: Encoder의 정보를 Decode에 전달
- » Backward pass: Encoder의 Gradient flow를 개선

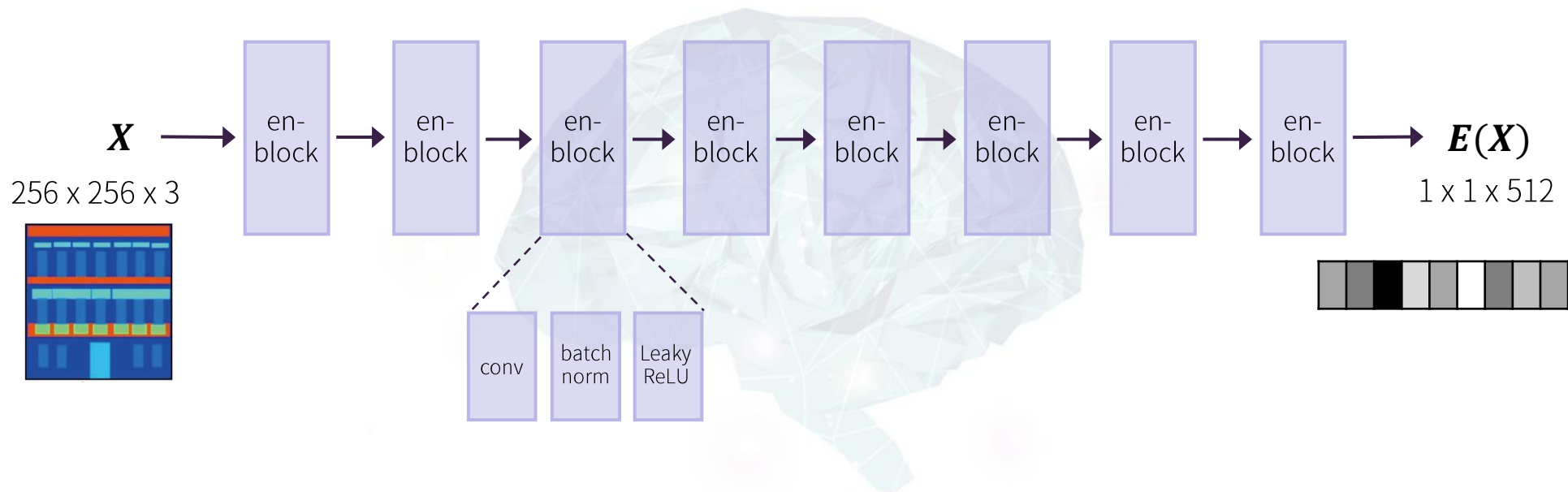


3. pix2pix의 구성 요소

Generator

U-net 구조를 이용한 Generator (1)

- ✓ 8개의 Encoder block
- ✓ 각 Encoder block은 Conv – Batch norm – Leaky ReLU로 구성
- ✓ $256 \times 256 \times 3 \rightarrow 1 \times 1 \times 512$

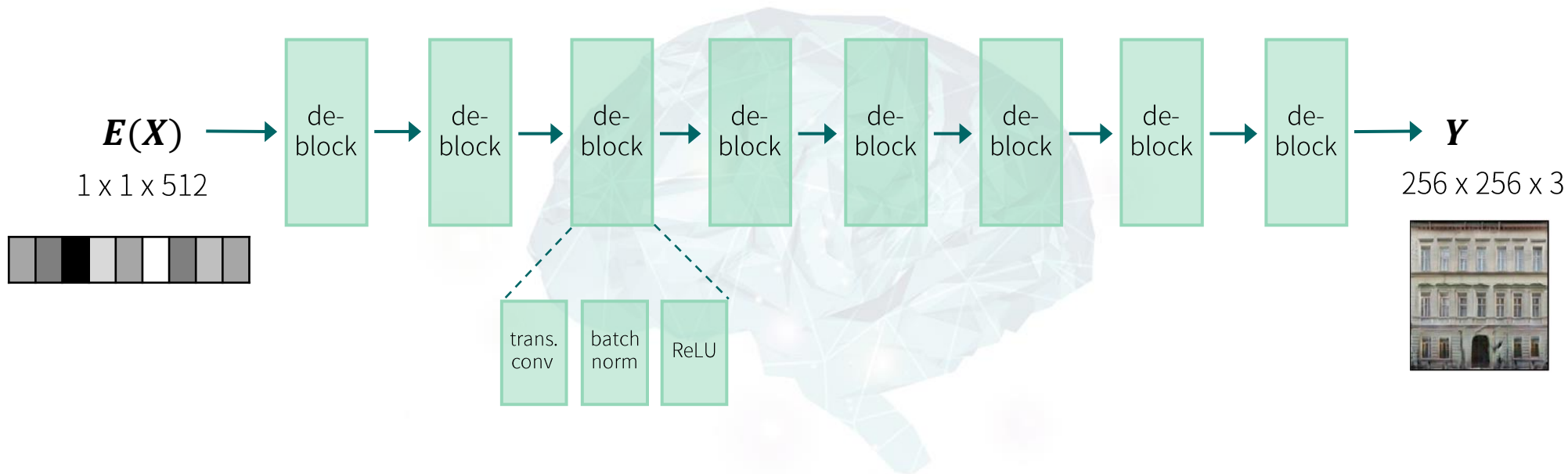


3. pix2pix의 구성 요소

Generator

U-net 구조를 이용한 Generator (2)

- ✓ 8개의 Decoder block
- ✓ 각 Decoder block은 Transposed conv – Batch norm – ReLU로 구성
- ✓ $1 \times 1 \times 512 \rightarrow 512 \times 512 \times 3$

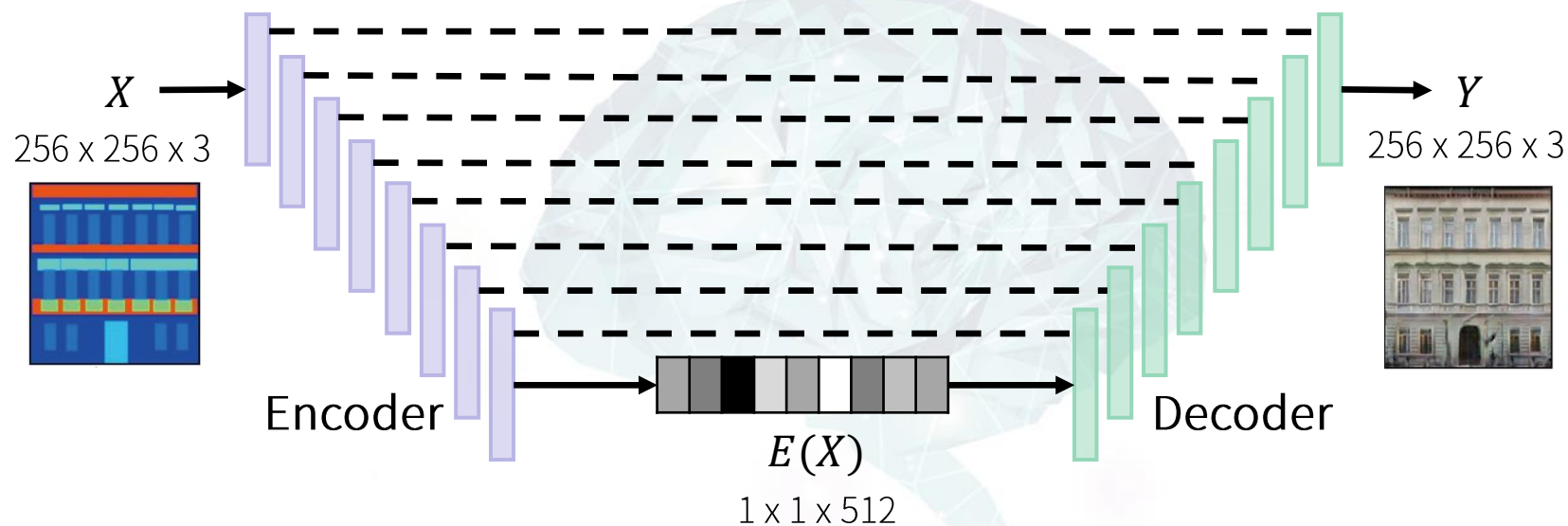


3. pix2pix의 구성 요소

Generator

U-net 구조를 이용한 Generator (3)

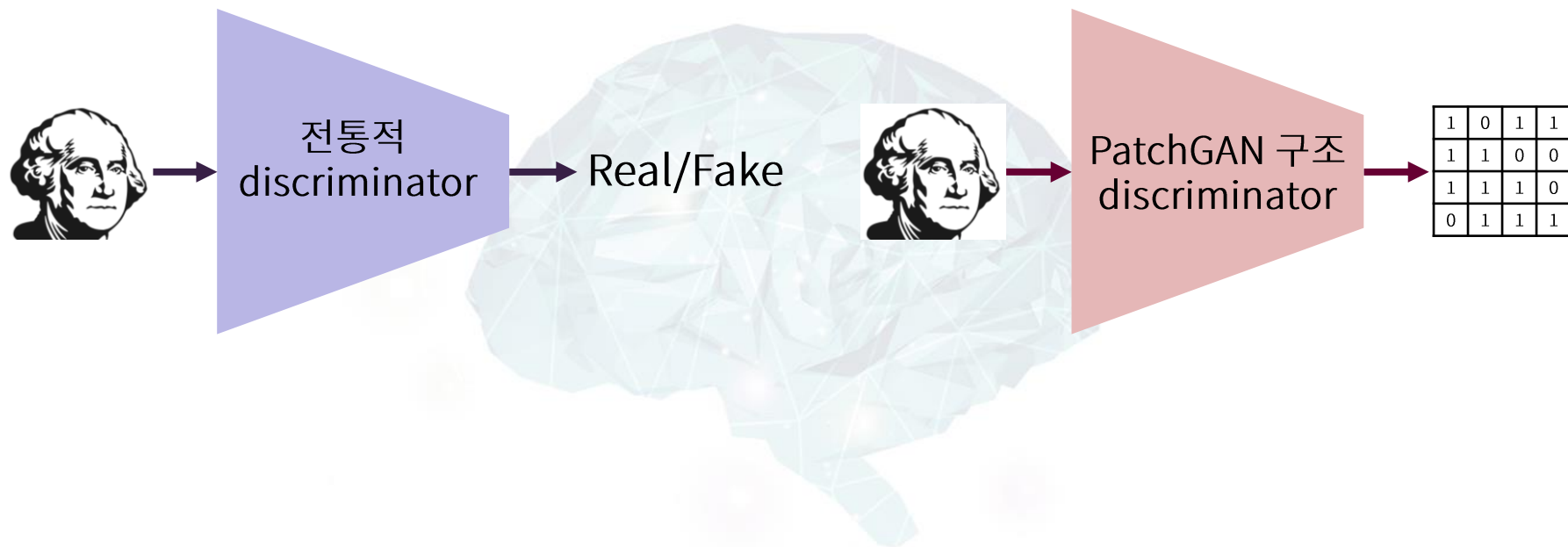
- ✓ 8개의 En-block을 가진 Encoder와 8개의 De-block을 가진 Decoder
- ✓ $512 \times 512 \times 3 \rightarrow 512 \times 512 \times 3$
- ✓ Skip connection



Discriminator

PatchGAN 구조를 사용

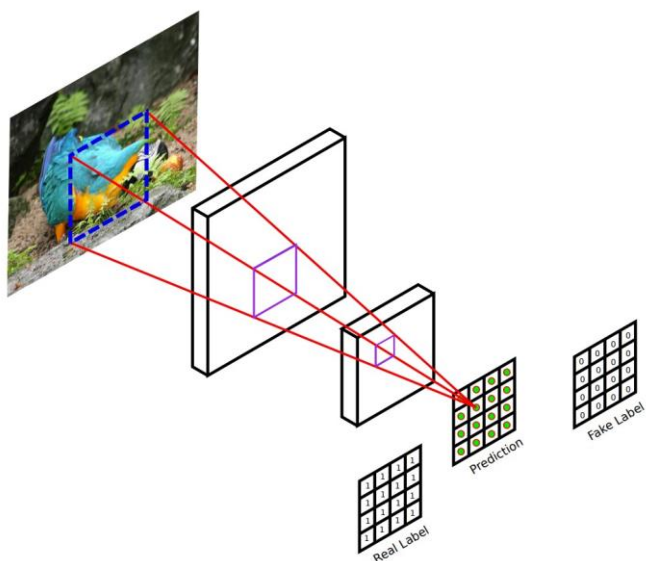
- ✓ 전통적 GAN에서는 Discriminator가 전체 영상에 대해서 Real/Fake를 판정함
- ✓ PatchGAN에서는 영상을 Patch로 분할하여 각 영역의 Real/Fake를 판정함



Discriminator

PatchGAN 구조를 사용

- ✓ 전통적 GAN에서는 Discriminator가 전체 영상에 대해서 Real/Fake를 판정함
- ✓ PatchGAN에서는 영상을 Patch로 분할하여 각 영역의 Real/Fake를 판정함
- ✓ 예) 4x4 Patch



Prediction의 값이 0에 가까우면 fake
Prediction의 값이 1에 가까우면 real

3. pix2pix의 구성 요소

loss 함수

- 전통적인 cGAN loss (adversarial loss, GAN loss) + pixel distance loss

Vanilla GAN의 loss 함수

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

CGAN의 loss 함수

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x/y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z/y)))]$$

pix2pix의 loss 함수

$$\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L_1}(G)$$

3. pix2pix의 구성 요소

loss 함수

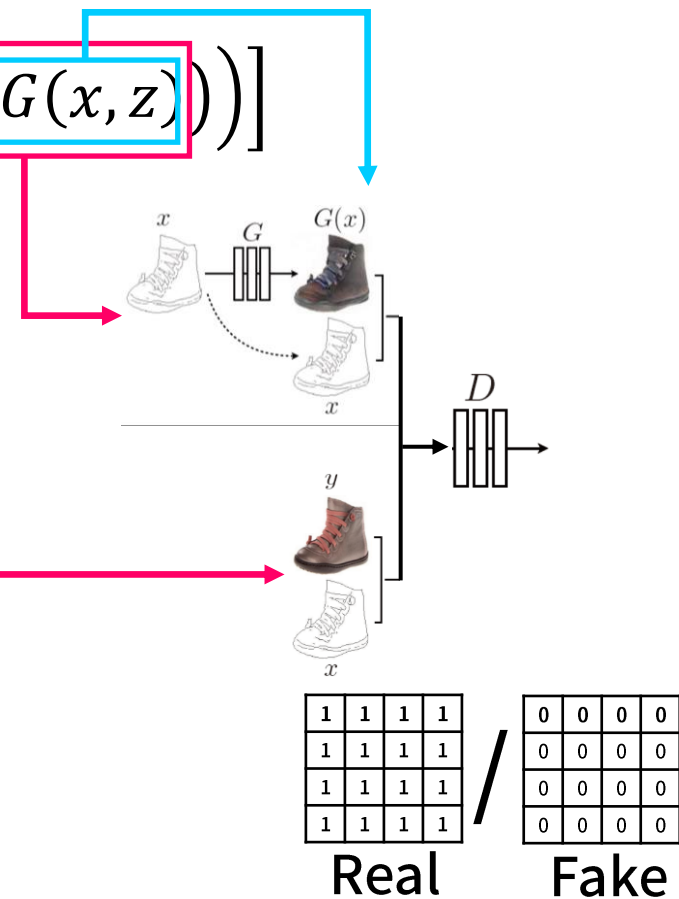
Adversarial loss ($\mathcal{L}_{cGAN}(G, D)$)

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

x : input

y : groundtruth output

$G(x)$: generated output



3. pix2pix의 구성 요소

loss 함수

Pixel distance loss ($\mathcal{L}_{L_1}(G, D)$)

- ✓ 생성된 영상 ($G(x, z)$)와 Groundtruth 영상 (y)와의 픽셀 간의 차이

$$\mathcal{L}_{L_1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1]$$

