

증강지능 연구실 발표자료

Generative Model A to Z

증강지능 연구실

황승현

# Contents

도입

01

발표의 목적과 필요한 배경지식

배경 지식

02

발표를 듣기 전 알아야할 기초 개념과  
Generative Model의 정의

Variational Autoencoder  
(VAE)

03

가변 오토인코더란 무엇일까?

Generative Adversarial Network  
(GAN)

04

적대적 생성 네트워크란 무엇일까?

Diffusion model

05

확산 모델이란 무엇일까?

# 배경 지식



Generative Model이란?

「  
입력한 데이터셋과  
유사한 새로운 데이터  
를 만드는 알고리즘」

생성 모델은 입력한 데이터셋과 유사한 새로운 데이터를 만드는 알고리즘. 생성 모델은 데이터의 분포, 특징을 학습하여 이 분포, 특징을 보고 새로운 샘플을 생성하는 것.

데이터 증강(늘리기), 새로운 이미지 / 텍스트 생성, 이상치 탐지에 사용

GAN, VAE 등의 모델이 있다.



# 어

[도입]

- Generative Model
- 발표의 목적
- 배경 지식

도입



Generative Model



Stable Diffusion AI

ChatGPT



Capabilities

uting in

Remembers what user said  
earlier in the conversation

May

n

Allows user to provide follow-  
up corrections

ChatGPT

## Generative Model

Generative Model(생성 모델)이란?

훈련 데이터셋과 유사한 새로운 데이터를 생성할 수 있는 머신러닝 알고리즘

데이터의 특징을 찾아 배우고, 입력한 데이터들 중 정말로 존재할 것 같은 데이터를 만들어내는 것

Stable Diffusion AI, ChatGPT 등

주로 쓰는 알고리즘으로는 VAE, GAN, Diffusion Model이 있다.

# 도입



## 발표의 목표

# Generative Model

## 알고 쓰자

요즘 유행하는 ChatGPT, 몇 달 전 우리 연구실에서 한 인공지능 미술전 모두 Generative Model을 기반으로 만든 서비스이다. 우리는 Generative Model을 쓰고 있지만, 정작 Generative Model이 무엇인지, 어떤 원리로 작동되는지 모른다. 알고 쓰자.

## 알아 보자

오늘 발표를 듣고 기억하면 다른 사람에게 ChatGPT가 어떻게 작동하는지 설명해줄 수 있지 않을까?  
연구실원이 Generative Model에 관심을 가졌으면 좋겠다. Generative Model을 알아 보자.

## Vision

Generative Model은 자연어 처리 분야에서도 많이 쓴다. 요즘 유행하는 ChatGPT가 자연어를 기반으로 하는 Generative Model이다. 하지만 오늘은 Vision만 알아 보자.



# 02

## [배경 지식]

- 인공지능이란?
- 용어 정의
- Generative Model



# 배경 지식

인공지능이란?



개념의 범위



인공지능이란?

## 인공지능

인공지능의 개념 범위

인공지능 > 머신러닝 > 딥러닝

인공지능

- 앨런 튜링: 생각할 수 있는 기계
- 스스로 학습할 수 있는 컴퓨터



# 배경지식

인공지능이란?

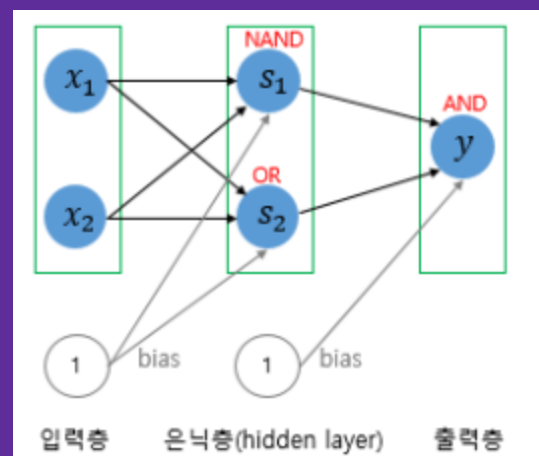
01

## 머신러닝

파라미터에 따라 동작하는 모델 만든다.

결과를 평가하여 더 좋은 결과를 가지는 모델을 위해 파라미터 수정

파라미터 수정: 학습

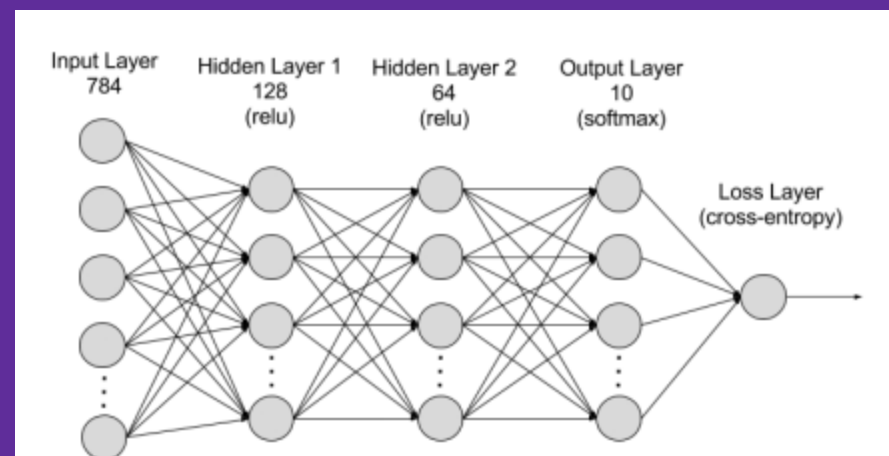


02

## 딥러닝

머신러닝의 입력층, 은닉층, 출력층 구조에서 은닉층이 매우 많은(Deep) 머신러닝이다.

대표적으로 CNN, RNN 등이 있다.



03

## 용어

데이터셋: 학습에 사용할 데이터

정확도: 모델이 데이터셋을 맞는 것 맞다고 하고, 틀린건 틀리다고 하는 것

오차: 정확도의 반대

손실 함수: 모델의 정확도를 평가하기 위한 함수

Generative Model에서는 원본 데이터셋과 얼마나 유사한지 측정한다.



# 03

## 「Variational Autoencoder」 (VAE)

- AutoEncoder
- VAE의 정의와 동작

# VAE

## AutoEncoder

# AutoEncoder

## 차원 축소

정의  
데이터를 표현하기 위해 사용하는 공간을 더 낮은 차원으로 바꾸는 것

데이터의 특징을 추출하는 것

데이터가 너무 많은 특성을 가지면 모델의 정확도가 떨어짐

특성의 수 증가 = 차원 증가

차원 축소 = 특성의 수 줄임 = 중요한 특징만 추출

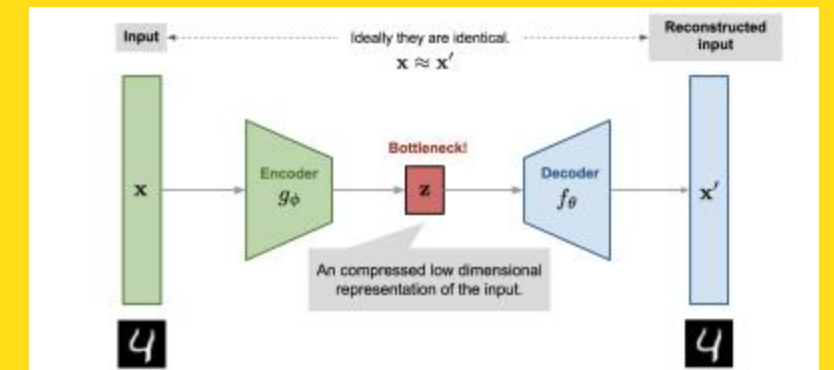
데이터에서 가장 좋은 특징을 찾는 방법

추출한 특징을 다시 복원하는 모델을 만든다.

## 인코더 디코더

인코더: 차원 축소하는 모델  
디코더: 차원 복원하는 모델  
신경망이 자동으로 생성하는 모델로 차원 축소한다.

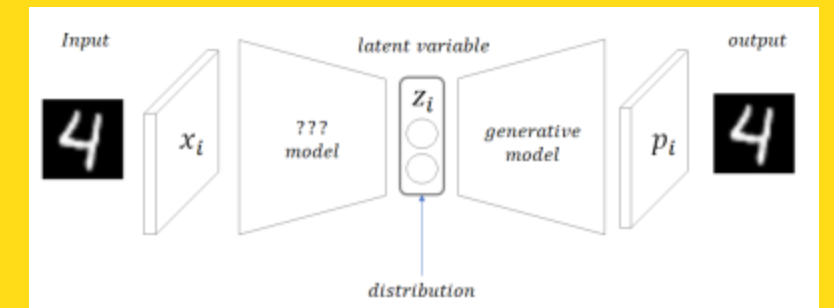
오토인코더는 원본 데이터에서 특징을 추출하고, 특징을 다시 원래 데이터로 복원하여 가장 특징을 잘 추출하는 방법을 찾는 것.



## 변형

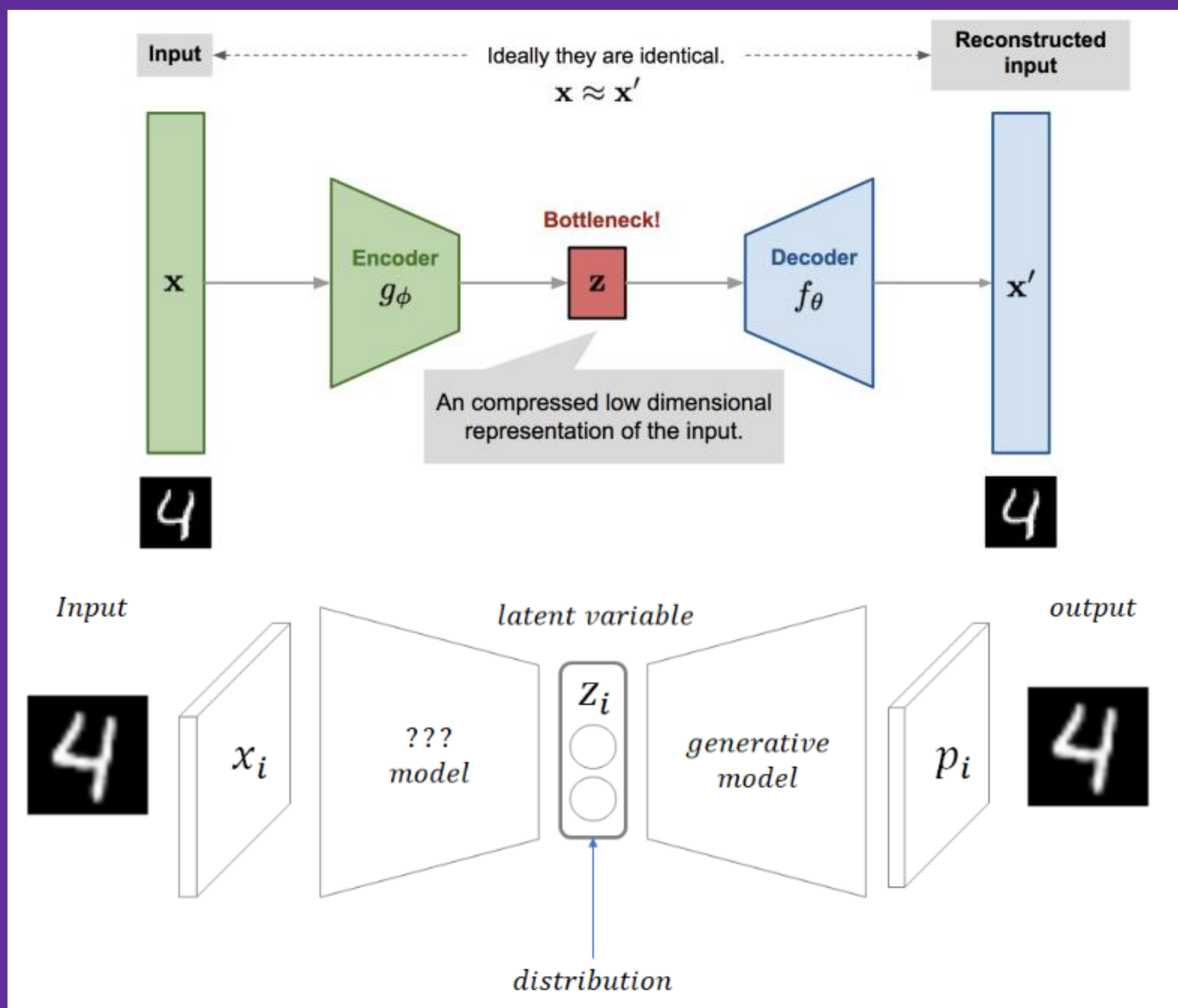
오토인코더의 목적은 특징을 가장 잘 추출하는 인코더를 만드는 것

만약 인코딩 디코딩을 확률적으로 한다면 어떨까?  
원본 데이터와 비슷하지만 다른 데이터가 나오지 않을까?



# VAE

VAE란 무엇인가?



# VAE

AutoEncoder는 한 점으로 인코딩한다. 그러나, VAE(변형 자동 인코더)는 "잠재 공간 분포" (distribution over the latent space)로 인코딩한다.

잠재 공간 분포 중 하나를 디코딩하여 오류를 계산한다. 오류는 reconstruction loss, regularization loss이다  
regularization loss는 일반적으로 사용하는 MSE 입력 데이터와 차이를 계산

regularization loss는 두 분포의 KL발산으로 계산하는데 posterior distribution이 prior distribution와 유사한지 측정하여, 유사하게 만들. 정규화가 제대로 안 되면 AutoEncoder가 과적합(분산이 작은 분포, 평균이 매우 다른 분포 반환)될 수 있다.

그리고 오류를 역전파하여 학습한다.



# 04

## 「Generative Adversarial Network (GAN)」

- GAN 정의와 동작



Generated Data

Discriminator

Real Data



FAKE

REAL



학습이 진행됨에 따라 생성기는 판별자를 속일 수 있는 출력을 생성하는 데 더 가까워집니다.



FAKE

REAL



마지막으로 생성기 학습이 잘되면 판별자가 진실과 거짓 사이의 차이점을 더 잘 알게 됩니다. 가짜 데이터를 실제 분류로 분류하기 시작하며 정확성이 낮아집니다.



REAL

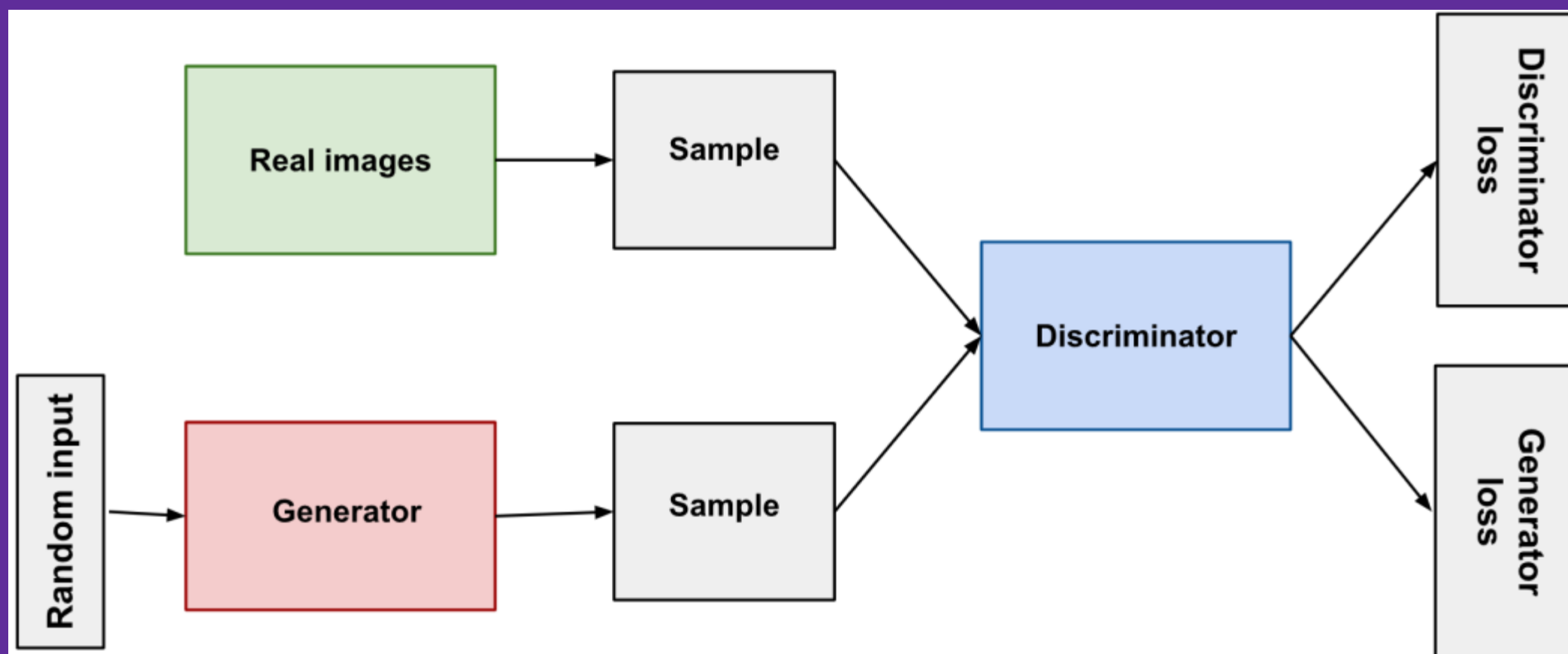
REAL





# GAN

GAN이란 무엇인가?



# GAN

GAN(적대적 생성 네트워크)은 generator(생성기)와 discriminator(판별기) 신경망 2개를 훈련시키는 것이다. GAN은 위조지폐에 비유하고는 한다. 위조지폐범은 더 정교한 위조지폐를 만들려고 하고, 경찰은 위조지폐를 더 잘 구분할 수 있게 훈련한다. 즉, generator의 목표는 실제 데이터와 구별할 수 없는 데이터를 만드는 것이며, discriminator의 목표는 주어진 데이터가 실제인지 생성되었는지를 식별하는 것이다.

생성기는 노이즈를 입력하여 유의미한 데이터를 출력한다. 그리고 판별기의 분류에 따라 그래디언트를 수정한다(역전파)

판별기는 생성기에서 만든 가짜 데이터와 실제 데이터를 분류하고, 잘못 분류한 경우에는 패널티를 적용한다. 그리고 이 오류를 역전파하여 생성기의 가중치를 수정한다.



# 05

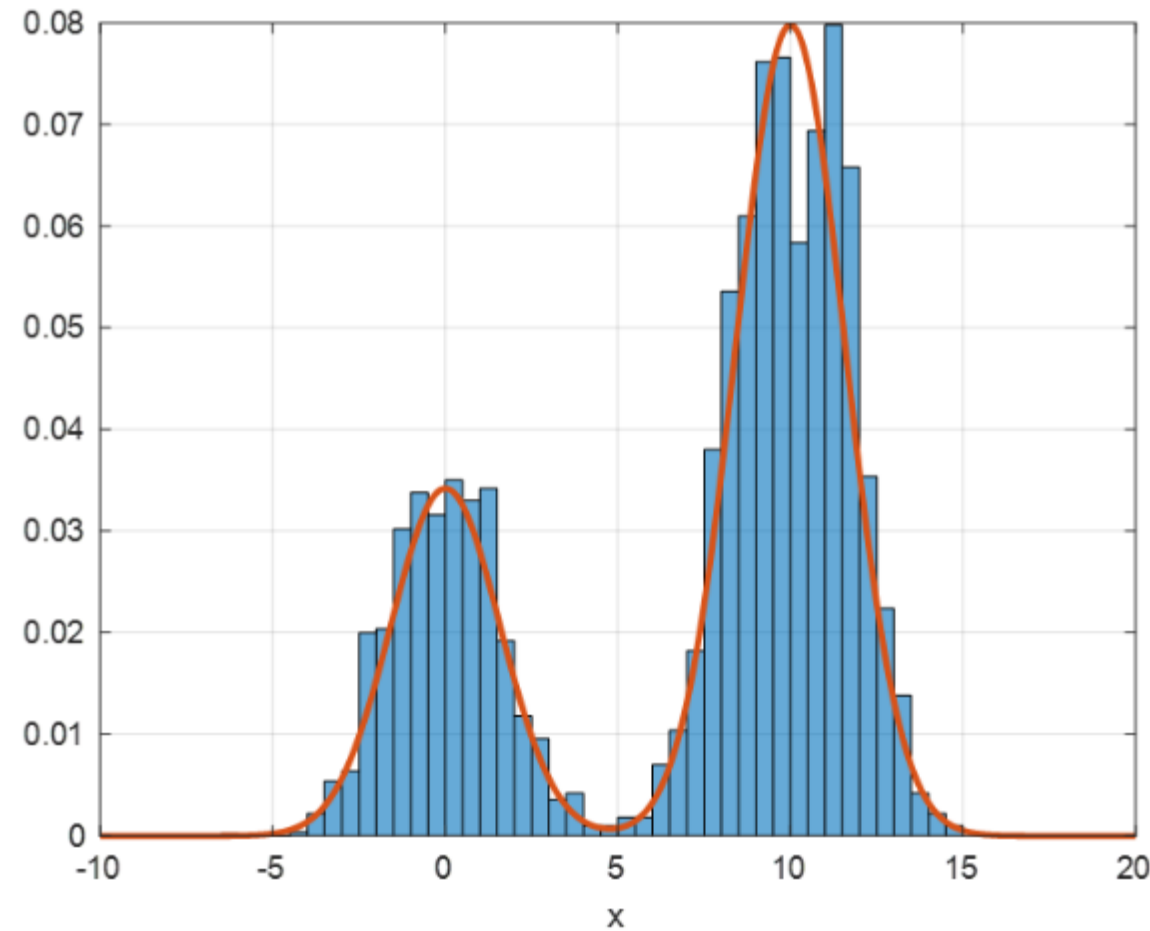
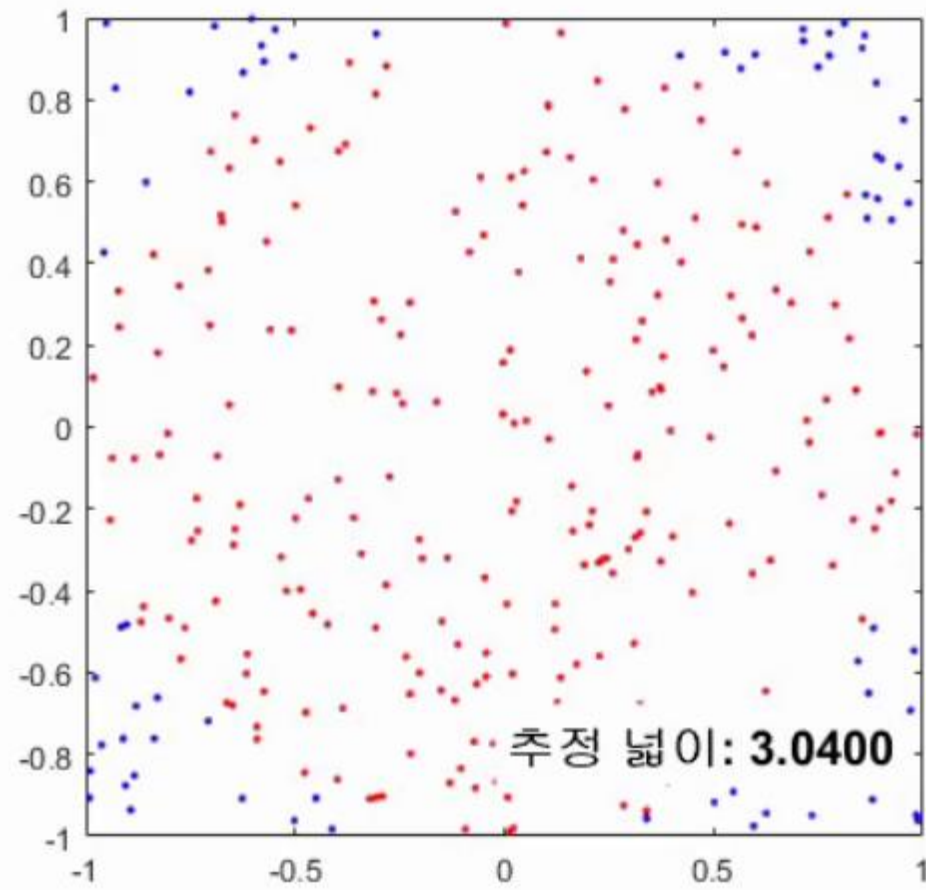
## [Diffusion Model]

- Markov Chain  
Monte Carlo
- Diffusion Model이란?

# Markov Chain Monte Carlo

MCMC

# MCMC



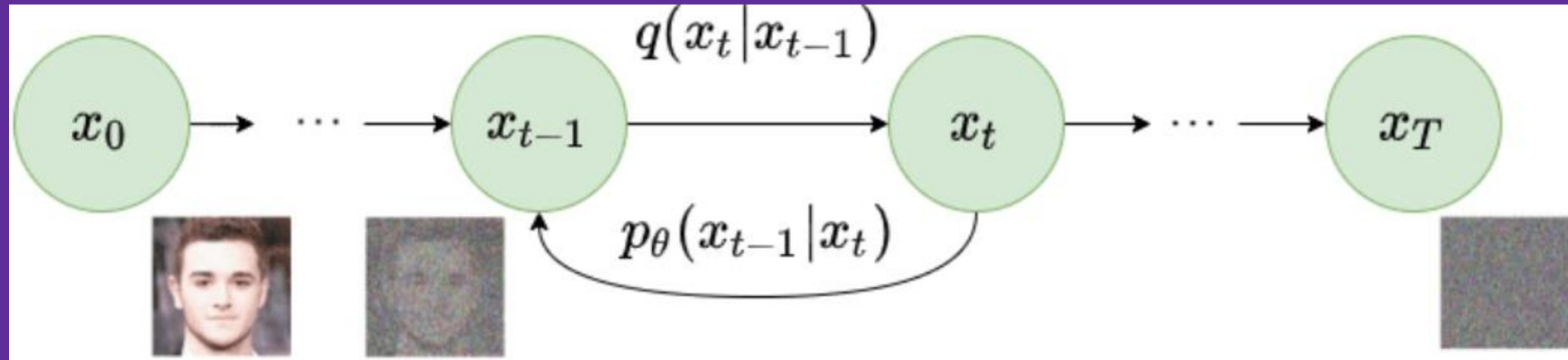
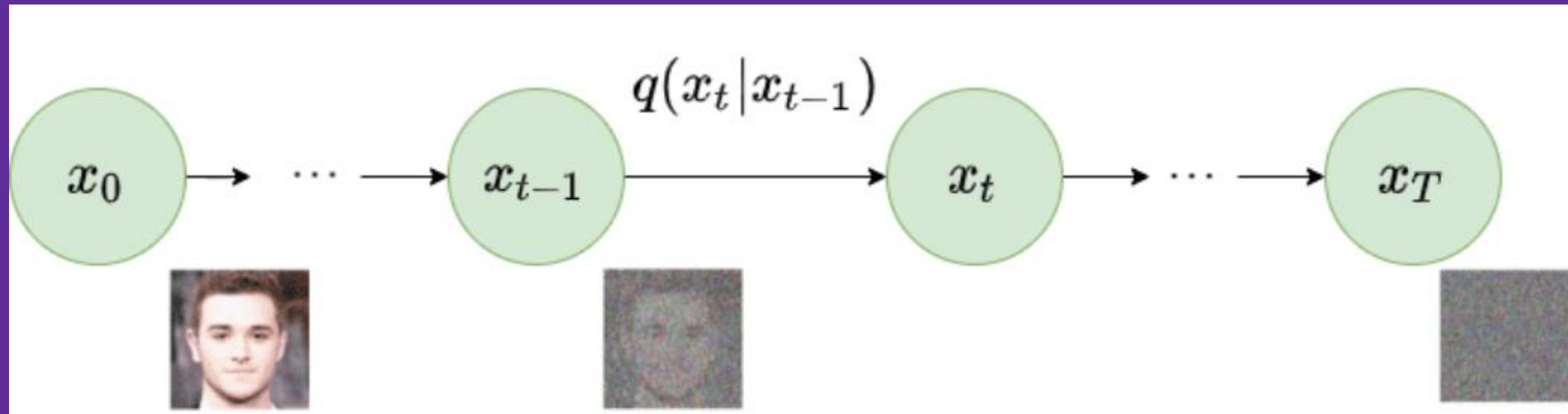
Markov Chain은 어떤 상태에서 다른 상태로 넘어갈 때, 바로 전 단계의 상태에만 영향을 받는 확률 과정을 의미한다.

Monte Carlo 통계적 수치를 얻기 위해 많은 시도를 거쳐 정답을 추정하는 것.

즉, 첫 샘플을 랜덤하게 선택한 뒤, 첫 샘플에 의해 그 다음번 샘플을 추천하는 것을 많이 반복하는 것.

# Diffusion Model

## 정의와 동작



# Diffusion Model

forward diffusion process :

이미지를 입력하고, 점진적으로 노이즈를 추가한다.

diffusing the feature representation

reverse diffusion process :

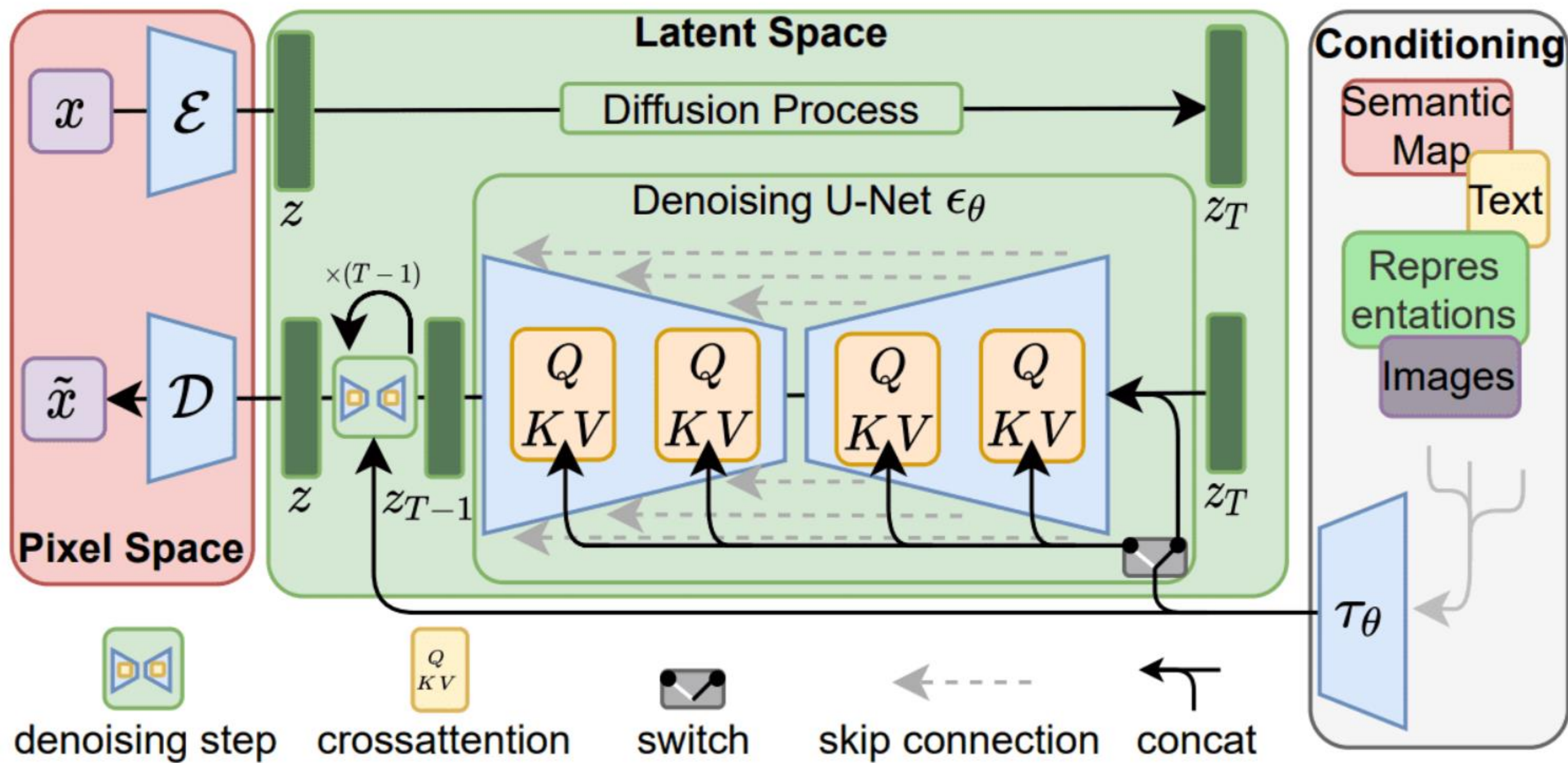
노이즈를 입힌 이미지를 원래 데이터로 복구하도록  
신경망을 훈련한다.

diffused representation back into the image

이 때 만들어지는 새로운 이미지는 입력 이미지와 유사하지만, 같지는 않다.

이 프로세스를 반복하여 학습하면, 다양한 새로운 이미지 생성하는 모델 개발







# 06

## 「정리」

- 발표 내용 요약
- 향후 계획



# 정리



## 발표 내용 요약

### Generative Model

#### VAE

- 데이터의 차원을 축소하고 (인코딩) 다시 확장하면서 (디코딩) 새로운 데이터 만들
- 인코딩, 디코딩을 확률적
- 입력과 유사하지만 약간 다른 데이터 생성

#### GAN

- Generator와 Discriminator로 이루어진 모델
- Generator 더 정교한 가짜 데이터를 생성
- Discriminator는 가짜 데이터를 더 잘 구분하기 위해 학습
- 입력과 유사하지만 약간 다른 데이터 생성

#### Diffusion Model

- forward diffusion process: 이미지에 노이즈 추가
- reverse diffusion process: 이미지에서 노이즈 제거 학습
- 프로세스 반복하면 입력과 유사하지만 약간 다른 데이터 생성

정리

향후 계획

## Recent Advances of Monocular 2D and 3D Human Pose Estimation: A Deep Learning Perspective

WU LIU and QIAN BAO, Explore Academy of JD.com, Beijing, China  
YU SUN, Harbin Institute of Technology, Harbin, China  
TAO MEI, Explore Academy of JD.com, Beijing, China

Estimation of the human pose from a monocular camera has been an emerging research topic in the computer vision community with many applications. Recently, benefiting from the deep learning technologies, a significant amount of research efforts have advanced the monocular human pose estimation both in 2D and 3D areas. Although there have been some works to summarize different approaches, it still remains challenging for researchers to have an in-depth view of how these approaches work from 2D to 3D. In this article, we provide a comprehensive and holistic 2D-to-3D perspective to tackle this problem. First, we comprehensively summarize the 2D and 3D representations of human body. Then, we summarize the mainstream and milestone approaches for these human body presentations since the year 2014 under unified frameworks. Especially, we provide insightful analyses for the intrinsic connections and methods evolution from 2D to 3D pose estimation. Furthermore, we analyze the solutions for challenging cases, such as the lack of data, the inherent ambiguity between 2D and 3D, and the complex multi-person scenarios. Next, we summarize the benchmarks, evaluation metrics, and the quantitative performance of popular approaches. Finally, we discuss the challenges and give deep thinking of promising directions for future research. We believe this survey will provide the readers (researchers, engineers, developers, etc.) with a deep and insightful understanding of monocular human pose estimation.

CCS Concepts: • **General and Reference** → **Surveys and Overviews**; • **Computing Methodologies** → **Artificial Intelligence - Computer Vision**;

Additional Key Words and Phrases: Human pose estimation, deep learning, 2D and 3D pose, monocular images

**ACM Reference format:**  
Wu Liu, Qian Bao, Yu Sun, and Tao Mei. 2022. Recent Advances of Monocular 2D and 3D Human Pose Estimation: A Deep Learning Perspective. *ACM Comput. Surv.* 55, 4, Article 80 (November 2022), 41 pages. <https://doi.org/10.1145/3524497>

This work was supported by the National Key R&D Program of China under Grand No. 2020AAA0103800. Authors' addresses: W. Liu, Q. Bao (corresponding author), and T. Mei, Explore Academy of JD.com, Beijing, China; emails: {liuwu1, baoqian, tmei}@jd.com; Y. Sun, Harbin Institute of Technology, Harbin, China; email: yusun@stu.hit.edu.cn. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2022 Association for Computing Machinery.  
0360-0300/2022/11-ART80 \$15.00  
<https://doi.org/10.1145/3524497>

ACM Computing Surveys, Vol. 55, No. 4, Article 80. Publication date: November 2022.

## 논문 리뷰

Recent Advances of Monocular 2D and 3D Human Pose Estimation: A Deep Learning Perspective

단안 카메라에서 2D, 3D 사람의 포즈 추정

2D, 3D 포즈 추정 주류 방식 요약, 본질적 연결 및 방법 분석, 데이터 부족 등 까다로운 사례에 대한 솔루션 분석, 모델 성능 정량적 분석, 유망한 연구 방향 제시

향후 맡게 될 과제 연구 및 연구 주제 선정

Recent Advances of Monocular 2D and 3D Human Pose Estimation : A Deep Learning Perspective

Wu Liu , Qian Bao , Yu Sun , Tao Mei

ACM Computing Surveys, Vol. 55, No. 4, Article 80. Publication date: November 2022.

QnA

질문이 있다면 말씀해주세요.