

# 생성 모델의 이해

## 1. 강의 개요

**교육 목표:** 오토인코더부터 확산 모델까지, 생성 모델의 역사를 이해하고 구현을 통해 원리를 이해할 수 있다.

오늘날 우리는 GPT로 글을 쓰고, 스테이블 디퓨전이나 미드저니로 그림을 그려내며, Instant NeRF로 3D 에셋을 만들어 활용합니다. 생성 모델이 만들어내는 결과물의 질은 갈수록 좋아지고, 다양한 방식으로 창작의 영역을 넓히고 있습니다. 이제는 일반인도 접근성 좋은 웹 데모와 프로그램을 통해 파라미터를 조절하여 이미지를 만들어낼 수 있지만, 생성 모델의 작동 원리나 유래에 대한 이해 없이는 룰렛을 돌리듯 반복 생성에 의존할 수밖에 없습니다. 또한, 스테이블 디퓨전, 달리, 칸딘스키 같은 여러 생성 툴 중에서 어떤 프로그램이나 알고리즘이 적합한지, 각각이 어떤 장점을 갖는지에 대한 명확한 구분도 어렵습니다.

이번 수업에서는 자가 회귀 모델부터 시작해 오토인코더, GAN, 확산 모델에 이르기까지 생성 모델의 발전 과정을 탐구하고, 각 모델의 작동 방식과 원리를 이해하는 것을 목표로 합니다. 이를 통해 현재 우리가 사용하는 생성 모델의 가능성과 한계를 파악하고, 앞으로 AI 생태계에서 지속적으로 발전할 생성 기술을 효과적으로 활용할 수 있는 기반을 마련하고자 합니다.

**교육 대상:** KAIST 학생 및 교직원, 일반인

**선수 학습:** Python 기초, 딥러닝 기초 및 공학수학에 대한 기본적 이해

## 2. 강의 내용

주제	시간	교육 내용
생성 모델 개요	1	생성 모델의 정의
	1	판별 모델(Discriminative model)과 생성 모델(Generative model)
	2	생성 모델의 목적 및 분류 : Explicit vs. Implicit, Tractable vs. Approximate
자가 회귀 모델 (Autoagressive Models)	2	자가 회귀 모델의 등장 배경과 확률적 모델링
오토인코더(AE)와 변분인코더(VAE)	1	오토인코더의 정의 및 배경 : 인코더와 디코더, 잠재 변수의 개념 알아보기
	1	VAE(변분 인코더)의 정의 및 배경 : AE(오토인코더)와 VAE의 차이로부터 VAE의 등장 배경 알아보기
	2	VAE의 구조 및 구현
적대적 생성 모델 (GAN)	1	GAN의 정의 및 배경
	1	GAN의 구조: 구분자와 생성기 : GAN의 기본 요소인 생성자와 구분자의 적용을 단계별로 이해함으로써, GAN에 담긴 직관을 습득할 수 있다.
	2	GAN의 학습 : GAN의 학습 방식을 이해하고, 구현을 통해 전체 구조를 학습한다.

확산 모델 이론	1	확산 모델의 정의 및 배경 (VAE to Diffusion)
	1	확산 모델의 구조 (확산 과정과 역확산 과정)
	2	확산 모델의 구현
확산 모델 응용	1	조건부 확산 모델의 이해
	1	LDM (Or, Stable Diffusion) : 최근 확산 모델의 대표격인 Stable Diffusion (or, latent diffusion model)에 대해 알아보고, 잠재 공간(Latent space) · CLIP · Attention에 대한 간단한 이해를 가질 수 있다.

### 3. 강의 세부 사항

**평가 방법:** 실습 과제 평가

**필수 강의 내용:** 오토인코더(AutoEncoder), 변분인코더(Variational AutoEncoder), 적대적 생성 모델(Generative Adversarial Model), 확산 모델(Diffusion Model)

**교육 환경:** Python 3.10

**추천 참고 서적:**

- Stanford CS236: Deep Generative Models  
: <https://deepgenerativemodels.github.io/syllabus.html>
- MIT 6.S978: Deep Generative Models  
: <https://mit-6s978.github.io/schedule.html>
- Understanding Diffusion Models: A Unified Perspective  
: <https://arxiv.org/pdf/2208.11970>

## 강의 세부 계획서:

# 1. 생성 모델 입문

목적: 생성 모델의 목적과 판별 모델과의 관계를 이해한다.

## 1.1. 생성 모델 (Generative Model) 이란 무엇인가?

- 생성 모델의 예시
- 생성 모델: 확률의 관점에서 세상을 바라보기

## 1.2. 왜 생성 모델인가?

- 생성 모델의 목표: 주어진 데이터셋  $D$ 에 접근하여 데이터 분포를 근사화하는 것.
- 생성 모델은 어떻게 학습하는가?: 확률적 모델링 (Probabilistic Modeling)에 대하여
  - 생성 모델은 데이터 생성 과정에서의 기본 분포를 가정하고 학습한다.
- 생성 모델 평가를 위한 두 가지 요소:
  - 1) 밀도 추정 (Density Estimation): 데이터가 공간에서 어떻게 분포하는가?
  - 2) 샘플링 (Sampling): 학습된 분포를 기반으로 새로운 샘플을 어떻게 생성할 수 있는가?

## 1.3. 판별 모델 (Discriminative Model) vs. 생성 모델 (Generative Model)

- 판별 모델과 생성 모델의 정의
  - 판별 모델 (Discriminative Model): 주어진 데이터 포인트에 대한 레이블을 예측
  - 생성 모델 (Generative Model): 데이터 전체의 결합 분포를 학습
- 생성 모델은 판별 모델이 될 수 있다; 판별 모델은 생성 모델이 될 수 있는가?

## 1.4. 딥 생성 모델 (Deep Generative Models)

- 딥러닝은 표현 학습 (Representation learning) 이다  
: 생성 모델은 확률 분포를 나타내는 방식을 학습한다.
- 핵심 구성 요소:
  - 공식화 (Formulation) / 표현 (Representation) / 목적 함수 (Objective Function) / 최적화 (Optimization) / 추론 (Inference)

## 1.5. 생성 모델로 실제 문제를 공식화하기

- 생성 모델은  $p(x|y)$ 를 다룬다
  - $x$ 와  $y$ 에 해당하는 것은 무엇인가?
  - 생성 모델 적용 예시
    - 자연어 대화 (Natural language conversation)
    - 텍스트-이미지/비디오 (Text-to-image/video)
    - 텍스트-3D 구조 생성 (Text-to-3D structure generation)
    - 클래스 조건부 이미지 생성 (Class-conditional image generation)
    - "비조건부" 이미지 생성 ("Unconditional" image generation)

- 이미지 캡션 생성 (Image captioning)

## 1.6. 강의 소개

## 2. 자가 회귀 모델 (Autoregressive Models)

목적: 자가 회귀 모델의 순차적 접근법과 응용 사례를 탐구한다.

### 2.1. 생성 모델의 분류 (Taxonomy of Generative Models)

- 명시적 밀도 모델 (Explicit Density Model) vs. 암시적 밀도 모델 (Implicit Density Model)
- 계산 가능한 밀도 (tractable) vs. 계산 불가능한 밀도 (intractable)

### 2.2. 자가 회귀 모델 (Autoregressive Model)의 등장 배경

- 확률적 모델링 (Probabilistic Modeling)에의 직관적 접근
  - 조건부 분포 (Joint distribution)를 사용한 결합 분포 모델링 (Conditional distribution modeling)

### 2.3. 자가 회귀 모델의 네트워크 아키텍처

- 자가 회귀 모델은 특정 아키텍처에 제한되지 않는다
- 주요 예시: RNN, CNN, 및 Attention
  - 예시) 텍스트 - GPT 모델, 이미지 - PixelCNN 모델

### 2.4. 자가 회귀 모델의 장점과 한계

- 장점: 텍스트, 오디오 등 순차적 데이터 처리에 효과적이다
- 한계: 고차원 데이터를 처리 시 비효율적이다

## 3. 변분 오토인코더 (Variational Autoencoders, VAEs)

목적: 잠재 변수 모델로서의 VAE와 확률적 딥러닝에서의 역할을 소개한다.

### 3.1. 잠재 변수 모델 (Latent Variable Models, LVMs)

- 잠재 변수란 무엇인가? 왜 사용하는가?
- 문제점: 계산 불가능한 주변 우도 (marginal likelihoods)

### 3.2. 변분 오토인코더 (VAE)의 등장 배경

- 오토인코더 (AE)에서 VAE로: 데이터를 분포로 인코딩
- ELBO (Evidence Lower Bound): 계산 불가능한 확률을 근사화

### 3.3.VAE가 데이터를 생성하는 방법

- 인코더: 입력 데이터를 잠재 분포로 매핑
- 디코더: 잠재 변수로부터 데이터 재구성

### 3.4.VAE의 장점과 응용

- 데이터 압축과 효율적인 확률적 모델링
- 응용: 이미지 생성, 이상 탐지

## 4. 생성적 적대 신경망 (Generative Adversarial Networks, GANs)

목적: 적대적 학습의 원리와 GAN의 독특한 훈련 역학을 탐구한다.

### 4.1.GAN의 등장 배경

- 왜 GAN이 도입되었는가?  
: 분포 간 차이를 측정하는 또 다른 방법이 있을까?

### 4.2.GAN의 기초

- 적대적 게임: 생성기(G) vs. 판별기(D)
- 손실 함수로서의 적대 관계 이용방법

### 4.3.GAN 훈련

- 실습: D-단계와 G-단계
- 문제점: 모드 붕괴, 불안정한 훈련

### 4.4.VAE와의 비교

- 암시적 밀도(Implicit density) 모델링 vs. 명시적 밀도(Explicit density) 모델링
- 샘플링 품질과 선명도

### 4.5.GAN의 변형 및 응용

- 조건부 GAN(cGAN), StyleGAN, CycleGAN
- 응용 분야: 이미지 생성, 도메인 변환

## 5. 확산 모델: 이론

목적: 확산 모델의 반복적 노이즈 기반 접근법과 안정성을 이해한다.

### 5.1.확산 모델 소개

- 역사적 배경: 왜 확산 모델인가?

- 기존 생성 모델(GAN, VAE)의 한계
- 분포 모델링 도구로서의 노이즈(noise)
- 확산 모델에 대한 직관적 이해: 열화(degradation) 과정을 역으로 학습하기 (denoising)

## 5.2.순방향 확산 과정 (Forward Diffusion Process)

- 순방향 과정: 데이터에 노이즈를 단계별로 추가
- 수학적 공식화
  - 노이즈 스케줄(Noise schedule) 정의
  - 마르코프 체인 가정(Markov chain assumptions)

## 5.3.역방향 과정 (Reverse Process)

- 노이즈가 추가된 입력에서 원본 데이터를 재구성하는 법
- 변분 공식화
  - 목적 함수와 리파라미터화 기법(Reparametrization trick)

## 5.4.확산 모델 훈련

- 노이즈 조건 스코어 네트워크 (Noise-conditioned Score Networks)
- 손실 함수
  - 순방향 및 역방향 과정 간의 KL divergence
  - 최대 우도 추정(Maximum Likelihood Estimation, MLE)과의 관계성

## 5.5.확산 모델의 장점과 과제

- GAN과 비교했을 때의 훈련 안정성
- 속도와 품질의 트레이드오프

# 6. 확산 모델: 응용

목적: 확산 모델의 응용과 실질적 영향을 탐구한다.

## 6.1.확산 모델의 주요 변형

- DDIM (Denoising Diffusion Implicit Models):
  - 더 적은 단계로 더 빠른 샘플링
    - DDIM은 어떻게 역방향 과정을 수정하는가?
- 스코어 기반 생성 모델 (Score-based Generative Models)
  - 점수 함수(score function)란 무엇인가?
  - 점수 매칭을 통해 확산 모델을 수정하는 방법
- Stable Diffusion
  - 잠재 확산(Latent Diffusion): 어떻게 확산 과정을 효율적으로 만드는가?
  - 이미지-이미지, 텍스트-이미지 생성으로의 응용

## 6.2. 다른 생성 모델과의 비교

- 확산 모델 vs. VAE: 표현력 vs. 샘플링 효율성
- 확산 모델 vs. GAN: 선명도 vs. 안정성

## 6.3. 다양한 분야에서의 응용

- 비전: 고해상도 이미지 생성, 인페인팅
- 텍스트-이미지 모델: DALL-E, Stable Diffusion
- 과학적 응용: 단백질 구조 생성, 물리 시뮬레이션

## 6.4. 미래 방향과 과제

- 샘플링 시간 단축: 확산 모델을 더 빠르게 만들 수 있을까?
- 멀티모달(Multi-modal) 도메인으로의 확장: 텍스트, 이미지, 오디오의 결합