



ComfyUI

ComfyUI란?

ComfyUI는 Stable Diffusion 모델을 위한 강력하고 모듈식의 GUI(Graphical User Interface) 및 백엔드 시스템

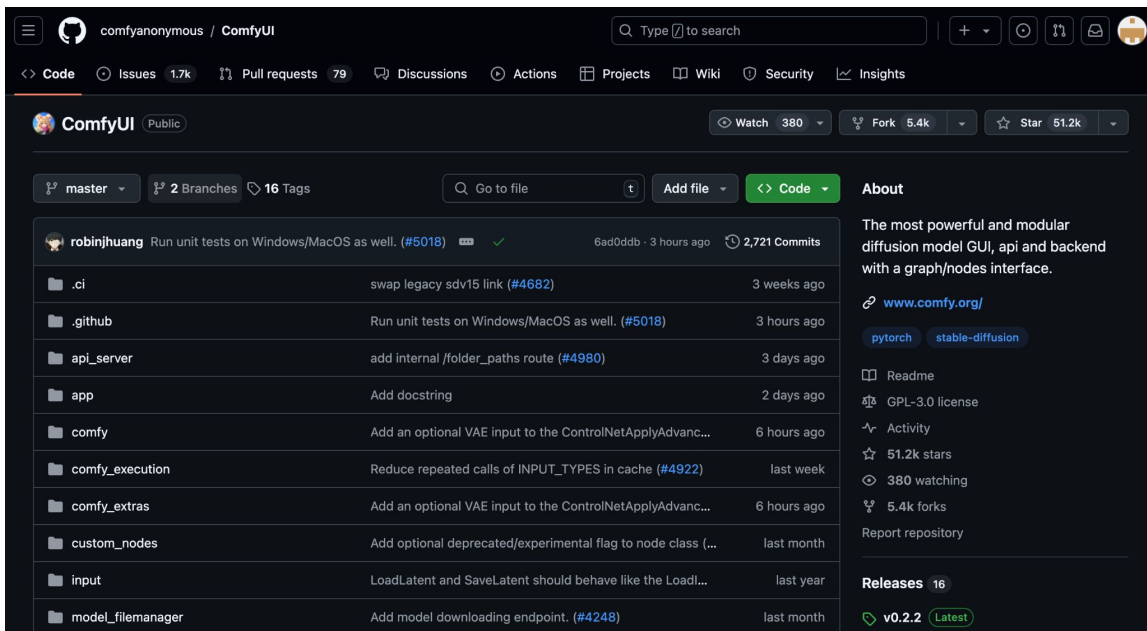
그래프, 노드, 플로우차트 기반의 인터페이스를 사용하여 복잡한 Stable Diffusion 작업 흐름을 설계하고 실행할 수 있게 해주는 것이 특징

ComfyUI의 주요 기능

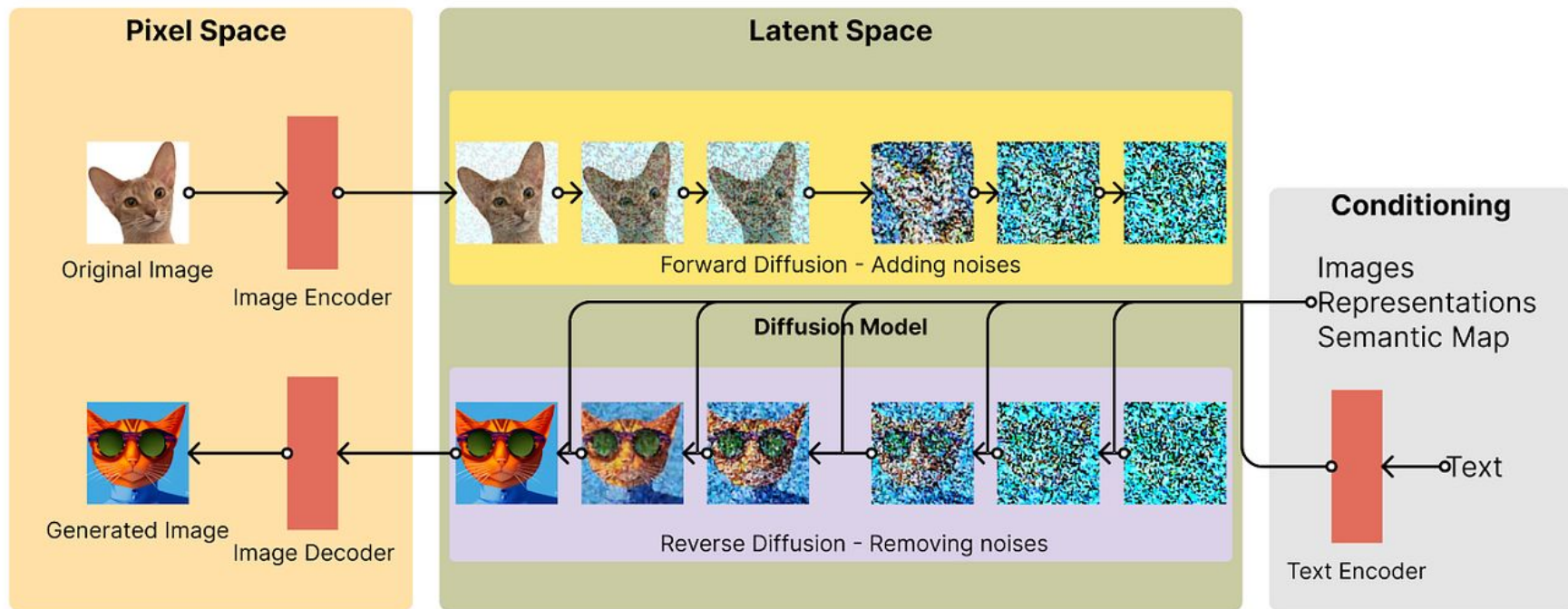
- 다양한 Stable Diffusion 버전을 지원하며, 비동기 큐 시스템을 통해 여러 작업을 동시에 처리
- 워크플로우 변경 사항이 있는 부분만 재실행하는 최적화 기능을 포함
- GPU 메모리가 3GB 미만인 GPU에서도 작동하도록 하는 `--lowvram` 옵션과 CPU만 사용하는 `--cpu` 옵션을 제공
- 다양한 모델과 체크포인트를 로드할 수 있으며, JSON 파일로 워크플로우를 저장하고 로드

Comfyui 설치

Github에서 다운로드 : <https://github.com/comfyanonymous/ComfyUI>



- ComfyUI_windows_portable 폴더 내의 ComfyUI 폴더로 이동
- NVIDIA GPU가 있는 경우 run_nvidia_gpu.bat을 더블 클릭하거나, CPU만 사용하는 경우 run_cpu.bat을 더블 클릭하여 ComfyUI를 실행

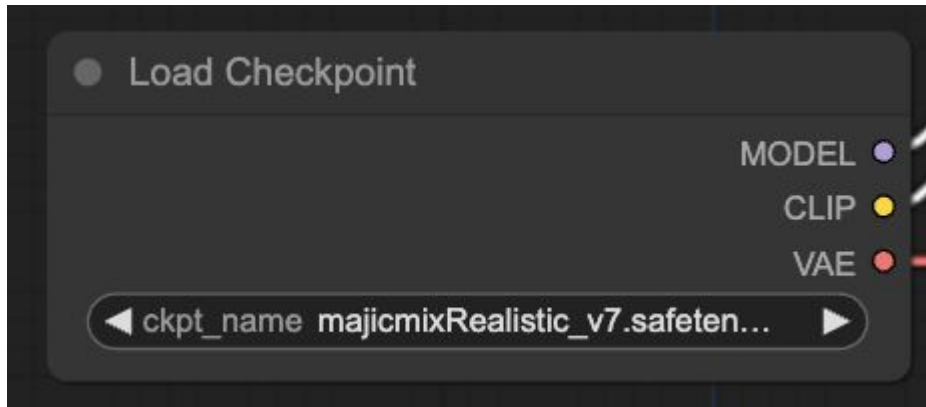


Load Checkpoint

Checkpoint란?

1. **Checkpoint**는 모델이 학습한 상태, 즉 모델의 가중치와 파라미터를 저장한 파일입니다.
2. 모델이 처음부터 학습되지 않고, 이전에 학습된 결과를 사용하기 위해 Checkpoint 파일을 불러옵니다.
3. 이를 통해 **모델을 다시 학습하지 않고도 학습된 모델을 사용**할 수 있습니다.

<https://civitai.com/> 에서 checkpoint 다운로드



Clip Text Encode

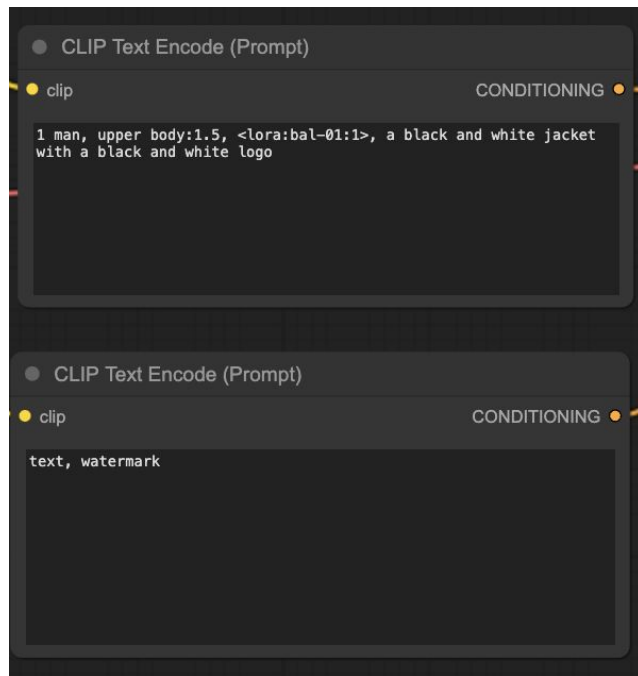
CLIP(Contrastive Language-Image Pretraining) 모델의 텍스트 인코더를 사용해 텍스트를 벡터로 변환하는 과정을 의미
텍스트와 이미지 간의 의미적 연관성을 찾는 핵심적인 단계
Stable Diffusion과 같은 이미지 생성 모델에서 텍스트 프롬프트를 이미지로 변환하는 데 중요한 역할

CLIP 모델이란?

CLIP은 OpenAI가 개발한 모델로, 텍스트와 이미지의 관계를 학습한 멀티모달 모델
CLIP은 두 가지 주요 인코더로 구성

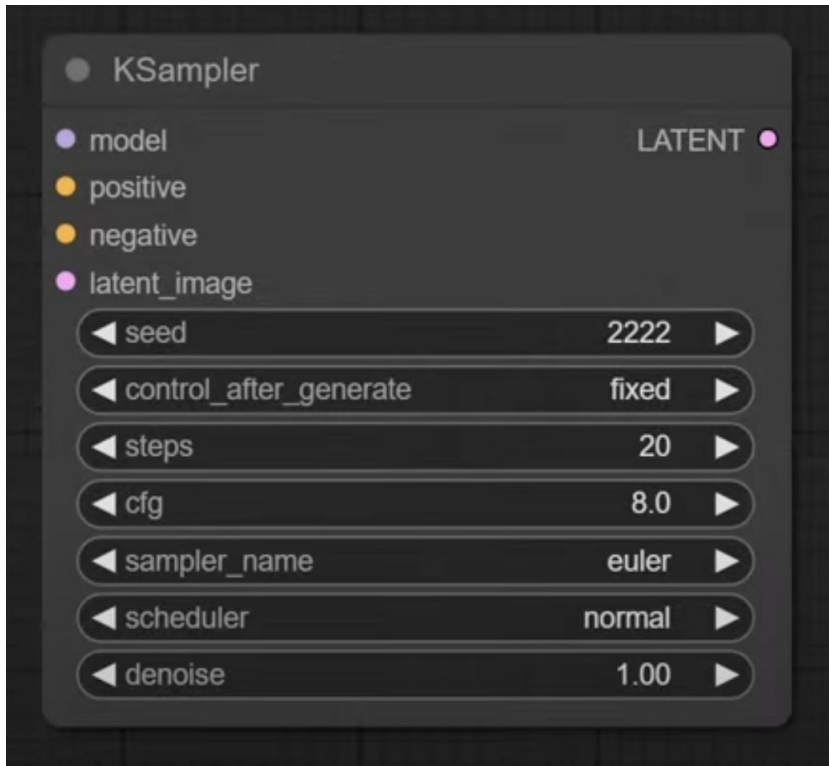
1. 텍스트 인코더 : 텍스트를 벡터(숫자의 집합)로 변환
2. 이미지 인코더 : 이미지를 벡터로 변환

이 두 인코더는 동일한 임베딩 공간에서 작동 하며, 이를 통해 텍스트와 이미지가 같은 의미 공간에서 비교
즉, 텍스트와 이미지의 유사도를 측정



Ksampler

KSampler는 이미지 생성 파이프라인에서 핵심적인 역할을 하는 모듈
주로 노이즈 제거 및 이미지 샘플링을 담당하며, 이미지의 품질, 스타일, 그리고 디테일에
영향



Sampling Method: DDIM, Euler A, PNDM 등의 알고리즘 선택.

Sampling Steps: 몇 단계에 걸쳐 샘플링할 것인지 설정.
보통 20~50 사이의 값을 많이 사용.

CFG Scale: 텍스트 프롬프트에 얼마나 강하게 반응할지를
결정. 보통 7~12 사이의 값을 많이 사용.

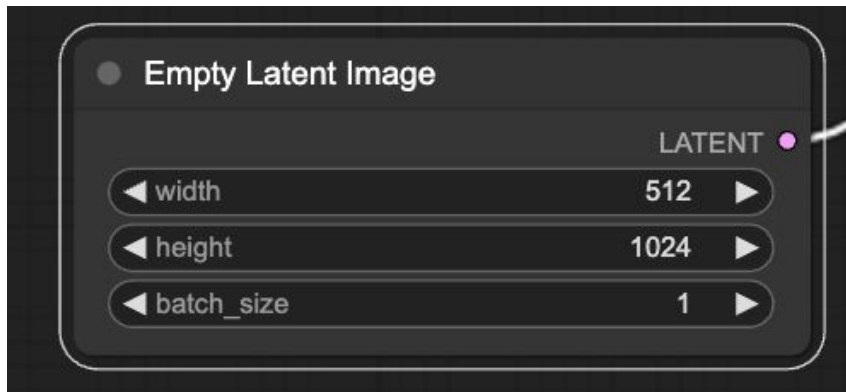
Seed: 이미지 생성의 무작위성을 제어하는 값.

Empty latent image

이미지 생성 파이프라인에서 초기 상태의 잠재 이미지 (Latent Image)를 정의하는 데 사용
이미지 생성 과정에서 노이즈나 초기 조건을 설정하기 위한 역할
Stable Diffusion 같은 모델에서는 **잠재 공간(Latent Space)**에서 이미지를 처리하고
생성하는데, 이 잠재 공간에 입력될 초기 상태의 이미지를 제공

Latent Image란?

Latent Image는 모델이 이미지 자체 대신 잠재 공간에서 처리하는 추상화된 형태의 이미지
Stable Diffusion과 같은 모델은 직접적으로 픽셀 기반 이미지 대신, 노이즈가 추가된 추상화된 형태(잠재 공간)에서
이미지를 생성하고 수정하는 과정을 거침

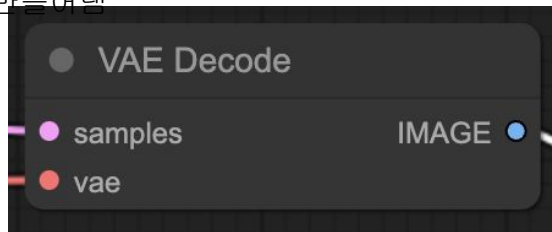


Vae Decode

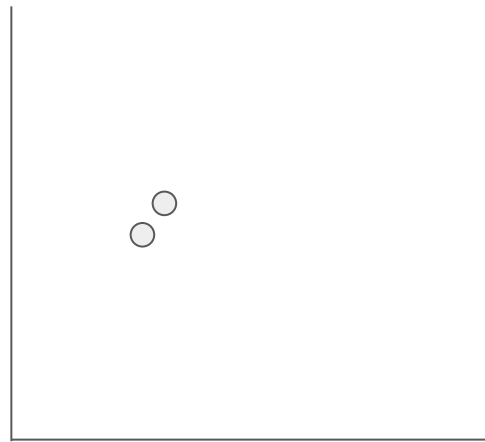
****Variational Autoencoder (VAE)****를 사용하여 **잠재 공간 (Latent Space)**에 있는 데이터를 실제 이미지로 변환하는 과정

VAE Decode의 역할

1. **잠재 공간에서 이미지 복원:**
 - 이미지 생성 과정에서 모델은 **노이즈가 있는 잠재 공간**에서 이미지를 다루고, 이 노이즈가 점점 줄어들면서 최종 잠재 이미지가 생성됩니다.
 - **VAE Decode**는 이 잠재 이미지를 **픽셀 기반의 실제 이미지로 변환**하는 과정
 - **Stable Diffusion**에서는 처음에 이미지가 **잠재 공간에서** 처리 되기 때문에, 최종적으로 이미지를 볼 수 있게 하려면 VAE를 사용해 잠재 공간의 데이터를 실제 이미지로 복원해야함
2. **압축된 정보의 복원:**
 - 잠재 공간에서 표현된 이미지는 압축된 형태이기 때문에, 원래의 고해상도 이미지를 복원하는 과정이 필요합니다. VAE Decode는 이를 수행하여, 잠재 공간의 데이터를 사람이 볼 수 있는 이미지로 생성
 - VAE는 **잠재 공간에 있는 추상적인 정보를 시각적으로 해석할 수 있는 형태로 변환**
3. **노이즈 제거 후 이미지 생성:**
 - **Stable Diffusion**과 같은 모델에서는 이미지가 **노이즈로부터 점진적으로 생성**됩니다. 이때 생성된 잠재 이미지에 마지막으로 VAE Decode를 적용하여, **최종 이미지를 노이즈 없는 상태로 만들어냄**



$$Y = ax + b$$



사과 -> 벡터

[23,324,6,43,86,11]