



오픈소스 SW 기여

프로젝트 제안서

박찬호 32181928

허찬용 32184939

Easy Diffuser

<https://github.com/Easy-Diffuser>

Contents 목차

I 프로젝트 배경 및 목표 01

프로젝트 배경	01
타겟 사용자	01
문제 상황	02
해결 방법	04
제안 기능	06

II 요구사항 07

플로우 다이어그램	07
유저 인터페이스	08
요구사항 명세	10

III 추진계획 11

간트 차트 및 역할 분담	11
소요 장비	12
위험 요인 및 대책	12

V 참고문헌 13

참고 문헌	13
-------	----

I 프로젝트 배경 및 목표

프로젝트 배경

Variational Diffusion Model (이하 VDM)이 등장하기 전 인공지능을 사용한 이미지 생성은 Generative Adversarial Network (이하 GAN)라는 적대적 생성 모델이 주도하였다(Dhariwal & Nichol, 2021). GAN은 높은 품질의 이미지를 빠르게 생성할 수 있기 때문에 다양한 분야에 적용되었다. 그러나 명시적인 확률 분포를 예측하지 않고 판별자(discriminator)를 속이는 방식으로 학습하기 때문에, 불안정한 학습과 낮은 다양성이라는 근본적인 한계를 가지고 있다.

반면 VDM은 명시적인 확률 분포를 근사하여 이미지의 높은 품질과 다양성을 함께 보장한다. 그러나 픽셀 공간에서 학습과 생성이 이뤄지기 때문에 높은 컴퓨팅 비용과 이에 따른 느린 생성 속도를 한계로 가지고 있다. 이 문제는 Latent Diffusion Model(이하 LDM)의 등장으로 해결된다. LDM은 latent 공간에서의 생성으로 낮은 컴퓨팅 비용을 요구하고 생성 속도를 향상시켰다(Rombach 외, 2022). 이후 LDM을 사전학습한 오픈소스 라이브러리 Stable Diffusion이 등장하게 되었고, 일반 PC 환경에서도 VDM을 사용한 이미지 생성이 가능하게 되었다.

타겟 사용자

일반 PC 환경에서의 diffusion 사용이 쉬워지면서 비전문가 사용자가 이미지를 생성하는 사례가 점차 늘어났다. 이러한 흐름은 관련 키워드의 구글 검색 통계에서 확인할 수 있다. 아래의 그래프(그림 1)은 Stable Diffusion이 출시된 2022년 8월을 기점으로 'Diffusion'(파란색)과 'Stable Diffusion'(빨간색)의 검색 횟수가 급격하게 증가함을 보인다(Google Trends, 2023).



그림 1

일반 PC 환경을 사용하는 비전문가 사용자가 이러한 흐름을 주도한다는 것을 확인할 수 있는 또 하나의 근거는 바로 관련 검색어 통계이다(Google Trends, 2023).

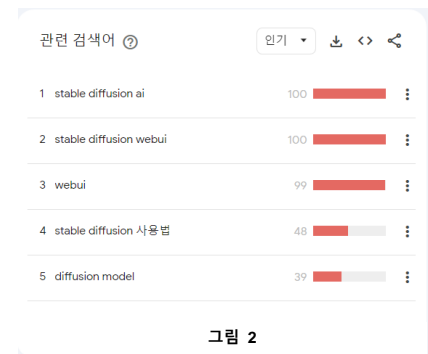


그림 2

먼저 'Stable Diffusion' 관련 검색어 중 인기 순 2 위와 3 위는 webui 이다. 관련 전문가가 api 나 모델 구현으로 이미지를 생성하는 것과 달리, WebUI(그림 3)는 웹사이트 상의 GUI 로 tag 와 파라미터를 간편하게 지정하여 생성할 수 있는, 비전문가를 위한 도구이다(Automatic1111/stable-diffusion-webui). 이를 통해 비전문가 집단에서 Stable Diffusion 을 사용한 이미지 생성에 많은 관심을 가지고 있음을 확인할 수 있다. 뿐만 아니라 'Stable Diffusion 사용법'이라는 검색어가 4 위에 위치하고 있다. 이 역시 비전문가 집단에서 새롭게 Stable Diffusion 을 배우는 사례가 증가하고 있음을 뒷받침한다.

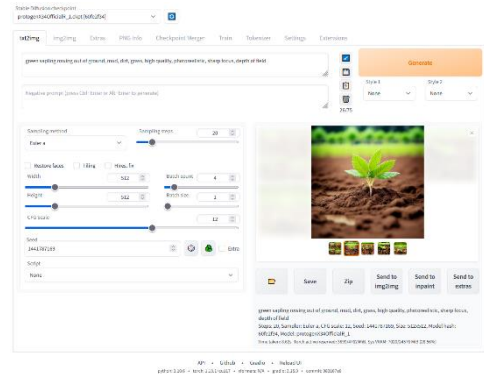


그림 3

위 근거를 통해 비전문가 대상의 적극적인 참여에 의해 Stable Diffusion 을 비롯한 VDM 의 생성 사례가 증가하고 있음을 추론할 수 있다.

Easy Diffuser 프로젝트는 이런 추론에 따라 비전문가 사용자의 Stable Diffusion 을 사용한 이미지 생성을 돕는 것을 목적으로 한다.

문제상황

Stable Diffusion 의 사용자는 태그의 조합과 매개변수 조정으로 생성 모델에 조건을 부여하고, 학습된 이미지 분포 내의 고품질 이미지를 샘플링 할 수 있다. 그러나 VDM 의 생성 결과물이 높은 다양성을 갖기 때문에, 사용자가 원하는 이미지를 생성하는 알맞은 조건은 좁은 범위에 존재한다. 따라서 PC 환경의 비전문가 사용자는 좁은 범위의 조건을 찾기 위해 필연적으로 반복된 시도를 하게 되고, 이러한 시도로 누적된 경험을 바탕으로 조건을 생성한다. 그런데 상기한 상황은 세가지 문제를 유발한다.

첫째로, diffusion 이미지 생성을 위해 사용자는 많은 시간을 소모해야 한다.

Stable Diffusion Text2Image PyTorch Speed		
Device	Precision	Seconds to Finish
NVIDIA A100 80GB PCIe	half	3.74
NVIDIA GeForce RTX 3090	half	4.83
NVIDIA RTX A6000	half	5.03
NVIDIA RTX A5500	half	5.05
NVIDIA GeForce RTX 3080	half	5.59
Quadro RTX 8000	half	5.93
NVIDIA A100 80GB PCIe	single	6.49
NVIDIA GeForce RTX 3090	single	7.96
NVIDIA RTX A6000	single	8.09
NVIDIA RTX A5500	single	8.55
Quadro RTX 8000	single	12.3

그림 4

Stable Diffusion 을 사용한 벤치마크 결과(Lambda, 2022)에 따르면 512 x 512 크기의 이미지 하나를 생성하기 위해 3 ~ 12 초의 시간이 소모된다(그림 4). 이러한 측정 결과는 전문가용

GPU를 온전히 점유하며, WebUI가 아닌 API를 사용하여 코드 상의 최적화를 수행하였을 때의 결과이다. 따라서 비전문가 사용자는 생성에 더 많은 시간이 소모될 수 있다.

이 실험에서는 단순한 생성 문구와 조건을 적용하였는데, 실제로는 복잡한 생성 문구와 모델 병합, upscaling 등의 조건으로 인해 생성 시간은 기하급수적으로 증가하게 된다. 또한 생성하고자 하는 이미지가 등장하는 빈도에 따라 반복횟수 역시 증가할 수 있다. 정리하면, 비전문가 사용자는 한 장의 이미지를 생성하기 위해 최소 12초 이상의 시간을 소모하고, 이 과정을 원하는 이미지가 나올 때까지 반복하게 된다.

만약 하나의 이미지를 생성하기 위해 100번의 반복을 수행한다고 가정하면, 벤치마크의 최대 시간 기준 1200초의 총 시간을 소모하게 된다. 따라서 세 장의 이미지를 생성하기 위해 1시간이 소요된다. 이는 사용자가 태그를 입력하고 조건을 설정하는 시간을 포함하지 않았기 때문에, 실제 이미지 생성 과정에서 사용자는 많은 시간을 소모한다는 것을 알 수 있다.

둘째로, 반복된 diffusion 이미지 생성은 많은 컴퓨팅 자원과 전력을 소모한다.

LDM은 latent 공간에서의 이미지 생성으로 막대한 자원을 소모하는 diffusion의 문제를 완화하였으나(Rombach 외, 2022), 여전히 많은 컴퓨팅 자원과 전력을 소모한다. 이러한 상황을 구체적으로 확인하기 위해 상황을 가정하여 전력을 계산한 결과는 아래와 같다.

먼저, 앞서 소개한 벤치마크에 사용된 GPU 중 비전문가 사용자에게 가장 높은 가격 접근성을 가진 RTX 3080을 사용한다고 가정한다. Diffusion 이미지 생성 시에 GPU를 100% 사용한다면, RTX 3080의 피크 전력이 340W이므로 시간당 340Wh의 전력이 소모된다(EcoEnergyGeek, 2022). 이는 인버터형 에어컨이 희망온도에 도달한 후 소모하는 시간당 전력(400Wh)과 맞먹는 수치이다(중앙 선데이, 2018). 이를 통해 diffusion 이미지 생성 과정에서 사용자가 많은 컴퓨팅 자원과 전력을 소모함을 알 수 있다.

마지막으로, 이미지 생성 과정의 경험이 사용자 간에 공유되기 어렵다.

먼저, 사용자는 반복된 생성을 통해 경험과 최종 결과물을 얻게 된다. 자발적인 공유를 희망하는 사용자가 존재할 때, 사용자의 반복된 생성으로 얻어진 최종 결과물은 어렵지 않게 공유될 수 있다. 사용자는 최종 결과물을 생성할 때 사용한 태그의 조합과 지정한 매개변수만 공유하면 된다.

그러나, 사용자가 생성을 반복하며 누적한 경험은 다른 사용자에게 공유되기 어렵다. 그 이유는 다음과 같다. 사용자의 경험을 공유하기 위해서는 중간 과정의 결과물들과 태그의 조합, 매개변수를 면밀히 기록 및 분석하고 이 과정에서 생기는 변화를 결과물의 변화로 정리해야만 한다. 그런데 이러한 과정은 자발적인 공유 의사와 사용자의 노력에 의존하며, 어떤 사용자가 생성 과정을 면밀히 정리했다고 하더라도 그 결과물에서의 추론은 사용자의 주관에 따르기 때문에 경험 그 자체가 온전히 공유되기 어렵다. 이는 아는 것과 이를 규칙으로 구성하는 것 사이에는 큰 간극이 존재하기 때문이다.

누적된 경험은 사용자의 생성과정에 중요한 역할을 한다. 이는 경험이 아래의 이유로 시간과 비용을 단축시키는 중요한 요소가 되기 때문이다. 만약 최종 결과물이 갖는 특성의 일부를 공유하는 결과물을 생성하고자 할 때, 최종 결과물을 생성하는 생성 조건에서 특성을 추출하는 일은 상당히 어려운 일이다. 왜냐하면 생성 조건이 어떻게 결과물에 영향을 주는지 알 수 없기 때문이다. 따라서, 최종 결과물 생성 과정의 누적된 경험이 공유된다면 최종 결과물에서 일부 특성을 분리하는 것이 비교적 쉬워진다. 그러나 누적된 경험이 공유되지 않기 때문에, 사용자는 최종 결과물로부터 역추적하여 일부 특성을 찾아야 한다. 이는 조건을 변화하며 생성을 반복해야 하는 일이기 때문에, 다시 첫번째와 두번째 문제로 이어진다.

해결 방법

이러한 문제를 해결하기 위해 아래의 서비스들이 사용되고 있다.

1. Civitai (<https://civitai.com/>)

Civitai 는 Stable Diffusion 이미지 생성 결과물과 태그 등의 조건을 공유하는 커뮤니티이다. 사용자들은 자발적으로 태그의 조합과 매개변수, 모델, 그리고 최종 결과물을 공유하며 Civitai 의 사용자는 Stable Diffusion 이미지 생성 시의 시행착오를 줄일 수 있다. 그러나 Civitai 의 문제점은 공유된 결과물을 적용하는 것이 사용자의 탐색에 의존한다는 데에 있다. Civitai 에 나열된 결과물에서 하나를 선택해 재생산하는 것은 어렵지 않으나, 원하는 이미지를 Civitai 에서 찾아 생성 조건을 가져오는 것은 온전히 사용자의 몫이다.

2. NovelAI 태그 생성기 (<https://novelai.app/>)

NovelAI 태그 생성기는 외모, 체형, 화풍, 장신구 등의 다양한 생성 태그를 정리하고, 이를 빠르게 복사하여 조합으로 정리할 수 있는 도구이다. 사용자는 NovelAI 태그 생성기를 통해 원하는 속성을 생성 조건으로 빠르게 적용할 수 있다. 그러나, NovelAI 태그 생성기는 한정된 범위의 널리 알려진 속성만을 태그로 제공하기 때문에, 이미지의 기본 형태를 구성하는 데에는 용이하지만 그 이상을 지원하지 못한다.

3. Stable Diffusion prompt typing assistant (<https://chrome.google.com/webstore/detail/stable-diffusion-prompt-t/dbjmlleeimjolifnjbmjpaooibaodhceg>)

Stable Diffusion prompt typing assistant는 Stable Diffusion의 태그를 추천해주는 chrome extension이다. NLP를 통해 다음 태그를 예측하도록 개발되었으며, 사용자는 태그 작성 시 태그를 추천받을 수 있다. 그러나 주어진 태그로 다음 태그를 예측하는 단순한 방식으로 태그를 추천하기 때문에, 사용자가 원하는 이미지의 특성을 명확히 반영하기 어렵다. 또한 Stable Diffusion을 개발한 Stability.ai에서 제공하는 유료 서비스(Dream Studio)에서 동작하기 때문에, 추가 비용 없이 PC 환경의 자원을 이용하는 사용자에게는 적합하지 않다.

4. Dreamer

Dreamer는 Stability.ai의 유료 Stable Diffusion 서비스 dream studio에서 사용할 수 있는 chrome extension이다. 사용자는 Dreamer를 사용하여 웹 페이지 상의 이미지를 빠르게 가져와 img2img 등의 생성 과정에 불러올 수 있다. 그러나 dreamer는 태그와 관련된 기능을 제공하지 않고, Dream studio라는 유료 서비스에서 사용할 수 있기 때문에 PC 사용자에게는 적합하지 않다.

위 내용을 정리하면 Stable Diffusion 이미지 생성의 문제상황은 아래와 같다.

1. Stable Diffusion 이미지 생성은 사용자의 많은 반복과 긴 생성 시간을 필요로 한다.
2. 반복된 Stable Diffusion 이미지 생성은 많은 컴퓨팅 자원과 전력을 소모한다.
3. Diffusion 이미지 생성 과정의 경험이 사용자 간에 공유되기 어렵다.

문제 1 과 2 는 반복된 이미지 생성이라는 시행착오에서 발생하는데, 이는 위의 서비스들이 모두 해결하고자 하는 목표이다. 기존 서비스들은 이 문제들을 어느정도 해결하고 있으나, 여전히 생성 조건의 탐색을 사용자의 몫으로 남겨두고 있다는 문제가 있다. 문제 3 은 누적된 경험을 모으고 분석하는 방법이 없다시피 한 현재의 상황에서 직접적으로 해결되기 어렵다. 그러나 문제 3 을 해결함으로써 얻고자 하는 장점이 최종 결과물로부터 개별 특성을 추론해내는 것임에서 착안하면, 경험에서 얻어지는 ‘특성에 대한 추론능력’을 근사하는 머신 러닝 모델을 학습하여 문제에 접근할 수 있다.

따라서 Easy Diffuser는 ‘Reference를 통한 이미지 생성 보조’라는 해결책을 제안한다.

사용자가 생성하고자 하는 이미지는 사용자가 밑바탕부터 구상한 것일 수도 있지만, 특정한 사진을 참고하여 구상한 것일 수도 있다. 게다가 밑바탕부터 구상한 이미지일지라도 인터넷에서 유사한 특성을 가진 이미지를 탐색하는 것은 역시 가능하다. 이와 같이 사용자의 생성을 보조하는 참고 이미지를 탐색하는 일은, 생성 조건을 탐색하는 것보다 많은 자료와 도구를 사용하여 탐색이 가능하므로, 사용자의 입장에서 훨씬 쉽게 접근이 가능하다. 따라서, Easy Diffuser는 Diffusion 생성 조건이 주어져 있지 않은 참고 이미지로부터 생성 조건을 추출하여 이미지 생성을 보조한다. 참고 이미지로부터 추출한 생성 조건은 사용자로 하여금 기반 조건이 되어주고, 덕분에 사용자는 적은 반복을 거쳐 원하는 이미지를 생성할 수 있다. 결론적으로, Easy Diffuser가 제공하는 ‘Reference를 통한 이미지 생성 보조’는 1, 2번 문제를 보다 잘 해결할 수 있으며, 3번 문제에서 근사하는 ‘특성에 대한 추론 능력’도 어느정도 해결할 수 있다.

제안 기능

Easy Diffuser 는 Reference 를 통한 이미지 생성 보조를 수행하기 위해 아래의 두 기능을 제안한다.

1. Reference 이미지로부터 생성 조건을 추출하는 기능

사용자는 Reference 이미지를 불러와 추출 기능을 작동시킨다. Easy Diffuser는 입력된 이미지를 OpenAI의 CLIP을 거쳐 텍스트로 변환하고, transformer 기반의 번역기를 사용해 태그 조합으로 번역을 수행한다. 결과물로 도출된 태그 조합은 사용자가 WebUI에서 txt2img의 입력으로 사용할 수 있다.

2. Reference 이미지를 입력하여 img2img 생성을 수행하는 기능

기존의 img2img 생성 기능은 사용자가 이미지를 찾고, 저장하고, 불러오는 등의 단계를 거쳐 동작하거나, txt2img로 생성한 이미지를 불러와 사용할 수 있었다. Easy Diffuser는 웹 페이지의 이미지를 바로 WebUI의 img2img 생성으로 불러올 수 있도록 간편한 메뉴를 지원한다.

II 요구사항

플로우 다이어그램

요구사항 명세를 통해 Easy Diffuser 가 제공할 기능을 구체화하기에 앞서 플로우 다이어그램으로 유저의 동작에 따른 작업과 기능의 변화를 흐름에 따라 구체화하였다. 그 흐름을 따라가면 아래와 같다.

먼저, 유저는 브라우저에서 WebUI 를 실행 이후 Easy Diffuser 서비스를 실행한다.

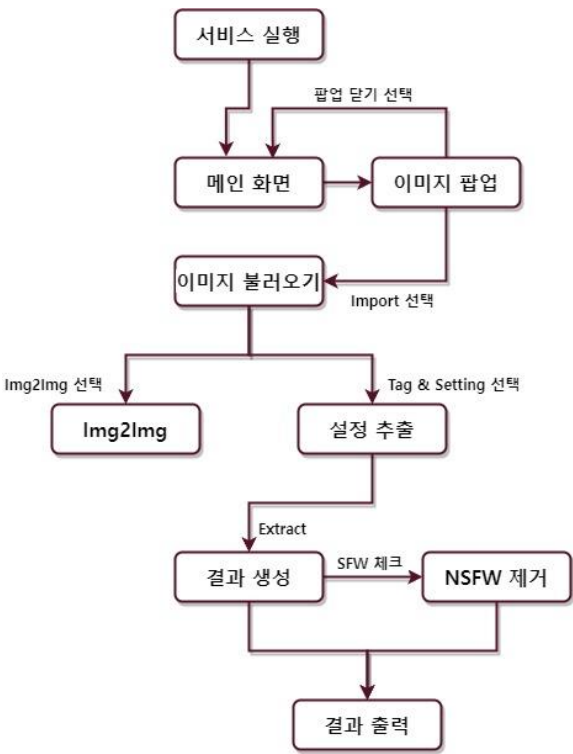
메인 화면이 브라우저 우측에 존재할 때, 유저는 웹 페이지에서 이미지를 선택하고 불러올 수 있다. 불러오기는 이미지 팝업을 사용해 유저의 확인을 거친다.

유저는 불러온 이미지를 img2img 나 설정 추출에 사용할 수 있다.

Img2img 를 선택한 후 import 버튼을 누르면 해당 이미지가 WebUI 로 이동한다.

Tag & Setting 을 선택한 후 Extract 버튼을 누르면 해당 이미지를 바탕으로 생성 조건을 추출한다. 만약 추출 시에 SFW 가 체크되어 있다면, 추출한 결과에서 선정적인 키워드 등 윤리성을 저해하는 단어를 제거한 후 결과를 출력한다.

이 과정을 기능 명세의 형태로 정리하면 아래와 같다.



기능명세서				
작성자/작성자		Easy Diffuser 팀 / 박찬호, 허찬용		
작성일/버전		2023.04.02/ v1.0		
비즈니스 요구사항	프로젝트 배경	높은 컴퓨팅 비용을 필요로 하는 이미지 생성이 Latent Diffusion Model의 등장으로 적은 컴퓨팅 비용으로 이미지를 생성할 수 있다.		
	프로젝트 목적	일반 대중이 Diffusion model을 사용하여 이미지 생성을 위해 필요한 어려운 Tag 조합, Reference를 통한 이미지 생성 한계를 해결에 도움을 줄 수 있다.		
	프로젝트 목표	사용자가 원하는 이미지를 정확히 생성을 위해 보조하는 프로그램을 설계하는 것이다.		
	성공기준	각각의 서비스에 대한 세부 기능들이 테스트 케이스들을 통과 하면 성공으로 간주한다.		
ID	구분	서비스(메뉴)	기능	상세기능
		서비스 실행		Chrome 검색창 우측에 있는 프로그램 버튼을 클릭하여 실행을 한다.
		익스텐션 창		서비스를 실행하면 Chrome 우측에 프로그램 화면이 생성된다. 생성된 화면에는 이미지를 넣을 수 있는 Image Box, 수행하고자 하는 작업을 선택하는 체크박스, 선정적인 태그를 제외시킬 수 있는 SFW 체크 박스가 있다. 해당 설정 박스들 아래에는 작업을 수행한 뒤 나오는 결과물들을 출력하는 Text Box가 존재한다.
		이미지 불러오기		사용하고자 하는 이미지 위에서 마우스 오른쪽 클릭을 한 뒤 메뉴창이 생성되면 "Easy Diffuser로 불러오기" 버튼을 눌러 이미지 박스로 사용하고 자 하는 이미지를 불러온다.
		이미지 팝업		이미지를 불러올 것인지 사용자에게 확인 받는 팝업이다. 사용자의 확인을 거친 후 이미지를 익스텐션 창의 이미지 칸에 불러온다.
		Img2Img		프로그램 화면에 있는 Img2Img 체크박스를 클릭을 하고 Extract 버튼을 클릭하면 사용자의 WEBUI 이미지 박스에 해당 사진을 보내준다.
		설정 추출		프로그램 화면에 있는 Tag&Setting 체크박스를 클릭을 하고 Extract 버튼을 클릭하면 해당 이미지와 관련된 Tag들이 생성된다.
		결과 생성		Extract 버튼을 누르고 생성된 Tag들을 Positive Tag와 Negative Tag로 분류된다.
		NSFW 제거		SFW(Safe for Work) 체크박스가 체크되어 있으면 선정적인 Tag들을 Positive와 Negative Tag에서 제외시켜준다.
		결과 출력		Positive, Negative prompt 텍스트 박스에 생성된 Tag 결과물들을 출력해준다.

유저 인터페이스

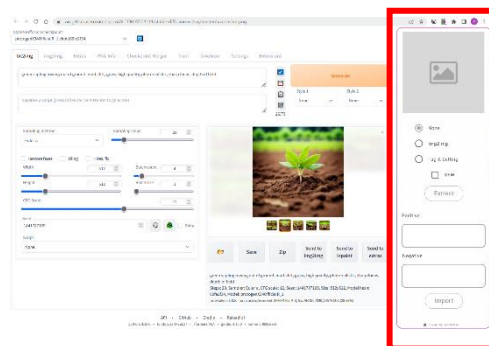
기능 명세를 바탕으로 인터페이스는 아래와 같이 구성된다.

서비스 실행



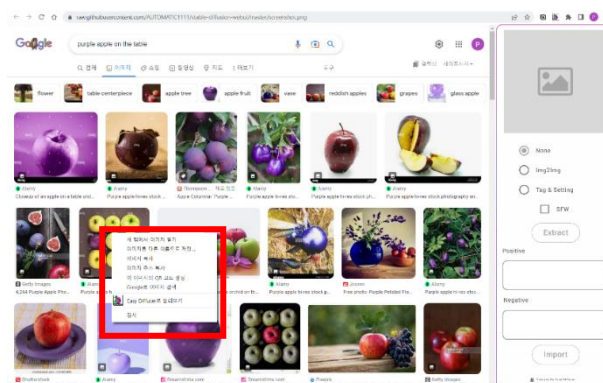
우상단의 Easy Diffuser 버튼을 통해 실행할 수 있다.

익스텐션 창



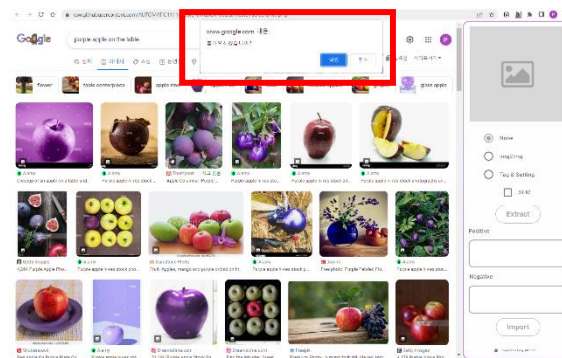
서비스 실행 시 우측에 익스텐션 창이 실행된다

이미지 불러오기



브라우저의 메뉴로 'Easy Diffuser 로 불러오기' 를 제공하고, 클릭 시 불러오기를 진행한다. 불러오기는 바로 다음에 등장하는 팝업 창에서의 확인을 거친다.

이미지 팝업

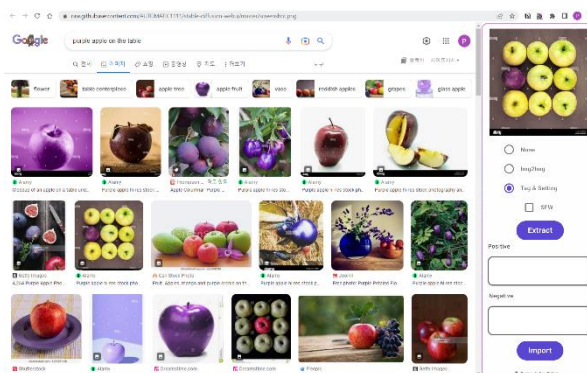


이미지 팝업은 불러오기 과정에서 사용자의 확인을 거친다.

Img2img

Img2img 는 WebUI 로 이미지를 보낸다. 동일한 익스텐션 화면에서 진행되므로 UI 는 생략한다.

설정 추출



설정 추출은 Tag & Setting 을 선택하고, Extract 버튼을 눌러 진행된다.

결과 생성 & NSFW 제거 & 결과 출력

결과 생성 및 NSFW 제거, 결과 출력은 사용자의 입장에서 화면 안에서 발생하는 한 번의 변화이므로, 묶어서 서술한다. 결과 생성은 Extract 버튼을 누른 후 내부적으로 태그를 생성하고, SFW 체크에 따라 선정성 있는 태그를 제거하고 결과를 출력하는 과정으로 진행된다. 출력된 태그는 클릭 가능한 별개의 버튼으로 존재하며, 사용자는 이 중 원하지 않는 태그를 해제하고 WebUI 로 불러올 수 있다.

요구사항 명세

위에서 설명한 흐름을 바탕으로 요구사항 명세를 통해 구체화한 결과는 아래와 같다.

요구사항정의서			
작성처/작성자		Easy Diffuser 팀 / 박찬호, 허찬용	
작성일/버전		2023.03.30/ v1.0	
비즈니스 요구사항	프로젝트 배경	높은 컴퓨팅 비용을 필요로 하는 이미지 생성이 Latent Diffusion Model의 등장으로 적은 컴퓨팅 비용으로 이미지를 생성할 수 있다.	
	프로젝트 목적	일반 대중이 Diffusion model을 사용하여 이미지 생성을 위해 필요한 어려운 Tag 조합, Reference를 통한 이미지 생성 한계를 해결에 도움을 줄 수 있다.	
	프로젝트 목표	사용자가 원하는 이미지를 정확히 생성을 위해 보조하는 프로그램을 설계하는 것이다.	
	성공기준	각각의 서비스에 대한 세부 기능들이 테스트 케이스들을 통과 하면 성공으로 간주한다.	

ID	구분	서비스(메뉴)	기능	상세기능	개발형태	출처(관련근거)	제약조건					특이사항
							사용자	사용시기	사전조건	사용조건	사후조건	
		Chrome Extension	이미지 불러오기 메뉴	메뉴창에 Extension을 사용하여 이미지를 불러올 수 있다. Extension 그림이 그려진 이미지 불러오기를 클릭하면 우측에 Extension 프로그램 이 보여진다. 프로그램이 켜진 뒤 상단 이미지 박스에 이미지가보여진다.	Json	신규개발	회원					
		Chrome Extension	선택 radio button	이미지가 불러져 프로그램이 활성화되면, None, Img2img, Tag & Setting radio button 3가지 버튼을 통하여 본인이 하고자 하는 작업을 선택할 수 있다. 디폴트 값은 None 이다.	Json	신규개발	회원					
		Chrome Extension	SFW 체크박스	SFW(Safe For Work) 체크박스를 선택할 경우 부적절한 이미지를 생성할 수 있는 태그를 제외하고 생성한다. Default 값은 uncheck이다.	Json	신규개발	회원					
		Chrome Extension	Extract 버튼	태그 추출을 실행하는 버튼이다. 프로그램 상단에 올라와 있는 이미지를 사용하여 해당 이미지와 비슷한 이미지를 생성할 수 있는 태그를 출력해 준다.	Json	신규개발	회원					
		Chrome Extension	Positive & Negative prompt	Extract 과정을 통해 생성된 Positive Tag와 Negative Tag를 구분하여 출력한다.	Json	신규개발	회원					
		Machine Learning	Clip(Connecting Text and Images)	Hugging Face나 Github에서 Pretrained된 Img2Text 모델을 사용하여 입력된 img를 사용하여 Text로 변환한다.	python	Hugging Face	관리자					
		Machine Learning	Tag & Setting translation	Clip 라이브러리를 사용하여 얻은 Text를 Transformer 모델에 입력하여 이미지 생성을 위한 Tag를 예측한다.	python	Hugging Face	관리자					
		Browser	이미지 불러오기	특정 이미지 위에서 마우스 오른쪽 클릭을 하면 chrome extension으로 이미지를 불러올 수 있다.	Json	신규개발	회원					

위에서 제시한 플로우 다이어그램, 기능 명세, 요구사항 명세는 모두 깃헙 레포지토리에서 관리하고 있다.

Ⅲ 추진계획

간트 차트 및 역할 분담

추진 계획 및 역할 분담은 간트 차트를 통해 정리하였으며, 그 내용은 아래와 같다.

Easy Diffuser Schedule

프로젝트 이름	Easy Diffuser	학과명	모바일시스템공학과
프로젝트 관리자	박찬호/허찬용	날짜	23년 3월 8일

WBS 번호	작업 제목	작업 소유자	시작일	마감일	기간	작업 완료 비율	1단계												2단계												3단계												4단계														
							1주				2주				3주				4주				5주				6주				7주				8주				9주				10주				11주				12주						
							월	화	수	목	금	월	화	수	목	금	월	화	수	목	금	월	화	수	목	금	월	화	수	목	금	월	화	수	목	금	월	화	수	목	금	월	화	수	목	금	월	화	수	목	금						
1	프로젝트 구성 및 착수																																																								
1.1	프로젝트 구성	박찬호	23년 3월 8일	23년 3월 14일	6	100%																																																			
1.1	아이디어 조사	박찬호, 허찬용	23년 3월 15일	23년 3월 21일	6	100%																																																			
1.3	실현 가능성 조사	박찬호, 허찬용	23년 3월 22일	23년 3월 28일	6	100%																																																			
1.3	요구사항 명세서 및 간트 차트 작성	허찬용	23년 3월 29일	23년 4월 4일	6	100%																																																			
1.3	제안서 작성	박찬호	23년 3월 29일	23년 4월 4일	6	100%																																																			
1.4	프로젝트 착수	박찬호, 허찬용	23년 4월 5일	23년 4월 6일	1	50%																																																			
2	코드 작성																																																								
2.1	메인화면 프론트 개발	박찬호	23년 4월 7일	23년 4월 12일	5	0%																																																			
2.2	이미지 팝업 및 불러오기 기능 개발	박찬호	23년 4월 13일	23년 4월 18일	5	0%																																																			
2.3	Tag 추출 모델 개발	박찬호, 허찬용	23년 4월 19일	23년 5월 9일	20	0%																																																			
2.3.1	Img2Text 모델 개발	박찬호, 허찬용	23년 4월 19일	23년 5월 9일	20	0%																																																			
2.3.1	Text2Tag 모델 개발	박찬호, 허찬용	23년 4월 19일	23년 5월 9일	20	0%																																																			
2.4	결과 Tag 출력 기능 개발	허찬용	23년 5월 10일	23년 5월 16일	6	0%																																																			
3	모델 평가																																																								
3.1	사용 후기를 통한 성능 평가	박찬호, 허찬용	23년 5월 18일	23년 5월 31일	13	0%																																																			
3.2	모니터링	박찬호, 허찬용	23년 5월 18일	23년 5월 31일	13	0%																																																			
3.3	모델 업데이트	박찬호, 허찬용	23년 6월 1일	23년 6월 14일	13	0%																																																			

위의 간트 차트는 깃헙 레포지토리에서 관리하고 있다.

소요 장비

Stable Diffusion 을 사용한 이미지 생성 및 머신러닝 모델 학습은 상당한 컴퓨팅 자원을 요구한다. 특히 GPU 가 중요한데, Easy Diffuser 의 개발에서는 연구실 서버의 Nvidia RTX 3090 을 사용하여 개발을 진행한다.

```
lspci | grep VGA
00:00.0 VGA compatible controller: ASPEED Technology, Inc. ASPEED Graphics Family (rev 41)
8a:00.0 VGA compatible controller: NVIDIA Corporation GA102 [GeForce RTX 3090] (rev a1)
```

이외의 추가적인 장비는 필요하지 않다.

위험 요인 및 대책

Easy Diffuser 의 개발 과정에서 발생할 수 있는 위험 요인과 그 대책은 아래와 같다

1. 컴퓨팅 자원 부족

컴퓨팅 자원이 부족하면 개발 속도를 느려질 수 있다. 그러나 현재 소요 장비에 명시한 자원은 고성능의 자원이고, 소규모 프로젝트에서 이 이상의 자원을 소모하는 것은 쉽지 않다. 따라서 이러한 문제가 발생할 경우, 추가 자원을 사용하는 것보다 번역 모델 등을 학습하는 비중을 줄이고 사전학습 된 모델의 비중을 늘릴 수 있다.

2. 생성 결과물의 정확도 부족

Easy Diffuser의 학습은 원본 생성 조건과 추론한 생성 조건 간의 차이를 작게 하도록 진행된다. 그러나 Stable Diffusion이 생성할 수 있는 데이터의 범위를 Easy Diffuser가 충분히 다루지 못할 경우, Easy Diffuser의 학습 정확도가 실제 생성 결과물의 정확도로 이어지지 않을 수 있다. 이에 대한 해결책으로, 온라인 학습을 진행할 수 있다. 데이터 셋에 대해 어느 정도의 정확도를 확보한 후, 오픈소스로 모델을 공개하여 오픈소스 커뮤니티의 데이터 추가와 참여를 통해 모델을 계속 업데이트 하여 Easy Diffuser가 다루지 못하는 범위의 데이터를 지속적으로 반영한다.

V 참고문헌

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34, 8780-8794.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10684-10695).

Google Trends. (n.d.). Google Trends. 4 월 2 일, 2023, <https://trends.google.com/trends/explore?date=today%205-y&geo=KR&q=Diffusion,Stable%20Diffusion&hl=ko>

Automatic1111/stable-diffusion-webui. (n.d.). Github. 4 월 2 일, 2023, <https://github.com/AUTOMATIC1111/stable-diffusion-webui>

All You Need Is One GPU: Inference Benchmark For Stable Diffusion. (10 월 5 일, 2022). Lambda. 4 월 2 일, 2023, <https://lambdalabs.com/blog/inference-benchmark-stable-diffusion>

RTX 3080 Power Consumption Guide. (5 월 5 일, 2022). EcoEnergyGeek. 4 월 2 일, 2023, <https://www.ecoenergygeek.com/rtx-3080-power-consumption/>

에어컨 하루 8 시간씩 켜면 전기료 8 만원 더 나온다. (7 월 28 일, 2018). 중앙선데이. 4 월 2 일, 2023, <https://www.joongang.co.kr/article/22841084#home>