그놈을 잡아라: 몽타주 생성 ai 프레임워크

캡스톤 디자인 1조

보고서 및 논문 윤리 서약

1. 나는 보고서 및 논문의 내용을 조작하지 않겠습니다.
2. 나는 다른 사람의 보고서 및 논문의 내용을 내 것처럼 무단으로 복사하지 않겠습니다.
3. 나는 다른 사람의 보고서 및 논문의 내용을 참고하거나 인용할 시 참고 및 인용 형식을 갖추고 출처를 반드시 밝히겠습니다.
4. 나는 보고서 및 논문을 대신하여 작성하도록 청탁하지도 청탁받지도 않겠습니다.

나는 보고서 및 논문 작성 시 위법 행위를 하지 않고, 명지인으로서 또한 공학인으로서 나의 양심과 명예를 지킬 것을 약속합니다.

학　　과 : 융합소프트웨어학부

과　　목 : 캡스톤디자인

담당교수 : 김대원

클 래 스 : 화 17:50~ 20:15

이　　름 : 노장현, 안미르, 안승연, 양채연, 황수진

I. 프로젝트 개요

1. 문제 정의

몽타주는 수사 과정에서 이용되는 시각적 자료 중 하나로 인물의 인상과 느낌을 표현한 스케치이다. 몽타주 제작은 CCTV에 기록되지 않은 용의자 검거 뿐만 아니라 실종 인물 찾기 등 시각적 기록이 없는 인물의 모습과 인상을 재구성할 때 꼭 필요한 작업이다. 한국의 수사 현장에는 2015년도에는 '폴리스케치'라는 3D 몽타주 제작 프로그램이 도입되어 몽타주 제작자가 진술자의 구체적 묘사를 바탕으로 얼굴형, 눈썹, 눈 등 데이터베이스에 저장된 다양한 이목구비를 골라 조합한 몽타주를 제작한다. 이는 구체적 정보를 조합하는 방식이기 때문에 전체적인 얼굴 인상에 대한 추상적인 진술은 직접적으로 반영하지 않는다는 한계가 있다. 또한 이목구비의 모양을 하나하나 찾아야 하기 때문에 시간과 노력이 적지 않게 든다는 단점이 있다.

2. 프로젝트 목표

본 프로젝트에서는 generative model을 이용하여 몽타주 제작자가 직접 얼굴의 이목구비를 조합할 필요 없이, 진술 텍스트를 입력하면 자동으로 이미지가 생성되는 프레임워크를 개발하고자 한다. 프로젝트 목표는 다음과 같다.

1) 진술 텍스트를 입력하면 자동으로 몽타주를 생성한다.
2) 구체적 특징 뿐 아니라 전체적인 인상 같은 추상적인 특징도 반영한다.
3) 30 초 이내에 몽타주를 생성한다.
4) 텍스트 부분 수정을 통해 이미지를 수정할 수 있다.

3. 서비스 대상

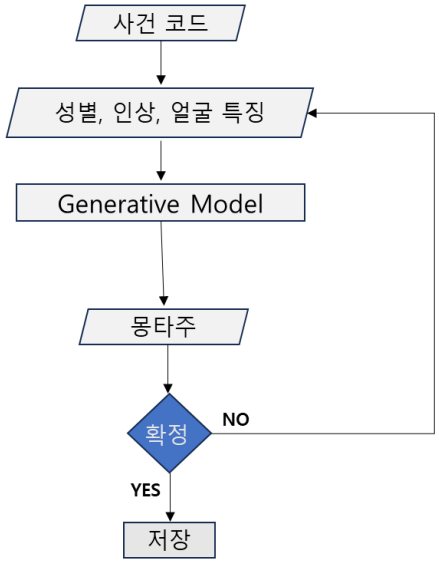
경찰, 검찰 등 범죄 수사기관의 몽타주 제작 전문가를 대상으로 한다.

4. 서비스 주요 기능

1) 몽타주 제작: 성별, 나이대, 얼굴형, 머리스타일, 이목구비, 인상과 느낌을 입력하면 몽타주를 생성한다. 입력 텍스트르 부분 수정하면 그에 따라 이미지를 수정한다.
2) 몽타주 관리: 제작된 몽타주를 데이터베이스에 저장하고 열람할 수 있다.

5. 서비스 시나리오

개략적인 서비스 시나리오는 다음과 같다.



II. 기획

1. 전문가 인터뷰

실제 현장에서의 몽타주 제작 및 관리, 법적 근거 등 효용성 측면에서의 근거를 마련하기 위해 몽타주 제작 전문가와의 인터뷰를 진행하였다. 그 결과 다음과 같은 정보를 얻을 수 있었다.

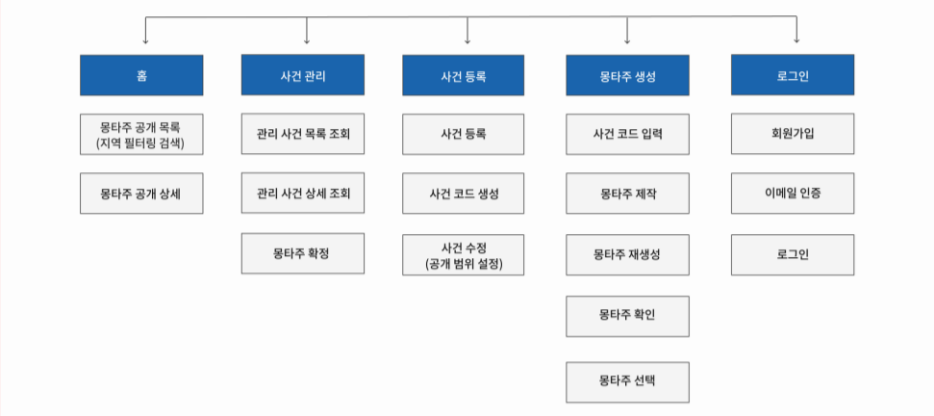
첫번째로 몽타주는 수사 목적으로 제작하기 때문에 아무나 그릴 수 없고 각 시도청에서 자격이 있는 사람만 제작할 수 있다. 두번째로 몽타주는 제작 의뢰가 들어온 사건에 대해서만 제작할 수 있다. 세번째로 몽타주 자체는 인상 표현이지 인물의 실제적 정보가 아니기 때문에 공개해도 신상정보 보호와는 관련이 없다. 네번째로 몽타주가 얼마나 잘 만들어졌는지 정량적 평가를 할 수 없다. 인상을 표현한 추상적인 스케치이기 때문에 실제 사진과의 얼굴 유사도 등의 평가지표를 적용할 수 없다. 그러나 정성적 평가는 할 수 있다.

위의 내용을 서비스 설계 및 평가 과정에 반영하였다.

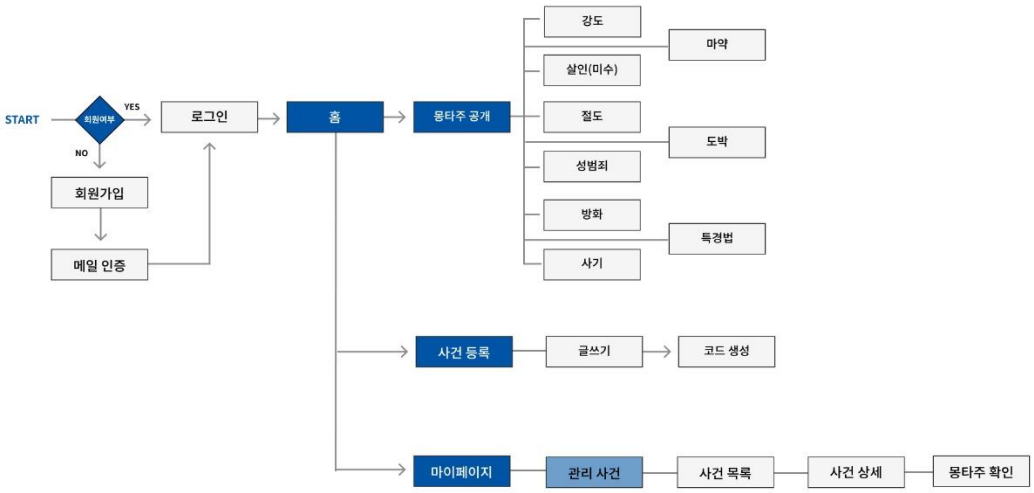
2. MVP

1) 권한 분리: 경찰이 사건을 등록하면 몽타주 제작자는 코드를 통해 사건에 접근
2) 몽타주 공개: 확정된 몽타주가 있는 사건은 공개 설정을 통해 시민들에게 공개
3) 사건관리(경찰): 사건을 등록, 수정하고 시민들에게 공개하는 일련의 과정
4) 몽타주관리(몽타주 제작자): 진술자로부터 입력받은 내용으로 몽타주를 생성하고 선택

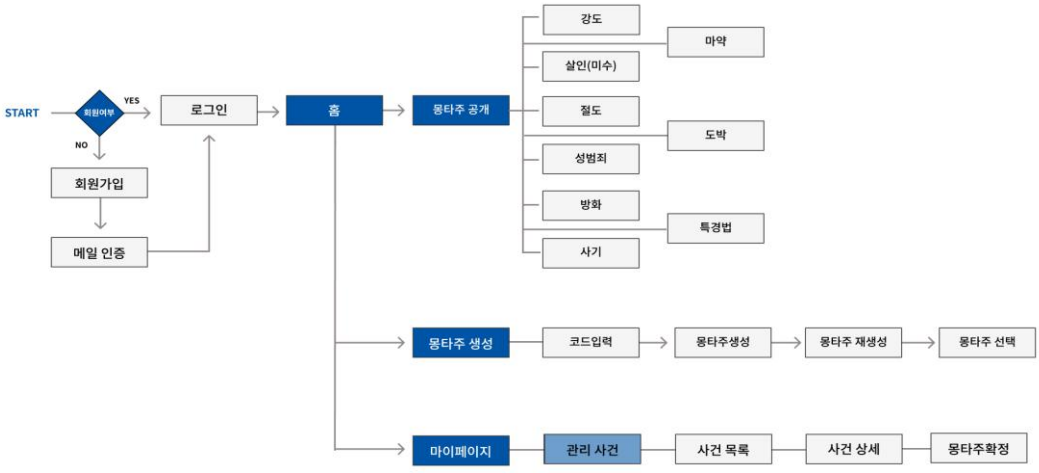
3. IA



4. Flow-chart: 경찰관 ver.



5. Flow-chart: 몽타주 제작자 ver.



III. 설계 및 개발

1. 사용 기술 스택

1) 데이터 파트: Huggingface, Fast API, AWS S3, EC2

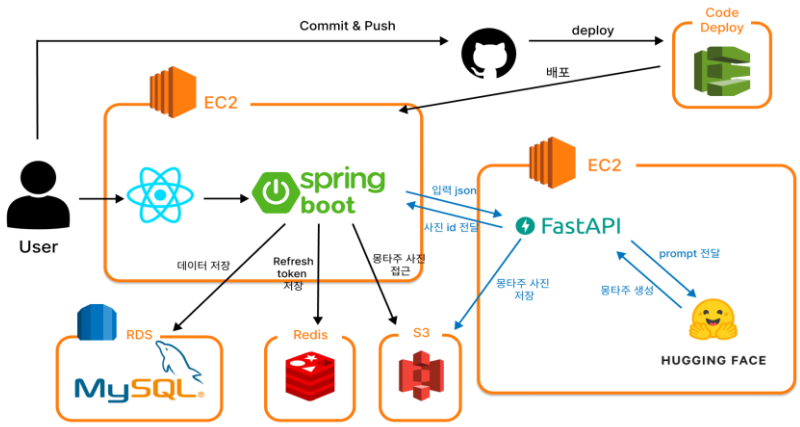
2) 백엔드 파트

a. Spring Boot 3.1.5, Java17, Gradle
b. MySQL, Redis
c. Github Actions
d. AWS S3, EC2, Code Deploy

3) 프론트엔드 파트

a. React.js + TypeScript
b. recoil, react-query
c. storybook, tailwindcss, eslint, husky

2. 시스템 아키텍처



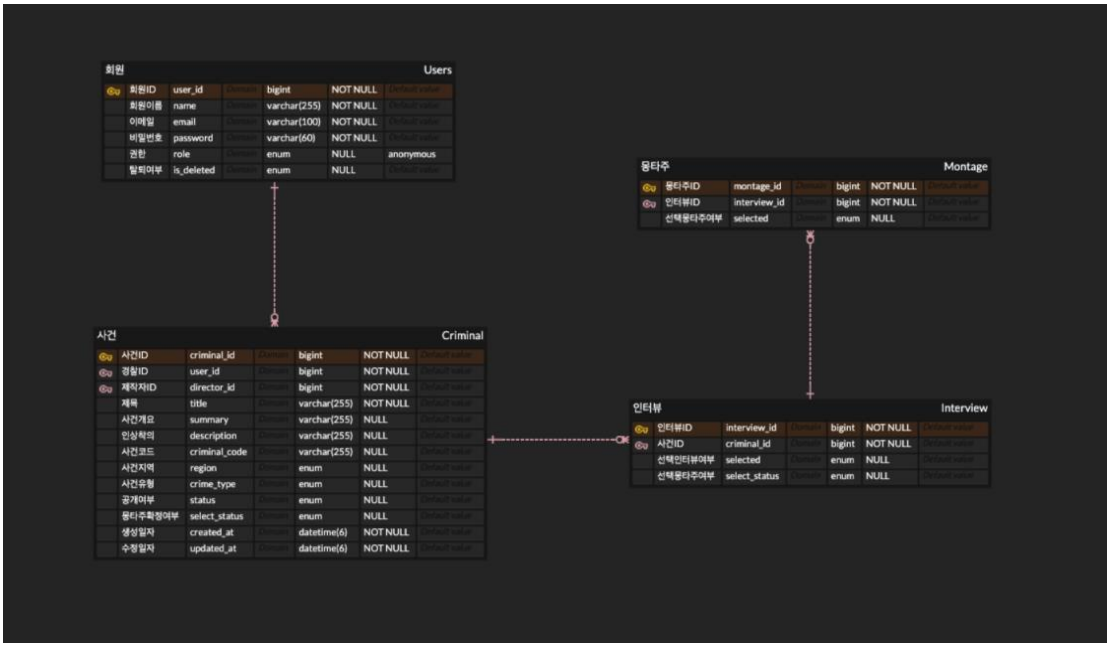
프론트엔드에서는 React.js(+Typescript)를 사용하였다. 전역 상태 관리를 위해 recoil을 사용했고, 아토믹 디자인 시스템을 도입하여 storybook을 활용하였다.

서버와의 통신을 위해 React Query를 사용했습니다. 코드의 품질과 일관성을 유지하기 위해 Tailwind CSS, ESLint, Husky를 도입하였다.

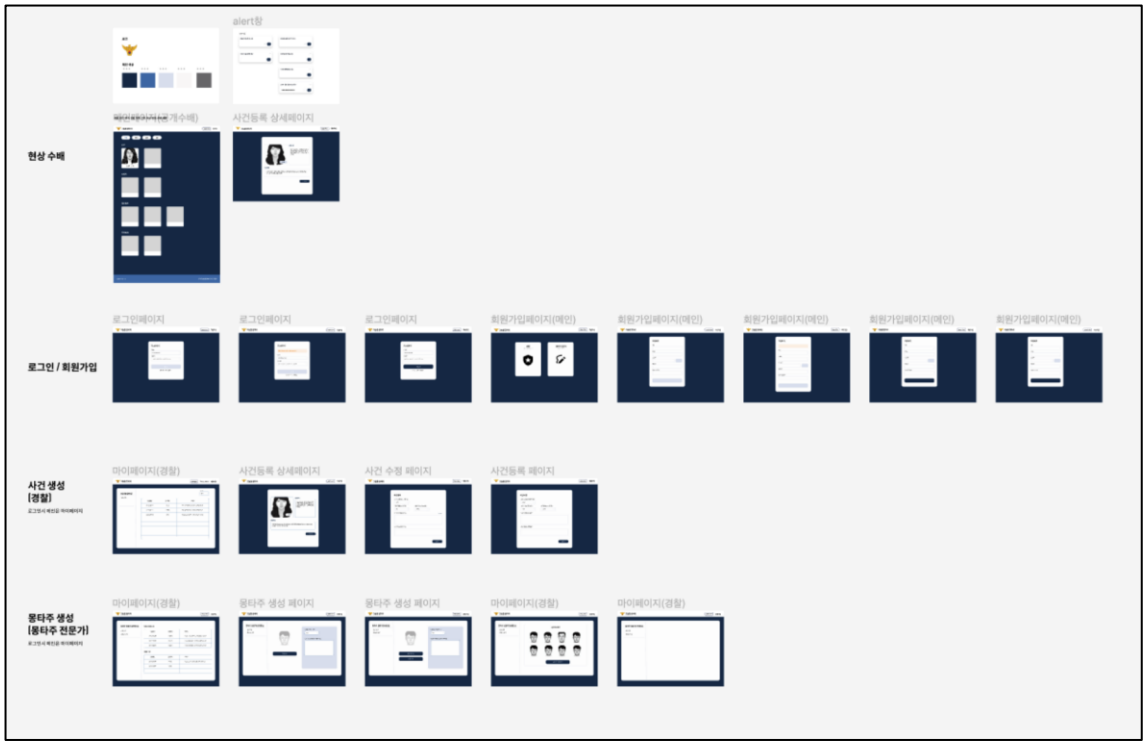
백엔드에서는 웹 서버를 위해 Spring boot로 서비스를 개발하였으며, 데이터베이스로는 MySQL을 사용하고 Refresh Token을 관리하기 위한 데이터베이스로는 Redis를 사용하였다.

또한, Open Feign을 이용해 데이터 모델 서버의 api를 호출할 수 있도록 구현했다. 추가로, Github Actions를 활용한 CI/CD 파이프라인을 구축하여 새로운 기능 개발시 빠르게 테스트 및 배포되도록 하였다.

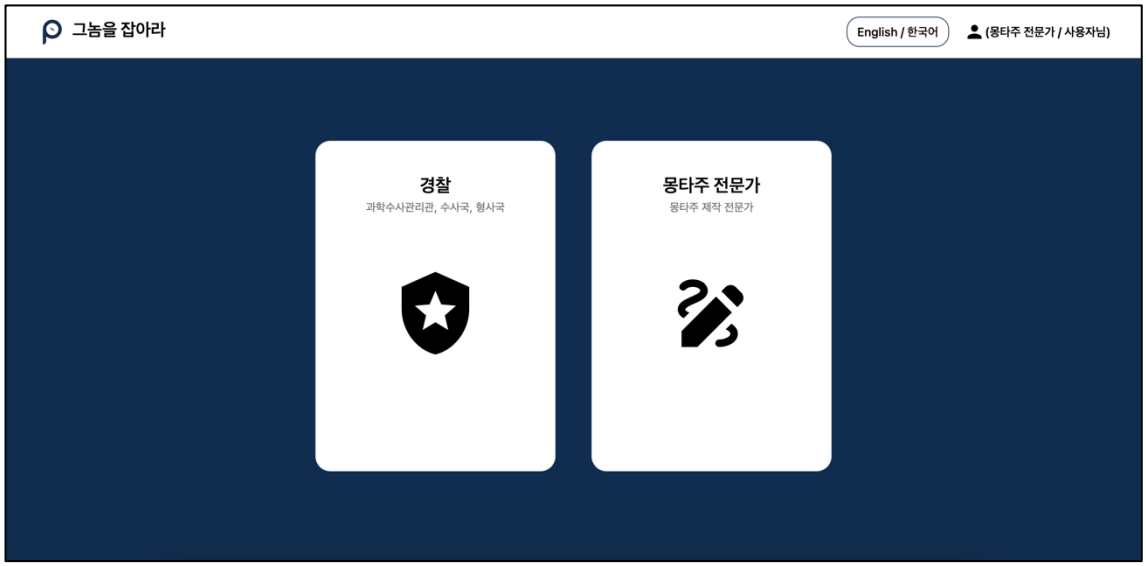
3. ERD



4. 화면 설계



1) 회원가입



회원가입은 경찰과 몽타주 전문가의 권한으로 나뉜다.

경찰은 사건을 등록하고, 사건의 공개범위를 설정할 수 있으며 자신이 담당한 사건목록을 조회할 수 있다.

몽타주 제작자는 경찰에게 부여받은 사건코드를 입력하여 몽타주 생성 페이지로 이동할 수 있다. 몽타주 제작자는 목격자의 진술을 바탕으로 몽타주 이미지를 생성한다. 기존에 생성했던 몽타주들을 불러와 진술자와 논의 후, 최종 몽타주를 확정하면 해당 몽타주가 사건에 자동 등록된다. 자신이 담당한 사건 목록을 조회할 수 있다.

2) 몽타주 공개 페이지

경찰은 사건 공개 여부를 결정하여 몽타주를 공개수배할 것 인지의 여부를 결정한다. 이때 공개된 몽타주들은 로그인 없이 일반시민도 접근해서 확인할 수 있다. 몽타주 공개페이지에서는 지역별로 공개수배된 몽타주를 확인할 수 있고, 각 범죄종류에 따라 모아볼 수 있다.

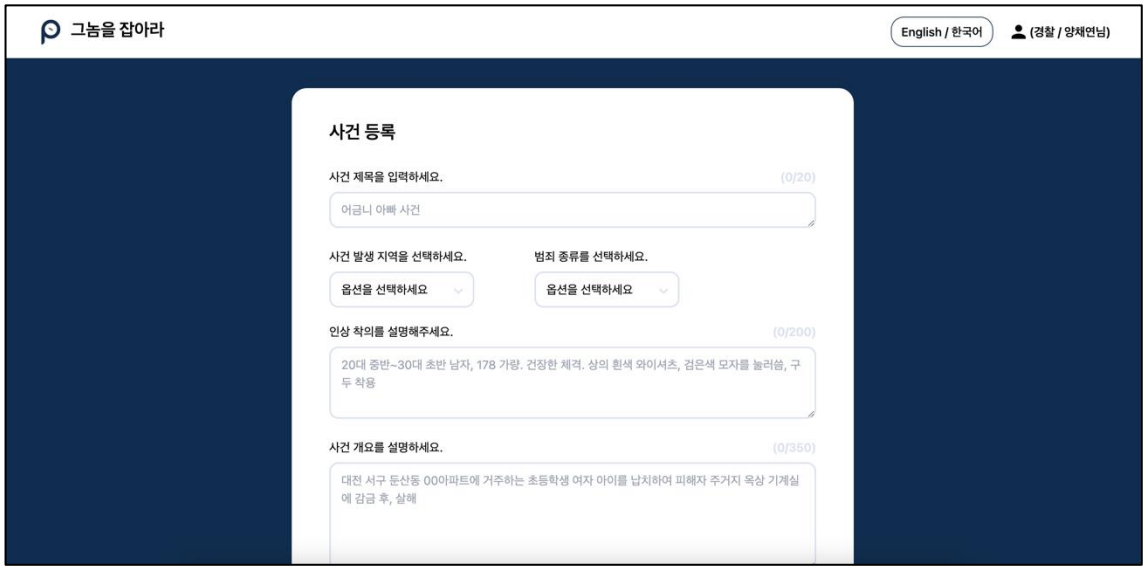
몽타주를 클릭하면 상세한 정보를 확인할 수 있다. 경찰 권한으로 접근했을 경우 사건의 공개범위, 사건 상세 내용에 대해 수정 가능하다.

3) 마이페이지

경찰은 자신이 담당했던 사건 목록, 사건 공개여부를 확인할 수 있다. 사건 번호를 클릭할 시 해당 사건의 상세페이지로 이동한다. 이때, 사건 내용 및 공개여부를 수정할 수 있다.

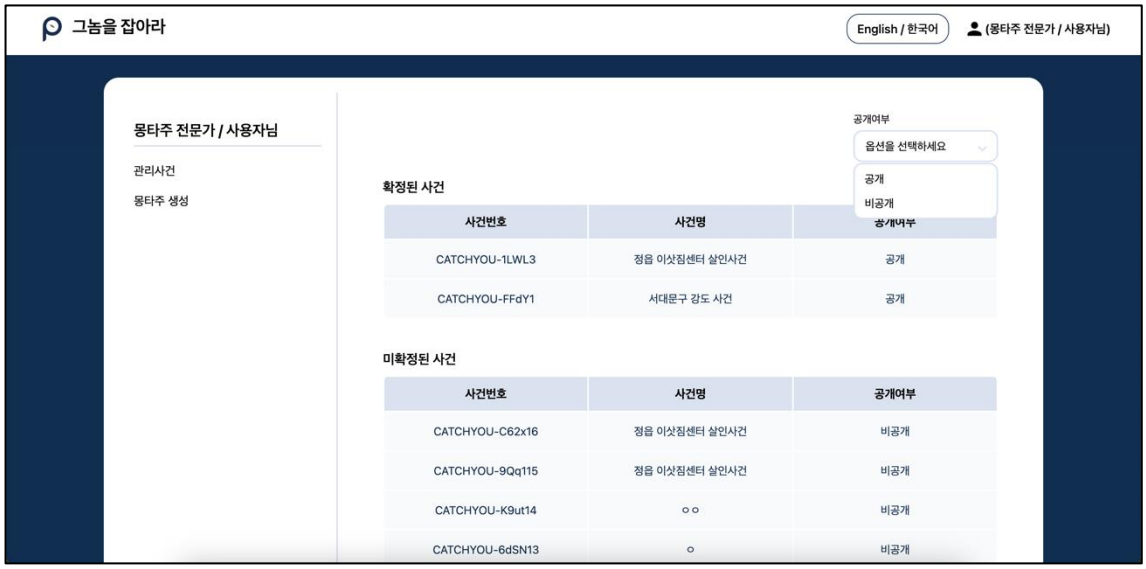
(‘마이페이지’ 계속)

사건 번호를 클릭할 시 해당 사건의 상세페이지로 이동한다. 제작자 배정여부와 게시글 공개 여부를 확인할 수 있다.



경찰은 마이페이지에서 사건을 등록할 수 있다.

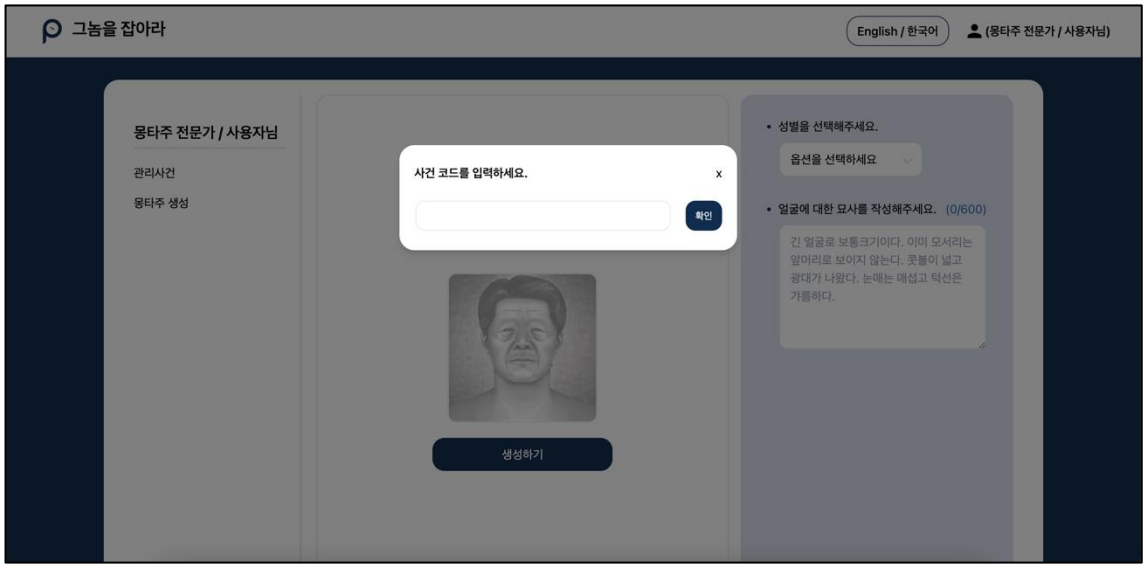
('마이페이지' 계속)



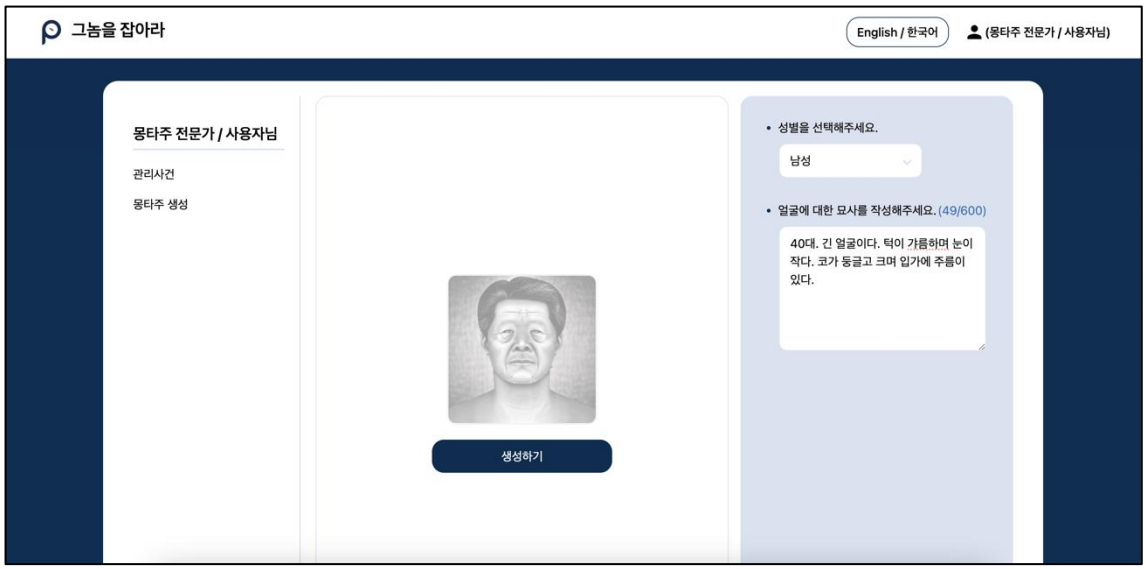
몽타주 제작자는 자신이 담당했던 사건 목록, 사건 공개여부를 확인할 수 있다.

사건 번호를 클릭할 시 해당 사건의 상세페이지로 이동한다. 사건에 한 번 확정된 몽타주는 수정이 불가능하다.

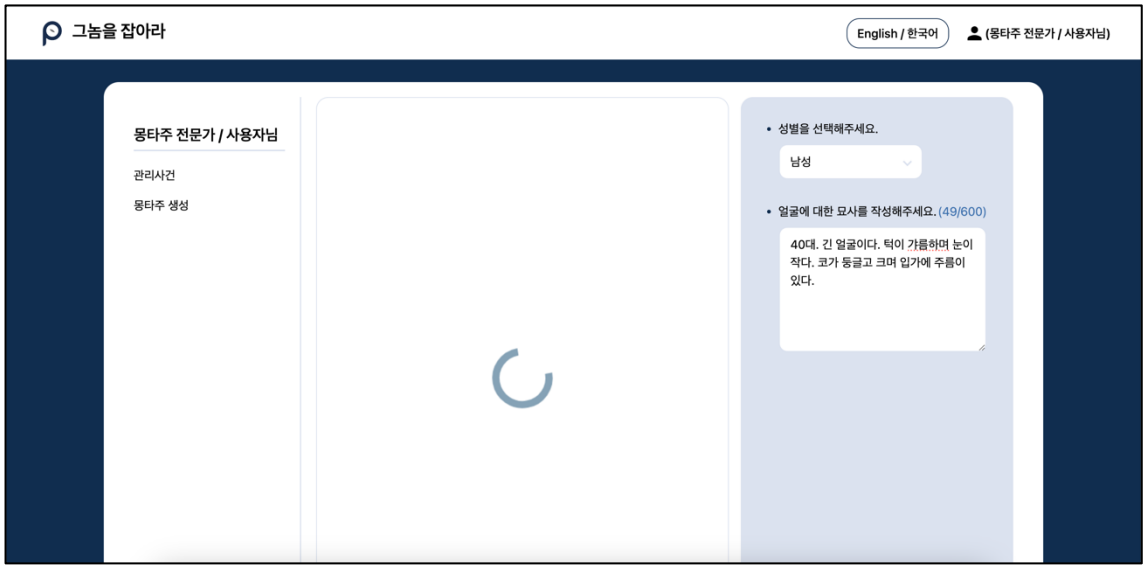
4) 몽타주 생성페이지

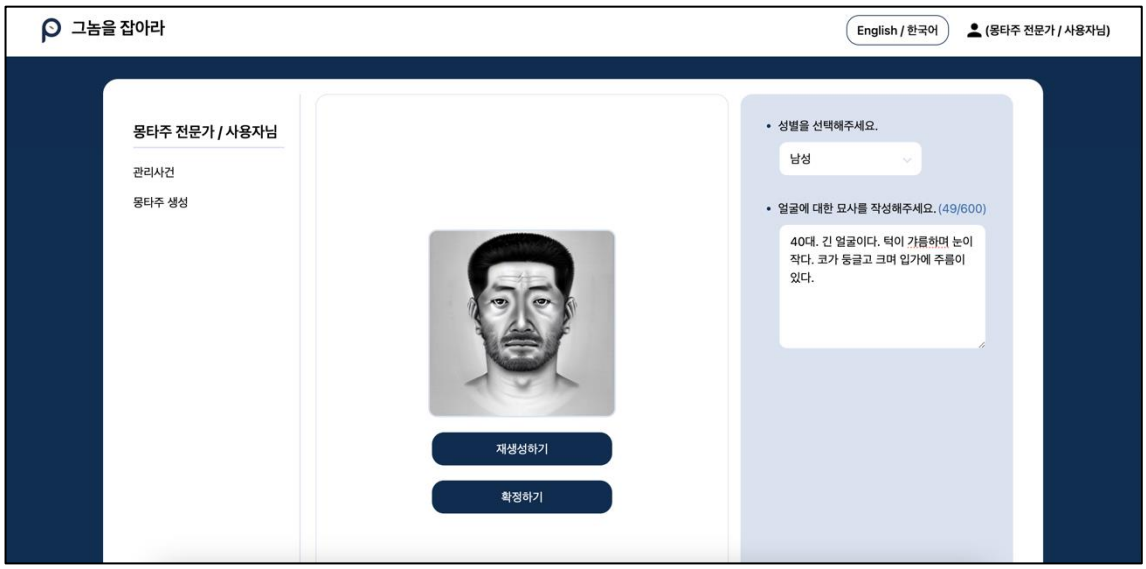


몽타주 제작자는 경찰에게 전달받은 사건 코드를 통해 해당 사건의 몽타주 생성 페이지에 접근한다.



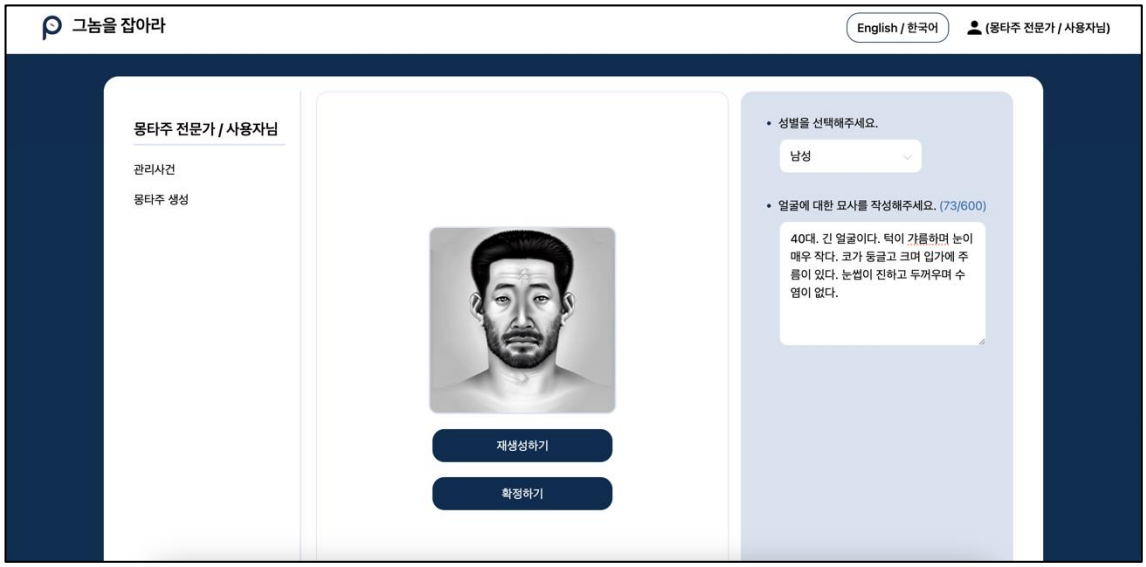
(‘몽타주 생성페이지’ 계속)





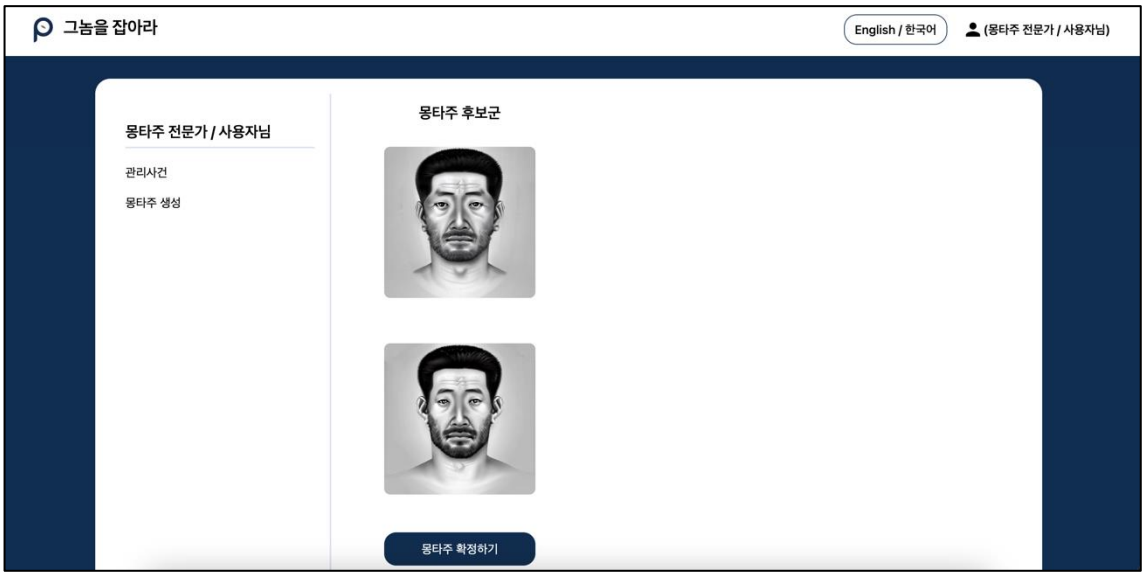
범인의 성별을 선택하고 얼굴에 대한 묘사를 텍스트로 입력한다. 생성하기 버튼 클릭 시 AI 모델이 몽타주를 생성한다.

(‘몽타주 생성페이지’ 계속)



몽타주 재생성을 원할 경우 기존의 텍스트를 수정해 다시 몽타주를 생성할 수 있다. 몽타주 생성을 끝내고 몽타주를 선택하고 싶을 시, 확정하기 버튼을 클릭한다.

5) 몽타주 투표페이지



지금까지 생성한 몽타주를 모두 모아 보여준다. 몽타주 후보군 중 최종적으로 가장 유사한 몽타주를 선택하고 최종 몽타주를 선택한다.

5. 요구사항 명세서

RQ-ID	화면명	요구사항명	권한	요구사항내용
RQ-001	회원 가입	중복 이메일 검사	일반	가입하려는 이메일로 이미 가입된 계정이 있는지 검사한다.
RQ-002		이메일 인증		가입하려는 이메일로 인증 코드를 보내 본인인증을 진행한다.
RQ-003		회원가입		가입 시 실명, 이메일, 비밀번호, 권한(경찰/제작자)를 입력받아 회원을 등록한다.
RQ-004	로그인	로그인	경찰/ 제작자	회원이 가입시 등록한 이메일, 비밀번호로 로그인을 진행한다.
RQ-006	몽타주 공개 페이지	몽타주 공개 글 조회	일반	몽타주 제작이 완료된 공개 사건에 대해서만 조회가 가능하다. 사건 발생 지역, 유형, 인상착의와 사건 개요 및 몽타주를 확인할 수 있다.
RQ-007		몽타주 공개 목록 조회		몽타주 제작이 완료된 공개 사건에 대해서만 조회가 가능하다. 조회시 디폴트 지역 값은 서울이며, 그 외 다른 여러 지역에 대해 필터링해 조회할 수 있다.
RQ-008	사건 관리	사건 등록	경찰	사건 발생 지역, 범죄 유형, 인상착의, 사건 개요를 작성해 사건을 등록한다. 등록 시 시민 공개 범위는 비공개만 가능하다.
RQ-009		사건 수정		등록한 사건에 대해 사건 발생 지역, 범죄 유형, 인상착의, 사건 개요 등을 수정할 수 있다. 수정 시 공개범위를 설정할 수 있으나 몽타주가 확정된 사건에 대해서만 공개로 설정할 수 있다.
RQ-010		사용자의 사건 글 조회		사용자가 등록하고 관리하고 있는 사건에 대해 글을 확인할 수 있다. 이 때, 생성된 사건 코드를 추가로 확인할 수 있다.
RQ-011		사용자의 사건 목록 조회		사용자가 등록하고 관리하고 있는 사건에 대해 목록을 확인할 수 있다.
RQ-012	사건 접근	사건 코드 검사	제작자	등록된 사건에 대해 부여된 사건 코드를 검사하면 제작자가 해당 사건에 접근할 수 있다. 만약, 이미 몽타주가 확정된 사건이거나 담당 제작자가 매칭된 사건일 경우 사건 코드가 올바른 코드일지라도 접근할 수 없다.
RQ-013	몽타주 생성	몽타주 제작		스크립트를 작성하고 생성하기 버튼을 누르면 몽타주가 생성되고 확인할 수 이쌀. 몽타주를 선택하기 전까지 스크립트를 추가하거나 수정함으로써 몽타주를 재생성할 수 있다.
RQ-014		몽타주 선택		현재까지 진술자의 진술 내용을 바탕으로 제작한 몽타주 중 하나를 선택한다. 선택시 현재까지 제작한 몽타주를 한 눈에 확인할 수 있다.
RQ-015	몽타주 관리	사용자의 사건 상세 조회		사용자가 등록하고 관리하고 있는 사건에 대해 글을 확인할 수 있다. 이 때, 생성된 사건 코드를 추가로 확인할 수 있다.
RQ-016		사용자의 사건 목록 조회		사용자가 담당하고 있는 사건에 대해 목록을 확인할 수 있다. 이 때, 몽타주가 확정된 사건/미확정된 사건으로 나누어 조회가 능하다.
RQ-017		몽타주 확정		몽타주 미확정된 사건에 대해 여러 진술자와 함께 선택한 몽타주들 중 하나를 확정할 수 있다. 확정시 현재까지 선택했던 몽타주를 한 눈에 확인할 수 있다.

6. 사용한 데이터셋

가상 인물 몽타주 데이터를 Ai Hub 에서 다운로드하였다. 이 데이터셋은 가상인물 8071명 각각의 이미지, 육안 관찰 인물 스케치, 설명문, 설명문 기반 몽타주 스케치로 구성되어 있다.

가상인물 이미지는 인물의 컬러 사진이다. 육안 관찰 인물 스케치는 흑백 사진에 가까운 세밀 스케치이다. 설명문은 인물의 얼굴 항목별 구체적 특징과 인상을 기술한 데이터이다. 설명문 기반 몽타주 스케치(이하 몽타주)는 인물의 몽타주이다. 설명문과 몽타주는 모두 묘사의 세밀한 정도에 따라 상, 중, 하 세 가지 버전이 존재한다. 따라서 8071명에 대해 24213개의 설명문-몽타주 페어가 존재한다.

본 프로젝트에서는 가상인물 이미지는 제외하고 육안 관찰 인물 스케치, 설명문 그리고 몽타주를 학습에 이용했다.

수집 (수량)		가공 (수량)			
가상인물 이미지	8071	(육안 관찰) 인물 스케치			8071
		설명문 (상)	8071	몽타주스케치(상)	8071
		설명문 (중)	8071	몽타주스케치(중)	8071
		설명문 (하)	8071	몽타주스케치(하)	8071

표1. 해당 데이터셋의 Ai Hub 페이지를 참고한 데이터 통계 표.

7. 사용한 인공지능 모델

입력 텍스트에 맞게 이미지를 생성하기 위해 생성 모델(generative model)을 쓰기로 하였다. 이를 위해 두 가지 generative model인 DALL-E와 Stable Diffusion을 각각 개발하고 성능을 비교하여 한 가지를 채택하였다.

1) DALL-E

DALL-E[1]는 transformer를 기반으로 하는 text-to-image 생성 모델이다. 이 모델은 text를 입력으로 받아 해당 내용을 설명하는 image를 생성할 수 있다. DALL-E를 학습하기 전, DALL-E의 image encoder로 사용될 VQ-GAN[2]을 먼저 학습하였다.

a) VQ-GAN 학습

- 학습 환경: NVIDIA GeForce RTX 3090 24GB GPU 두 대
- hyperparameter: VQ-GAN 논문[2]에서 제시한 기본 설정을 따라 학습함

b) DALL-E 학습

- 학습 환경: NVIDIA GeForce RTX 3090 24GB GPU 한 대
- Text Encoder: hugging face[1]에서 제공하는 klue/roberta-large[3]
- Image Encoder: 앞서 학습한 VQ-GAN
- hyperparameter: 아래 표와 같음

Hyperparameter	
batch size	24
learning rate	3.0e-5
text sequence length	256
depth	16
attention type	full

표2. DALL-E 모델 학습에 사용한 hyperparameter

표2의 hyperparameter를 포함한 그 외의 조건들은 모두 고정하고 epoch 단위로 모델 성능을 확인하며 학습을 세 단계로 나누어 진행했다. 이 과정에서는 특별한 성능 평가 지표를 사용한 것이 아니라 여러 샘플을 보며 판단하였다.

epoch	input image	input text
0~9	인물 스케치, 몽타주 상, 중, 하	인상
10~12	인물 스케치, 몽타주 상	인상
13~19	인물 스케치, 몽타주 상	특징

표3. DALL-E 모델 학습 단계별로 조정한 input

처음에는 인물 스케치와 모든 몽타주 데이터를 활용하여 학습을 시도하였으나, 낮은 품질의 몽타주 중, 하 데이터로 인해 오히려 모델의 성능까지 하락하는 문제가 발생하였다. 이에 따라 이후 학습 단계에서는 인물 스케치와 몽타주 상의 데이터만을 입력 이미지로 사용하였다. 이로써 낮은 품질의 데이터로 인한 모델의 성능 하락을 방지하고자 하였다. 또한, 모델이 단순히 인물의 인상만을 학습하는 것이 아니라, 특징까지 학습함으로써 구체적인 묘사도 반영하는 모델을 구축하고자 노력하였다.

---

[1] 데이터 사이언스와 머신 러닝을 위한 오픈 소스 플랫폼

2) Stable Diffusion

Stable Diffusion 은 Latent Diffusion Model(LDM)[4] 의 버전들 중 하나이다. LDM 은 text to image 생성 모델이다. Diffusion Model 을 메인 아키텍처로 가지고 있고, 그 앞뒤에 Auto Encoder 와 Decoder 를 추가로 가지고 있다. 이 프로젝트에서는 사전 학습된 Stable Diffusion 모델을 fine tuning 하여 사용했다. 컴퓨팅 자원 절약을 위해 사전 학습된 모델의 일부 파라미터를 조정하여 fine tuning 을 수행하는 Low-Rank Adaptation of Large Language Models(LoRA)[5] 방식을 따랐다.

- 학습 환경: Google Colab 의 NVIDIA T4 GPU, 16GB Memory, 500GB SSD
- Base Model: huggingface 에 공개된 Bingsu/my-korean-stable-diffusion-v1-5
- hyperparameter: 아래 표와 같음.

Hyperparameter	
random flip	FALSE
center crop	TRUE
train batch size	1
gradient accumulation steps	4
learning rate	1.00E-04
learning rate scheduler	"cosine"

표4. LoRA를 적용한 Stable Diffusion 학습 hyperparameter

위의 hyperparameter를 포함한 그 외의 조건들은 모두 고정하고 VQ-GAN 학습 시와 마찬가지로 몇 epoch 단위로 모델 성능을 확인하며 학습을 세 단계로 나누어 진행했다. 이 과정에서는 특별한 성능 평가 지표를 사용한 것이 아니라 샘플 여러 가지를 뽑아 판단하였다. 성능이 더 이상 개선되지 않는다고 판단하였을 때 학습을 중단하였다.

epoch	image	input text	resolution
0 - 8	몽타주 상, 중, 하	인상	256
9 - 23	몽타주 상	특징	256
23 - 33	몽타주 상	인상 + 특징	512

표5. Stable Diffusion 학습 단계별로 조정한 파라미터.

IV. 모델 성능 평가 및 선정

DALL-E, Stable Diffusion 모델의 성능을 비교하기 위해 test data 로부터 임의로 9개의 텍스트를 뽑았다. 구체적 텍스트 입력과 추상적 텍스트 입력의 경우를 종합적으로 평가하기 위해 3개는 구체적인 표현만으로, 다른 3개는 추상적인 표현만으로, 마지막 3개는 구체적 표현과 추상적 표현의 혼합으로 구성하였다. test data의 개수는 당초 30개로 구상하였으나 사용자 평가 시 설문자의 피로를 우려하여 9개로 축소했다.

다음은 성능 평가에 이용된 샘플 이미지로, 두 모델에 제시된 텍스트를 각각 입력하여 생성하였다. 그림 1은 구체적 텍스트 입력 결과 이미지이고, 그림 2는 구체적, 추상적 표현 혼합 텍스트 입력 결과 이미지이다.

그림 1. 구체적 텍스트 입력 결과 이미지(좌 DALL-E, 우 Stable Diffusion)

"남성. 50대. 얼굴은 둥글고 턱은 둥근형이다. 광대가 나왔다. 짧은 머리이고 눈썹이 흐리고 미간은 넓다. 작은 눈에 코가 크고 입술은 얇다. 눈가주름이 있다. 팔자주름이 있다."

그림 2. 구체적, 추상적 표현 혼합 텍스트 입력 결과 이미지 (좌 DALL-E, 우 Stable Diffusion)

"남성. 30대. 얼굴은 둥글고 턱은 둥근형이다. 볼이 통통하다. 짧은 머리이고 왼쪽가르마를 탔다. 눈썹이 흐리고 미간은 좁다. 인중은 길다. 팔자주름이 있다. 깔끔하게 내려온 짧은 머리와 굴곡없는 동그란 얼굴은 깔끔하고 자기 관리를 잘하는 사람으로 보이며 표정이 차분해 보여 진지하고 꼼꼼한 이미지로도 느껴진다. 매사에 침착하고 성실한 사람으로도 보인다."

평가 지표로는 다음의 두 방법을 사용했다.

1. TIFA(인용 삽입)

TIFA는 text-to-image 모델의 성능을 평가하기 위한 지표로 생성된 이미지에 텍스트가 얼마나 잘 반영되었는지를 0부터 1의 스케일로 나타낸다. TIFA는 입력 텍스트로부터 채점 기준이 될 질문들을 생성한다. 그리고 생성된 이미지를 인식하여 각 질문의 요구사항을 충족하는지를 1 (예)또는 0 (아니오)으로 채점하고 점수들의 평균을 최종점수로 제출한다.

2. 사용자 평가

사용자 평가는 생성된 이미지에 텍스트가 얼마나 잘 반영되었는지를 사용자가 직접 1부터 5의 스케일로 답변한다. 설문지 형식으로 구성하였으며 참여인원은 26명이다. 결과는 다음과 같았다.

	TIFA	사용자 평가
StableDiffusion	0.50 / 1	3.57 / 5
DALLE	0.53 / 1	2.92 / 5

표6. 두 모델의 성능 평가 결과

TIFA 로는 DALLE 가 근소한 차이로 앞서는 듯 했으나 사용자 평가에서는 Stable Diffusion 이 유의미한 차이로 더 나은 결과를 보였으며 GPU 환경에서 30초 안에 이미지 생성이 가능하였다. 이에 따라 Stable Diffusion 을 최종적으로 채택하였다.

다음은 본 프로젝트에서 fine tuning 을 완료한 Stable Diffusion의 이미지 생성 결과 예시이다. 프롬프트 텍스트의 일부가 수정되었을 때 이미지에 반영되는 모습을 보여준다. 그림 3-1 의 프롬프트를 기준으로 일부 단어만 수정하여 이미지를 생성하였다.

그림 3-1. *"남성. 40대. 얼굴은 사각형이고 턱은 각진형(사각)이다. 광대가 나왔다. 짧은 머리이고 왼쪽가르마를 탔다. 눈썹이 진하고 작은 눈에 코가 크고 인중은 길다. 입이 작다. 팔자주름이 있다. 관리하지 않은 헤어나 피부에서 상당한 피로감이 보인다. 째려보는 듯 째진눈에서 날카로운 통찰을 할 것 같은 느낌이 있고, 형사또는 건축의 총괄을 할 것 같은 차분하면서 치밀한 성격의 이미지이다. "*

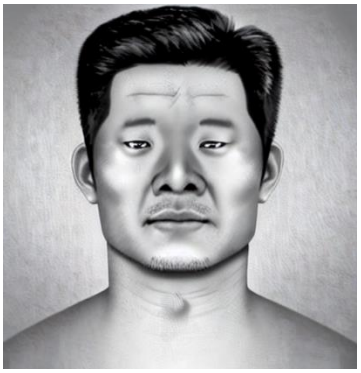


그림 3-2. *"40대" → "20대"* 로 수정

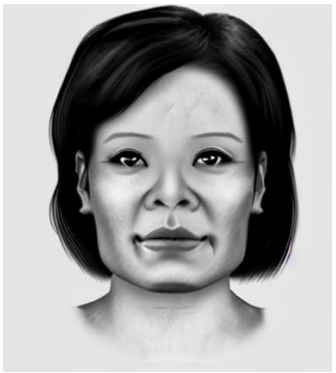


그림 3-3. *"남성" → "여성"* 으로 수정

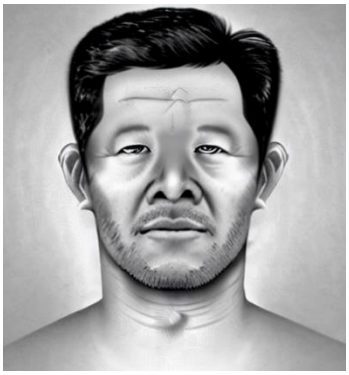


그림 3-4. 얼굴 *"사각형"→"계란형", 턱 "각진형" → "둥근형"* 으로 수정

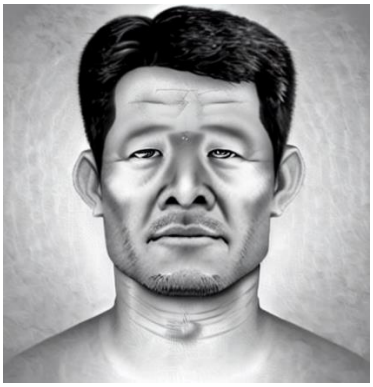


그림 3-5. *"왼쪽가르마" → "오른쪽가르마"*로 수정

그림 3-6. *"쌍꺼풀이 있다"* 추가

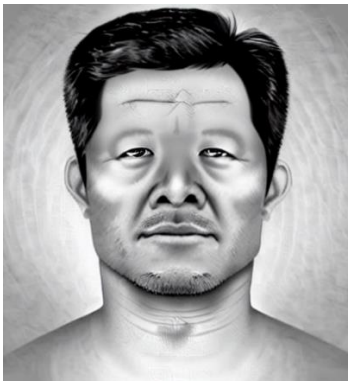


그림 3-7. *"눈썹이 진하고" → "흐리고"* 로 수정

이러한 Stable Diffusion 모델을 가지고 구현을 완료하였다. 다음은 몽타주 생성 화면을 테스트 해 본 모습 이다.

그림 4. 구현 완료 후 몽타주 생성 테스트 화면

V. 결론

본 프로젝트에서는 Generative Model을 활용하여 입력 텍스트로부터 몽타주를 생성하는 프레임워크를 개발하였다. 이를 위해 DALL-E와 Stable Diffusion을 기반으로 하는 두 가지 text to image 모델을 개발하고 각각의 성능을 비교 분석하였다. 그 결과, Stable Diffusion 모델이 전반적으로 더 효과적인 성능을 보였다.

본 프로젝트를 통해 개발된 몽타주 제작 프로그램은 한국 수사 현장의 몽타주 제작 과정과 같이 회상 원리에 기반하되, 구체적인 진술 뿐만 아니라 추상적인 진술도 반영할 수 있는 가능성을 제시하였다. 또한, 30초 이내에 몽타주 생성을 가능하게 함으로써 효율성을 향상시킬 수 있는 가능성을 제시하였다. 따라서 본 연구는 진술을 다각적으로 반영하여 몽타주 제작에 도움을 줄 수 있는 방법을 제시한 점에서 의의가 있다. 이는 기존의 몽타주 제작 방식의 장점을 유지하면서도 범죄 수사에 보다 효과적으로 활용될 수 있는 몽타주 제작 프로그램을 개발하는 데 기여할 것으로 기대된다.

그러나 Generative Model의 특성상 때로 무너진 이미지가 생성될 수 있는 점, 긴 텍스트의 경우 모든 내용을 완벽하게 반영하지 못할 수 있는 점, 텍스트 수정에 따른 이미지 수정 기능이 완벽하지 않다는 점에서 한계가 있다.

VI. 참고문헌

[1] Ramesh, Aditya, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. "Zero-shot text-to-image generation." In International Conference on Machine Learning, pp. 8821-8831. PMLR, 2021.

[2] Esser, Patrick, Robin Rombach, and Bjorn Ommer. "Taming transformers for high-resolution image synthesis." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 12873-12883. 2021.

[3] Park, Sungjoon, Jihyung Moon, Sungdong Kim, Won Ik Cho, Jiyoon Han, Jangwon Park, Chisung Song et al. "Klue: Korean language understanding evaluation." arXiv preprint arXiv:2105.09680 (2021).

[4] Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. "High-resolution image synthesis with latent diffusion models." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10684-10695. 2022.

[5] Hu, Edward J., Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. "Lora: Low-rank adaptation of large language models." arXiv preprint arXiv:2106.09685 (2021).