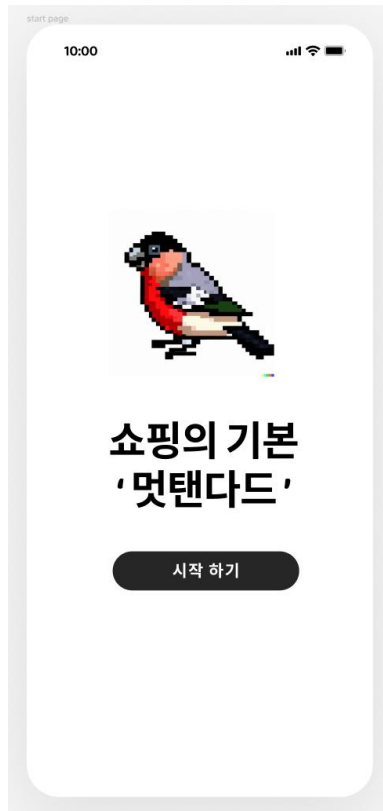


멋탠다드

CV-2조

팀 멋쟁이 | 김성한, 박수영, 정호찬, 이채원, 이다현

INDEX



1. Project Overview

- 팀원소개
- 주제 선정 배경, 서비스 기대효과
- 서비스 디자인 스케치(Figma)

2. Data

- 검색용 데이터, 학습용 데이터

3. Model

- Segmentation
- Retrieval
- Virtual-Try-On
- 경량화

4. Service Architecture

- Service Architecture
- Service Architecture (Seg, Retrieval, Vit-On)

5. Result / Conclusion

- Demo
- Self-assessment
- Todo list

1. Project Overview

-
1. 팀소개
 2. 주제 선정 배경 및 기대효과
 3. 프로젝트 타임라인

팀원 소개 및 역할



김성한

- 전체 서비스 아키텍처 설계 / Retrieval BE & DB 설계 / CLIP 모델 경량화



박수영

- Retrieval, Try-on FE / Virtual Try-On BE 구현



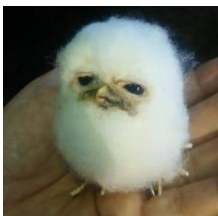
정호찬

- Data 크롤링 / Virtual Try-On 모델링



이채원

- Segmentation BE/FE 구현 / SAM모델 경량화



이다현

- Segmentation BE/FE 구현 / SAM 모델 경량화/ Image 유사도 측정 모델 비교

1.1 Project Overview – 주제 선정 배경



1. 옷 정보를 물어보거나
2. 직접 검색해보거나
3. 정보가 부족해 포기하거나

1.1 Project Overview – 주제 선정 배경



페이지를 찾을 수 없습니다

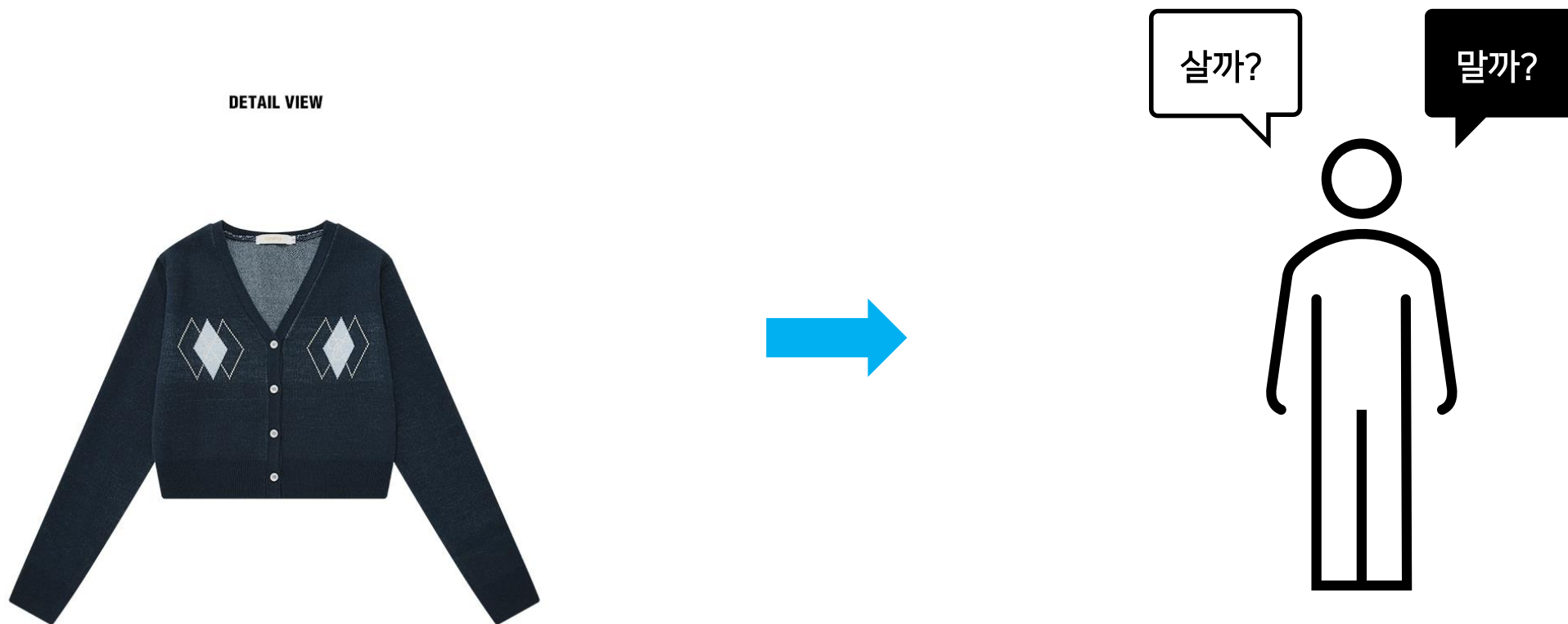
페이지의 주소가 잘못 입력되었거나,
변경 혹은 삭제되어 요청하신 페이지를 찾을 수 없습니다.
입력하신 페이지 주소를 다시 한번 확인해 주세요.

무신사 홈

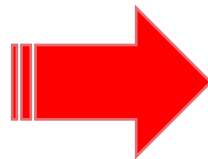
이전 페이지



1.1 Project Overview – 주제 선정 배경



1.1 Project Overview – 주제 선정 배경

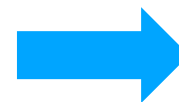


 *Meotandard* (멋탠다드)

유사 상품 검색



가상 피팅



SNS에서 발견한 옷에 대한 정보 부재와 온라인에서의 스타일 판단 어려움을 해결하는 서비스.

✓ 기대 효과

사용자, 소비자 측면

찾고 싶은 옷과 유사한 옷을 비교적 쉽게 찾을 수 있으며, 가상 피팅을 통해 자신의 스타일과 어울리는지 판단 가능

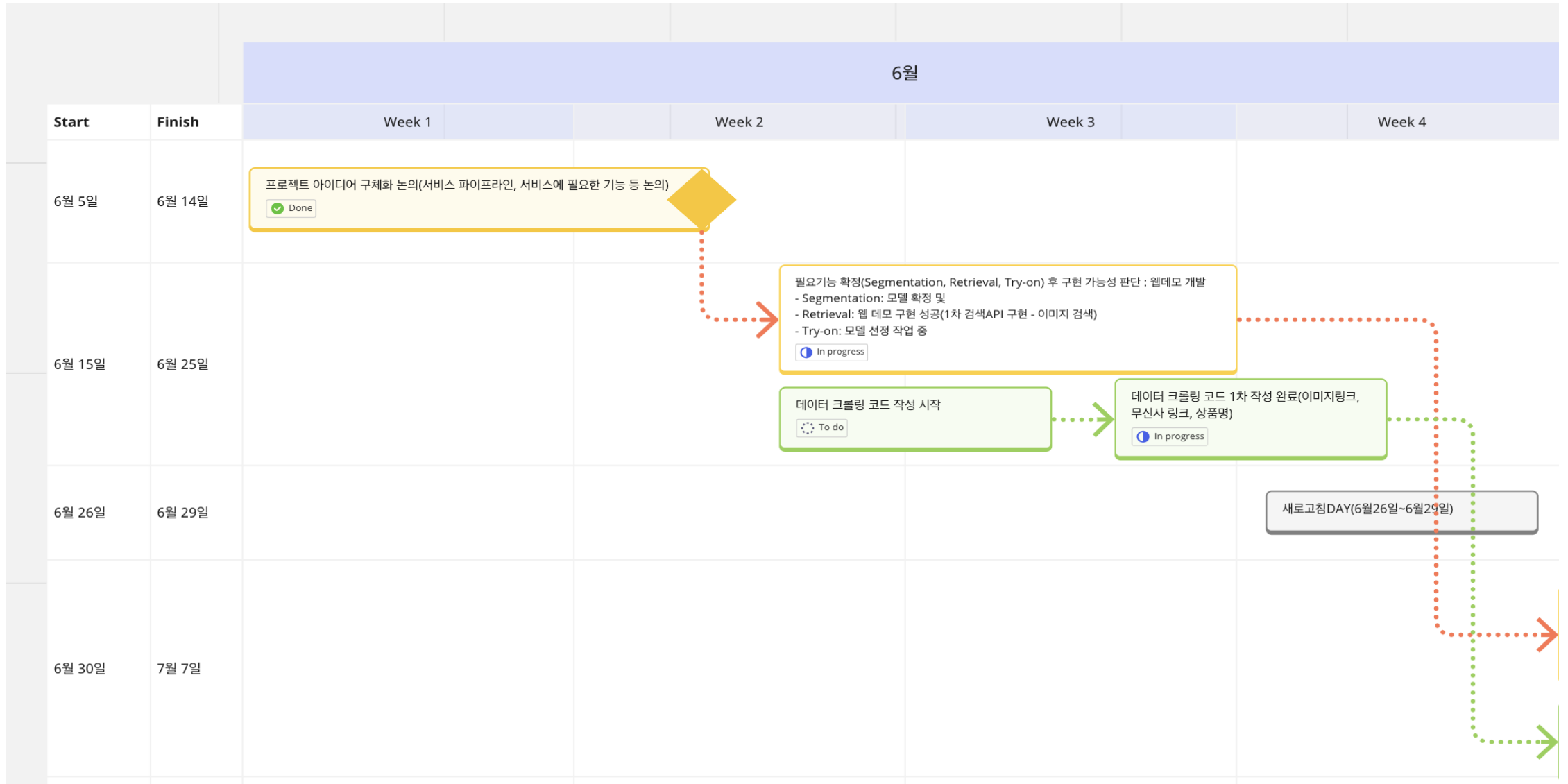
→ 온라인 환경에서 원하는 옷을 합리적으로 구매해 만족도 상승

공급 업체(온라인 쇼핑몰 등) 측면

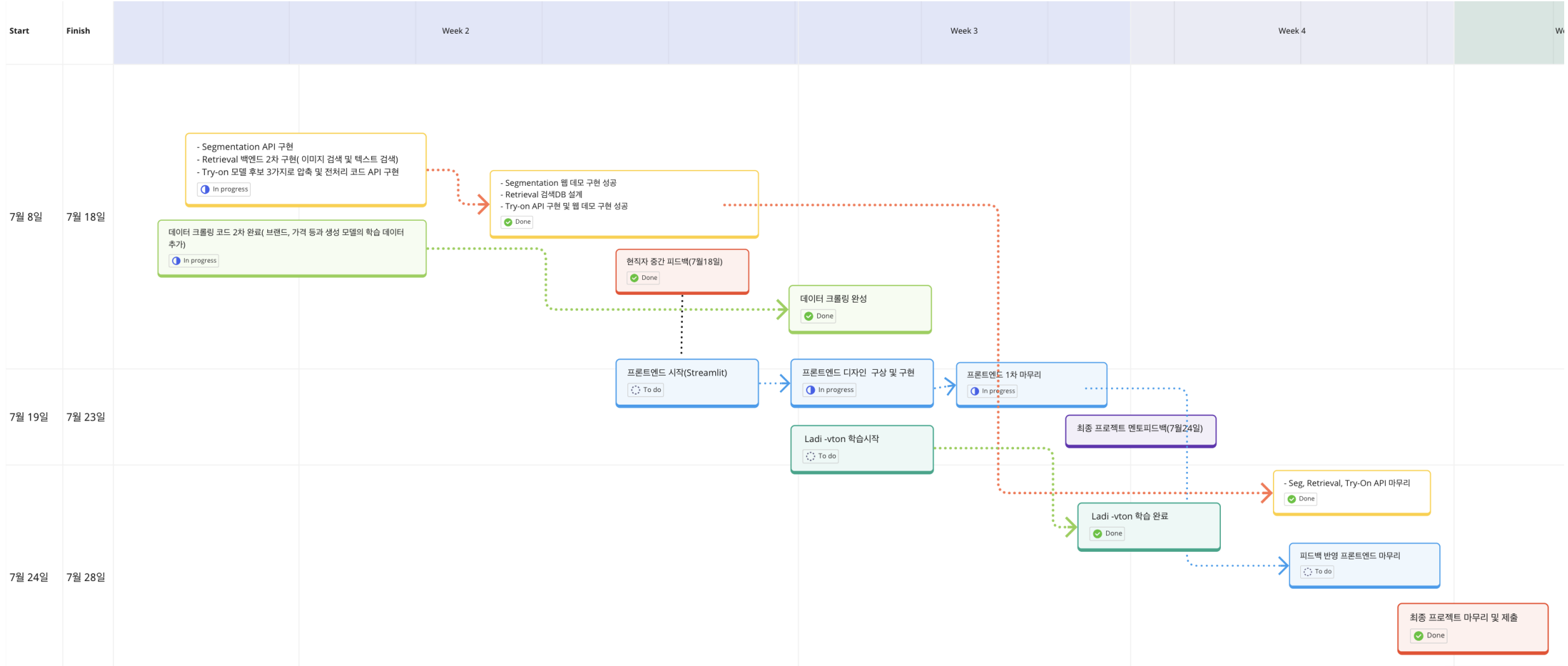
→ 유사 상품 검색 : 품질로 인한 고객 이탈 방지, 체류시간 증대

→ 가상피팅 : 구매에 대한 의사 결정이 어려웠던 잠재적인 **고객들을 확보**할 수 있을 것으로 예상함.

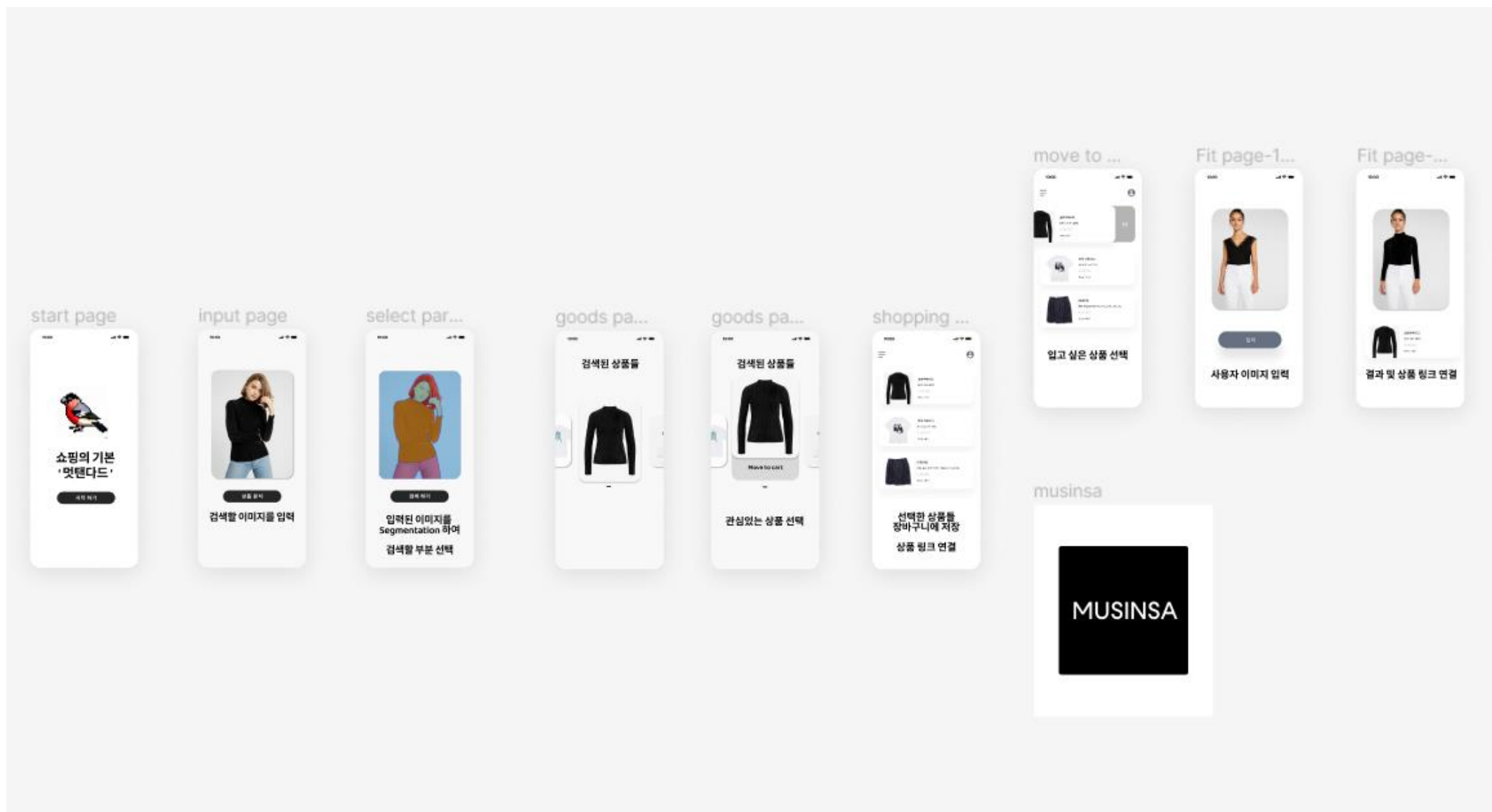
프로젝트 타임라인(1)



프로젝트 타임라인(2)



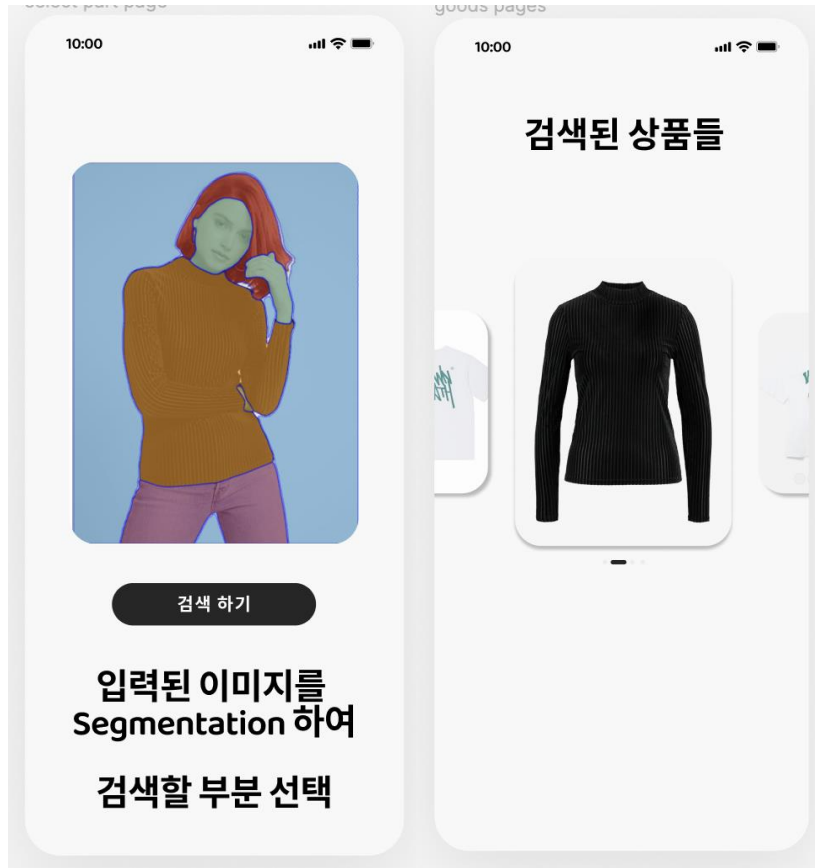
1.3 Usecase Diagram – 서비스 디자인 스케치(Figma)



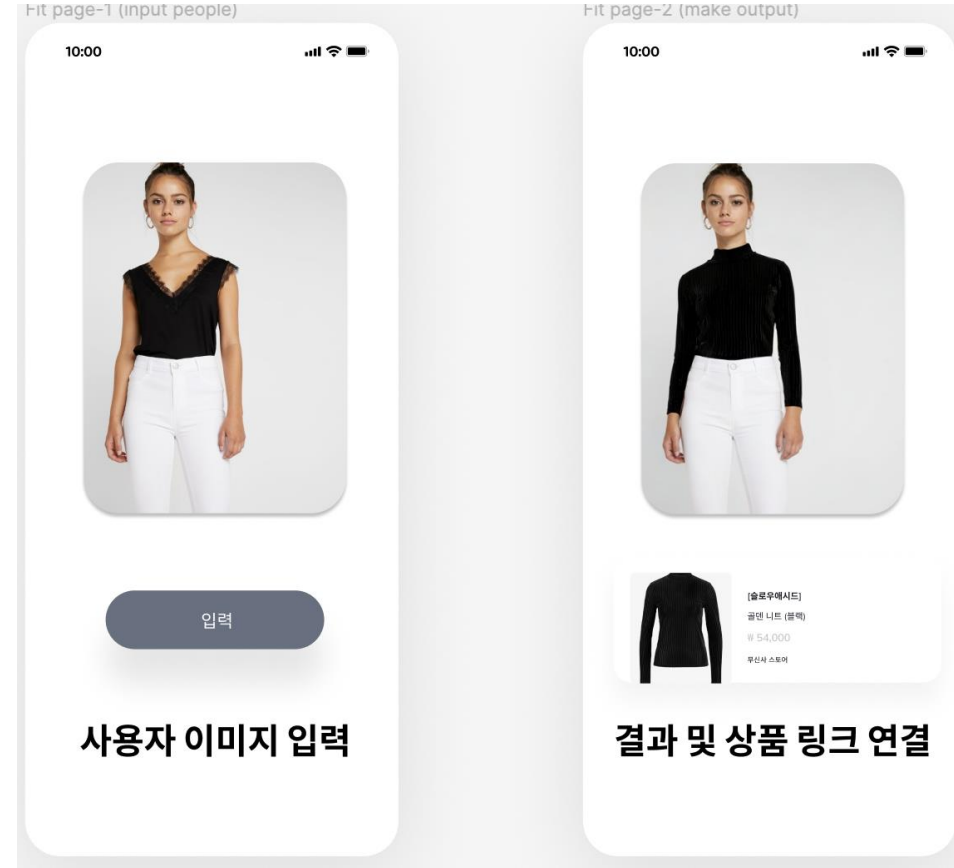
1.3 Usecase Diagram – 서비스 디자인 스케치(Figma)



1.3 Usecase Diagram – 서비스 디자인 스케치



Segmentation & Retrieval(검색)



Virtual Try-On

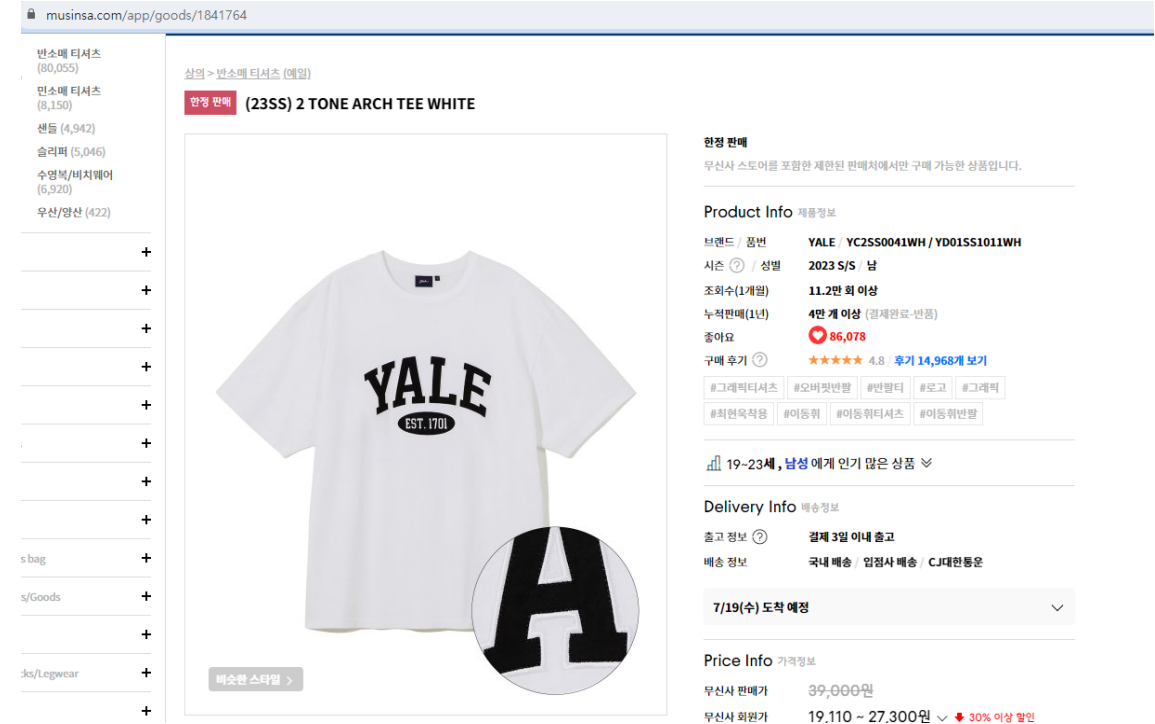
2. Data

-
1. 검색용 데이터
 2. 생성모델 학습용 데이터

2.1 Data - 검색용 데이터

1. 검색용 데이터

수집한 데이터	- 상/하의 카테고리에 있는 상품 사진 및 정보
수집방법	- 무신사 카테고리별 추천순으로 수집 - 상의 36000장, 하의 9000장
데이터 수집 이유	- 이미지 또는 텍스트로 검색된 상품 결과 제시
수집한 데이터 구성	- 상품 번호, 상품명, 상품 이미지 & 링크, 브랜드, 가격, 태그
라벨링 자동화	- 수집한 데이터 중 1000장을 수동 라벨링하여 라벨링 자동화 모델(별첨 2.)을 학습 - 이를 활용하여 제품의 생성 가능여부를 자동으로 라벨링



2. 학습용 데이터

수집한 데이터

- 모델 사진 및 해당 모델이 착용하고 있는 제품사진

수집방법

- 무신사 '코디 숏', '브랜드 스냅' 데이터 크롤링
- 상의-하의 카테고리에 포함되는 약 25000장 데이터 수집

데이터 수집 이유

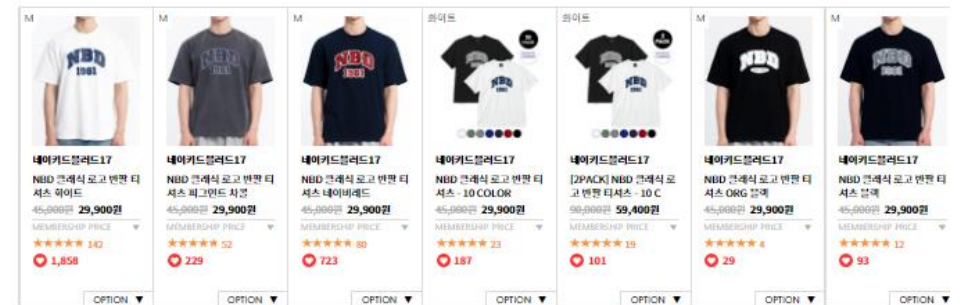
- 생성모델 추가학습용

학습 데이터셋 구성

- 자동 라벨링 모델을 활용하여 25000장 데이터에서 생성 가능한 이미지를 자동 라벨링하여 학습가능한 6000장 데이터셋 구성



Musinsa Store 스토어 관련상품

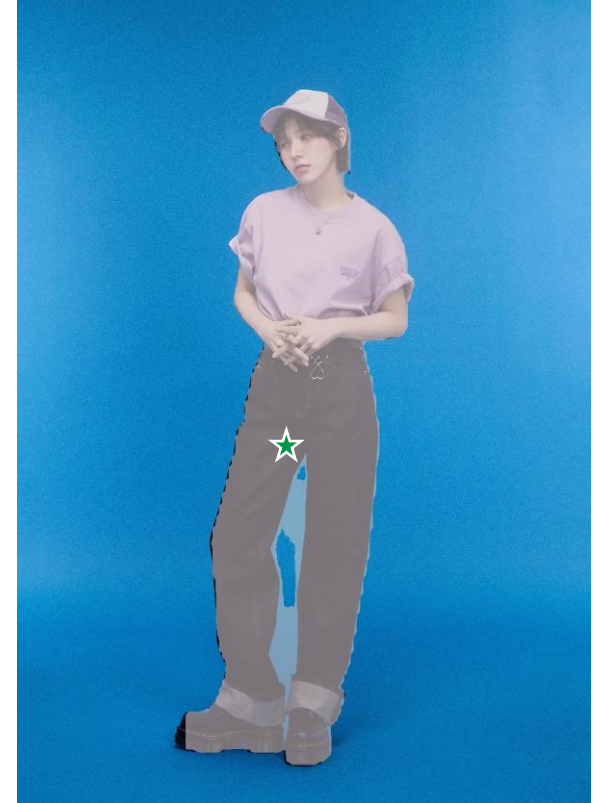


3. Model

-
1. Segmentation
 2. Retrieval
 3. Virtual Try-On
 4. 경량화

3.1 Model - Segmentation - 실험

- 모델 후보 : SAM(Segment Anything Model), Fast SAM
- 정성 평가기준 : 사용자가 타겟으로 하는 하나의 의류만 정확하게 segment 했는가?



(실험1) 상의 - 좌표(450,450)

(실험2) 하의 - 좌표(450, 750)

- 같은 좌표에 대한 Segmentation 결과 비교(실험 예시1, 2)

3.1 Model - Segmentation - 모델 선정

- Segment Anything Model (SAM) 채택이유
 - 검색 품질에 큰 영향을 미침 → Inference 속도보다 **높은 정확도** 중요
 - 사용자가 마우스로 클릭한 point prompt에 해당하는 Segment mask결과 중 score가 가장 높은 상위 3가지 제시 → **사용자에게 선택권 부여**

Mask 1, Score: 0.979



Mask 2, Score: 0.993



Mask 3, Score: 0.990



Mask 추출
좌표 추출



Crop



최종 output

3.1 Model – Segmentation - 모델 선정

Model	장점	단점
1. SAM	<ul style="list-style-type: none">- Ambiguity 특성을 반영해 3가지의 segment mask가 생성되므로 선택 가능- 타 모델들에 비해 정확성이 높음	<ul style="list-style-type: none">- ImageEncoder가 Transformer 기반의 vit 모델이므로 FastSAM보다 속도가 느림
✓ 최종 모델		
2. SAM(onnx)	<ul style="list-style-type: none">- SAM 모델을 onnx 변환과 quantization을 통해 경량화(Inference 1s → 0.5s)	
3. Fast SAM	<ul style="list-style-type: none">- 인퍼런스 시간 SAM에 비해 50배 단축	<ul style="list-style-type: none">- 정확성이 SAM보다 떨어짐.

3.2 Model - Image Retrieval - 실험

- Embedding 모델 후보 : Multi Modal - CLIP, ResNet50, EfficientNet b3
- 정성 평가기준 : 특징이 뚜렷한 옷의 사진을 넣었을 때 유사도 기준(Cos_sim) 기준 상위 10개의 결과에 대해서 색만 다른 같은 상품이 검색이 되는가?
- 테스트 데이터셋: 15개 / 9000개



CLIP - 5 / 10



EfficientNetb3 - 2 / 10



ResNet50 - 1/10



3.2 Model - Image Retrieval - 모델 선정

- CLIP 채택이유
 - 유사 상품 검색 성능 비교 → CLIP > EfficientNet b3 > Resnet50 : 유사 상품 검색 품질 우수
 - 패션 도메인 파인 튜닝 CLIP → CLIP보다 성능 우수 (10/10) → [패션 도메인에 파인튜닝된 오픈소스 모델](#) 도입
→ 추후 한국 의류 상품으로 파인튜닝 가능
 - 멀티모달 모델로 검색 시 **텍스트 활용** → 사용자의 **검색 의도 반영** 가능(기존 이미지 검색 서비스 개선점)



CLIP - 5 / 10



패션 도메인 파인 튜닝 CLIP - 10 / 10

3.4.1 Project Status – Virtual Try-On

- 모델 후보 : HR-VITON, GP-VITON, Ladi-VTON
- 정량 평가기준 : VITON-HD dataset에 대한 Metric 평가

Model	LPIPS ↓	SSIM ↑	FID ↓
HR-VITON	0.061	0.878	9.90
GP-VITON	0.0799	0.8939	9.197
Ladi-VTON	0.049	0.928	13.40

<LPIPS, SSIM, FID 등의 여러 Metric을 활용하여 모델 평가>

LPIPS :

- VGG 모델의 $n(\text{num_layers})$ 만큼 통과시켜 만들어진 생성 이미지와 실제 이미지 activation map 간의 차이를 구한다.
- 값이 낮을수록 생성 이미지와 실제 이미지가 비슷하다는 것을 의미한다.

SSIM :

- 생성된 이미지와 실제 이미지의 휘도, 대조, 구조를 이용하여, RGB 각 채널 별 합을 통해서 이미지의 유사도를 측정한다.
- 0 ~ 1 혹은 -1 ~ 1 사이 값을 가지며, 1에 가까울수록 두 이미지가 유사하다는 것을 의미한다.

FID :

- Inception V3의 feature extractor로 생성된 이미지와 실제 이미지의 activation map을 추출, map 간의 확률 분포 차이를 구한다.
- 값이 낮을수록 이미지가 유사함을 의미한다.

3.3 Model - Virtual Try-On - 실험

- 비교 데이터셋 : VITON-HD Dataset
- 모델 후보 : HR-VITON, GP-VITON, Ladi-VTON
- 정성 평가기준 : 생성된 사진이 자연스러운가?

HR-VITON



GP-VITON



Ladi-VTON



- 좌측 : 원본 모델 사진, 우측 : 상품을 입힌 생성 사진
- 팔이 드러나는 원본 모델 사진에 긴 팔 옷을 합성시킨 결과

3.3 Model - Virtual Try-On - 실험

Musinsa Dataset - Ladi-VTON



<제품>



<원본>



<합성>



<제품>



<원본>



<합성>



<제품>



<원본>



<합성>


3.3 Model – Virtual Try-On --모델 선정

Model	장점	단점
1. HR-VITON	<ul style="list-style-type: none">• 추가학습 가능	<ul style="list-style-type: none">• GAN 모델 기반 -> 생성 결과가 Diffusion에 비해 부자연스러움• 소매 부분같이 그리기 어려운 부분 존재• Multi category 지원 X• 비용이 높은 전처리 과정 필요
2. GP-VTON	<ul style="list-style-type: none">• 추가학습 가능• HR- VTON과 비교했을 때 소매 부분같이 어려운 부분도 잘 생성• Multi category 지원	<ul style="list-style-type: none">• GAN 모델 기반 -> 생성결과가 Diffusion에 비해 부자연스러움
✓최종 모델 3. Ladi-VTON	<ul style="list-style-type: none">• Stable Diffusion 기반 : 생성결과가 자연스러움• Multi category 가능	<ul style="list-style-type: none">• 최종 결과물의 자연스러움을 위해서 복잡한 로고나 무늬를 인위적으로 변화시킴• 높은 inference 비용

3.3 Model – Virtual Try-On – 모델 선정

Viton 모델 학습에 사용된 데이터 셋

1) Viton – HD dataset

- Ladi-vton 모델의 학습에 사용 ( [Link](#))

2) Dresscode dataset

- Ladi-vton 모델의 학습에 사용 ( [Link](#))

3) 크롤링한 Musinsa Dataset

- (모델, 상품)으로 구성된 6000여 개의 pair data 수집
- VITON-HD dataset 과 Dresscode dataset으로 학습된 모델의 fine-tune에 활용

3.4 Model - 경량화

웹 데모 Inference 속도	Torch		Onnx
SAM	1s 이내	약 50%감소 ➡	0.5s 이내
CLIP →Text, Image 단독 검색	4 ~ 6s(평균5s)	약 80% 감 ➡	모델 로드 시 : 1.5s 이내 모델 로드 후 : 1s 이내
CLIP →Text & Image 검색	10 ~ 12s(평균11s)	약 86.3% 감 ➡	모델 로드 시 : 3s 이내 모델 로드 후 : 1.5s 이내

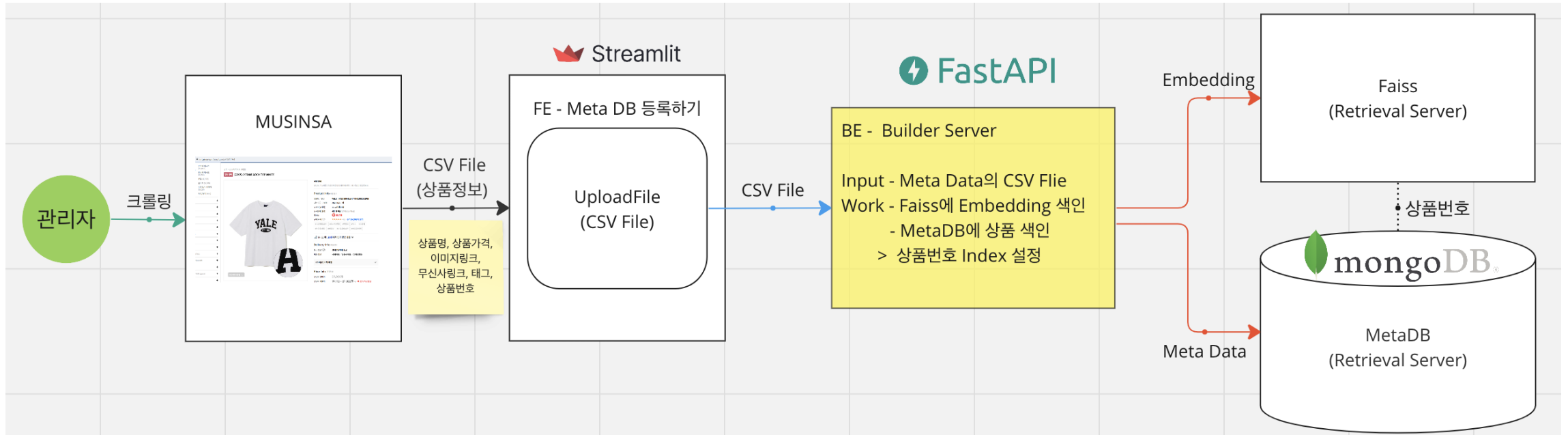
⇒ 성능 유지 But Inferenec 속도 개선 효과 큼!

⇒ CPU와 GPU에서 Inference 속도 유사함!

4. Service Architecture

-
1. Management DB
 2. Seg & Retrieval
 3. Virtual Try-On

4.1 Service Architecture – Management DB



무신사 크롤링 CSV 파일 업로드로
다음의 과정들을 한 번에 처리

- FAISS에 전처리된 임베딩 벡터 저장
- MongoDB에 임베딩 벡터 인덱싱 & 상품 정보 등록
- 36000건 등록 5분 소요

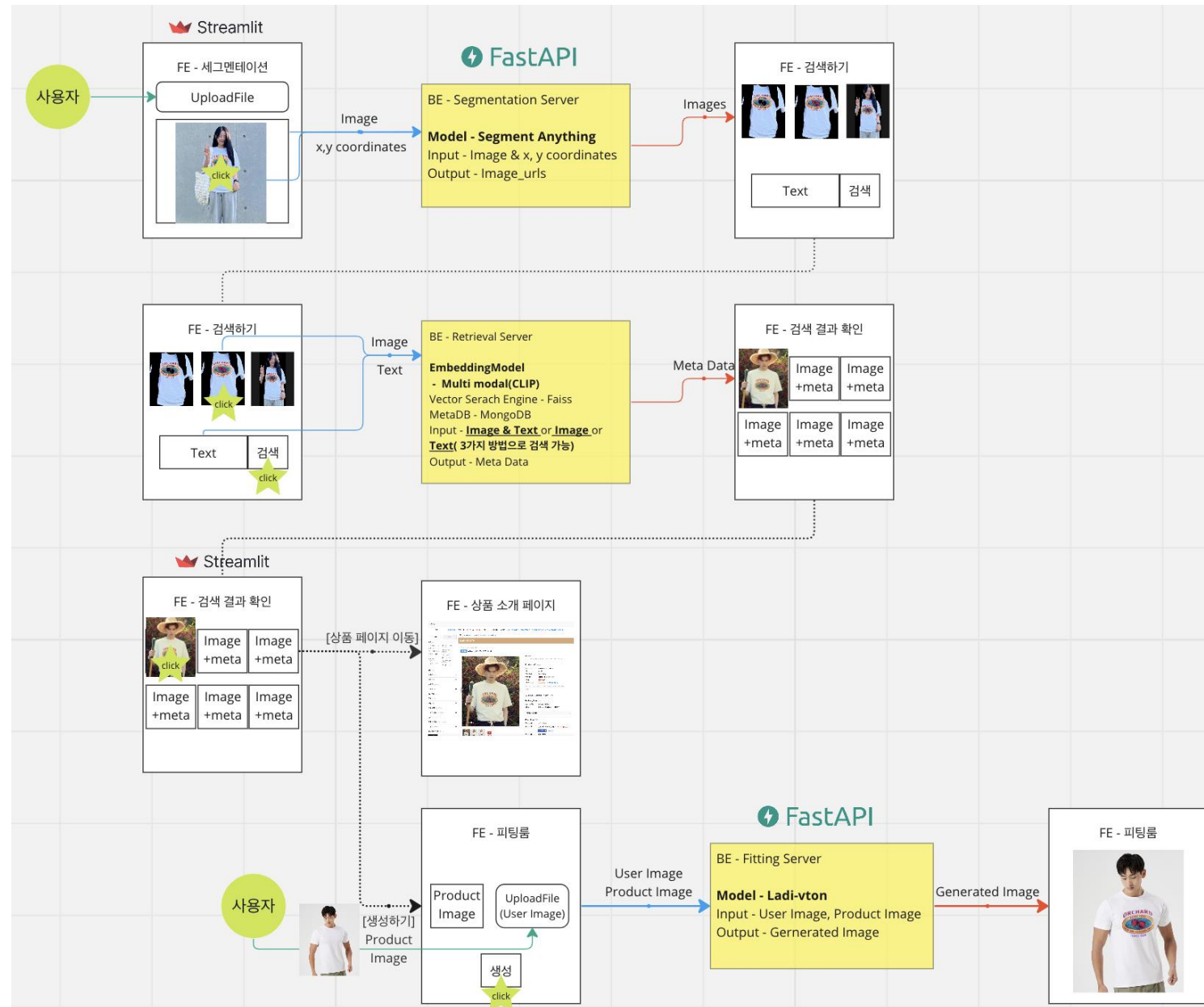
FAISS 선정 이유

- 오픈소스
- 다양한 레퍼런스
- 데이터 수에 따른 다양한 Index 설정
 - 100만장 이하 : 전탐색 : L2 사용
 - 100만장 이상 : Clustering 활용

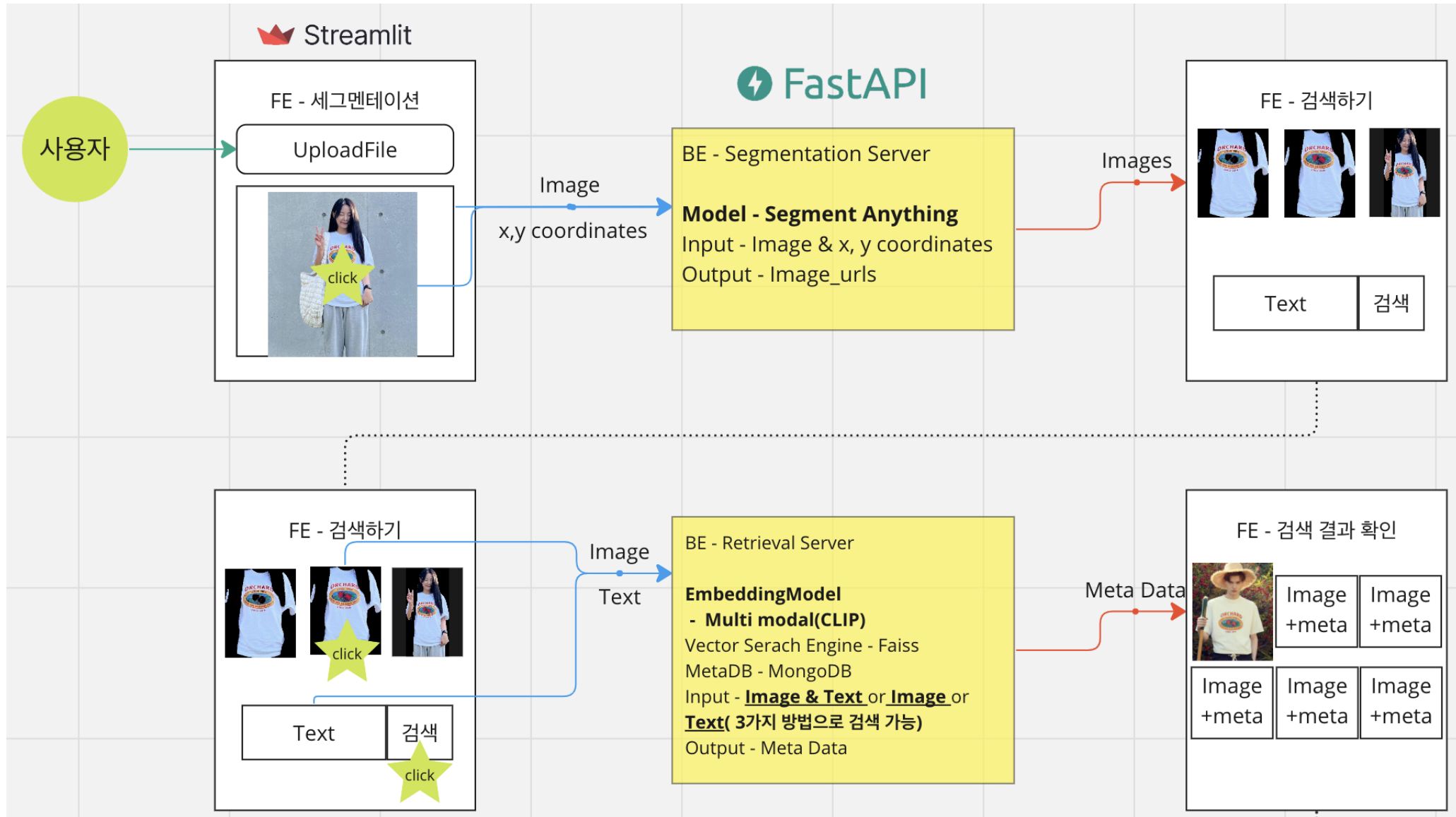
MongoDB 선정 이유

- 유연한 스키마 구조
- 간단한 쿼리(상품번호)와 빠른 조회
- 모델링 방식: 임베디드 방식

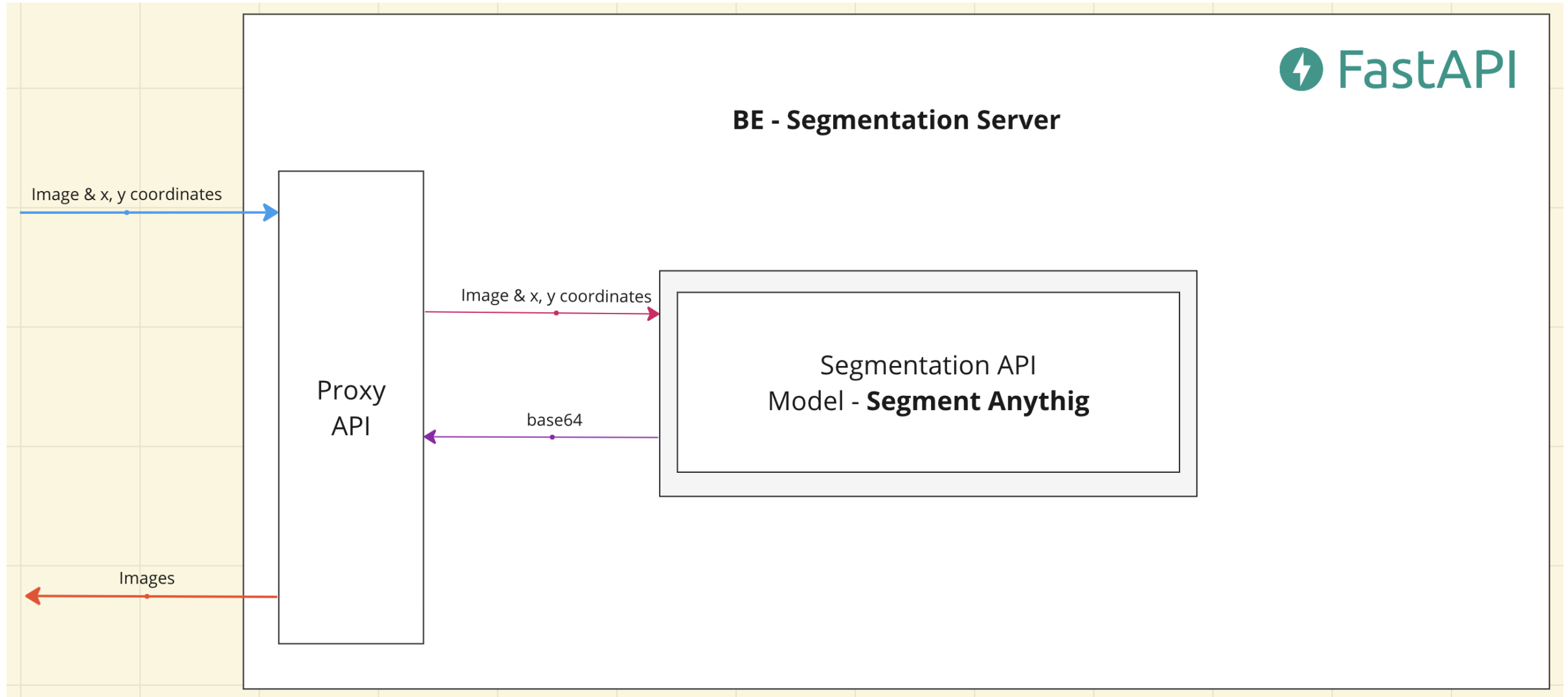
3.2 Service Architecture



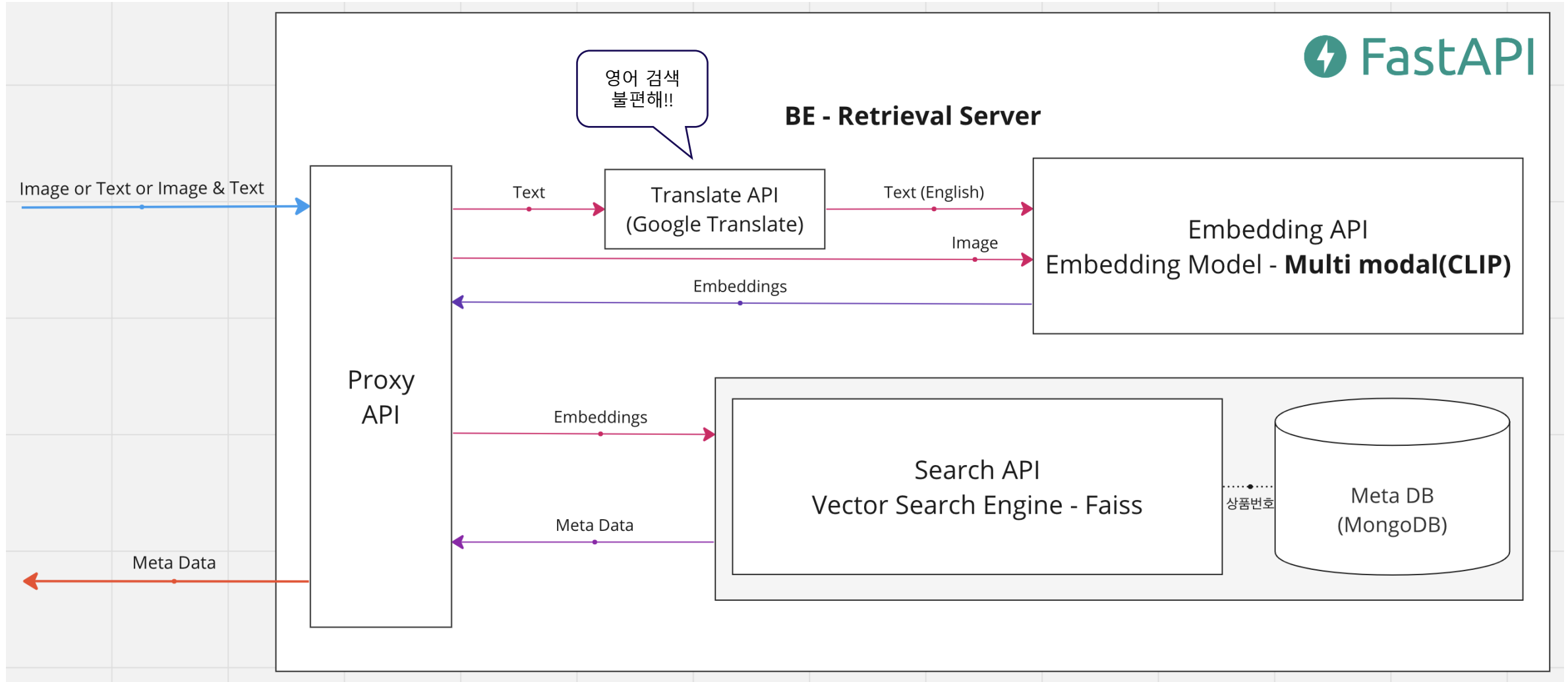
3.3 Service Architecture – Segmentation & Retrieval (Input & Output 중심)



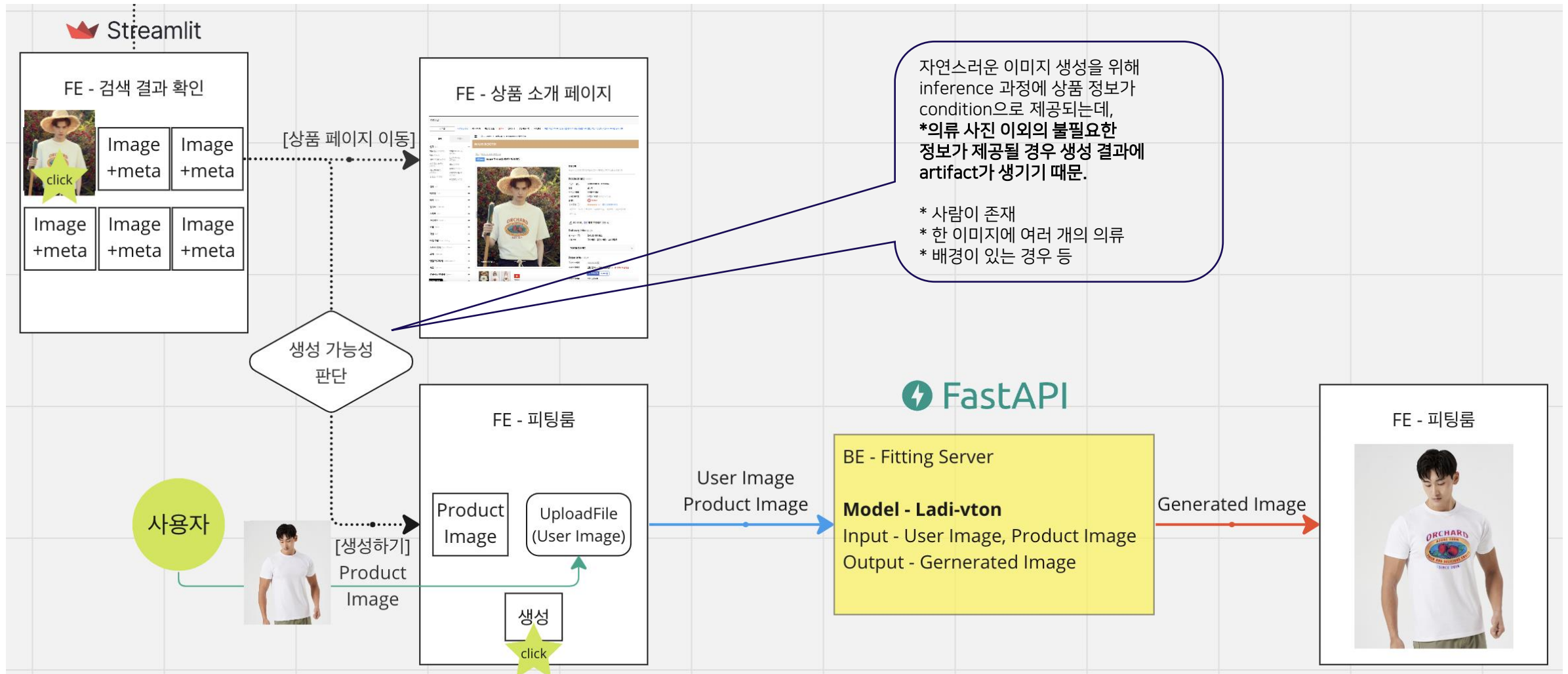
3.3.1 Project Status - Segmentation



3.3.2 Project Status – Retrieval



3.4 Service Architecture – Virtual Try-On (Input & Output 중심)(별첨1. 자세한 생성API 구조)



5. Result / Conclusion

1. Demo
2. Self-assessment
3. Todo list

5.2 Result / Conclusion – Self-assessment

① Segmentation

- 정성적 평가로 가장 정확한 이미지가 score와 Segmentation 결과의 score가 가장 높은 이미지가 동일하지 않다.
- 의류 데이터로 추가 파인튜닝을 하지 못한 점.

② Image Retrieval

- DB에 저장된 데이터가 적을 때는 검색 성능이 불완전함.
- 세그멘테이션의 영향을 크게 받아 세그멘테이션의 성능이 중요함.
- DB에 들어갈 상품들을 주기적으로 업데이트할 수 있어야함.
- 대규모서비스시 예상되는 문제 및 개선방향 :
 - 임베딩 벡터 저장 비용 → quantization 통한 벡터 feature size 줄이기
 - 검색속도 이슈 → FAISS : 전탐색보다는 클러스터링을 이용한 검색
 - 이미지 저장 이슈 → Cloud Storage 사용

③ Try - on

- 데이터 관점 : 공개된 데이터 및 수집가능한 데이터가 한정적임.
- 데이터 전처리 관점 : 생성에 필요한 데이터를 전처리하는 과정이 까다로움. 전처리 과정에 또다른 딥러닝 모델을 필요로 함.
- 생성 성능 관점 : 전처리된 데이터가 생성 성능에 영향을 미침. 생성 가능한 의류 카테고리가 제한적. 원본 상품 이미지의 프린팅이나 로고 등의 디테일한 부분에 대한 변형이 심함.
- 서비스 관점 : 전처리 및 생성 모델을 서빙하기 위해 고성능의 서버 컴퓨터가 필수적.

5.3 Result / Conclusion – Todo list

① 회원가입 / 로그인 기능

회원별로 검색한 옷을 장바구니에 저장하고 가상피팅할 수 있게 한다.

② Try-On 모델 성능 고도화

우선 코드 레벨의 튜닝을 통해 사전학습된 try-on 모델로 안정적인 생성 성능 확보를 시도하고, 이후에 직접 수집한 데이터로 재학습하는 것을 고려하고 있습니다.

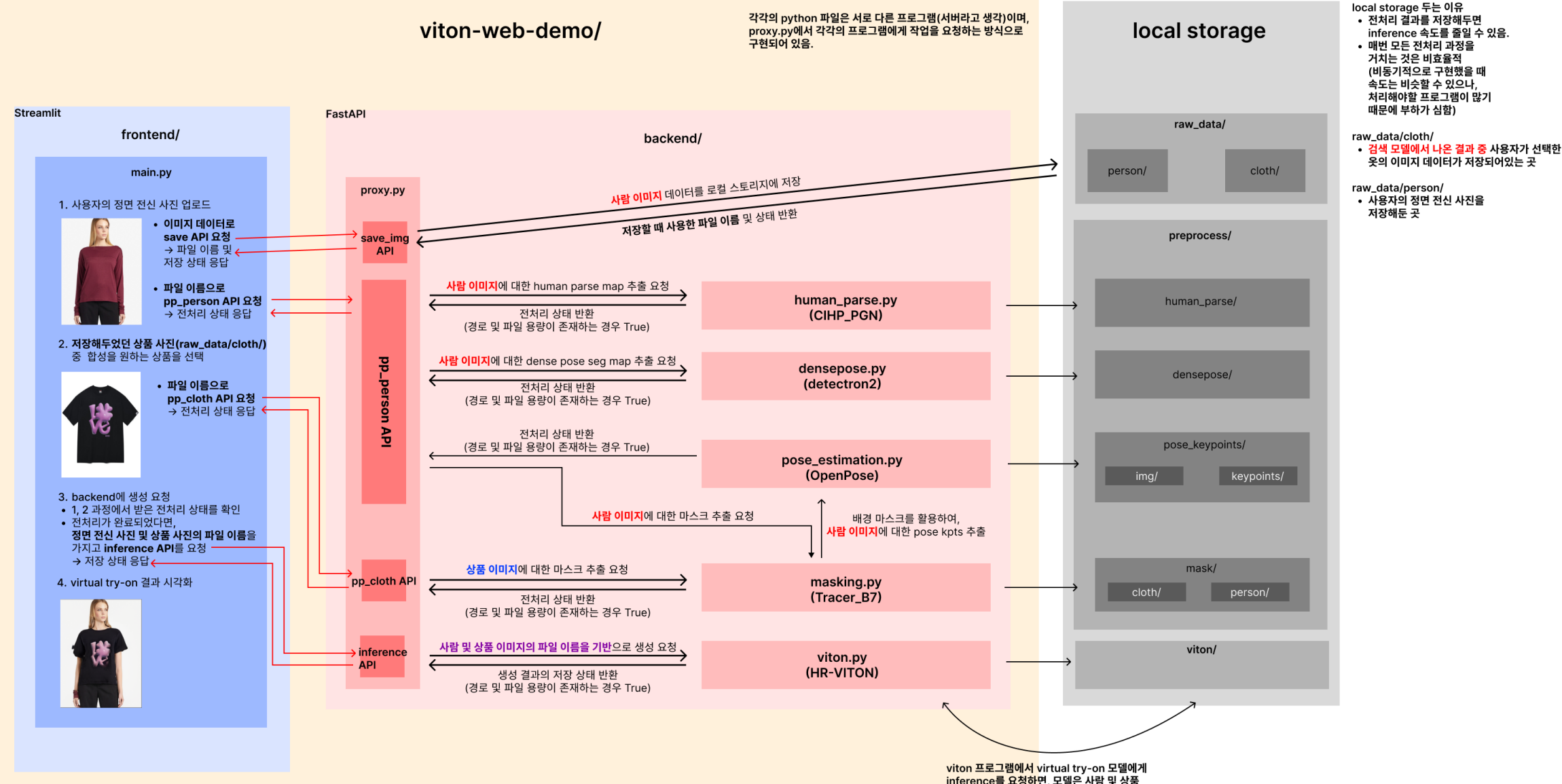
③ 생성 서비스 인퍼런스 속도 향상

서비스를 사용할 때 속도로 인한 불편함이 없도록 전체적인 코드를 리팩토링 및 최적화하는 것을 고려하고 있습니다.

별첨

-
1. 자세한 Virtual Try-On 의 BE구조
 2. 생성 가능성 판단 모델

별첨1. Virtual Try-On



별첨2. 생성 가능 판단 모델

별첨 2.1 생성 가능 판단 모델이 필요한 이유

검색용 데이터 유형



제품사진

✓활용가능한 데이터



제품일부

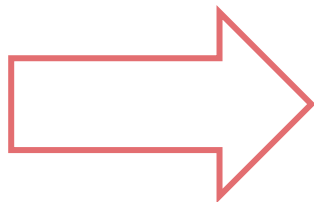


모델 사진



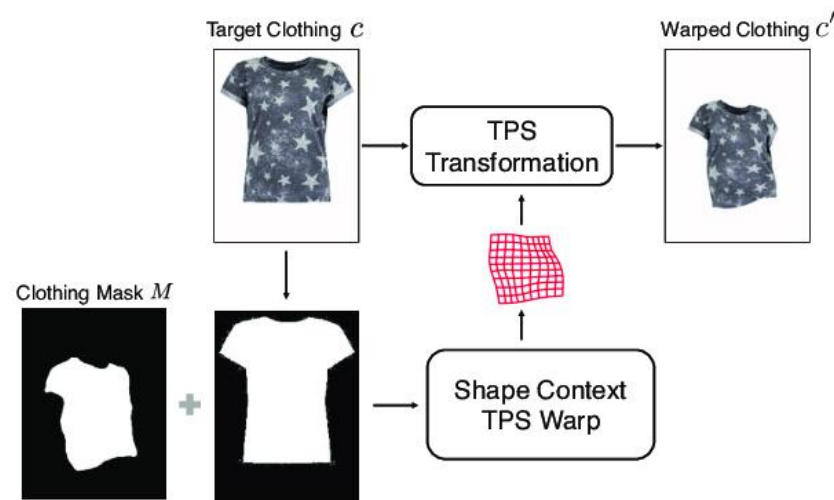
여러 개의 제품사진

생성모델 Warping 과정의
활용할 데이터는 상품사진 유형



따라서 이를 판단 할
추가 Label 필요!!

Ladi-vton



상품사진을 해당 모델의 몸에 맞게 Warping 하는 과정

별첨2. 생성 가능 판단 모델

별첨 2.1 생성 가능 판단 모델이 필요한 이유



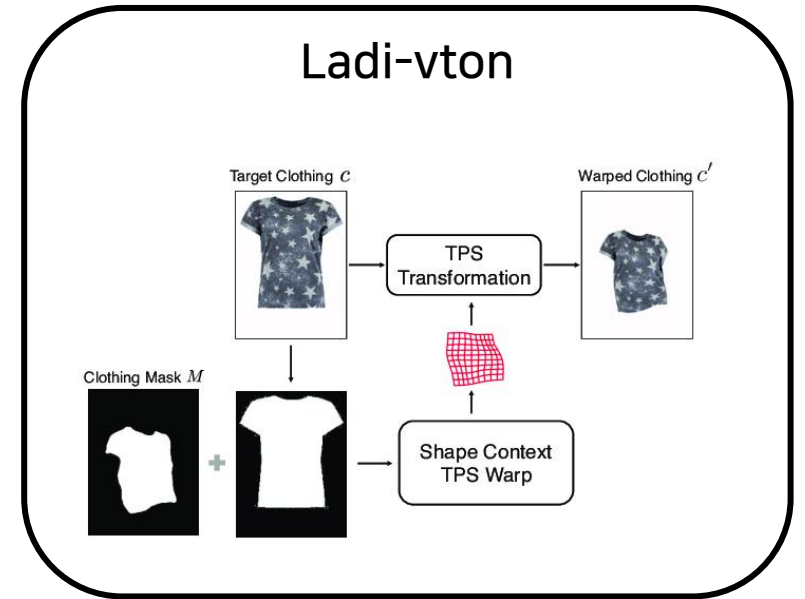
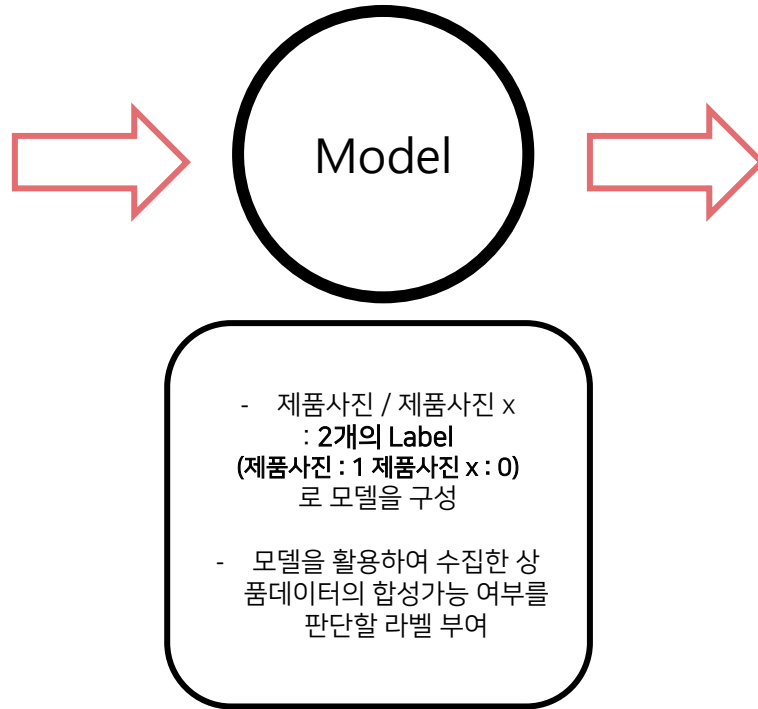
- 상품 사진 + 원본 모델 사진 -> **try-on** 결과 이미지
- 상품 사진에 불필요한 정보가 존재하는 경우, 합성 결과에 예상치 못한 artifact가 생김
(모델의 팔에 문신이 없음에도, 상품 이미지의 우하단 정보가 생성에 영향을 미쳐 무늬가 생김)

별첨2. 생성 가능 판단 모델

별첨 2.2 생성 가능 판단 모델 구성



수동으로 라벨링하면 많은 시간이 소요



별첨2. 생성 가능 판단 모델

별첨 2.2 생성 가능 판단 모델 학습 데이터

학습 데이터 Label



제품 사진 : 1



제품 사진 X : 0

- 수집한 검색용 데이터의 약 1000장 정도를 수동으로 Labeling 진행
- 상/하의 500장씩 데이터셋 구성
- 이를 활용하여 8/2로 Train/Val set 생성 (상/하의 balance 맞춰 구성)
- 또한 추가적으로 Test를 하기위해서 500장 정도 수동 Labeling을 하여 Test set 구성
- Test set은 상/하의 250장씩 구성

별첨2. 생성 가능 판단 모델

별첨 2.2 생성 가능 판단 모델 학습 및 평가

Base Model

Efficientnet V2

Densenet 201

Mobilenet V3

Mobile vit V3

- 추후에 데이터 등록 기능에 Serving을 고려하여 Inference 비용이 크지 않은 모델들로 선정
- 4개의 모델 모두 10 Epochs 이내로 Val Accuracy가 98%에 수렴
- 따라서 Test Data를 활용하여 Confusion Matrix 계산하여 추가적으로 모델을 판단
- 특히, Serving 관점에서 보면 Model의 예측결과가 제품사진이라고 판단한 결과에 제품사진이 아닌 경우가 적은 것이 중요하다고 판단
- 따라서 정밀도(Precision)관점으로 모델을 평가
- Test Data 결과
 - Efficientnet V2 : (ACC : 98.5, Precision : 0.99)
 - DenseNet 201 : (ACC : 98.0, Precision : 0.92)
 - Mobilenet V3 : (ACC : 98.3, Precision : 0.92)
 - Mobile Vit V3 : (ACC : 98.4, Precision : 0.93)
- 최종 모델 선택 : ACC와 Precision이 가장 높은 Efficientnet V2를 선택, 추후 정성평가를 진행했을 때도 Efficientnet V2 결과가 가장 좋았다

별첨2. 생성 가능 판단 모델

별첨 2.2 생성 가능 판단 모델 개선 방안

① 추가 카테고리 학습

현재, 상/하의 카테고리에만 모델이 학습 되어 있기 때문에 추후 아우터, 모자 등 다양한 카테고리 학습이 필요해 보인다

② 추가 학습데이터 Labeling

현재 1000개의 데이터만을 활용하여 모델을 학습하였기 때문에 추가 Labeling을 하여 학습데이터를 확보 하면 모델의 성능이 더욱 향상될 것으로 보인다

③ 데이터 등록 서비스와 연결

현재 데이터 등록 서비스의 경우 생성 가능성 판단 모델을 오프라인으로 inference하여 결과를 csv로 저장하여 등록하여하는데 생성 가능 판단 모델의 API를 만들어서 데이터 등록 서비스와 연결하면 서비스가 향상될 것으로 보인다

End of Document Thank You.