

---

# 이미지와 리뷰를 통한 의류 추천

13기 - 9팀

팀원

강선민, 이수현, 전형준, 정승기, 한유성

# 목차


01. 프로젝트 개요
02. 프로젝트 팀 구성 및 역할
03. 프로젝트 수행 절차 및 방법
04. 프로젝트 수행 결과
05. 자체 평가 의견

# 01. 프로젝트 개요

## ▶ 프로젝트 주제 및 선정 배경

### "카·몸무게·나이까지 본다" 디테일에 강한 온라인몰 리뷰

기사등록 2022/09/03 13:00:00  
최종수정 2022/09/03 13:14:39


 [가](#)

#### 기사내용 요약

11번가 "성별, 연령, 키, 몸무게 등 기입 요청"  
이달 중 화장품..유아동 카테고리에도 적용  
해지스닷컴 '입어봤어요'...스타일링 팁도 제공

### 실패없는 구매 척도, 리뷰가 돈된다...차별화가 곧 매출

기사등록 2022/09/11 10:00:00

 [가](#)

#### 기사내용 요약

리뷰 커머스 시대...제품정보 제공 역할 넘어서  
이커머스 업계, 리뷰 경쟁력 높이는 방안 고심  
실구매자 검증, 항목 세분화...구매경험도 제공

# 01. 프로젝트 개요

## ▶ 프로젝트 주제 및 선정 배경(기획의도 등)

한경 생활·문화

"스타 공항패션의 비밀이 궁금해?"...대세 스타들의 출국 패션!



[사진]레드벨벳 슬기, '완벽한 공항패션'

OSEN

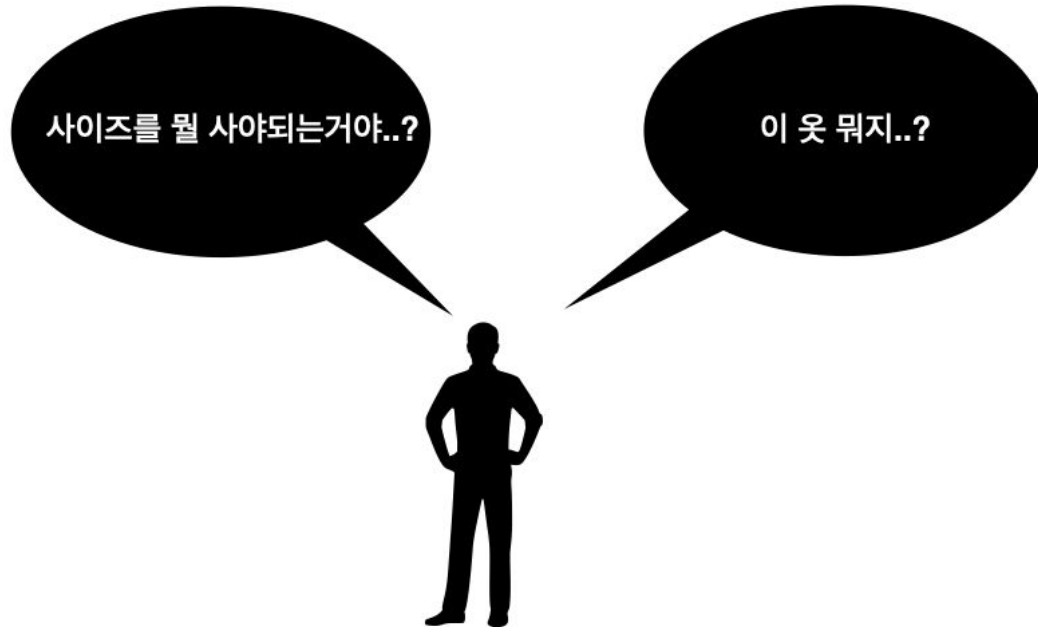
입력 2022.09.11 08:55



# 01. 프로젝트 개요

---

## ▶ 프로젝트 개요



### 리뷰 데이터를 통한 사이즈 추천 서비스

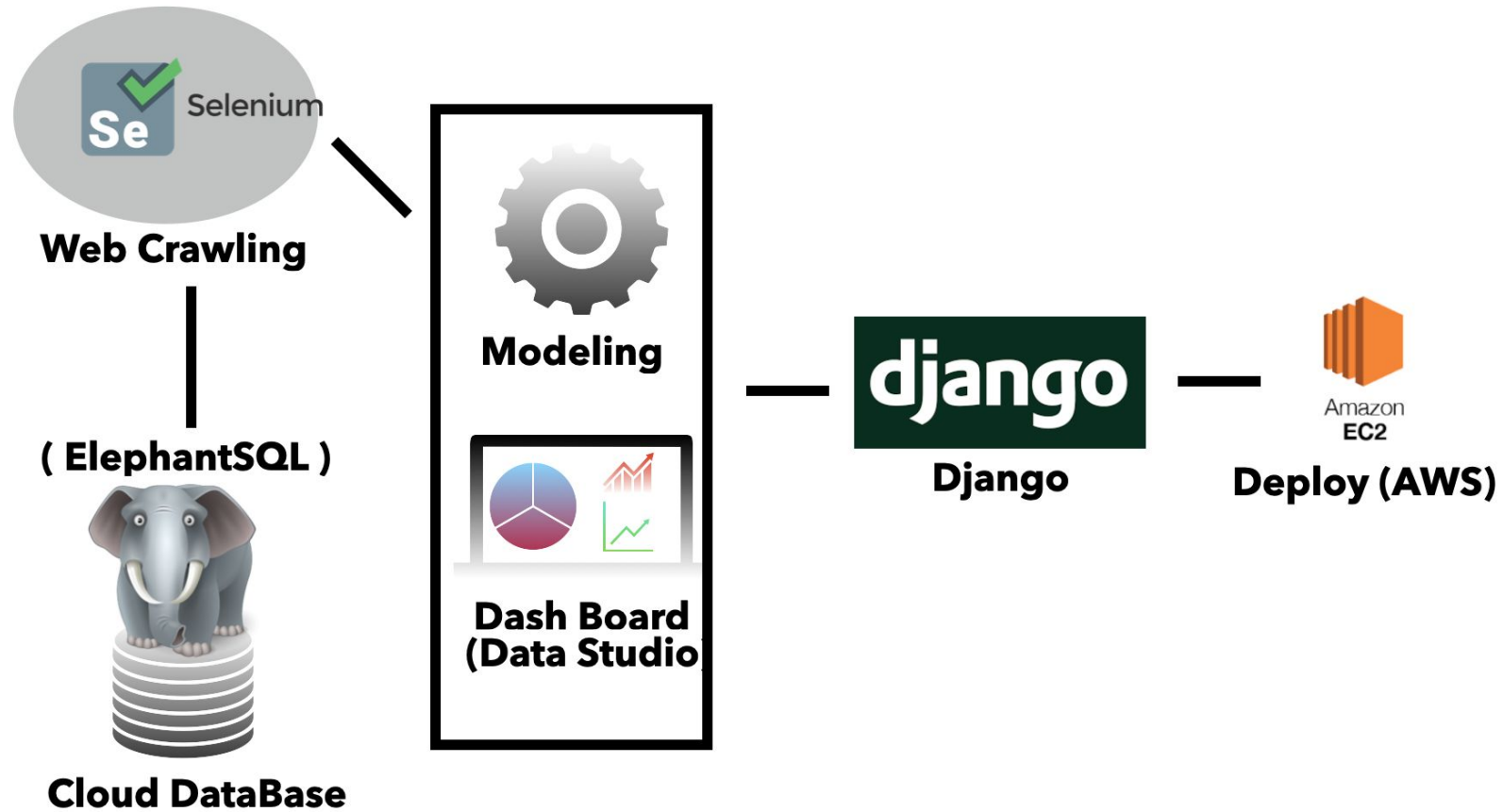
- ML 모델을 이용한 모델링
- Cloud DB와 연동

### 이미지를 이용한 의상 추천 서비스

- DL 모델을 이용한 모델링
- Cloud DB와 연동

# 01. 프로젝트 개요

## ▶ 프로젝트 구조



## 02. 프로젝트 팀 구성 및 역할

---

훈련생	역할	담당 업무
정승기	팀원	<ul style="list-style-type: none"><li>▶ 데이터 수집 및 정제</li><li>▶ <i>Object detection</i> / 사이즈 추천 모델 구현</li></ul>
전형준	팀원	<ul style="list-style-type: none"><li>▶ 웹 서비스 개발 및 DB 연동</li><li>▶ Docker, AWS EC2를 통한 배포 진행</li></ul>
이수현	팀원	<ul style="list-style-type: none"><li>▶ 웹 서비스 개발 및 DB 연동 (<i>Django</i>)</li><li>▶ 이미지 데이터 라벨링</li></ul>
강선민	팀원	<ul style="list-style-type: none"><li>▶ 클라우드 DB 설계 및 서비스 연동</li></ul>
한유성	팀원	<ul style="list-style-type: none"><li>▶ 데이터 정제 및 자동화</li></ul>

### 03. 프로젝트 수행 절차 및 방법

구분	기간	활동	비고
사전 기획	▶ 08/31(수) ~ 09/02(금)	▶ 프로젝트 기획 및 주제 선정 ▶ 기획안 작성	▶ 아이디어 선정
데이터 수집	▶ 09/02(금) ~ 09/07(수)	▶ 필요 데이터 및 수집 절차 정의 ▶ 외부 데이터 수집	▶ 팀별 일일보고 실시
데이터 전처리	▶ 09/03(토) ~ 09/07(수)	▶ 데이터 정제 및 정규화 ▶ 데이터 라벨링 ▶ DB 연동 및 데이터 적재	▶ 팀별 일일보고 실시
모델링	▶ 09/05(월) ~ 09/09(금)	▶ 모형 구현	▶ 팀별 최종보고 실시
서비스 구축	▶ 09/06(화) ~ 09/11(일)	▶ 웹 서비스 개발 ( Django ) ▶ AWS EC2를 통한 배포 진행	▶ 최적화, 오류 수정
총 개발기간	▶ 08/31(수) ~ 09/11(일)(총 2주)	-	-



## 04. 프로젝트 수행 결과

결과 제시 ① WebApplicationServer

### ▶ WAS 소개 (Django)

	Django	Flask
편리성	MVT 구조로 이루어져 있고, 대부분의 기능이 이미 개발이 완료된 상태	자유롭게 작성이 가능하나 하나하나 설정을 해주어야 함
정보	다양한 커뮤니티, 정보가 존재	Django에 비해 적은양의 정보
보안	자체적으로 여러 보안기능이 적용 되어있음	사용자가 설정을 해주어야함
데이터관리	ORM을 통해 DB를 쉽게 관리 가능 admin page를 이용해서 생성, 삭제가 용이함	사용자가 설정 해주어야 하는 부분이 상대적으로 많음

개발용도, 상황에 맞게 사용하면 되지만 본 프로젝트에서는 대부분의 기업에서 사용하는 Django의 구조를 이해하고자 선택

## 04. 프로젝트 수행 결과

결과 제시 ② 서비스별 동작 방식

---



### 서비스 동작 방식

리뷰를 통한 사이즈 분석

- 사용자의 데이터를 입력 받아 리뷰데이터를 기반으로 훈련된 ML모델에 입력
- 이후 모델에서 나온 Output(Size)와 사용자의 입력 데이터를 기반으로 상위 제품을 DB에서 추천

데이터를 입력 받은 이후 “POST”를 누르게 되면 Size 와 Url을 내부적으로 얻음

- 입력 받은 데이터는 DB에 저장되어 이후 ML Model의 적용 고려

해당 url을 통해서 웹페이지에 들어가서 크롤링을 진행하여 제품명과 이미지 정보를 가져옴

## 04. 프로젝트 수행 결과

결과 제시 ② 서비스별 동작 방식

---



### 서비스 동작 방식

#### 이미지를 통한 추천모델

- 사용자로부터 이미지 파일을 받고, 해당 이미지를 DL 모델의 input으로 전송
  - 이미지를 받을 때는 Drag and Drop 기능도 구현
- 모델에서는 해당 input을 통해 Object detect를 진행하고, 결과 값을 DB에 보내어 적절한 제품 중 상위 제품을 추천
  - Object detect를 통해 상의,하의 하나씩 신뢰도가 가장 높은 라벨을 선택

이미지가 들어오면 임시저장, 해당 이미지파일로 부터 데이터(DL 결과값)를 모두 얻은 이후에는 메모리 관리를 위해

#### 자동삭제

- 모델을 통해 얻은 결과(상의,하의)을 가지고 DB에 쿼리문을 보내고 각 이미지에 대한 url주소를 가져옴

해당 url을 통해서 웹페이지에 들어가서 크롤링을 진행하여 제품명과 img 정보를 가져옴

## 04. 프로젝트 수행 결과

데이터 정제 및 자동화

### Web scraping & Crawling

	Selenium	Scrapy	Request, Urllib
속도	느림	빠름	빠름(조건부)
기능	보통	많음	적음
난이도	쉬움	보통	쉬움



Selenium



Scrapy

urllib3

데이터의 갱신을 목적으로한 수집에서는 속도 보다는 도커,Django 연동이 좋은 Selenium 채택

## 04. 프로젝트 수행 결과

결과 제시 ① 활용된 기술/ 장점

### ▶ ElephantSQL(PostgreSQL 호스팅 서비스) 소개

- ElephantSQL은 여러 관계형 데이터베이스 중 PostgreSQL 데이터베이스의 호스팅 서비스
  - 데이터베이스 호스팅 서비스는 웹서버와 데이터베이스 서버가 같은 물리적 장치에 있는 것이 아닌 웹서버와 데이터베이스 서버를 분리하여 효율적으로 데이터베이스를 활용할 수 있도록 도와주는 서비스
  - 쉽게 설명하자면, 데이터베이스만을 위한 서버를 구축  
(ElephantSQL은 클라우드 PostgreSQL 데이터베이스를 제공)
- 1) 클라우드DB의 가장 큰 장점은 여러 장소, 분리된 환경에서 하나의 데이터베이스를 사용할 수 있다는 점
  - 2) 로그 분석 및 리스크 트래킹 등등 다양한 별도 서비스를 제공받을 수 있다는 것 역시 큰 장점



## 04. 프로젝트 수행 결과

결과 제시 ② Django + PostgreSQL 연동

---



### Django project + PostgreSQL

- Django에 기본적으로 다른 DBMS와는 달리 속도면에서 장점이 있다고 알려진 `sqlite`라는 DBMS가 내장되어 있음
- 하지만 이번 프로젝트에서는 클라우드 DB를 통한 분리된 환경에서의 작업을 구현하기 위해 ElephantSQL을 통해 만들어진 PostgreSQL 클라우드 DB와 Django를 연동하여 DB시스템을 구축



## 04. 프로젝트 수행 결과

결과 제시 ①데이터 수집

### ▶ 학습 데이터 소개 (Train/dev set)

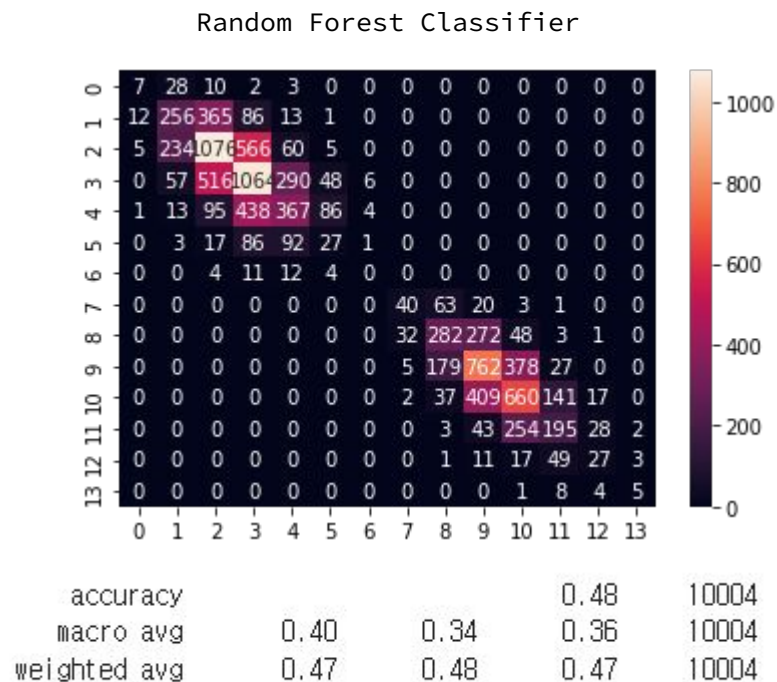
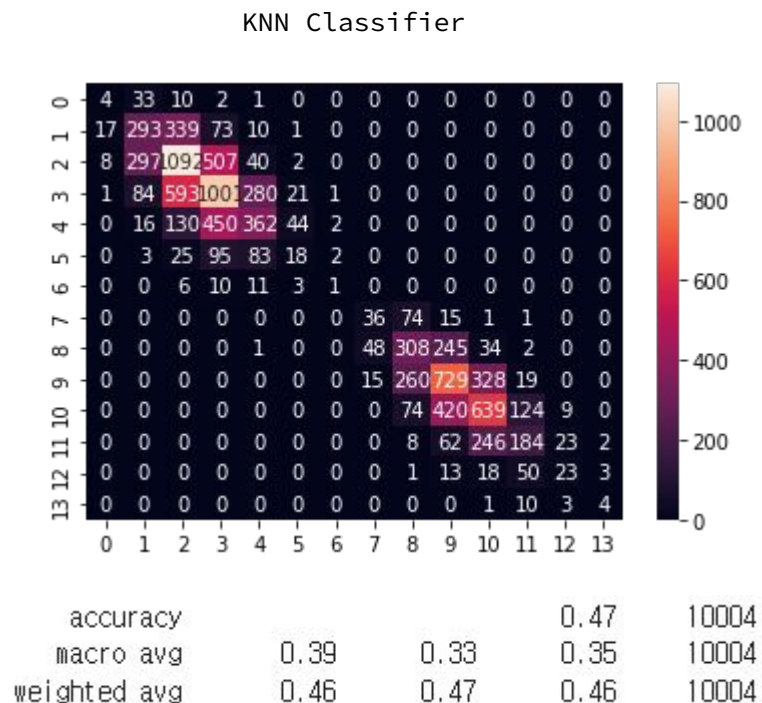
	customer_gender	customer_height	customer_weight	customer_size	major	sub
0	0.0	175.0	75.0	9.0	2	12
1	0.0	178.0	75.0	9.0	2	12
2	0.0	176.0	88.0	10.0	2	12
3	0.0	173.0	65.0	8.0	2	12
4	0.0	172.0	75.0	10.0	2	12
5	0.0	174.0	70.0	10.0	2	12
6	0.0	171.0	58.0	9.0	2	12
7	0.0	171.0	68.0	10.0	2	12
8	0.0	177.0	68.0	9.0	2	12
9	0.0	178.0	68.0	10.0	2	12

무신사 사이트 내에서 필요할 것으로 판단 된 것은 모두 수집 후 모델링에 필요한 데이터만 정제

## 04. 프로젝트 수행 결과

결과 제시 ②모델 평가 및 개선

### ▶ KNN Classifier vs Random Forest Classifier



두개의 모델을 비교했을때 큰 차이는 없으나 조금 더 성능이 좋은 Random Forest Classifier를 사용



## 04. 프로젝트 수행 결과

### 결과 제시 ①데이터 수집

#### Scraping



Summer 심플룩의 대표아이템,남자 ...  
menknow.tistory.com



♥질리지않는 나염반팔티 코...  
lookpin.co.kr



남자 오버핏 반팔티 코디 이쁘게...  
m.blog.naver.com



여름에도 예쁜 반팔티코디♥  
lookpin.co.kr



여름 반팔티 코디 블랙, 화이트 ...  
m.post.naver.com



여름 반팔티 코디 블랙, 화이트 무...  
m.post.naver.com



남자 반팔티 추천 브랜드 디스커...  
blog.naver.com



Unique한 style로 정리된 남자 베이...  
menknow.tistory.com



방 on Instagram: "아우터 더 ...  
pinterest.com



남자 스트라이프 반팔티 이쁘게 입는 법과 여름 코디 ...  
youtube.com



대한민국 남자 코디 - 여름 코디...  
ko-kr.facebook.com



2018 남자 여름 반팔 긴팔 티셔...  
tksyun.tistory.com



대한민국 남자 코디 - 여름 코디를...  
ko-kr.facebook.com



여름에 뭐 입지? 반팔티셔츠 추천 with 코...  
the-caker.co.kr



티셔츠 넣어입기@회색 반팔티 코디 : 네...  
m.blog.naver.com



남자 검정 반팔티 코디 대중 입어도 예쁘...  
m.blog.naver.com



더보이 THEBOY] 2017...  
tksyun.tistory.com



여름에 뭐 입지? 반팔티셔...  
the-caker.co.kr



musinsacom on Twitter: "#레...  
twitter.com



진정한 오버핏 반팔 코디 :)  
lookpin.co.kr



코디세트]오버핏 반팔티+커빙 청...  
tomonari.co.kr



키작은남자 봄이 빛나는...  
m.under70.kr



Summer Basic Style, 남자 검은색 반팔티 ...  
menknow.tistory.com



2019 남자 여름 코디 검정 반팔티는 필수지 : 네...  
m.blog.naver.com



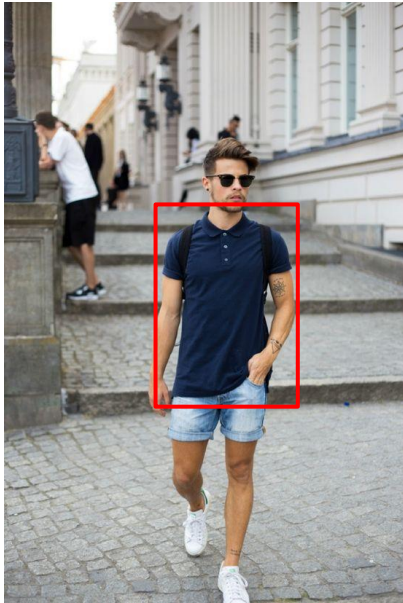
윙윙 실컷 무지 여름 남자 반팔 ...  
brich.co.kr

셀레니움을 이용하여 구글에서 데이터를 수집하고 추가로 무신사, 유니클로 등 쇼핑몰에서 수집

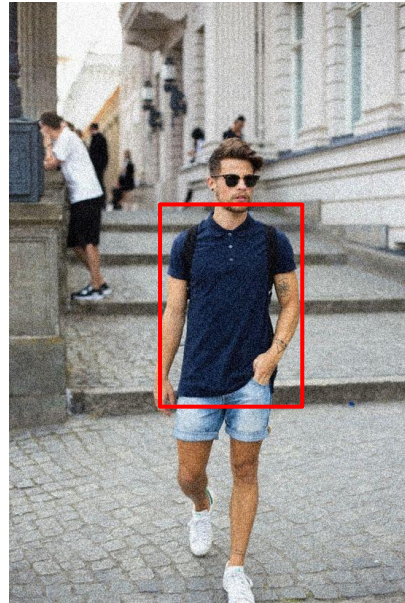
## 04. 프로젝트 수행 결과

결과 제시 ①데이터 수집

### ▶ Data augmentation(데이터 증강)



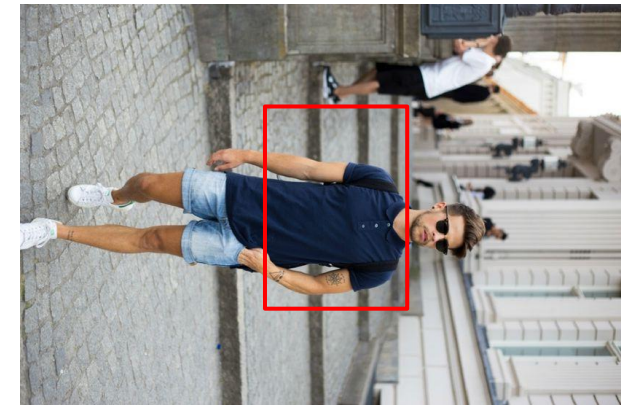
원본 이미지



노이즈 효과



사이즈 조절



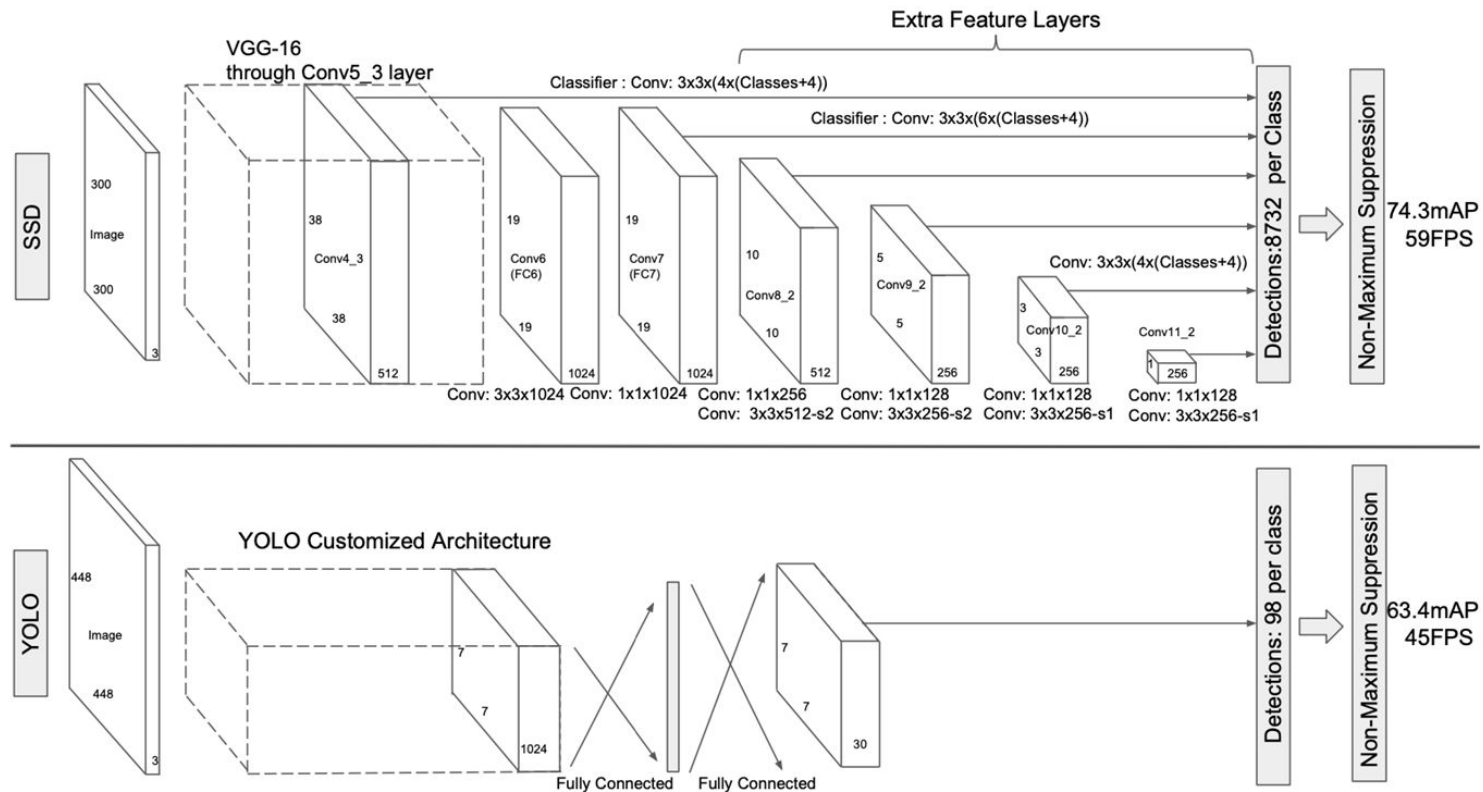
이미지 회전

데이터가 부족하여 이미지 증강기법을 사용, 이미 바운더리와 라벨링이 끝난 데이터를 다시 사용하기 위해 노이즈 사용

## 04. 프로젝트 수행 결과

결과 제시 ②모델 선정

### Yolo vs SSD



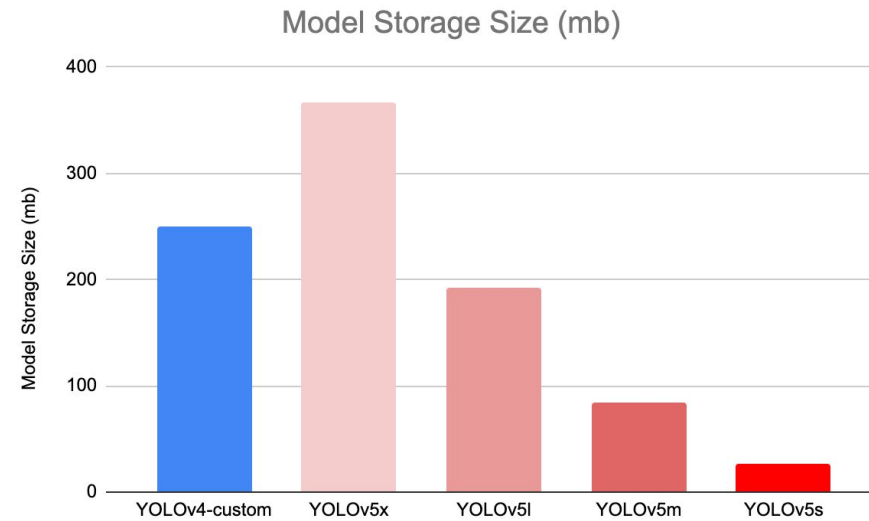
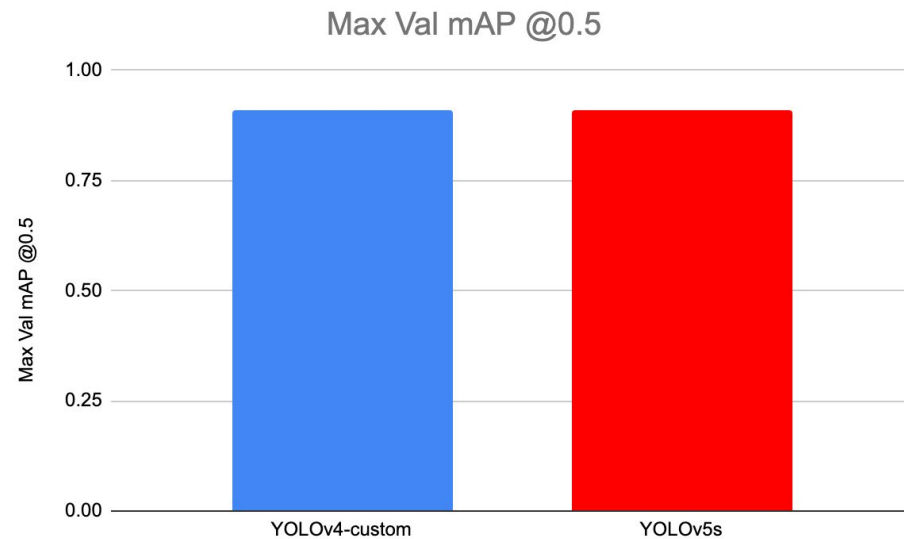
정확도는 SSD가 비교적 우수하나 속도에선 Yolo가 우수, 따라서 사용자의 편의를 생각하여 속도가 우수한 Yolo 선택



## 04. 프로젝트 수행 결과

결과 제시 ②모델 선정

### ▶ Yolov4 vs Yolov5

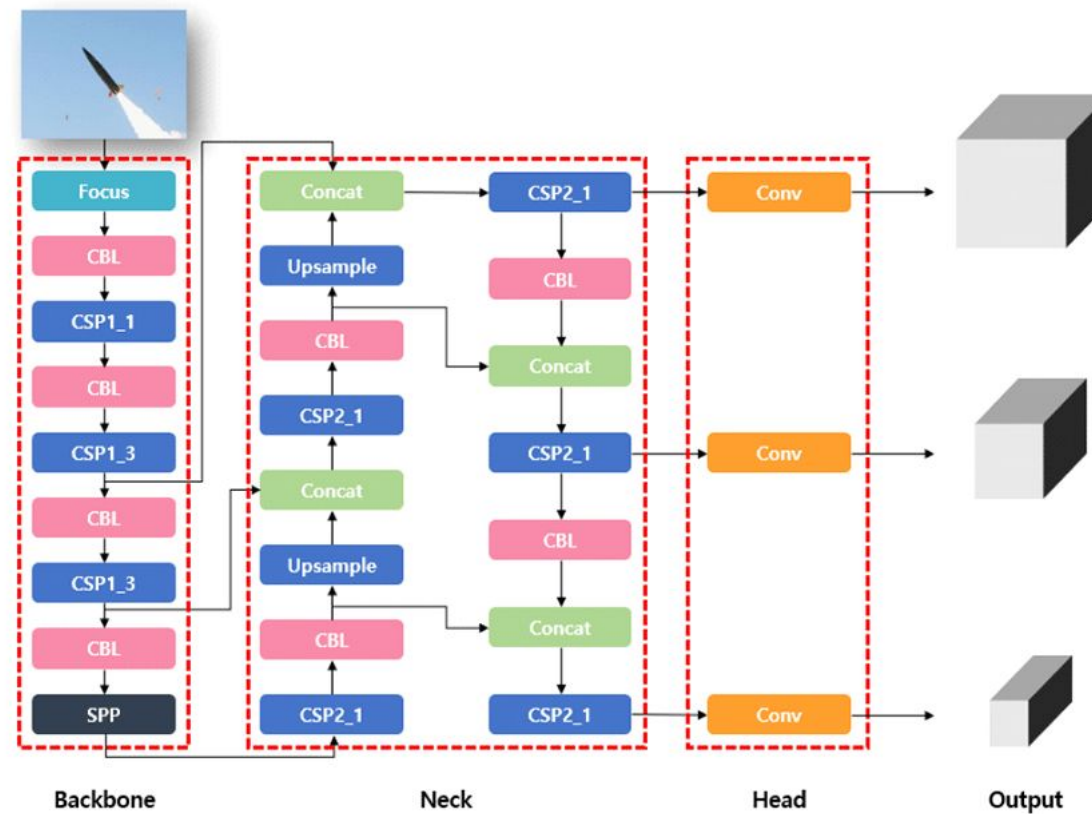


mAP(mean average precision)은 비슷하나 용량은 더 작은 YOLOv5s 선택

## 04. 프로젝트 수행 결과

결과 제시 ②모델 선정

### ▶ Yolov5 구조

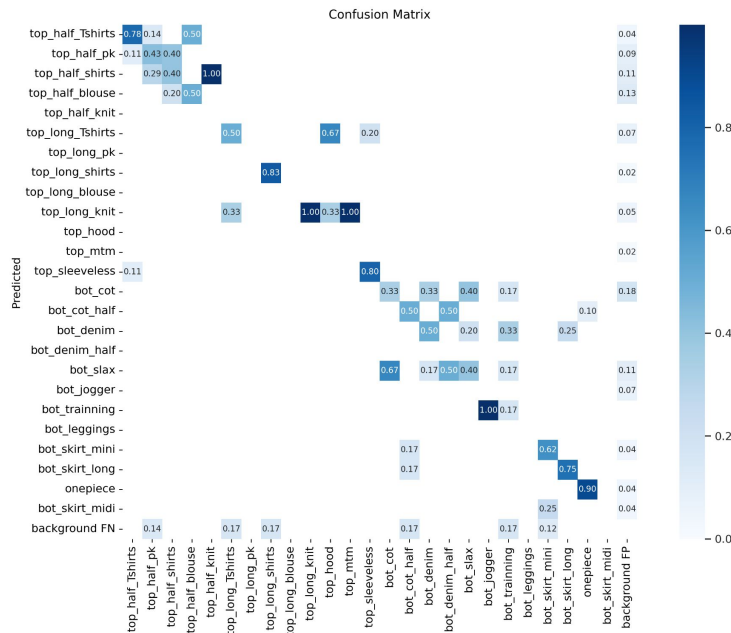


백본(backbone), 넥(neck), 헤드와 같이 3종류 부분으로 이루어짐

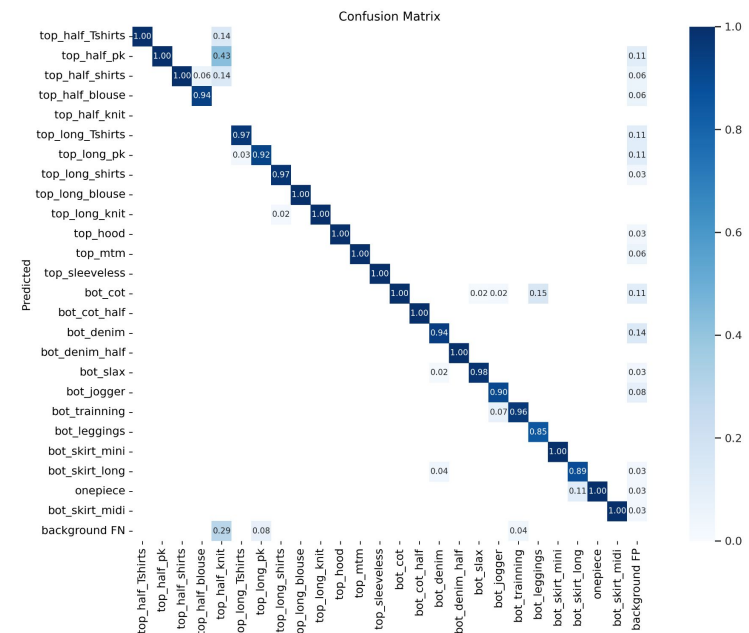
## 04. 프로젝트 수행 결과

### 결과 제시 ③모델 평가 및 개선

#### Confusion Matrix



초기 모델



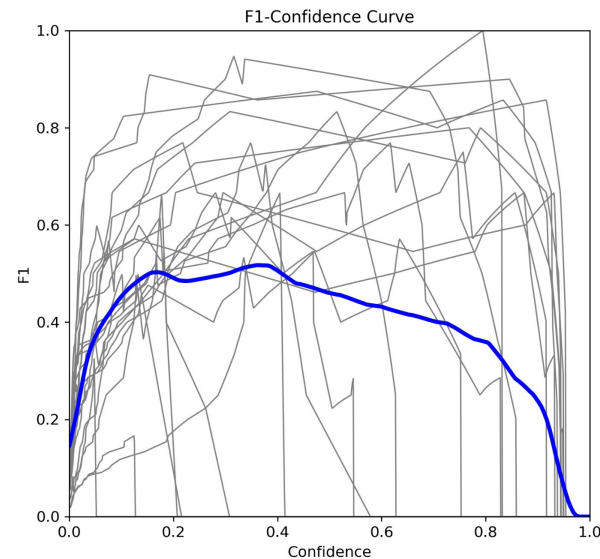
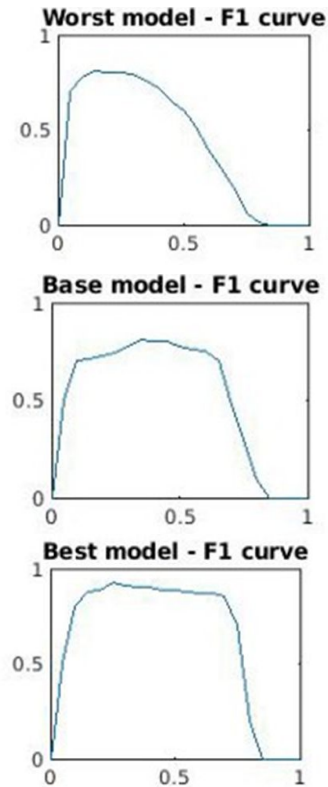
최종 모델

데이터를 추가할때 confusion matrix를 통해 부족한 데이터가 무엇인지 파악하여 추가

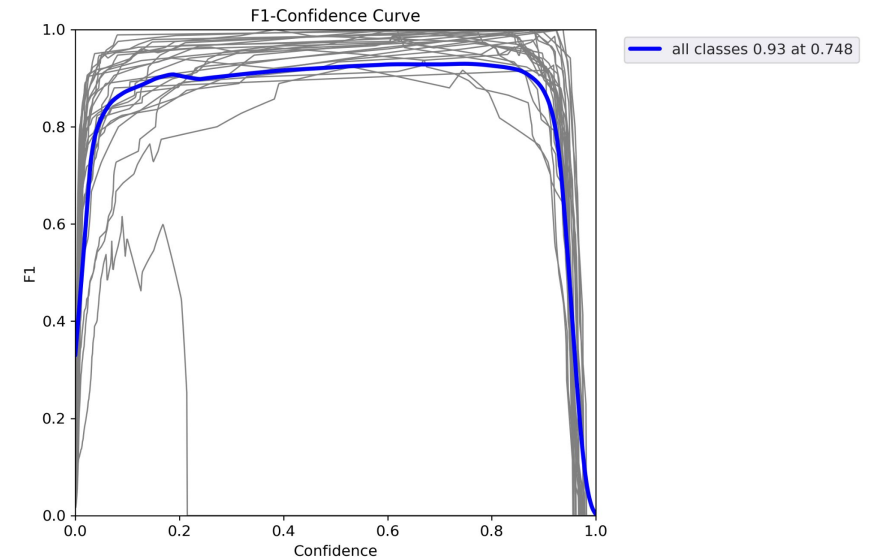
## 04. 프로젝트 수행 결과

결과 제시 ③모델 평가 및 개선

### F1-Confidence Curve



초기 모델



최종 모델

초기 모델은 Worst model에 가깝지만 최종 모델은 best model과 유사함

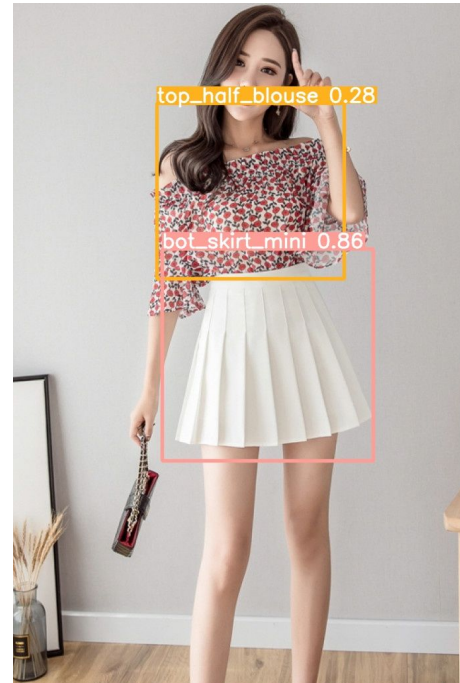
## 04. 프로젝트 수행 결과

결과 제시 ③모델 평가 및 개선

### Image Test



비슷한 디자인이  
훈련 안된 경우



비슷한 디자인을  
훈련 시킨 경우



포즈로 인해 원피스를  
상,하의로 구분



유사한 포즈의 데이터를  
훈련 시킴

디자인, 포즈 등 여러 요인들로 잘 인식 못하던 객체들은 다양한 데이터가 추가 될 수록 성능이 향상



## 04. 프로젝트 수행 결과

결과 제시 ① WebApplicationServer

---

### ▶ Docker 소개

Docker는 어플리케이션을 신속하게 구축, 테스트 및 배포하기 위해 사용하는 소프트웨어 플랫폼

Docker는 Container라는 표준화된 유닛으로 패키징하는데 여기에 라이브러리, 시스템 도구, 코드, 런타임 등 소프트웨어 실행에 필요한 모든 것이 포함

Docker를 사용하면 환경에 구애받지 않고 애플리케이션을 신속하게 배포 및 확장할 수 있고 코드가 문제없이 실행될 것이라고 확신



## 04. 프로젝트 수행 결과

### 결과 제시 ① WebApplicationServer

#### ▶ Docker를 사용하는 이유



##### 더 많은 소프트웨어를 더 빨리 제공

Docker 사용자는 평균적으로 Docker를 사용하지 않는 사용자보다 7배 더 많은 소프트웨어를 제공합니다. Docker를 사용하면 필요할 때마다 격리된 서비스를 제공할 수 있습니다.



##### 운영 표준화

작은 컨테이너식 애플리케이션을 사용하면 손쉽게 배포하고, 문제를 파악하고, 수정을 위해 롤백할 수 있습니다.



##### 원활하게 이전

Docker 기반 애플리케이션을 로컬 개발 시스템에서 AWS의 프로덕션 배포로 원활하게 이전할 수 있습니다.



##### 비용 절감

Docker 컨테이너를 사용하면 각 서버에서 좀 더 쉽게 더 많은 코드를 실행하여 사용률을 높이고 비용을 절감할 수 있습니다.



## 04. 프로젝트 수행 결과

결과 제시 ① WebApplicationServer

---

### ▶ AWS EC2 소개 및 장점

AWS 클라우드에서 확장 가능한 컴퓨팅 용량을 제공하는 서비스

AWS EC2를 사용하면 하드웨어에 선투자할 필요가 없으므로 더 빠르게 애플리케이션을 개발하고 배포할 수 있음

원하는 수의 가상 서버를 구축하고 보안 및 네트워킹을 구성하며 스토리지를 관리가능

확장 또는 축소를 통해 요구사항을 변경하거나 사용량 스파이크 처리가 가능해 트래픽을 예측해야할 필요성이 줄어듦



## 04. 프로젝트 수행 결과

### 결과 제시 ① WebApplicationServer



#### Docker와 AWS EC2를 통한 배포 과정

1. 프로젝트 Github에 업로드
2. Dockerfile 작성
3. AWS EC2 로그인 후 키페어 생성
4. AWS EC2 인스턴스 생성 (Ubuntu)
5. 터미널에서 ssh와 키페어를 이용해 EC2 인스턴스 접속 후 Docker 설치
6. 인스턴스 내부로 Dockerfile 이동 후 Docker image 빌드
7. 빌드된 image로 Docker Container 생성 및 Run (8000번 port로 진행했으나 변경 가능)
8. 연결할 도메인 구매
9. AWS Route 53 호스팅 영역 생성 (구매한 도메인 등록)
10. 생성된 호스트 영역에서 레코드 생성
11. 생성된 레코드에서 확인 가능한 도메인 네임서버 도메인 구매한 서비스에서 연결
12. 포트포워딩 (80 -> docker container port)
13. AWS EC2 인스턴스 보안 -> 인바운드 규칙에서 80번 포트 설정



## 04. 프로젝트 수행 결과

결과 제시 ⑤ 시연 동영상

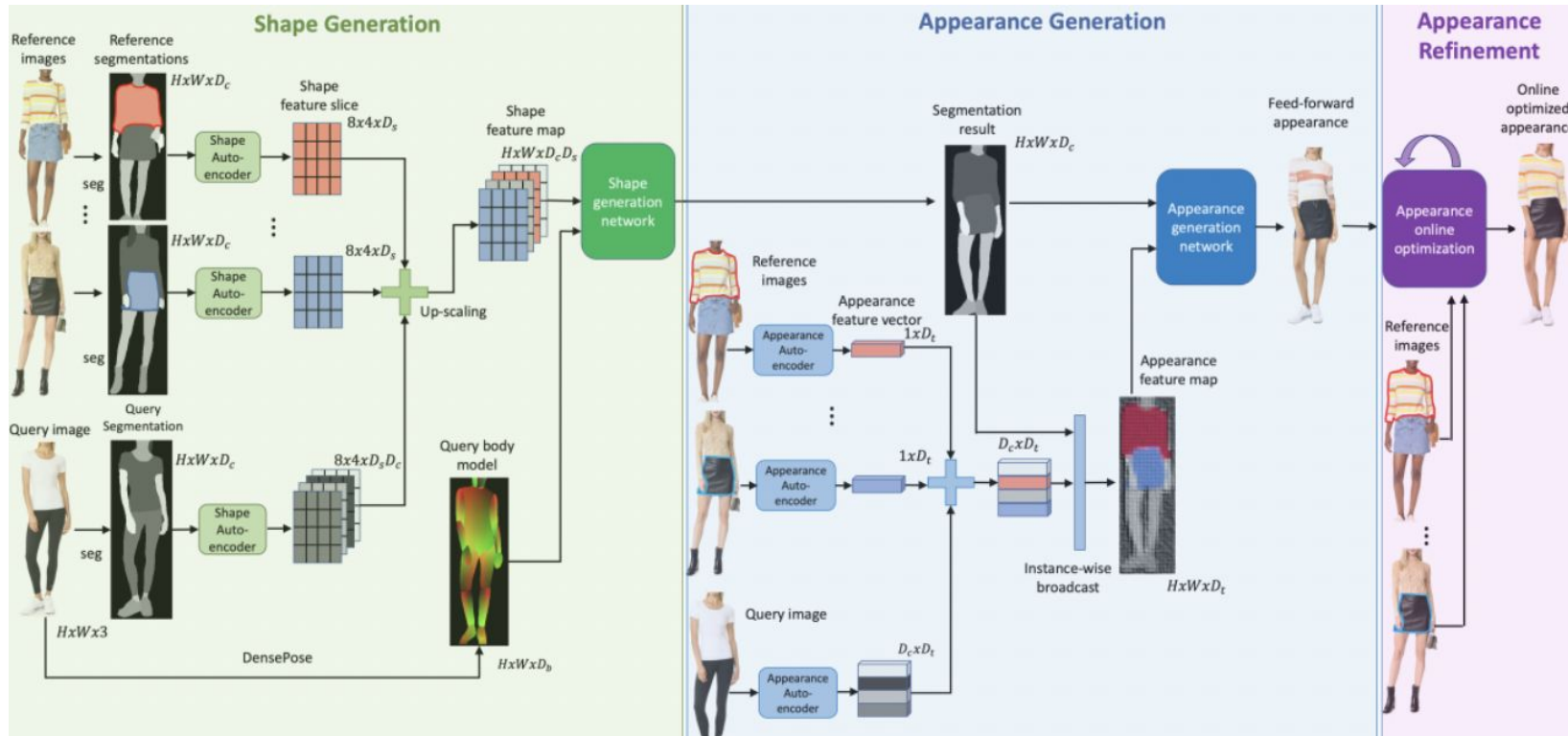
---

※시연 동영상을 넣을 경우의 화면 구성 예시, 동영상은 별도 제출 가능



## 05. 자체 평가 의견

### 전체적인 프로젝트 보완점 및 아쉬운점



Stylegan2 Model을 사용하여 가상핏팅을 구상. 하지만 시간적, 학문적으로 부족하여 접근하기 어려움

## 05. 자체 평가 의견

### ▶ 전체적인 프로젝트 보완점 및 아쉬운점



시티보이 빅오버 옥스포드 셔츠  
\_Royal Blue



구매

레귤러 핏 데님 팬츠 [미디엄 인디  
고]



구매

컬러(color), 핏(fit) 까지 분류하는 모델링을 했으면 사용자들에게 보다 높은 신뢰도를 얻었을 것

## 05. 자체 평가 의견

### ▶ 개인별 회고

훈련생	아쉬운점
정승기	데이터 수집 시 사용자들의 구매기록과 의상에 대한 지식(핏, 컬러 등)이 있었으면 조금 더 정교한 모델이 완성 되었을 것
전형준	로컬에서는 .env를 사용해서 보안성을 높였으나,docker 이미지 빌드 시 적용하는 방법을 찾지 못해 적용하지 못함으로써 보안상 취약점이 존재
이수현	Django의 강점 중 하나인 ORM 사용을 해보지 못해 아쉬운 점이 남음 단순 쿼리문을 통해서 결과를 도출하는 것 보다 GAN 모델 등을 통해 이미지 유사도를 이용했으면 어땠을까 하는 아쉬움이 남음
강선민	DJango 와 ElephantSQL을 연동한 후, 컨백션의 제한과 용량 부족으로 작업이 일시적으로 원활하지 못했던 점이 아쉬움
한유성	단일 라이브러리가 아닌, Scrapy 나 BeautifulSoup을 활용하였으면 더욱 효율적인 데이터 수집이 되었을 것



감사합니다.