

# 인공지능 데이터 구축·활용 가이드라인

## — K-Fashion 이미지 AI 데이터 구축 분야 —

인공지능 데이터 구축	주관기관	Opinion LIVE
	참여기관	weary NEXT TREND
	참여기관	ai.m
	참여기관	 이화여자대학교 EWHHA WOMANS UNIVERSITY
	참여기관	KRIFi 한국패션산업연구원
가이드라인 작성	오피니언라이브	조인호
	웨얼리	김사현
	에이아이닷엠	최예림
	한국패션산업연구원	김윤영
가이드라인 버전	Ver 1.04 2020. 11. 24	

# 목 차

<b>1. 데이터 구축 개요 .....</b>	<b>1</b>
<b>2. 임무 정의 .....</b>	<b>1</b>
2.1 임무 정의 .....	1
2.2 데이터 구축 유의사항 .....	1
<b>3. 획득·정제 .....</b>	<b>2</b>
3.1 원시데이터 선정 .....	2
3.2 획득·정제 절차 .....	2
3.3 획득·정제 기준 .....	2
3.4 획득·정제 조직 .....	2
3.5 획득·정제 도구 .....	2
<b>4. 어노테이션/라벨링 .....</b>	<b>2</b>
4.1 어노테이션/라벨링 절차 .....	2
4.2 어노테이션/라벨링 기준 .....	3
4.3 어노테이션/라벨링 조직 .....	4
4.4 어노테이션/라벨링 도구 .....	4
<b>5. 검수 .....</b>	<b>5</b>
5.1 검수 절차 .....	5
5.2 검수 기준 .....	5
5.3 검수 조직 .....	5
5.4 검수 도구 .....	6
5.5 기타 품질관리 활동 .....	6
<b>6. 활용 .....</b>	<b>6</b>
6.1 활용 모델 .....	6
6.1.1 모델 학습 .....	6
6.1.2 서비스 활용 시나리오 .....	6
6.2 접근 .....	6
6.3 유지보수 .....	6
붙임1 인공지능 데이터 명세서 양식 .....	33
붙임2 인공지능 데이터 명세서 작성 예시 .....	35

# 1. 데이터 구축 개요



그림 6. 데이터 구축 개요

## 1.1 기관별 업무

기관명	주요업무
(주)오피니언라이브	- 저작도구 등 시스템 개발 및 관리 - 인공지능 응용서비스 프론트 개발
주식회사 웨얼리	- 패션 레이블 선정 및 레이블링 작업 관리 - 원천데이터 확보 및 클라우드 소싱 - 원천데이터 기준 설정, 품질관리
주식회사 에이아이닷엠	- 인공지능 서비스 기획 및 모델 개발 - AI 인력에 대한 교육
이화여자대학교	- 패션 스타일 정의 - 이미지 어노테이션 - 이미지 레이블링
한국패션산업연구원	- 데이터 및 레이블링 결과 검수

## 1.2 데이터 구축 프로세스

### 1.2.1 원천데이터 확보

#### 1.2.1.1 원천데이터 정의

- 원천데이터는 실용성을 확보하기 위하여 여성복(20~30대)을 대상으로한 이미지 데이터
- 패션 스타일, 색상, 디자인 등 패션전문가 그룹의 의견에 따라 분류된 다양한 패션 데이터

#### 1.2.1.2 원천데이터 확보

- 제조사, 유통사 등 패션 관련 기업들의 데이터 중 저작권이 확보된 이미지를 구매하여 확보

### 1.2.1.3 원천데이터 지식재산권 검토

- 인공지능 전문 법무법인 등을 통한 지식재산권 검토

### 1.2.1.4 원천데이터 제공

- 수집된 데이터를 향후 AIHub 등을 통해 레이블링 정보와 함께 제공

## 1.2.2 이미지 레이블링

### 1.2.2.1 레이블링 속성 개발

- 스타일, 카테고리, 아이템, 컬러 속성, 프린트 속성 등 패션전문가 그룹을 통한 레이블링 속성 선정
- 레이블링 속성 개발 시 아이템, 패턴, 컬러 등 객관적 기준이 명확한 속성 외에 스타일 등 객관적 기준이 명확하지 않은 속성에 대하여 명확한 스타일 정의 개발
- 레이블링 속성 중 스타일 정의 개발 시 패션 관련 학과 교수 및 현직 실무자 등 전문가들 로 이루어진 자문단을 구성하여 스타일 분류 기준 정립
- 스타일 분류를 포함한 모든 레이블링 속성 분류 기준은 레이블링 작업자들이 숙지 할 수 있도록 충분한 사전 교육 진행

### 1.2.2.2 레이블링 진행

- 패션전문가 그룹이 지정한 레이블링 속성을 대상으로한 저작도구 개발
- 레이블러의 등급과 권한을 통한 그룹별 레이블링 진행
- 일반인과 준전문인력, 전문인력을 통한 3단계 레이블링 진행

### 1.2.2.3 레이블링 결과 검수

- 레이블링 우수인력을 통한 1차 검수 진행
- 전문인력과 전문가 그룹을 대상으로 2차 검수 진행

## 1.2.3 AI 모델 및 서비스 개발

### 1.2.3.1 AI 모델 개발

- 인공지능 학습용 데이터 구축에 따른 데이터 검토 및 응용서비스 구축 가이드를 위한 모델 개발 진행

### 1.2.3.2 AI 학습 데이터 설계

- 패션 인공지능 학습용 데이터 외 추가로 필요한 데이터가 있는지를 확인하고 해당 데이터를 포함한 인공지능 학습용 데이터 설계

## 1.2.4 응용서비스 개발

### 1.2.4.1 응용서비스 기획 및 개발

- 인공지능 학습용 데이터 구축에 따른 데이터 검증 및 데이터 활용 가이드를 위해 응용 서비스 기획 및 개발 추진
- 구축된 학습용 데이터와 모델을 접목한 응용서비스 기획

### 1.2.4.2 응용서비스 제공

- 구축된 학습용 데이터와 응용서비스용 인공지능 모델을 웹페이지 형태로 AIhub 등으로 누

구나 쉽게 활용할 수 있도록 제공

## 1.2.5 품질 검증

### 1.2.5.1 구축된 AI 데이터 품질 검증 및 인증

- 한국정보통신기술협회(TTA)에서 마련한 AI 데이터 구축 분야별 표준/가이드라인을 통해 구축 단계별 품질검증 실행
- 데이터 전문가로 구성된 외부 자문단을 위촉하여 데이터 구축 과정의 단계별 품질 점검 사항에 대한 전문가 의견을 바탕으로 데이터 품질 확보 활동

## 2. 임무 정의

### 2.1 임무 정의

#### 2.1.1 데이터 구축 배경

##### 2.1.1.1 국내 패션 관련 데이터의 부재

##### 2.1.1.2 의류 업체의 상품 이미지 데이터 체계적 구축 및 판매 방안 필요

#### 2.1.2 해결 과제

##### 2.1.2.1 패션 트렌드 연구개발을 위한 기초 데이터로 공개된 학습용 데이터 구축

##### 2.1.2.2 한국형 AI 학습용 데이터 구축

##### 2.1.2.3 지능형 패션산업 실현을 위한 AI 활용 이미지 데이터 확보

#### 2.1.3 데이터 구축 목적

##### 2.1.3.1 다양한 패션 이미지 원본 데이터에 대한 속성정보 추출 및 레이블링을 통하여 검증된 K-Fashion 이미지 AI 데이터셋 구축 및 배포

##### 2.1.3.2 구축된 K-Fashion 이미지 AI 데이터셋을 기반으로 레이블 추천방식(label recommendation)을 도입한 레이블러와 스타일 예측 모델 개발 및 배포

##### 2.1.3.3 저작도구와 스타일 예측 모델을 이용한 스타일 예측 및 추천 서비스 구현 및 API 배포

### 2.2 데이터 구축 유의사항

#### 2.2.1 데이터 구축 유의사항

##### 2.2.1.1 패션 이미지 데이터를 확보하는 과정에서 원본 및 이미지 데이터셋 공개와 재사용에 제한이 없도록 초상권 및 저작권 문제를 완전히 해결한 원천데이터 확보

필수 해결 항목	해결 방안
저작권	<ul style="list-style-type: none"><li>• 원천 데이터 저작권 및 사용권이 확보된 업체 또는 개인과 이미지 사용 동의서 작성 후 사용</li></ul>
초상권	<ul style="list-style-type: none"><li>• 원천 데이터에 포함된 모델의 얼굴은 삭제 또는 모자이크로 초상권 문제 해결</li></ul>

- 2.2.1.2 단순한 색상, 모양 등 특정 속성정보로만 구성된 기존 이미지 데이터셋과 다른 원문 데이터의 다양성 추구
- 2.2.1.3 다중 레이블과 디테일 레이블을 포함하는 데이터셋 구성
- 2.2.1.4 패션 이미지의 다양성 확보를 통해 인식 알고리즘의 성능 향상 추구

### 3. 획득·정제

#### 3.1 원시데이터 선정

##### 3.1.1 데이터 종류

3.1.1.1 시각지능 학습에 활용이 가능한 여성복(여성 20~30대) 데이터 선정

3.1.1.2 사람이 실제 착용한 실제 사진을 원천데이터로 선정

- 픽토그램, 일러스트로 만들어진 데이터가 아닌 실제 의상을 촬영한 이미지 데이터 포맷



3.1.1.3 학습이 용이하도록 다양한 스타일의 패션 이미지 활용

3.1.1.4 초상권, 저작권이 해결된 패션 이미지 데이터(마네킹, 크롭이미지)

3.1.1.5 데이터 포맷 : jpeg, jpg, png, tiff 등 레스터 이미지 파일

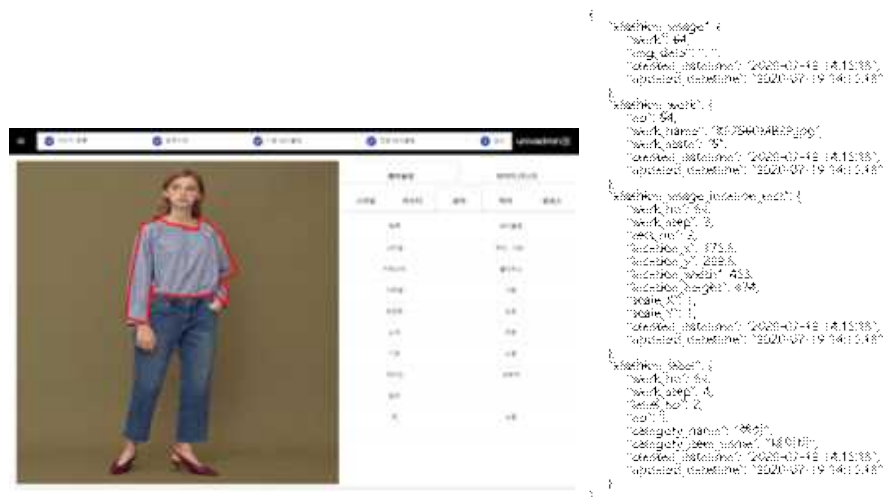


그림. 패션 이미지 데이터 및 파일 구조 예시

##### 3.1.2 데이터 확보방안

###### 3.1.2.1 직접확보

- 스튜디오 촬영으로 데이터 직접 확보
- 간접확보로 구축된 스타일별 데이터 수량 중 확보된 데이터 수량이 부족한 스타일에 대한 직접 촬영 확보로 학습 데이터 구축

### 3.1.2.2 간접확보

- 제조사, 유통사, 패션 관련 기업들의 데이터 중 저작권이 확보된 이미지 데이터 구매
- 여성 쇼핑몰 110개 업체에서 스타일별 데이터 수량 분배하여 구축



그림 10. 원천데이터 구매 이미지 샘플

### 3.1.2.3 데이터 규모

- 학습이 가능한 저작권과 초상권 문제가 해결된 확보 데이터(이미지 파일) 120만건
- 시맨틱 세그멘테이션을 통한 전문레이블(연구용 정보 포함) 데이터 120만건 구축

## 3.2 획득·정제 절차

### 3.2.1 데이터 획득·정제 단계별 흐름도



### 3.2.2 각 데이터에서 학습용 패션 이미지만 추출

3.2.2.1 데이터 중복 방지를 위하여 직접 의류를 디자인하고 제작하는 쇼핑몰을 대상으로 원시 데이터 확보

3.2.2.2 획득한 데이터의 중복 및 패션 이미지가 아닌 데이터 확인 후 제거

3.2.3 시맨틱 세그멘테이션 데이터를 활용하여 패션 이미지 데이터 추출(클라우드소싱 및 자체개발)




### 3.3 획득·정제 기준

3.3.1 레이블링 효율성을 위하여 원천데이터에서 중복 및 패션과 무관한 이미지 제거

3.3.1.1 원천 데이터(이미지) 권장 기준

이미지 크기	300px x 300px ~ 1080px x 1080px
이미지 해상도	72dpi ~ 300dpi
이미지 형식	JPG, JPEG, PNG, GIF, BMP, TIF, TIFF, WEBP
이미지 비율	1:2 ~ 2:1
이미지 형태	모델 및 마네킹 착용 형태 (모델 얼굴은 크롭 또는 모자이크)

3.3.1.2 원천 데이터(이미지) 제한 기준

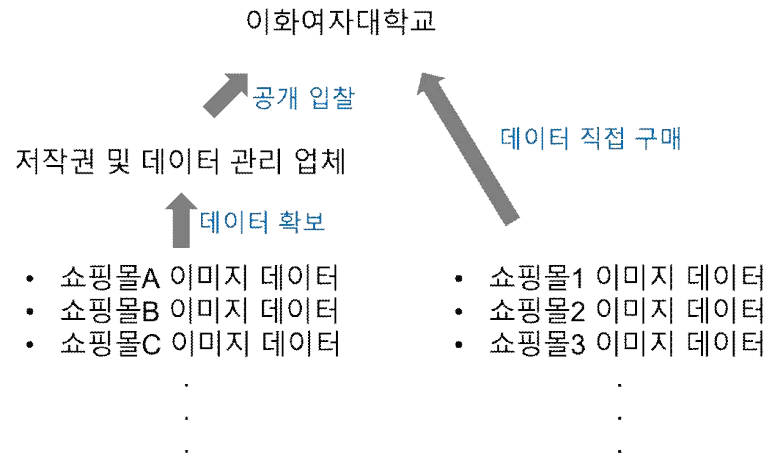
유형	예시1	예시2
워터마크		
콜라주		
잘림		
겹침		
제품겹침		



색상오류		
색상필터		
흐림/아웃 포커싱		
부분확대/ 축소		
자세변형		
저해상도		

3.3.2 모델의 얼굴이 노출된 이미지의 경우 레이블링 작업시 크롭 작업 진행

### 3.4 획득·정제 조직



획득·정제 단계	담당 조직	역할	참여 인원
1단계	쇼핑몰 A,B,C ...	• 원천 패션 데이터 전달	100여개 업체
2단계	데이터 공급 업체	• 확보된 데이터 중 제한 기준 데이터 분류 • 데이터 정제 조직 전달	공급 업체 내부 분류 (2~4인)
3단계	데이터 정제 조직	• 초상권 삭제, 데이터 리사이징 • 권장 데이터 취합 후 DB 저장 • 레이블링 조직 전달	3~4인
4단계	레이블링 조직	• 권장 기준 미달 데이터 1차 사전 검수	일반 레이블링 작업자 200명

### 3.5 획득·정제 도구

#### 3.5.1 패션 이미지 간접확보

3.5.1.1 제조사, 유통사 등 패션관련 기업들의 데이터 중 저작권이 확보된 이미지 구매

3.5.1.2 확보된 이미지를 저작도구를 통하여 1차 정제 수행(not fashion 이미지 선별)

## 4. 데이터 저작도구 활용

### 4.1 데이터 가공 및 검수의 효율성을 위해 웹 형태의 가공 도구 개발 및 가이드라인 제공

4.1.1 이미지 레이블링 작업 중 레이블러들의 일관성 있는 작업을 위한 가이드라인 제공

4.1.2 레이블러 관리 푸시 버튼을 통해 진척률, 오류율 등을 제공하여 용이한 관리

4.1.3 작업 유형에 일관성 있는 작업을 진행 할 수 있도록 유형별 가이드라인 제공

4.1.4 알고리즘 기반의 레이블 추천 기술을 통하여 레이블러의 작업을 지원

4.1.5 리커트 척도를 활용한 검수 및 평가를 통하여 불량 작업자 및 검수자 체크

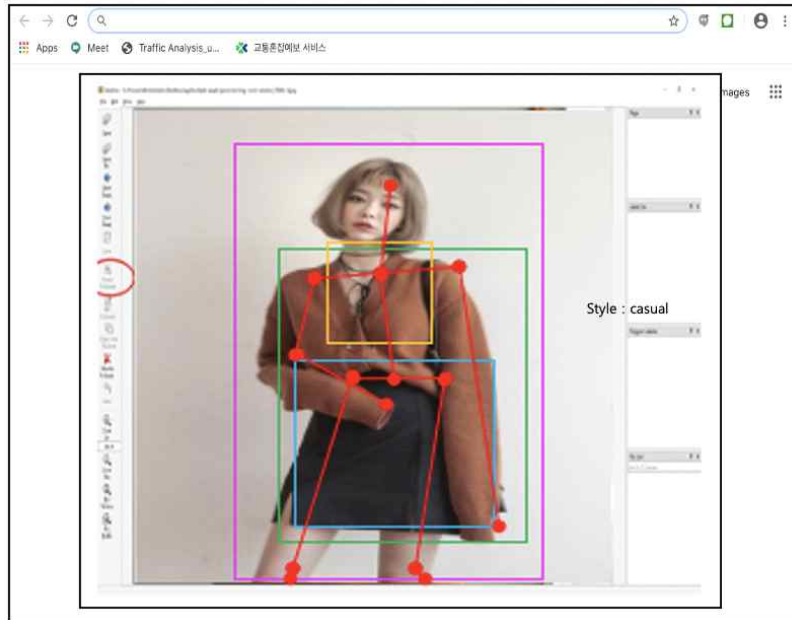


그림 35 클라우드소싱을 위한 웹기반 검수 도구 화면(예시)

## 5. 어노테이션/라벨링

### 5.1 어노테이션/라벨링 절차



- 이화여대 의류학과 석박사(패션전문가그룹), 마이스터고(일반그룹), 클라우드소싱(일반그룹) 참여자 교육, 레이블, 평가
- 패션전문가그룹은 주관적 판단이 요구되는 패션 스타일 다중레이블링 업무 수행 (소요 시간, 정확성 모니터링)
- 정확한 판단이 어려운 데이터의 경우 여러 연구자가 동일한 작업을 수행하는 cross checking 을 통해 합의

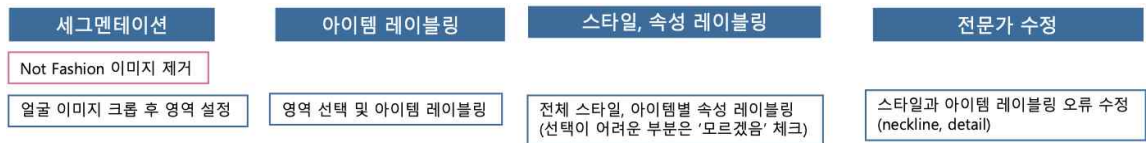


그림. 이미지 레이블링 가이드라인 구조

## 5.2 어노테이션/라벨링 기준

### 5.2.1 데이터 패션 스타일 분류 체계

#### 5.2.1.1 계층구조를 활용한 10가지 베이직 스타일, 23가지 하위 스타일을 정의하여 분류



그림 38. 스타일 속성 체계

#### 5.2.1.2 패션 스타일 속성 분류

상위 스타일	하위 스타일	속성 특징(12속성중 4가지)			
		(1)카테고리	(2)아이템	(3)컬러 속성	(5)프린트 속성
클래식	클래식	탑, 코트, 재킷, 팬츠, 스커트	샤넬 재킷, 트렌치 코트, 셔츠, 셔츠 드레스, 가디건	버건디, 카멜, 네이비, 브라운, 그레이, 딥톤	체크, 헤링본, 스트라이프, 무지, 하운드투스
	프레피	탑, 코트, 재킷, 팬츠, 스커트, 드레스	블레이저, 폴로 셔츠, 케이블니트, V넥 스웨터, 더플 코트, 가디건, 재킷, 테니스 스커트	네이비, 블루, 화이트, 버건디	체크, 헤링본, 스트라이프,
매니시	매니시	탑, 코트, 재킷, 팬츠	체스터필드코트, 박스코트, 셔츠, 베스트, 슬렉스, 핀턱팬츠	그레이, 그린, 브라운, 블랙, 다크 톤, 그레이쉬 톤	기하학, 스트라이프
	툼보이	탑, 코트, 재킷, 점퍼, 팬츠, 점프수트	루즈핏 팬츠, 티셔츠, 청바지, 오버롤즈	블랙, 브라운, 그린, 다크 톤	스트라이프, 체크, 헤링본
엘레강스	엘레강스	탑, 코트, 재킷, 스커트, 드레스	롱 드레스, 롱 스커트 블라우스, 플레어 스커트, 머메이드 드레스	퍼플, 그레이, 베이지, 크림, 그레이 톤, 네	무지, 체크, 스트라이프

	소피스트케이트	탑, 코트, 재킷, 팬츠, 스커트, 드레스	H라인 스커트, 블라우스, H라인 스커트, 펜슬 스커트, 드레이프 드레스	이비 그레이, 스카이블루, 모노톤, 딥톤	무지, 체크, 스트라이프
	글래머러스	탑, 코트, 재킷, 팬츠, 스커트, 드레스	타이트 드레스, 블라우스, 플레어 스커트, 드레이프 자켓	그레이, 골드, 실버, 크림, 네이비	무지, 애니멀
에스닉	에스닉	탑, 재킷, 팬츠, 스커트, 드레스	드레스, 루즈핏 블라우스, 판초	레드, 그린, 옐로우, 딥톤, 비비드 톤	페이즐리, 플로럴
	히피	탑, 코트, 스커트, 드레스	로브, 케이프, 가디건, 청바지, 스웨터, 홀터넥 드레스, 선 드레스	베이지, 블루, 레드, 오렌지, 와인	타이다이, 플로럴, 호피, 지브라
	오리엔탈	탑, 코트, 재킷, 팬츠, 스커트, 드레스	스탠드 칼라 재킷, 차이나 칼라 재킷, 노칼라 재킷, 하렘팬츠, 로브	블랙, 골드, 빨강, 그린, 퍼플, 딥 톤	플로럴
모던	모던	탑, 코트, 재킷, 팬츠, 스커트, 드레스	셔츠, 박스코트, 이레귤러 스커트, 스트레이트 팬츠	블랙, 화이트, 그레이, 실버, 네이비, 블루, 그레이쉬 톤	기하학, 옵아트, 큐비즘, 무지
	미니멀	탑, 코트, 재킷, 팬츠, 스커트, 드레스	노칼라재킷, 슈미즈 드레스, 슬리브리스, 싱글코트	화이트, 블랙, 모노톤, 그레이쉬 톤	스트라이프
내추럴	내추럴	탑, 코트, 재킷, 팬츠, 스커트, 드레스	니트탑, 스웨터, 싱글코트, 와이드 팬츠, 슈미즈 드레스, 가디건	베이지, 카키, 브라운, 그레이쉬 톤, 소프트 톤	체크, 스트라이프
	컨트리	탑, 코트, 재킷, 팬츠, 스커트, 드레스	니트탑, 스웨터, 싱글코트, 와이드 팬츠, 슈미즈 드레스, 가디건	베이지, 카키, 브라운, 그레이쉬 톤, 소프트 톤	체크, 아가일
	리조트	팬츠, 스커트, 드레스	루즈핏 블라우스, 롱 드레스, 와이드 팬츠, 슬리브리스	베이지, 카키, 브라운, 핑크, 스카이블루, 소프트 톤	플로럴, 스트라이프
로맨틱	로맨틱	탑, 코트, 재킷, 스커트, 드레스	A라인 드레스, 플레어 스커트, 티어드 스커트, 블라우스, 프린세스 드레스, 피터팬 칼라 셔츠	옐로우, 핑크, 블루, 퍼플, 페일 톤, 페일 톤, 소프트 톤	체크, 무지, 도트, 스트라이프, 플로럴
	섹시	탑, 팬츠, 스커트, 드레스	슬림 드레스, 미니 드레스, 오프숄더 탑, 숏 팬츠, 슬림 탑, 뷔스티에, 브라탑, 원숄더 드레스	블랙, 화이트, 레드, 실버, 골드	호피, 지브라
스포티	스포티	탑, 코트, 재킷, 점퍼, 팬츠, 점프수트	티셔츠, 블루종, 스타디움, 트레이닝팬츠, 아노락, 맨투맨, 조거 팬츠, 후드탑, 후드집업	레드, 블루, 그린, 그린, 비비드 톤	레터링, 로고, 그래픽, 도트

	애슬레저	탑, 코트, 재킷, 점퍼, 팬츠	요가 팬츠, 크롭탑, 브라탑, 조거 팬츠, 레깅스, 트레이닝 팬츠, 윈드 브레이커	블랙, 화이트, 베이지, 그린	로고, 그래픽,
	밀리터리	탑, 코트, 재킷, 점퍼, 팬츠, 점프수트	블루종, 야상, 피코트, 카고팬츠	카키, 브라운, 블랙, 베이지, 딥 톤	카무플라주
문화	뉴트로	탑, 코트, 재킷, 점퍼, 팬츠, 스커트, 드레스, 점프수트	오버사이즈 핏 재킷, 크롭탑, 부츠컷 팬츠, 벨보텀/플레이 팬츠	네온 컬러, 그레이쉬 톤	레터링, 로고, 그래픽, 도트
	힙합	탑, 점퍼, 팬츠	점퍼, 캐주얼 상의, 청바지	블랙, 화이트, 비비드 톤, 다크 톤	레터링, 무지, 해골
	키치/키덜트	탑, 코트, 재킷, 점퍼, 팬츠, 스커트, 드레스, 점프수트	티셔츠, 니트탑, 스커트, 드레스, 오버롤즈	옐로우, 퍼플, 오렌지, 레드, 비비드 톤	플로럴, 체크, 기하학, 스트라이프
	맥시멈	탑, 코트, 재킷, 점퍼, 팬츠, 스커트, 드레스	다양한 아이템의 레이어드	레드, 블루, 그린, 옐로우, 네이비, 퍼플, 비비드 톤	무지, 플로럴, 체크, 스트라이프, 타이다이, 체크, 플로럴, 호피, 지브라
	핑크/로커	탑, 코트, 재킷, 팬츠	스웨터, 셔츠, 후리스, 가디건, 데님 팬츠, 치노팬츠	블랙, 화이트, 카키, 그레이, 블루	무지, 해골, 그래픽
캐주얼	캐주얼	탑, 코트, 재킷, 팬츠, 스커트, 드레스	티셔츠, 점퍼, 니트탑, 스웨터, 청바지, 조거 팬츠, 가디건, 집업	블랙, 화이트, 레드, 블루, 그린, 옐로우, 네이비, 비비드 톤	무지, 스트라이프, 그래픽, 로고
	놈코어	탑, 코트, 재킷, 팬츠	스웨터, 셔츠, 후리스, 가디건, 배기 팬츠, 시가렛 팬츠	블랙, 화이트, 그레이쉬 톤	체크, 스트라이프

## 5.2.2 데이터 속성 분류 체계

5.2.2.1 FGI(focus group interview)를 통하여 상위 속성 종류를 정의하고 세부 속성 항목을 구조화

5.2.2.2 11가지 상위 속성 종류 분류

5.2.2.3 1,000여가지 하위 속성 항목 중 비슷한 항목을 그룹화하고 부정확한 속성 항목들을 구조화하여 분류 체계 구축

5.2.2.4 카테고리 속성 정의 : 카테고리는 상품 이미지 태깅 결과의 패션 상품을 분류하는 상위 체계의 속성

Category(EN)	Category(KO)	Category(EN)	Category(KO)
tops	탑	jumpers	점퍼
blouses	블라우스	paddings	패딩
casual-tops	캐주얼상의	jeans	청바지
knitwear	니트웨어	pants	팬츠
shirts	셔츠	skirts	스커트
vests	베스트	dressess	드레스
coats	코트	jumpsuits	점프수트
jackets	재킷	swimwear	수영복

#### 5.2.2.5 컬러 속성 정의 : 상품의 색상 정보를 고객들의 사용도가 높은 색상으로 제공

Attribute(EN)	Attribute(KO)	Attribute(EN)	Attribute(KO)
black	블랙	khaki	카키
white	화이트	mint	민트
grey	그레이	blue	블루
red	레드	navy	네이비
pink	핑크	skyblue	스카이블루
orange	오렌지	purple	퍼플
beige	베이지	lavender	라벤더
brown	브라운	wine	와인
yellow	옐로우	neon	네온
green	그린	gold	골드

#### 5.2.2.6 디테일 속성 정의 : 상품을 구성하고 있는 부분적인 특징을 디테일이라는 속성값으로 제공

Attribute(EN)	Attribute(KO)	Attribute(EN)	Attribute(KO)
beads	비즈	fur-trimmed	퍼트리밍
buttoned	단추	glitter	글리터
cable	니트짜매기	hood	후드
chain	체인	lace-up	레이스업
cut-off	컷오프	padding	패딩
double-breasted	더블브레스티드	patchwork	패치워크
drop-shoulder	드롭숄더	peplum	페플럼
embroidery	자수	pleats	폴리츠
frill	프릴	pockets	포켓
fringe	프린지	pompom	폼폼
flare	플레어	puff	퍼프
quilting	퀼팅	tassel	태슬
ribbon	리본	zip-up	집업
roll-up	롤업	sash	띠
ruffle	러플	destroyed	디스트로이드
shirring	셔링	low-waist	드롭웨이스트
slit	슬릿	buckle	버클
spangle	스팽글	cut-out	컷아웃
stitch	스티치	x-strap	X스트랩
stud	스터드	unbalanced	비대칭

shirring	셔링		
----------	----	--	--

#### 5.2.2.7 프린트 속성 정의 : 상품에서 자주 쓰이는 패턴을 프린트라는 속성값으로 제공

Attribute(EN)	Attribute(KO)	Attribute(EN)	Attribute(KO)
check	체크	floral	플로럴
stripe	스트라이프	lettering	레터링
zigzag	지그재그	skull	해골
leopard	호피	tie-dye	타이다이
zebra	지브라	gradation	그라데이션
dot	도트	solid	무지
camouflage	카무플라주	graphic	그래픽
paisley	페이즐리	Hound's tooth	하운즈 투스
argyle	아가일	gingham	강엄

#### 5.2.2.8 소재 속성 정의 : 상품을 구성하고 있는 소재를 소재감이라는 속성값으로 제공

Attribute(EN)	Attribute(KO)	Attribute(EN)	Attribute(KO)
fur	퍼	knit	니트
mouton	무스탕	lace	레이스
suede	스웨이드	linen	린넨
angora	앙고라	mesh	메시
corduroy	코듀로이	fleece	플리스
sequin/glitter	시퀸/글리터	neoprene	네오프렌
denim	데님	silk	실크
jersey	저지	spandex	스판덱스
tweed	트위드	jacquard	자카드
velvet	벨벳	leather	가죽
vinyl/pvc	비닐/PVC	cotton	면
wool/cashmere	울/캐시미어	chiffon	시폰
synthetic/polyester	합성섬유		

#### 5.2.2.9 소매기장 속성 정의 : 상품의 소매 기장 정보를 속성값으로 제공

Attribute(EN)	Attribute(KO)	Attribute(EN)	Attribute(KO)
sleeveless	민소매	3/4th-sleeve	7부소매
short-sleeve	반팔	long-sleeve	긴팔
Cap-sleeve	캡		

#### 5.2.2.10 차) 넥라인 속성 정의 : 상품의 넥라인 정보를 속성값으로 제공

Attribute(EN)	Attribute(KO)	Attribute(EN)	Attribute(KO)
round-neck	라운드넥	square-neck	스퀘어넥
u-neck	유넥	collarless	노카라
v-neck	브이넥	hood	후드
halter-neck	홀터넥	turtleneck	터틀넥
off-shoulder	오프숄더	Boat neck	보트넥
one-shoulder	원 숄더	sweetheart	스위트하트

#### 5.2.2.11 카라 속성 정의 : 상품의 카라 정보를 속성값으로 제공

Attribute(EN)	Attribute(KO)	Attribute(EN)	Attribute(KO)
---------------	---------------	---------------	---------------



shirt-collar	셔츠칼라	peterpan-collar	피터팬칼라
bow-collar	보우칼라	notched-collar	너치드칼라
sailor-collar	세일러칼라	standup-collar	차이나칼라
shawl-collar	술칼라	tailored-collar	테일러드칼라
Polo-collar	폴로칼라	band collar	밴드칼라

5.2.2.12 핏 속성 정의 : 상품의 핏 정보를 속성값으로 제공 (핏 속성 정보는 상의와 팬츠류에서 제공되어지는 속성)

Attribute(EN)	Attribute(KO)	Attribute(EN)	Attribute(KO)
normal	노멀	skinny	스키니
loose	루즈	wide	와이드
oversized	오버사이즈	tight	타이트

5.2.2.13 세이프 속성 정의 : 상품의 세이프 정보를 속성값으로 제공

Attribute(EN)	Attribute(KO)	Attribute(EN)	Attribute(KO)
peplum	페플럼	mermaid	머메이드
asymmetrical	비대칭	bell-bottom/flare	벨보텀/플레어
unknown	알수없음	boots-cut	부츠컷
pencil	펜슬	tapered	테이퍼드
a-line	A라인	straight	스트레이트
h-line	H라인		

하) 실루엣 속성 정의 : 상품의 실루엣 정보를 속성값으로 제공

Attribute(EN)	Attribute(KO)	Attribute(EN)	Attribute(KO)
X	아우터글래스	A	A
H	스트레이트	T/Y	T/Y
O	벌크		

### 5.3 어노테이션/라벨링 조직

#### K-패션 AI데이터 레이블링 그룹 기준

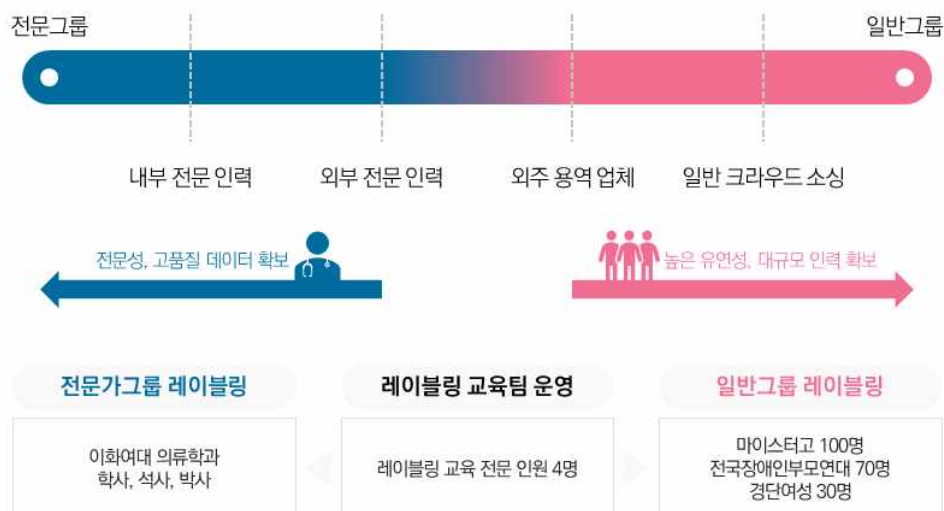


그림 39. 레이블링 참여 인력 구분

#### 5.3.1 어노테이션/라벨링 교육 및 훈련계획

5.3.1.1 이미지 레이블링 작업 중 레이블러들의 일관성 있는 작업을 위한 가이드를 통한 교육 실시

5.3.1.2 작업 유형에 일관성 있는 작업을 진행 할 수 있도록 유형별 가이드를 통한 교육 실시

## 5.4 어노테이션/라벨링 도구

### 5.4.1 도구 개요

- K-Fashion 이미지 데이터 제작 도구는 기존 배포된 데이터 가공 도구가 아닌 패션 전문가 그룹과 인공지능 전문가 그룹의 협업을 통해 구성한 어노테이션 도구이다.
- 사업 종료 후, 패션 관계자 및 인공지능 관계자, 그리고 일반인이 사용할 수 있도록 사용자 중심의 UI/UX로 구현하였고, 패션 이미지에 대한 속성 추천 기능을 추가 구현하였다.



그림 40. 어노테이션 및 라벨링 화면

### 5.4.2 주요 도구 기능

#### 5.4.2.1 이미지 자동 전처리 기능

- 이미지를 업로드하여 패션 이미지 여부를 판단하여 자동으로 삭제하는 기능과 패션 이미지에서 패션 이미지에서 필요없는 부분을 자동으로 크롭하는 기능을 제공함

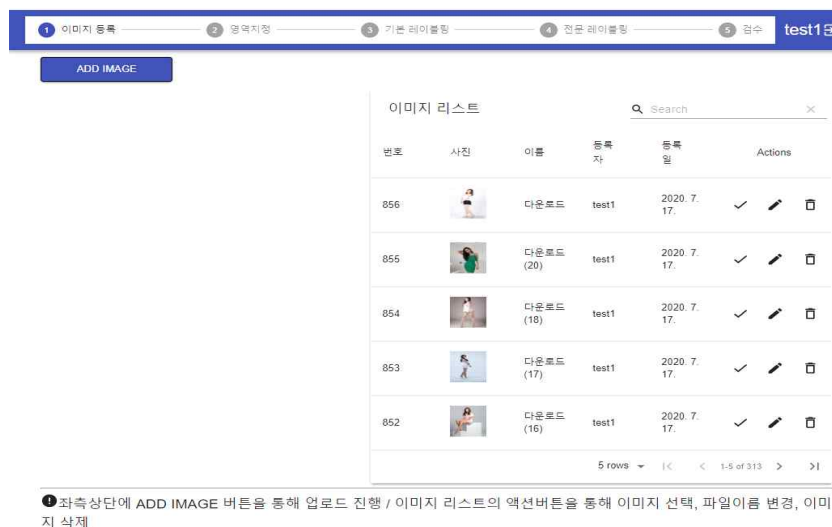


그림 41. 이미지 업로드 화면

#### 5.4.3 패션 아이템과 스타일에 대한 인식 및 후보 추천 기능

- 레이블러가 레이블링을 수행할 때 기 학습된 패션 이미지를 통해 패션 아이템과 스타일을 인식 및 후보리스트를 추천하는 기능을 제공함

#### 5.4.4 패션 이미지에 대한 패션 스타일 다중 선택

- 패션 이미지에 대한 레이블을 패션 전문가 그룹을 통해 구성하고, 패션 스타일을 레이블링 다중선택을 제공함

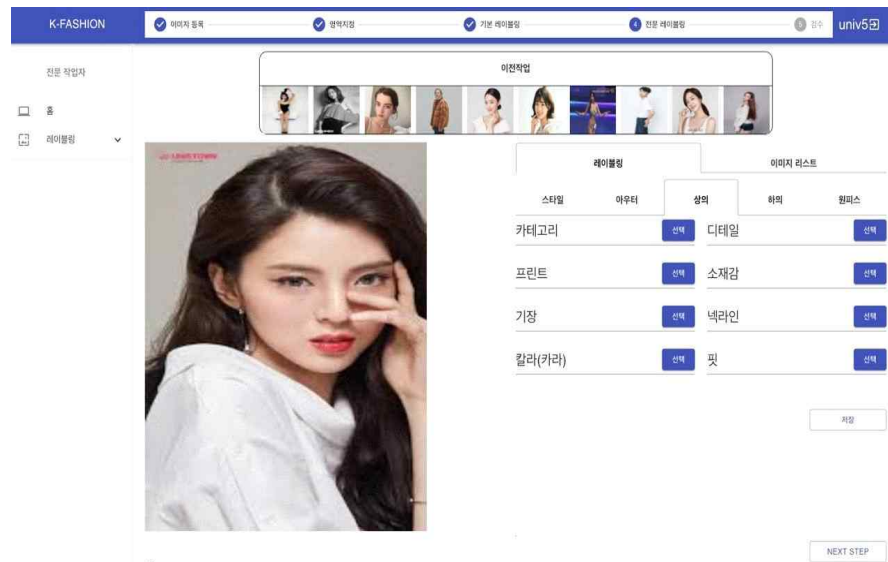


그림 42. 패션 스타일 레이블 다중 선택 화면



그림 43. 패션 스타일 다중 선택 화면

#### 5.4.5 레이블러 등급과 권한을 통한 사용자 관리

- 사용자간 레이블러 등급과 권한을 구분하여 체계적인 사용자 관리를 통해 안정적이고 효율적인 레이블링 수행을 지원함

아이디	이름	작업진도	작업지정
test01	test01	100% 0 / 0	작업지정
univ_00	univ_00	100% 0 / 0	작업지정
univ02	univ02	50.0% 1 / 2	작업지정
univ_01	univ_01	100% 0 / 0	작업지정

그룹 회원 리스트의 작업지정 버튼을 통해 해당 작업자에게 작업분배

그림 44. 관리자 등급 화면 중 작업지정 페이지

#### 5.4.6 어노테이션/라벨링 도구 배포계획

##### 5.4.6.1 중소, 벤처기업 등 외부 이미지 수집, 구축할 수 있도록 구축 관리도구 개방, 대외 제공

- 구축 완료된 이미지 데이터셋을 AI 테스트 결과와 함께 공개하여 다양한 AI 기반의 이미지 데이터 활용 기반을 제공함
- 필요한 경우 AI 개발 방법론 및 개발 프로세스를 세미나 개최 및 논문 발표 등 다양한 방법으로 공개
- 한국정보화진흥원에서 운영하고 있는 AI허브를 통해 패션 이미지 데이터셋과 스타일 예측 서비스를 개방하여 활용 활성화

### 5.5 교육 및 훈련계획

#### 5.5.1 저작도구를 사용한 대상 이미지의 전처리 및 레이블링에 대한 교육 수행

#### 5.5.2 이화여대 의류학과 석박사(패션전문가그룹), 마이스터고(일반그룹), 클라우드소싱(일반그룹) 참여자 교육 및 레이블링 및 평가

##### 5.5.2.1 마이스터고의 데이터 전공 학생들을 대상으로 데이터 분석 교육 제공 후 데이터 전처리 및 레이블링 작업 수행

- 대상인원 : 마이스터고 10곳의 데이터 전공 학생들 각 30명으로 총 300명
- 교육장소 : 지역별 참여기관이 보유한 장소를 활용, KT-경기대 빅데이터센터(경기도), 한국패션산업연구원(대구/경북)

## 6. 검수

### 6.1 검수 절차

#### 6.1.1 데이터 품질 확보를 위해 2단계 정제 및 레이블링과 3단계 데이터 전수 검수를 진행

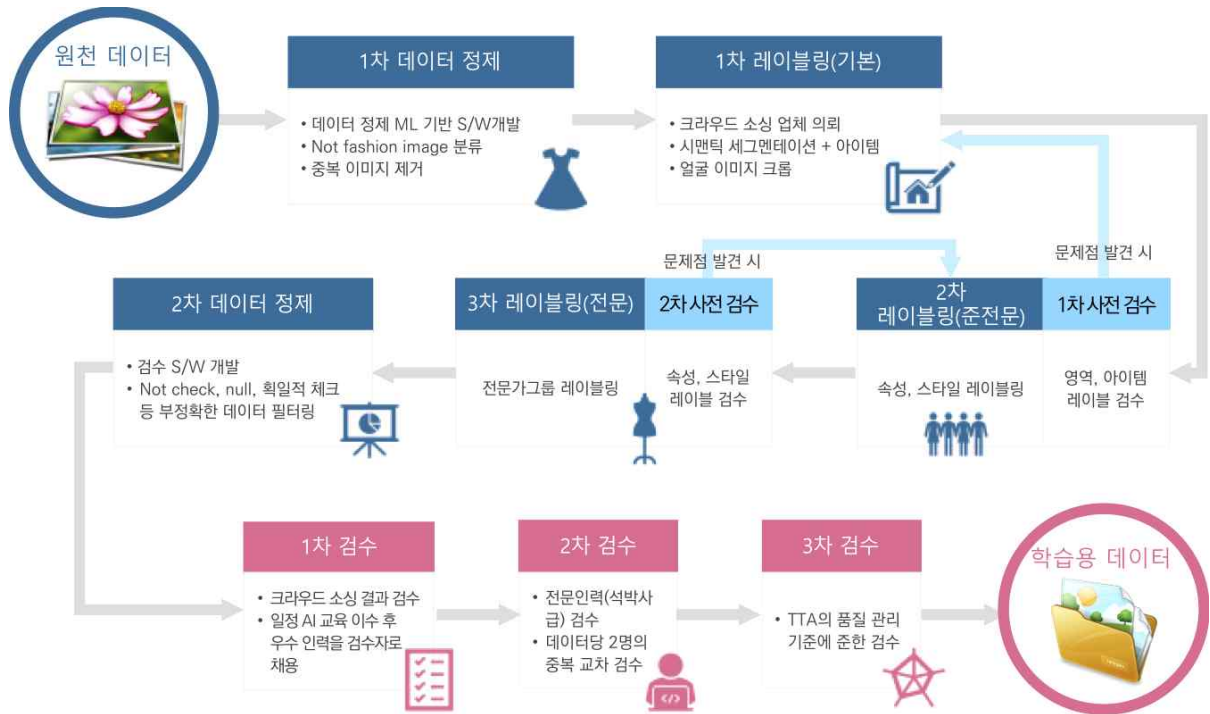


그림 45. 검수 시스템

## 6.2 검수 기준

- 6.2.1 1차 사전 검수 : 일반 레이블링 작업자들의 레이블링 작업전 육안으로 해당 데이터의 세그멘테이션 확인 (문제점 발견 시 1차 레이블링(세그멘테이션) 단계로 돌려보냄)
- 6.2.2 2차 사전 검수 : 전문가 레이블링 작업자들의 레이블링 작업전 육안으로 일반 레이블링 작업자들이 체크한 기본 속성 레이블링 확인 (문제점 발견 시 2차 레이블링 단계로 돌려보냄)
- 6.2.3 1차 검수 : 클라우드 소싱 레이블링 작업자 대상으로 일정 AI 교육 이수 후 우수 인력을 검수자로 채용
- 6.2.4 2차 검수 : 전문인력 대상으로 데이터당 2명의 검수자가 중복 교차 검수

## 6.3 검수 조직

검수 단계	검수 구성원	역할	참여 인원
1차 사전 검수	클라우드 소싱 레이블러	• 권장 기준 미달 원천 데이터 전수 검수	200명
2차 사전 검수	이화여대 전문가 레이블러	• 권장 기준 미달 원천 데이터 전수 검수 • 기본 속성 레이블링 확인	50명
1차 검수	클라우드 소싱 우수 레이블러	• 불량 레이블링 확인 분류 (오기, 일괄 표기, 미표기 등)	10명
2차 검수	한국패션산업연구원 전문가	• 학습 데이터 전수 교차 검수	2명

## 6.4 검수 도구

#### 6.4.1 저작도구의 검수 기능을 통한 레이블링 데이터에 대한 검수 진행

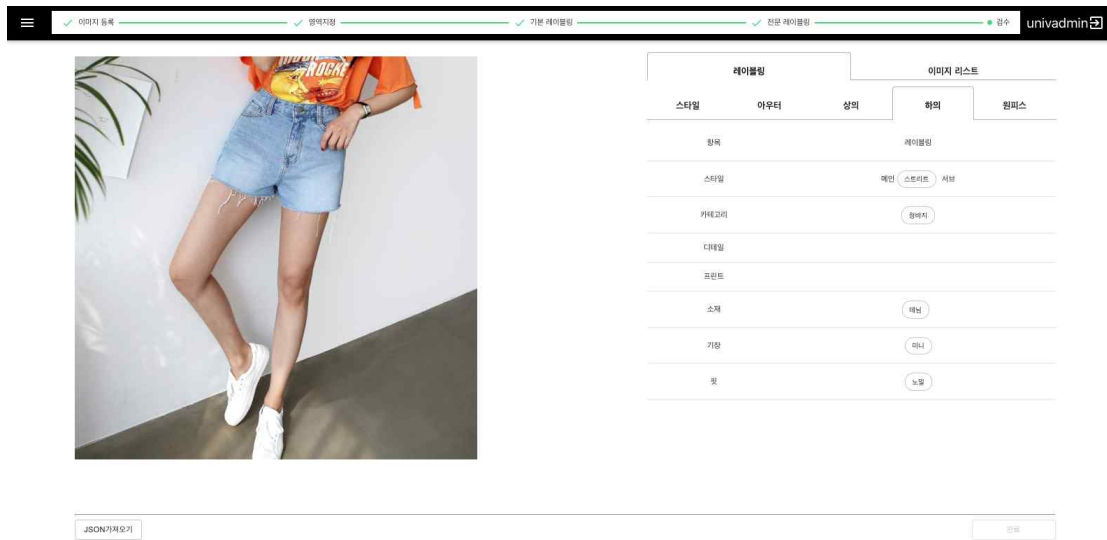


그림 46. 검수 기능을 통한 레이블링 데이터 검수 화면

생산공정	품질 관리활동	설명
검수 도구	작업 도구 정의	가공도구를 통한 검수 작업 수행 가능성 판단 및 대안 마련
	도구의 검수 기능 정의	학습용 데이터 제작 도구의 작업물 오류 확인 및 수정할 대상, 관련 자료 수집 및 저장 방안 확인
	도구 내 통계 기능 확인	학습용 데이터 제작 도구에 탑재된 통계 기능을 활용하여 관리 및 분석
	도구의 유효성 확인	검수자의 피드백을 추적 관리 방안 마련

### 6.5 기타 품질관리 활동

#### 6.5.1 AI 학습용 데이터 품질관리 및 검증 방안

##### 6.5.1.1 가공 전 전처리 시스템을 활용한 일관성 있는 품질 방안

- 참여기업 웨얼리에서 보유하고 있는 레이블링 인력을 활용하여 원본에 전처리 과정을 진행하여 효율적인 작업 방안 마련 (패션이미지 분리 및 모델의 얼굴이 포함된 데이터에 대한 크롭 등)
- 전처리 과정을 통하여 이미지 레이블링 작업에 효율성과 시간 단축이 가능하며 다양한 결과물로 모델 개발 활용

##### 6.5.1.2 이미지 레이블링 작업의 정량적 평가 방법

- 이미지 데이터((120만건)에 대한 전수 검사

##### 6.5.1.3 이미지 레이블링 작업의 정성적 평가 방법

- 전문가 그룹을 통한 이미지 데이터에 대한 평가를 진행
- 정성적인 평가 방법으로 1~5점까지 점수를 부여하는 리커트 척도 방법이 있음

### 정성평가 5점 척도 예시

점수	설명
1	이미지와 레이블 정보가 모두 이상
2	이미지와 아이템 정보가 다름
3	이미지와 스타일 정보가 다름
4	이미지와 레이블 정보가 약간의 차이를 보임
5	이미지와 레이블링 정보가 일치

#### 6.5.1.4 가공 후 검수 시스템을 통한 품질 확보



- 전문가 검수
  - ✓ 가공 경험이 있고, 1차 검수 결과 우수 가공자를 검수자로 확보함
  - ✓ 레이블링 이미지 검토 : 정량적 평가 방법을 통한 이미지 레이블링 규칙 준수 여부 확인
  - ✓ 1차 검수를 통하여 확보된 여유 인력과 패션 관련 준전문가를 통한 검수를 진행함으로써 레이블링 데이터의 품질을 확보
  - ✓ 전문가 교차 검수
  - ✓ 우수 검수자로 선정된 검수자와 전문가의 검수 결과 상호 교차 검수를 통한 데이터셋 최종 확인

#### 6.5.2 사업 진행 단계별 품질 관리 활동

6.5.2.1 인공지능 학습용 데이터를 구축함에 있어 전체 공정 단계를 기획, 설계, 수집, 정제 및 가공, 검수, 검수 후 단계로 구분하여 각 단계별로 데이터 품질을 유지 및 개선하기 위한 활동을 수행함

##### 6.5.2.2 데이터셋 정제 및 가공 단계의 품질 관리

- 확보된 이미지를 가공 과정의 프로세스, 작업 가이드, 오류 개선 활동을 중심으로 품질관리 활동을 수행함

#### 데이터셋 정제 및 가공 단계의 품질 관리 활동

생산공정	품질 관리 활동	세부설명	체크
정제/가공	정제/가공 프로세스 일정 정의	정제/가공 일정과 순서 결정	
	정제/가공 규정 표준화안 수립	데이터 구축 가이드 라인 및 각 프로세스 매뉴얼 확인	
	프로세스 입출력 요구사항 정의	가이드라인 상, 입출력 데이터의 형태 확인	
	정제/가공 프로세스 별 매뉴얼 문서화	가이드라인 프로세스별 매뉴얼 상세 확인	

	작업 요소 확인	가이드라인 상 각 데이터셋을 위한 가공 요소 정의	
	정제/가공 과정 지표 관리	작업분류, 작업량, 정확도에 관리 지표 확인	
	정제/가공 프로세스 내 자원 가용성 확인	예산 세부 계획, 인력 참여 계획, 추진일정 확인	
	프로세스 일관성 확보	정제/가공 절차의 목표 실현성과 일관성 확인	
	작업자 관리 방안	작업자 교육 방안 및 관리 방안 확보	
	오류 개선 절차 확인	정제/가공 프로세스 내에 오류 또는 작업자의 의견을 관리자와 공유할 수 있는 방안 확보	
	리스크 예방 활동	검수 프로세스 및 주기적 작업자 간담회 활동	
가공 조직 체계	가공 조직 체계 마련	작업자/검수자/책임자의 정제/ 가공 조직 구성 확인	
	조직 권한 및 책임 설정	정제/가공, 검수 팀의 책임 및 권한을 확인	
	조직 내 인력 규모 확인	수행계획서 상 업무 규모에 맞는 인력 참여 계획을 확인	
	조직 내 역할 분담	정제/가공 프로세스의 작업자/검수자/책임자의 구분과 역할분담을 확인	
	교육 훈련 방안 마련	작업 가이드라인 제공 및 교육/학습 과정 진행	
정제/가공 도구	작업 도구 정의	이미지 저작 도구 기능 정의	
	도구 대체 가능성 정의	데이터 특성에 따른 대체 가공 도구 현황 파악	
	규격 범용성 정의	가이드라인 상 해당 도구를 활용한 프로세스 및 규격 준수 데이터 포맷 확인	
	도구 내 통계 기능 확인	학습용 데이터 제작 도구에 탑재된 통계 기능을 활용하여 관리 및 분석 기능 활용 확인	
	도구의 유효성 확인	작업자의 피드백에 따라 지속 확인 및 갱신 확인	

#### 6.5.2.3 데이터셋 검수단계의 품질 관리

- 가공된 데이터를 검수하는 단계에서 프로세스, 도구, 조직 체계를 검토하여 일정한 품질의 데이터셋 구축 결과를 확인함

##### 데이터셋 검수단계의 품질 관리 활동

생산공정	품질 관리 활동	세부설명	체크
검수 프로세스	검수 프로세스 일정 정의	검수 프로세스 일정 및 순서 결정	
	검수 프로세스 내 커뮤니케이션 방안 확보	이미지 데이터 제작 도구를 통해 작업자-검수자 간의 상호작용 확인	
	검수 규정 표준화안 수립	검수 프로세스에 대한 규정이나 표준 등의 기준과 방법, 매뉴얼 등 결정	
	검수 과정 지표 관리	사업추진계획 상 작업량 및 작업 정확도 지표 제시	
	프로세스 입출력 요구사항 정의	입출력 데이터의 형태 확인	
	검수 프로세스 내 자원	예산 세부 계획, 인력 참여 계획, 추진일정 확인	



	가용성 확인		
	검수 방법 유효성 정의	검수 결과 오류 통계 및 수정사항 추적 기록 방법 확보	
	검수자 표준화 방안	- 지속적 강의 및 트레이닝 운영 - 관리자/작업자 교육 방안 수립 - 관리자/작업자 오류 전달 사항 및 의사소통 방안 확보	
	검수 방법의 적정성 판단	- 검수 프로세스 단계별 세부 수행 계획 등 수립 - final audit을 통한 검수 작업의 오류 방지 계획 수립	
검수 조직체계	검수 조직 체계 마련	검수자/관리자/책임자의 검수 조직 구성 확인	
	조직 권한 및 책임 설정	정제/가공, 검수 팀의 책임 및 권한을 확인	
	조직 작업 수행 가능성 확인	작업 규모 대비 조직 역량 확인	
	조직 내 인력 규모 확인	업무 규모에 맞는 인력 참여 계획을 확인	
	인력 관리 방안 수립	작업분류, 작업량, 정확도에 관리 지표 확인 및 책임자 지정	
검수 도구	작업 도구 정의	가공도구를 통한 검수 작업 수행 가능성 판단 및 대안 마련	
	도구의 검수 기능 정의	학습용 데이터 제작 도구의 작업물 오류 확인 및 수정할 대상, 관련 자료 수집 및 저장 방안 확인	
	도구 내 통계 기능 확인	학습용 데이터 제작 도구에 탑재된 통계 기능을 활용하여 관리 및 분석	
	도구의 유효성 확인	검수자의 피드백을 추적 관리 방안 마련	

#### 6.5.2.4 데이터셋 검수 후 단계의 품질 관리

- 최종 결과물의 전달 과정에 대한 확인

##### 데이터셋 검수 후 단계의 품질 관리

생산공정	품질 관리 활동	세부설명	체크
산출물 제공	산출물 접근 절차 및 방법	K-Fashion 데이터셋, 프로토타입 모델 및 시범 서비스 등 산출물 전달 방법 사전 정의	
	산출물 접근 권한 설정	K-Fashion 데이터셋, 프로토타입 모델 및 시범 서비스 등 산출물 접근 권한 사전 정의	
	산출물 접근 도구 제공	산출물 전달 방법 사전 정의	

#### 6.5.2.5 외부 전문 검증기관을 통한 품질 검증 실행

- 한국정보통신기술협회(TTA)에서 마련한 AI 데이터 구축 분야별 표준/가이드라인을 통해 구축 단계별 품질검증 실행
- 데이터 전문가로 구성된 외부 자문단을 위촉하여 데이터 구축 과정의 단계별 품질 점검 사항에 대한 전문가 의견을 바탕으로 데이터 품질 확보 활동을 확인

### 6.5.3 AI 학습용 데이터의 정량적 품질 관리

#### 6.5.3.1 정량적 성능 모니터링 프로세스를 통한 K-Fashion AI데이터의 품질 검증

- (목적) 구축된 K-Fashion AI데이터의 품질 검증을 위해 패션 아이템 속성 분류 모델의 정량적 성능에 기반하여 지속적인 피드백 제공
- ImageNet 이미지 인식대회(ILSVRC)에서 공개한 모델 중 이미지 분류에 우수한 성능을 보이는 ResNet모델 사용

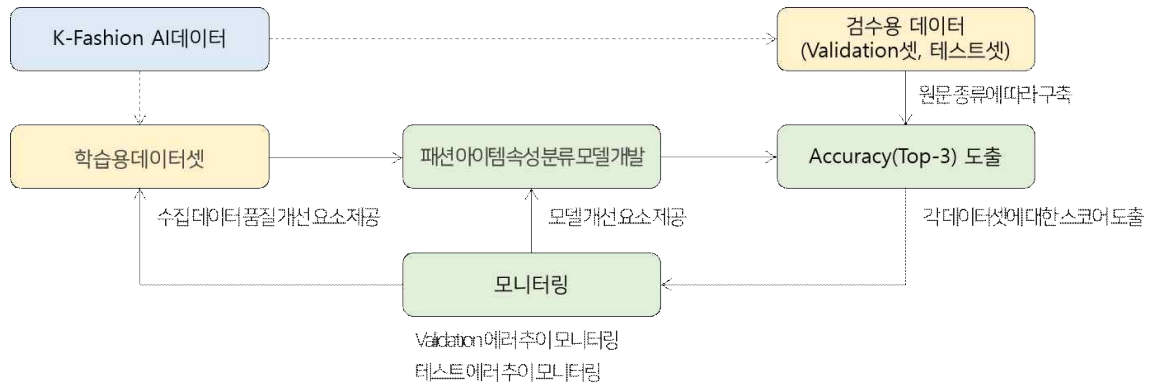


그림 48. AI모델 기반 데이터 품질 모니터링 프로세스

- 패션 아이템 속성에 따른 벤치마크 데이터셋 구축
- 대표적 정량적 성능 평가 척도인 Accuracy(Top-3)를 활용하여 K-Fashion AI데이터로 학습한 모델의 성능을 측정하여 간접적으로 데이터셋의 품질을 평가함
- 패션 아이템 속성 분류 모델의 성능은 DeepFashion으로 학습한 모델 대비 테스트셋 데이터 기준 약 5% 향상을 목표로 함
- 지속적으로 학습용 데이터셋과 벤치마크 데이터셋의 Accuracy(Top-3)를 모니터링하여 데이터의 현시점의 품질을 파악
  - ✓ 학습 에러, 검증 에러, 테스트 에러를 지속적으로 모니터링 함
  - ✓ 학습 에러는 학습 데이터셋에서 발생한 오차를, 검증(validation) 에러는 구축된 K-Fashion AI데이터 중 일부에 대한 에러를, 테스트 에러는 벤치마크 데이터셋으로 측정한 오차를 의미함
- 모니터링 중 문제가 발생 시 이에 대한 레이블 종류 및 모델에 대한 피드백을 제공하여 고품질의 AI데이터 구축과 모델에 대한 개선 작업을 실시할 수 있도록 함
  - ✓ 구축하는 AI데이터의 양이 증가함에 따라 다양한 결함 유형에 예측되는 변화를 지속적으로 모니터링하여 결함 변화의 패턴에 따라 예측되는 문제를 점검할 필요가 있음
    - ☞ AI데이터의 품질의 문제
    - ☞ AI데이터 양의 문제
    - ☞ 과적합 등 모델의 설계 및 학습 관련 이슈
    - ☞ (예시1) 데이터 양이 증가함에 따라 테스트 에러의 감소를 기대하지만, 그렇지 않은 경우 구축된 데이터 품질에 대한 확인 필요
    - ☞ (예시2) 검증 에러는 감소하지만 테스트 에러가 증가한다면, 구축된 데이터의 다양성이 부족한지에 대한 확인 필요

## 7. 활용

## 7.1 활용 모델

### 7.1.1 모델 학습

#### 7.1.1.1 아이템 속성 분류기

- 학습 알고리즘

- ✓ 패션 아이템 특성상 하나의 아이템에 대하여 다양한 특성(성별, 계절감, 형태, 색상등)이 존재하며, 하나의 아이템에 대하여 자동으로 특성을 분류하는 것이 목표
- ✓ VGG, ResNet등 이미지넷 이미지 인식대회(ILSVRC)의 아키텍처 기반으로 모델 생성
- ✓ 아이템 속성 분류기는 이미지넷 이미지 인식대회에서의 학습 가중치를 일정 부분 전의하고, K-Fashion AI데이터의 아이템 속성을 분류하도록 학습하는 형식의 구성

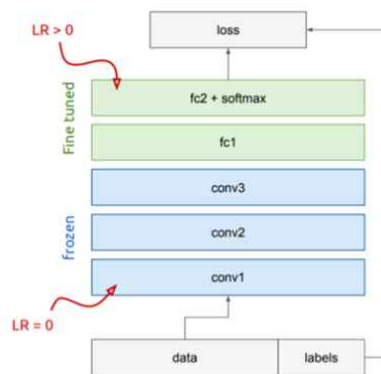


그림 49 아이템 속성 분류기 학습 구조

- 학습데이터 구축

- ✓ 아이템 속성 분류기는 패션 이미지와 학습 하고자 하는 속성 정보를 모델의 입력값으로 사용
- ✓ 아이템 속성 분류기의 입력 형식으로 데이터를 구성하기 위해 각 이미지 속성에 해당하는 패션 이미지형식으로 구성
- ✓ 아이템 속성 분류기의 원활한 학습을 위해 이미지 데이터를 80 : 10 : 10의 비율로 학습 세트, 검증 세트, 평가 세트로 분할 하여 파일 생성

```
├─ train (folder)
│   ├── casual (folder)
│   │   ├── imagefile01.jpg
│   │   ├── imagefile02.jpg
│   │   └── ...
│   ├── elegant (folder)
│   │   ├── imagefile01.jpg
│   │   ├── imagefile02.jpg
│   │   └── ...
│   └── ...
└─ ...
```

그림 50 아이템 속성 분류기의 입력 형식 예시

- 모델 학습 코드를 입력하여 학습 진행

```
python3 training_model.py
```

```
Epoch 1/50
```

```
1/64 [.....] - ETA: 20:19 - loss: 1.7939 - top1_acc: 0.2031 - top2_acc: 0.5234 - top3_acc: 0.8594
2/64 [.....] - ETA: 10:12 - loss: 1.6664 - top1_acc: 0.3555 - top2_acc: 0.6016 - top3_acc: 0.8086
3/64 [>.....] - ETA: 6:50 - loss: 1.5400 - top1_acc: 0.4583 - top2_acc: 0.6641 - top3_acc: 0.8229
```

그림 51 아이템 속성 분류기의 학습 예시

- 학습 환경 및 파라미터 설정

- ✓ 패션 아이템 탐지 모델은 Python3 기반의 Keras를 통해 개발되었으며 Linux, MacOS 및 Windows에서 작동 가능
- ✓ 빠른 병렬 연산을 통한 딥러닝 모델 학습 및 실험 시간 단축을 위해 GPU 환경에서의 학습 권장
- ✓ 학습하고자 하는 데이터와 컴퓨터 사양에 맞춰 모델의 학습 파라미터를 조정

## 7.1.2 스타일 분류기

### 7.1.2.1 학습 알고리즘

- 하나의 패션 이미지는 여러 아이템의 매치에 따라서 다양한 스타일을 가질 수 있어 패션 이미지에 대한 다중 스타일 label이 고려된 스타일 분류가 목표
- 이를 딥러닝 분야에서는 딥러닝 분야에서는 하나의 레이블로 분류하는 multi-class classification, 여러 레이블로 분류하는 multi-label classification이라 명명
- 이미지넷 이미지 인식대회(ILSVRC) 2015에서 Top-5 오류율 3.6%로 1등을 차지한 ResNet의 기본 구조는 합성곱레이어와 풀링 레이어를 번갈아가며 쌓는 전통적인 구조를 가지는 VGGNet과 비슷하지만, shortcut이 추가된 구조임
- 보통 신경망이 깊어지면 사라지는 경사 문제로 학습이 잘 진행되지 않는 문제가 존재하지만 ResNet의 깊이는 152층으로 매우 깊음에도 불구하고 입력값이 중간을 건너뛰고 전달 되게 하는 shortcut으로 이 문제를 해결함

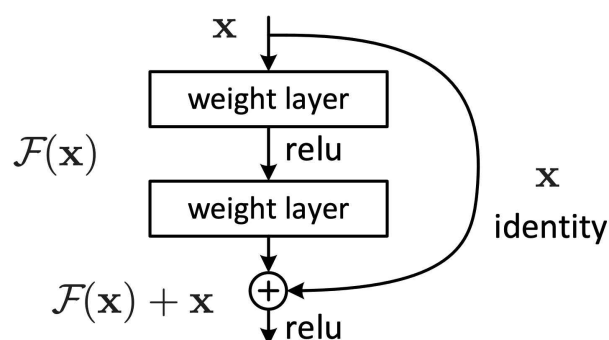


그림 52 ResNet의 shortcut 구조

- Multi-label일 경우에는 label 간 의존도가 발생하기 때문에 이를 고려한 Graph Convolutional Network (GCN)을 스타일 분류 모델과 같이 적용하여 성능 개선
- Label 간 의존도는 한 이미지 내에서 label A가 존재할 때 label B도 존재할 조건부 확률로 구하고 이를 전체적으로 label 간의 correlation matrix로 표현하여 활용함

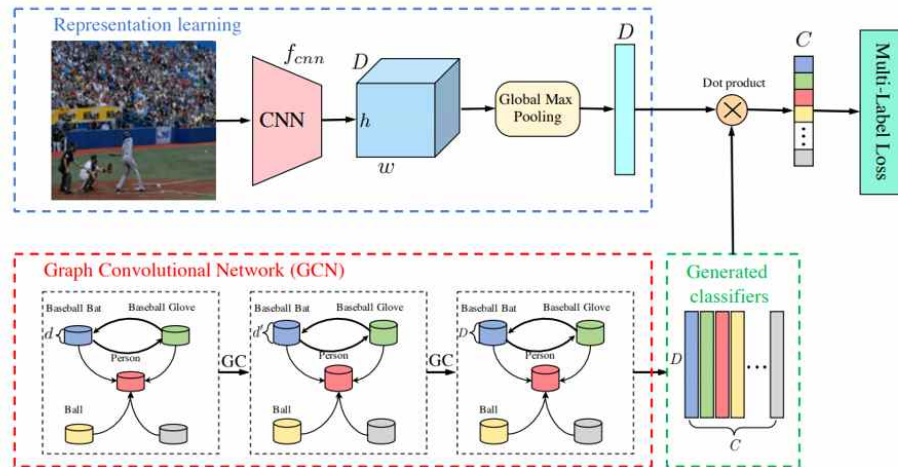


그림 53 스타일 분류 모델의 학습 구조

### 7.1.2.2 학습데이터 구축

- 스타일 분류기는 패션 이미지와 학습 하고자 하는 속성 정보 및 레이블 간 의존도를 모델의 입력값으로 사용
- 스타일 분류기의 원활한 학습을 위해 이미지 데이터를 80 : 10 : 10의 비율로 학습 세트, 검증 세트, 평가 세트로 분할 하여 패션 이미지와 각 이미지의 속성 정보 파일 생성

```
# custom_train_data

[{"file_name": "style_B0zF4pDayPe_3.jpg", "labels": [0]},
{"file_name": "style_B0zF35CJqk0_0.jpg", "labels": [0, 4, 5]},
{"file_name": "style_B0zF36FAGhu_0.jpg", "labels": [2, 3]},
...]
```

그림 54 스타일 분류기의 입력 형식 예시

- 스타일 레이블 간의 correlation matrix 파일 생성

```
python3 custom_label_dependency.py

{'nums': array([1119, 214, 215, 198, 401, 332, 162], dtype=int64),
'adj': array([[ 0, 12, 16, 44, 121, 118, 73],
[ 12,  0, 25,  8, 22, 27,  1],
[ 16, 25,  0, 14, 20, 10,  0],
[ 44,  8, 14,  0, 27, 11,  0],
[121, 22, 20, 27,  0, 67,  4],
[118, 27, 10, 11, 67,  0, 24],
[ 73,  1,  0,  0,  4, 24,  0]], dtype=int64)}
```

그림 55 Label 간 의존도를 나타내는 correlation matrix 생성 예시

- 레이블에 대한 정보를 Word2vec으로 사전 훈련된 임베딩 벡터로 치환한 파일 생성

```
python3 custom_word2vec.py

[[-0.18273    0.3894    0.41375    ...  0.57125    0.23673   -0.57446   ]
 [-0.030501  -0.2208    0.27621    ... -0.06464    0.24228    0.1277    ]
 [ 0.41866    0.1983    0.79477    ... -0.19799   -0.41475    0.47639   ]
 ...
 [-0.29585   -0.7609    0.19858    ...  0.39118   -0.14669    0.075624  ]
 [-0.26281   -0.2569    0.0031634 ...  0.16767    0.19906    0.44566   ]
 [ 0.055092  -0.11982   -0.054864 ...  0.42143    1.0199     0.22701   ]]
```

그림 56 Word2vec으로 임베딩 된 레이블에 대한 정보 예시

- 생성한 데이터를 입력 값으로 컴퓨터 사양에 맞춰 모델의 학습 파라미터를 조정 후 학습 진행

```
python3 training_graph_classifier.py

Training: 100%|██████████| 379/379 [04:05<00:00, 1.54it/s]
Test: 100%|██████████| 82/82 [00:33<00:00, 2.46it/s]
Test: Loss 8.4227 mAP 35.732
OP: 0.1867 OR: 0.9996 OF1: 0.3146 CP: 0.1868 CR: 0.9993 CF1: 0.3147
OP_3: 0.3206 OR_3: 0.7391 OF1_3: 0.4472 CP_3: 0.2716 CR_3: 0.6324 CF1_3: 0.3800
save model checkpoint/coco/checkpoint.pth.tar
*** best=35.732
lr: [1.e-06 1.e-03]
Training: 35%|███████| 131/379 [01:15<02:13, 1.86it/s]
```

그림 57 스타일 분류기의 학습 예시

### 7.1.2.3 학습 환경 및 파라미터 설정

- 패션 아이템 탐지 모델은 Python3 기반의 PyTorch를 통해 개발되었으며 Linux, MacOS 및 Windows에서 작동 가능
- 빠른 병렬 연산을 통한 딥러닝 모델 학습 및 실험 시간 단축을 위해 GPU 환경에서의 학습 권장
- 학습하고자 하는 데이터와 컴퓨터 사양에 맞춰 모델의 학습 파라미터를 조정

```

parser = argparse.ArgumentParser(description='WILDCAT Training')
parser.add_argument('--data', default='data/coco/custom2',
                    help='path to dataset (e.g. data/)')
parser.add_argument('--image-size', default=448, type=int,
                    help='image size (default: 224)')
parser.add_argument('-j', '--workers', default=4, type=int,
                    help='number of data loading workers (default: 4)')
parser.add_argument('--epochs', default=50, type=int,
                    help='number of total epochs to run')
parser.add_argument('--lr', '--learning-rate', default=0.001, type=float,
                    help='initial learning rate')
parser.add_argument('--lrp', '--learning-rate-pretrained', default=0.001, type=float,
                    help='learning rate for pre-trained layers')
parser.add_argument('--momentum', default=0.9, type=float,
                    help='momentum')
parser.add_argument('--weight-decay', '--wd', default=1e-4, type=float,
                    help='weight decay (default: 1e-4)')
...

```

그림 58 스타일 분류기의 학습 파라미터 예시

### 7.1.3 시멘틱 세그멘테이션 모델

#### 7.1.3.1 학습 모델

- Attribute Detection 모델(1) - FashionNet
  - ✓ FashionNet은 이미지 전체 정보를 활용하는 전역 특성과 패션 아이템 정보만을 고려하는 지역 특성을 모두 고려하는 모델
  - ✓ 패션 아이템은 Region of Interest(RoI) 풀링과 같은 방식을 사용하여 위치를 탐지
  - ✓ 전역 특성의 경우 일반적인 합성곱 신경망(CNN)을 사용하여 특성 도출
  - ✓ 상기 두 특성을 모두 고려하기 위하여 두 특성을 연결하여 최종 결과 산출
  - ✓ 안정적인 학습을 위해 패션 아이템 탐지는 큰 가중치와 작은 가중치를 번갈아가면서 사용
- Attribute Detection 모델(2) - Mast R-CNN
  - ✓ 기본적으로 물체의 위치를 사각형 형태로 찾아내는 모델
  - ✓ 추가적으로 이미지의 각 픽셀이 해당 물체인지 아닌지를 판별하는 세그멘테이션 업무를 수행
  - ✓ 보다 정확한 세그멘테이션 성능을 얻기 위하여 기존의 RoI 풀링이 가지는 한계를 완화한 RoIAlign을 새롭게 제안
  - ✓ 제안하는 RoIAlign는 합성곱 신경망의 중간 산출물인 특성지도에서 픽셀의 정보를 가중합하는 구조
- 추천 모델(1) - BERT4REC
  - ✓ 한 개의 패션 이미지에 포함된 상품들을 하나의 패션 패키지로 간주
  - ✓ 패션 패키지 상품에서 임의로 몇 개의 아이템을 Masking 수행
  - ✓ Embedding 과정에서 이미지의 특징 뿐만 아니라, 카테고리, 텍스트 정보 등을 함께 활용하여, 풍부한 정보를 반영함
  - ✓ 상품들의 조합을 기반으로 Masked Item을 예측하고, 실제 이미지에서의 상품이 추천될 순위를 통해 모델 구축
  - ✓ Compatibility Score를 통해, 유사상품 추천 혹은 조합의 적합성을 검증
- 추천 모델(2) - GraphSage
  - ✓ 패션 이미지 노드와 관련 상품 노드를 구별하여 이미지-상품 관계를 학습
  - ✓ 이미지 Segmentation을 진행하지 않고 메타 정보와 전체이미지를 통해 학습
  - ✓ 패션 이미지에는 여러개의 상품이 포함되고, 상품 또한 여러개의 이미지에 포함되는 이분할

구조의 그래프 생성

- ✓ 최종적으로, 이미지와 연결되지 않지만 유사성이 높은 상품을 추천

#### 7.1.3.2 권장 학습 분배량

- 구축된 데이터를 훈련 집합, 검증 집합, 테스트 집합으로 분배함
- 훈련 집합은 모델을 학습하는 데 사용됨
- 검증 집합은 모델의 하이퍼 파라미터 조합을 탐색하는 데 사용되며 검증 집합을 통해 모델 성능을 측정함으로써 가장 좋은 성능을 보인 최적의 파라미터 조합을 찾을 수 있음
- 각 데이터 집합은 8 : 1 : 1의 비율로 분배하는 것을 권장함

#### 7.1.3.3 학습 적용 방법

- 소스 코드 다운로드
  - ✓ 모델과 학습 파일이 있는 소스코드를 다운 받음
  - ✓ 각 모델들은 파이썬 라이브러리인 torch를 이용해 구현되었으며 이를 통해 학습 및 검증, 테스트를 모두 수행함으로 물체의 위치를 사각형 형태로 찾아내는 모델을 생성
- 모델 선택
  - ✓ 상기의 학습 모델들 중 상황에 맞는 모델 선택
  - ✓ 모델을 담고 있는 main.py 파일의 실행을 통해 해당 모델을 선택
- 하이퍼파라미터 설정
  - ✓ 하기 두 가지 방식을 통하여 모델의 하이퍼 파라미터를 지정

##### ◆ Argparse 이용

- 이 경우 파이썬 내부의 argparse module을 사용하게되며 config.py에 기본적인 파라미터들 내용을 포함
- 사용자가 파라미터 변경을 원할시에는 학습 수행시 직접 변경

##### ◆ 직접 수정

- 기본적인 하이퍼 파라미터를 담고있는 config.py 파일을 직접 수정하여 사용자가 원하는 하이퍼 파라미터로 지정이 가능

#### 7.1.3.4 학습 수행

- Linux기반 터미널에서 소스 코드를 다운 받은 폴더로 이동
- 이후 python main.py 명령어를 통해 학습 수행이 가능함
- Argparse를 이용하여 하이퍼 파라미터를 변경하는 경우 위의 단계에서 python main.py --batch\_size=64 와 같이 변경하고자 하는 파라미터를 입력
- 검증과 테스트 또한 하이퍼 파라미터에 포함되어 있으므로 상황에 맞게 설정하여 코드를 실행

### 7.1.4 서비스 활용 시나리오

#### 7.1.4.1 패션 이미지 자동 태깅 서비스

- 1단계: 패션 아이템 탐지 모델 및 아이템 속성 분류기 학습
  - ✓ 자동으로 분류할 패션 아이템의 속성을 정의함
  - ✓ 패션 이미지에 대하여 아이템을 탐지할 수 있는 아이템 탐지 모델을 사전에 학습시킴



- ✓ 정의된 속성을 바탕으로 K-Fashion AI데이터로 속성 분류기를 사전에 학습시킴
- 2단계: 이미지 파일(폴더)의 입력
  - ✓ 사용자가 자동으로 분류할 패션 이미지 혹은 패션 이미지로 구성된 폴더를 입력함
- 3단계: 입력 이미지에 태깅
  - ✓ 1단계에서 학습된 아이템 탐지 모델은 2단계에서 입력된 이미지 내 아이템을 인지하고 아이템 종류를 분류함
  - ✓ 분류된 아이템을 사전에 학습된 속성 분류기(색상, 패턴, 성별 등)에 입력하여 각 속성에 대한 태그 정보가 주어짐
- 4단계: 태깅 결과 노출 및 output file의 제공
  - ✓ 입력된 이미지와 그 이미지에 대한 태그 정보가 출력되고 함께 저장됨
  - ✓ 사용자가 관리하기 편한 형태(excel, csv)로 산출물이 제공됨
  - ✓ 형태는 이미지에 레이블 정보 태깅, 정리된 테이블 형식의 엑셀 파일 등 다양하게 제공

#### 7.1.4.2 트렌드 모니터링을 위한 패션 이미지 검색 서비스

- 1단계: 패션 아이템 탐지 모델 및 아이템 속성 분류기 학습
  - ✓ 패션 이미지가 유사하다고 판단할 때 고려할 속성을 정의함
  - ✓ 패션 이미지에 대하여 아이템을 탐지할 수 있는 아이템 탐지 모델을 사전에 학습시킴
  - ✓ 정의된 속성을 바탕으로 K-Fashion AI데이터로 속성 분류기를 사전에 학습시킴
- 2단계: 텍스트 혹은 패션 이미지 입력
  - ✓ 사용자가 검색할 텍스트(태그에 해당) 또는 패션 이미지를 입력함
- 3-1단계(텍스트 입력 시): 태그에 해당하는 패션 이미지 검색
  - ✓ 2단계에 입력된 태그를 기반으로 주어진 태그에 해당하는 이미지 추출
- 3-2단계(이미지 입력 시): 입력된 이미지와 유사한 이미지 검색
  - ✓ 입력된 이미지를 1단계의 모델을 이용하여 태깅
  - ✓ 추출된 태그를 포함하는 이미지 filtering
  - ✓ Filtering 된 이미지와 입력된 이미지 간의 유사도 도출
- 4단계: 추출된 이미지 노출 및 저장 기능 제공
  - ✓ 추출된 이미지를 3-1단계의 결과물의 경우 random하게, 3-2단계의 결과물의 경우 유사도 순으로 노출
  - ✓ 노출되는 이미지에 저장 버튼을 두어 필요한 경우 영구 활용이 가능

#### 7.1.4.3 매칭 아이템 추천 서비스

- 1단계: 추천을 위한 사용자 survey 진행
  - ✓ 패션 아이템 추천에 의미 있다고 판단되는 개인 속성을 정의함
- 2단계: 패션 아이템 탐지 모델 및 패션 아이템 추천 모델 학습
  - ✓ 패션 이미지에 대하여 아이템을 탐지할 수 있는 아이템 탐지 모델을 사전에 학습시킴
  - ✓ 패션 이미지와 아이템 간 유사도를 계산하여 선호하는 아이템을 추천하는 패션 아이템 추천 모델을 사전에 학습시킴
- 3단계: 사용자의 패션 아이템 입력
  - ✓ 사용자가 선호하는 패션 아이템을 입력함
- 4단계: 입력된 아이템과 같이 스타일링 하고자 하는 패션 아이템 검색
  - ✓ 패션 아이템 추천 모델을 통해 해당 아이템과 추천하고자 하는 아이템이 표현되어있는 이미지 추출

- ✓ 추천된 이미지로부터 패션 아이템 탐지 모델을 통해 패션 아이템 도출
- 5단계: 추천 결과 노출 및 저장 기능 제공
  - ✓ 사용자 입력과 추천된 아이템이 표현된 이미지에 대해 적합도를 토대로 상위 N개 노출
  - ✓ 상위 N개의 이미지를 1단계에서 정의한 개인 속성을 기준으로 filtering
  - ✓ 노출되는 이미지에 저장 버튼을 두어 필요한 경우 영구 활용이 가능

## 7.2 접근

## 7.3 유지보수

데이터 이름	K-Fashion 이미지 AI 데이터																																					
데이터 포맷	jpeg +Json file 쌍																																					
활용 분야	패션 분야																																					
데이터 요약	한국적 특성을 반영한 패션 이미지 데이터																																					
데이터 출처	동대문 상인 110개 업체, 소상공인 쇼핑몰 30개 업체, 동대문 소상공인 이미지																																					
데이터 이력	배포버전	V1.0																																				
	개정이력	신규																																				
	작성자/ 배포자	(주)오피니언라이브																																				
데이터 구성	<p>패션 이미지 데이터는 스타일별로 구분되어 있으며 패션 영역과 속성, 스타일 정보를 인식 및 도출할 수 있도록 레이블링 구성</p> <table><tr><th>대분류</th><th>세부 속성</th></tr><tr><td>상의 카테고리</td><td>탑, 블라우스, 티셔츠, 니트웨어, 셔츠, 브라탑, 후드티</td></tr><tr><td>하의 카테고리</td><td>청바지, 팬츠, 스커트, 레깅스, 조거팬츠</td></tr><tr><td>아우터 카테고리</td><td>코트, 재킷, 점퍼, 패딩, 베스트, 가디건, 쥬업</td></tr><tr><td>원피스 카테고리</td><td>드레스, 점프수트</td></tr><tr><td>컬러</td><td>블랙, 화이트, 그레이, 레드, 핑크, 오렌지, 베이지, 브라운, 옐로우, 그린, 카키 ... 실버</td></tr><tr><td>디테일</td><td>비즈, 단추, 니트팍배기, 체인, 컷오프, 블브레스티드, 드롭숄더, 자수, 프릴, 프린지 ... 퍼프</td></tr><tr><td>프린트</td><td>체크, 스트라이프, 지그재그, 호피, 지브라, 도트, 카무플라주, 패이즐리, 아가일 ... 믹스</td></tr><tr><td>소재</td><td>퍼, 무스탕, 스웨이드, 헤어니트, 코듀로이, 시퀸, 데님, 저지, 니트 ... 스판덱스</td></tr><tr><td>상의 기장</td><td>크롭, 노멀, 롱</td></tr><tr><td>하의 기장</td><td>미니, 니렌스, 미디, 발목, 맥시</td></tr><tr><td>아우터 기장</td><td>크롭, 노멀, 하프, 롱, 맥시</td></tr><tr><td>원피스 기장</td><td>미니, 니렌스, 미디, 발목, 맥시</td></tr><tr><td>소매기장</td><td>민소매, 반팔, 캡, 7부소매, 긴팔</td></tr><tr><td>넥라인</td><td>라운드넥, 유넥, 브이넥, 홀토넥, 오프숄더, 원숄더, 스퀘어넥, 노카라, 후드 ... 스위트하트</td></tr><tr><td>칼라</td><td>셔츠칼라, 보우칼라, 세일러칼라, 솔칼라, 폴로칼라, 피터팬칼라, 너치드칼라 ... 밴드칼라</td></tr><tr><td>상의,아우터,원피스</td><td>타이트, 노멀, 루즈, 오버사이즈</td></tr><tr><td>하의</td><td>스키니, 노멀, 와이드, 루즈, 벨보텀</td></tr></table>		대분류	세부 속성	상의 카테고리	탑, 블라우스, 티셔츠, 니트웨어, 셔츠, 브라탑, 후드티	하의 카테고리	청바지, 팬츠, 스커트, 레깅스, 조거팬츠	아우터 카테고리	코트, 재킷, 점퍼, 패딩, 베스트, 가디건, 쥬업	원피스 카테고리	드레스, 점프수트	컬러	블랙, 화이트, 그레이, 레드, 핑크, 오렌지, 베이지, 브라운, 옐로우, 그린, 카키 ... 실버	디테일	비즈, 단추, 니트팍배기, 체인, 컷오프, 블브레스티드, 드롭숄더, 자수, 프릴, 프린지 ... 퍼프	프린트	체크, 스트라이프, 지그재그, 호피, 지브라, 도트, 카무플라주, 패이즐리, 아가일 ... 믹스	소재	퍼, 무스탕, 스웨이드, 헤어니트, 코듀로이, 시퀸, 데님, 저지, 니트 ... 스판덱스	상의 기장	크롭, 노멀, 롱	하의 기장	미니, 니렌스, 미디, 발목, 맥시	아우터 기장	크롭, 노멀, 하프, 롱, 맥시	원피스 기장	미니, 니렌스, 미디, 발목, 맥시	소매기장	민소매, 반팔, 캡, 7부소매, 긴팔	넥라인	라운드넥, 유넥, 브이넥, 홀토넥, 오프숄더, 원숄더, 스퀘어넥, 노카라, 후드 ... 스위트하트	칼라	셔츠칼라, 보우칼라, 세일러칼라, 솔칼라, 폴로칼라, 피터팬칼라, 너치드칼라 ... 밴드칼라	상의,아우터,원피스	타이트, 노멀, 루즈, 오버사이즈	하의	스키니, 노멀, 와이드, 루즈, 벨보텀
대분류	세부 속성																																					
상의 카테고리	탑, 블라우스, 티셔츠, 니트웨어, 셔츠, 브라탑, 후드티																																					
하의 카테고리	청바지, 팬츠, 스커트, 레깅스, 조거팬츠																																					
아우터 카테고리	코트, 재킷, 점퍼, 패딩, 베스트, 가디건, 쥬업																																					
원피스 카테고리	드레스, 점프수트																																					
컬러	블랙, 화이트, 그레이, 레드, 핑크, 오렌지, 베이지, 브라운, 옐로우, 그린, 카키 ... 실버																																					
디테일	비즈, 단추, 니트팍배기, 체인, 컷오프, 블브레스티드, 드롭숄더, 자수, 프릴, 프린지 ... 퍼프																																					
프린트	체크, 스트라이프, 지그재그, 호피, 지브라, 도트, 카무플라주, 패이즐리, 아가일 ... 믹스																																					
소재	퍼, 무스탕, 스웨이드, 헤어니트, 코듀로이, 시퀸, 데님, 저지, 니트 ... 스판덱스																																					
상의 기장	크롭, 노멀, 롱																																					
하의 기장	미니, 니렌스, 미디, 발목, 맥시																																					
아우터 기장	크롭, 노멀, 하프, 롱, 맥시																																					
원피스 기장	미니, 니렌스, 미디, 발목, 맥시																																					
소매기장	민소매, 반팔, 캡, 7부소매, 긴팔																																					
넥라인	라운드넥, 유넥, 브이넥, 홀토넥, 오프숄더, 원숄더, 스퀘어넥, 노카라, 후드 ... 스위트하트																																					
칼라	셔츠칼라, 보우칼라, 세일러칼라, 솔칼라, 폴로칼라, 피터팬칼라, 너치드칼라 ... 밴드칼라																																					
상의,아우터,원피스	타이트, 노멀, 루즈, 오버사이즈																																					
하의	스키니, 노멀, 와이드, 루즈, 벨보텀																																					
어노테이션 포맷	<p>- 대분류 10가지, 세부속성 186가지, 스타일 23가지 레이블링 정보</p> <p>- 전문영역에서의 활용성을 고려한 스타일 및 세부속성 분류 구성</p>																																					
데이터 통계	데이터 구축 규모	연구용 레이블링 정보 포함 120만건																																				

	데이터 분포	//데이터의 특성(클래스, 속성) 분포를 확인할 수 있는 통계정보
	기타 활용 통계	//관련 연구를 통해 도출한 통계 정보(유사 데이터와 비교 등)가 있을 경우 작성
기타 정보	대표성 (Coverage)	//데이터가 특성 모집단의 어느 정도를 표현하는지 작성 ex) 전체 지역 범위 중 몇 개의 도시 대상
	독립성	//데이터가 원시데이터에 의존하는지 확인하여 표기 ex) 법률 개정, 민감 정보, 법적 문제 등
	유의사항	//데이터 배포 시 파급효과, 데이터 활용 시 유의사항 등 기술
	관련 연구	//해당 데이터를 활용한 논문 등 관련 연구가 있는 경우 표기

## □ 데이터 포맷



레이블링

이미지 리스트 ( 26 / 1252 )

스타일	아우터	상의	하의	원피스
text labeling data				
항목 전체선택				
카테고리	선택	디테일	선택	
브라탑		데슬		
프린트	선택	소재감	선택	
강염		레이스		
기장	선택	넥라인	선택	
노멀		원슬더		
칼라(카라)	선택	핏	선택	
보우할라		노멀		

- 데이터 JSON 형식 추출

```
{
  "이미지 정보":{
    "이미지 식별자":16566,
    "이미지 높이":1066,
    "이미지 파일명":"u_14859239888517400_1493917978.jpg",
    "이미지 너비":800
  },
  "데이터셋 정보":{
    "파일 생성일자":"2020-09-18 17:07:48",
    "데이터셋 상세설명":{
      "렉트좌표":{
        "아우터":[{}],
        "하의":[
          {
            "X좌표":210.5,
            "Y좌표":329.5,
            "가로":322,
            "세로":375
          }
        ],
        "원피스":[{}],
        "상의":[
          {
            "X좌표":180.5,
            "Y좌표":15.5,
            "가로":325,
            "세로":370
          }
        ]
      }
    }
  }
}
```

```

},
"폴리곤좌표":{
  "아우터":[{}],
  "하의":[
    {
      "X좌표1":222.0,
      "Y좌표1":522.0,
      ...
      "X좌표5":256.0,
      "Y좌표5":705.0
    }
  ],
  "원피스":[{}],
  "상의":[
    {
      "X좌표2":273.0,
      "Y좌표4":69.0,
      ...
      "X좌표38":385.0,
      "Y좌표38":506.0
    }
  ]
},
"라벨링":{
  "스타일":[
    {
      "스타일":"리조트",
      "서브스타일":"컨트리"
    }
  ],
  "아우터":[{}],
  "하의":[
    {
      "기장":"발목",
      "색상":"블랙",
      "카테고리":"팬츠",
      "디테일":[
        "리본"
      ],
    },
    "소재":[
      "울\캐시미어"
    ],
    "프린트":[
      "아가일"
    ],
    "핏":"루즈"
  ]
},
"원피스":[{}],
"상의":[
  {
    "기장":"롱",
    "색상":"블루",
    "카테고리":"브라탑",
    "옷깃":"차이나칼라",
    "디테일":[
      "리본"
    ],
    "소매기장":"긴팔",
    "소재":[
      "실크"
    ],
    "프린트":[
      "하트"
    ]
  ]
}

```

```

    ],
    "넥라인":"브이넥",
    "핏":"루즈"
  }
]
}
},
"파일 번호":16566,
"파일 이름":"u_14859239888517400_1493917978.jpg"
}

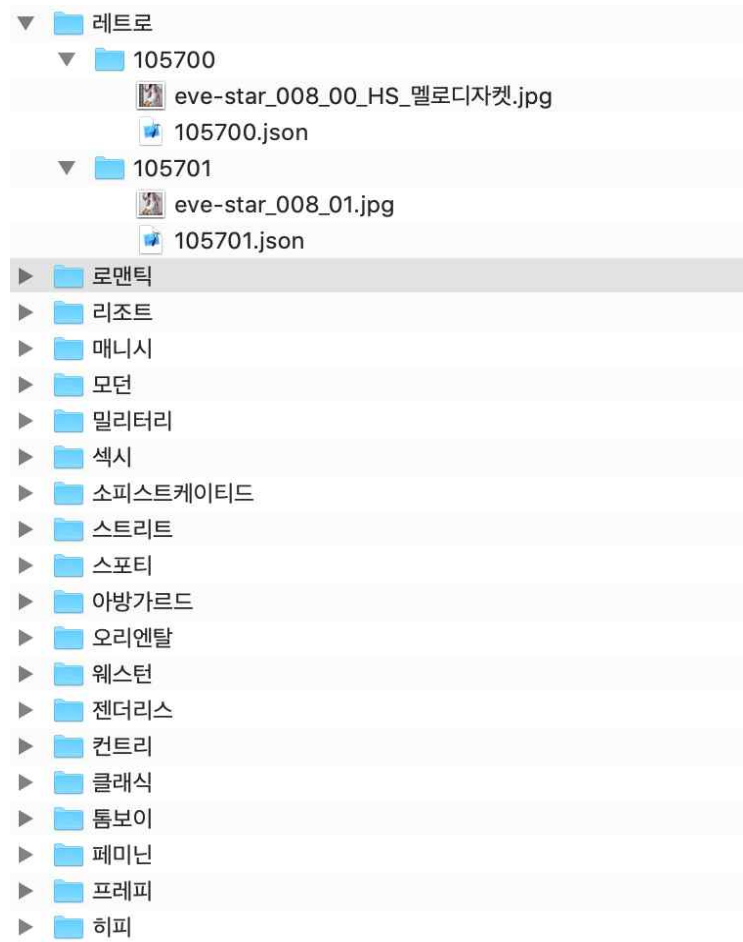
```

## □ 데이터 구성

이미지의 스타일에 따라 폴더로 구분하여 데이터 정렬

[스타일이름] - [데이터 고유번호(work\_no)] - [이미지이름.jpg]

- [이미지의 해당하는 데이터.json]



## □ 어노테이션 포맷

No	항목		길이	타입	필수여부	비고
	한글명	영문명				
1	이미지 테이블	image_table		Object		
1.1	이미지 경로 및 이름	img_path	100	String	Y	
1.2	생성일자	create_datetime	50	Datetime		

1.3	갱신일자	update_datetime	50	Datetime		
2	이미지 좌표값	image_location_rect	50	List	Y	
2.1	영역카테고리	rect_no	11	Number		1.아우터, 2.상의, 3.하의, 4.원피스
2.2	x좌표값	location_X	11	Number	Y	
2.3	y좌표값	location_Y	11	Number	Y	
2.4	width값	location_width	11	Number		
2.5	height값	location_height	11	Number		
2.6	Field	scale_X	11	Number		
2.7	Field2	scale_Y	11	Number		
2.8	생성날짜	create_datetime	50	Datetime		
2.9	수정날짜	update_datetime	50	Datetime		
3	이미지 좌표값	image_location_polygon		Object		
3.1	좌표값	no	11	Number		
3.2	x축 좌표값	location_X	11	Number	Y	
3.3	y축 좌표값	location_Y	11	Number	Y	
3.4	좌표가 찍힌 순서	location_seq	11	Number	Y	
3.5	생성날짜	create_datetime	50	Datetime		
3.6	수정날짜	update_datetime	50	Datetime		
4	이미지의 속성정보	category		Object		
4.1	속성명	category_name	64	Number	Y	
4.2	생성단계	category_step	11	Number		레이블링 단계
4.3	생성날짜	create_datetime	50	Datetime		
4.4	수정날짜	update_datetime	50	Datetime		
5	카테고리 속성	category_item		Object		
5.1	생성된 아이템 번호	category_item_name	64	String	Y	
5.2	생성날짜	create_datetime	50	Datetime		
5.3	수정날짜	update_datetime	50	Datetime		
6	이미지 레이블	label		Object	Y	
6.1	카테고리 아이템 속성 번호	category_item_no	11	Number		
6.2	생성날짜	create_datetime	50	Datetime		
6.3	수정날짜	update_datetime	50	Datetime		

## □ 데이터 통계

### ○ 데이터 구축 규모



- 최종 인공지능 데이터: K-Fashion 이미지 120만건, 800\*800 jpeg 이미지 120만 장, 어노테이션 json파일 120만건
- > 800\*800 해상도 착장 이미지, 약 180GB
- > polygon ,semantic segmentation 및 라벨 부착, 120만건의 annotation 수행

- 구축 데이터셋 정보

구분	사진수	데이터 출처	비고
패션 이미지	120만장	<ul style="list-style-type: none"> <li>- 동대문상인 기반 촬영 소싱 기업</li> <li>- 온라인 패션 유통 플랫폼(소상공인 쇼핑몰 30개 업체 연계)</li> <li>- 동대문 소상공인 이미지 직접 구매</li> </ul>	

- 구축목표 수량

구분	구축내용			운영인력 (명)	운영기간	비고
	분류	소분류	구축량			
최종 120만장	세그멘테이션	폴리곤	120만장	400명	10. 15~11. 20	
	일반 레이블	속성 레이블			10. 15~11. 20	
	전문가 레이블	속성 레이블		60명	07. 01~11. 20	

## ○ 데이터 분포

- 구축 데이터 이미지 수: 120만장
- 스타일별 데이터 수 및 비율

클래스	개수	비율	클래스	개수	비율
스트리트	63,575	48.07%	섹시	2,119	1.60%
리조트	22,228	16.81%	매니시	1,823	1.38%
페미닌	21,170	16.01%	레트로	1,658	1.25%
로맨틱	19,719	14.91%	오리엔탈	1,443	1.09%
모던	16,563	12.52%	키치/키덜트	1,291	0.98%
클래식	9,346	7.07%	밀리터리	1,285	0.97%
소피스트케이티드	8,891	6.72%	힙합	681	0.51%
컨트리	8,851	6.69%	아방가르드	606	0.46%
젠더리스	4,616	3.49%	프레피	489	0.37%
히피	3,649	2.76%	웨스턴	362	0.27%
스포티	3,068	2.32%	핑크	216	0.16%
툼보이	2,459	1.86%			

\* 120만건중 10%(12만건)에 대한 분포

## ○ 기타 활용 통계

//관련 연구를 통해 도출한 통계 정보(유사 데이터와 비교 등)가 있을 경우 추가/예시 (출처: Know What You Don't Know: Unanswerable Questions for SQuAD- Pranav Rajpurkar et al. 2018)