

인공지능 데이터 구축·활용 가이드라인

- 생활 및 거주환경 VQA -

인공지능 데이터 구축	주관기관	(주)유클리드소프트  유클리드소프트
	참여기관	공주대학교  공주대학교 KONGJU NATIONAL UNIVERSITY
	참여기관	(주)심스리얼리티  심스리얼리티 SIMSREALITY
가이드라인 작성	(주)유클리드소프트	나현우 과장
가이드라인 버전	Ver 1.1 2021-02-08	

목 차

1. 데이터 명세 정보	1
1.1 데이터 정보 요약	1
1.2 데이터 포맷	1
1.3 어노테이션 포맷	2
1.4 데이터 구성	2
1.5 데이터 통계	2
1.6 원시 데이터 특성	3
1.7 기타정보	6
2. 데이터 구축 가이드	7
2.1 데이터 구축 개요	7
2.2 문제정의	8
2.3 획득·정제	10
2.4 어노테이션/라벨링	12
2.5 검수	15
2.6 활용	18

1. 데이터 명세 정보

1.1 데이터 정보 요약

데이터 이름	생활 및 거주 환경 VQA	
활용 분야	어린이, 노인 등 생활 안전 관련 서비스	
데이터 요약	어린이, 노인, 일상, 실내 환경 등 일상 생활 속 장면을 대상으로 시각정보 기반 질의응답(Visual Question Answering, VQA) 기술연구를 위한 학습용 데이터셋 구축	
데이터 출처	직접 수집 및 가공(클라우드 소싱 활용)	
데이터 이력	배포버전	Ver.1.0
	개정이력	신규
	작성자/ 배포자	나현우 과장

1.2 데이터 포맷

{

```
"info": {
  "year": 2020,
  "version": 1,
  "description": "VQA 모델 트레이닝 데이터셋",
  "contributor": "유클리드소프트 연구소",
  "url": "www.euclidsoft.co.kr",
  "date_created": "2020-12-07 00:00:00"
},
"data_type": "개인 일상생활",
"data_subtype": "test",
"images": [
  {
    "image_id": 76159,
    "image": "000000076159.jpg",
    "category": "가구",
    "weather": "맑음"
  }
],
"license": {
  "name": "Creative Commons Attribution 4.0 International License",
  "url": "http://creativecommons.org/licenses/by/4.0/"
}
}
```

}

- 질문/답변 JSON

```
"data_type": "개인 일상생활",
"data_subtype": "test",
"questions": [
  {
    "question_id": 81001001,
    "image_id": 81001,
    "question": "도르에 있는 자들은 몇 대입니까?",
    "answer": "2",
    "answer_type": "궁정",
  }
],
```

1.3 어노테이션 포맷

No	항목	설명	길이	타입	필수여부	예시
1	info	데이터셋 정보		object		
1-1	year	생성년도	4	int		2020
1-2	version	버전	2	int		1
1-3	description	데이터셋 설명	100	string		VQA 검증 모델
1-4	contributor	관련자	100	string		유클리드소프트
1-5	url	데이터셋 URL	100	string		http://www.euclidsoft.co.kr
1-6	date_created	데이터셋 생성일자		datetime		2020-12-08 12:00:00
2	data_type	데이터셋 구분	100	string	○	개인 일상생활
3	data_subtype	데이터셋 하위구분	100	string	○	train
4	images	이미지 정보		object		
4-1	image_id	이미지 식별키		int	○	10001
4-2	image	이미지 파일명	200	string	○	10001.jpg
4-3	category	카테고리	50	string		실내
4-4	weather	날씨	50	string		맑음
5	question	질문 정보		object		
5-1	question_id	질문 식별키		int	○	10001001
5-2	image_id	이미지 식별키		int	○	10001
5-3	question	질문 텍스트	500	string	○	의자는 몇 개인가요?
5-4	answer	정답 텍스트	100	string	○	3
5-5	answer_type	정답 유형	50	string	○	공정

1.4 데이터 구성

개인 일상생활_test	2020-12-03 오후 7:01	파일 폴더	
개인 일상생활_train	2020-12-03 오후 7:07	파일 폴더	
개인 일상생활_validate	2020-12-03 오후 7:09	파일 폴더	
개인 일상생활_test_images.json	2020-12-03 오후 6:59	JSON File	1,602KB
개인 일상생활_train_images.json	2020-12-03 오후 6:59	JSON File	4,676KB
개인 일상생활_validate_images.json	2020-12-03 오후 6:59	JSON File	1,566KB

1.5 데이터 통계

1.5.1 데이터 구축 규모

◆ 시각정보기반 질의응답(Visual Question Answering, VQA) 기술연구를 위한 학습용 데이터셋은 이미지와 이미지에 연관된 다수의 질문과 대담으로 구성된다.

◆ 시각정보기반 질의응답 학습용 데이터셋 구축은 VQA 대상이 되는 이미지 데이터 수집, 확보한 이미지에 대한 질문과 답변 데이터의 생성 등 고품질의 인공지능 학습용 데이터셋을 구축하는 공정에 대한 품질관리 가이드라인을 제시한다.

과제명	주요 내용	데이터 수집 방법	데이터 구축량	데이터 형식
한국인의 생활 환경에 맞는 VQA AI 학습데이터	개인 생활 환경, 노인 생활 환경, 어린이 생활 환경, 실내 가구/가전 배치 환경	클라우드소싱 (직접 수집)	이미지: 100만장	jpeg
			텍스트 QA: 700만건	json

1.5.2 데이터 분포

KVQA데이터 구축 대상	실내 가구 및 가전기기 VQA	개인 일상생활 VQA	노인 생활환경 및 위험상황 이미지에 대한 VQA	어린이 생활환경 및 위험상황 이미지에 대한 VQA
목표이미지	25만장 (3D)	25만장	25만장	25만장
카테고리 (슈퍼헤더)	총 12개 카테고리 가구, 가전, 공공장소, 매장, 물건, 사무공간, 생활, 휴게공간, 실내, 음식, 개인공간, 여가공간 (각 카테고리별 3% 이상 확보)	총 12개 카테고리 가전(가구), 매장, 골목, 공공장소, 물건, 생활, 실내, 야외, 음식, 여가공간, 변화가, 차량 (각 카테고리별 3% 이상 확보)		
QA 텍스트 (질문+답변)	175만건 (이미지당 평균 7건의 질의응답)	175만건 (이미지당 평균 7건의 질의응답)	175만건 (이미지당 평균 7건의 질의응답)	175만건 (이미지당 평균 7건의 질의응답)

1.6 원시데이터 특성

1.6.1 대상분류

- ◆ 원시데이터 수집 대상 : 다양한 질의응답을 생성할 수 있는 실제 생활환경 촬영 이미지
 - 클라우드 워커를 고용하여 해당 데이터 및 분류에 맞는 이미지 직접 촬영 후 업로드
 - 총 4종의 데이터로 분류 (어린이 생활 및 거주환경 25만장, 노인 생활 및 거주환경 25만장, 개인 일상생활환경 25만장, 실내 가전 및 가구 배치 환경 25만장)
 - 밝기, 선명도, 흔들림, 노이즈 등으로 객체 검출이 불가능하게 하는 이미지는 배제
 - 개별 이미지가 서로 독립성 유지 : Hash상 동일하거나 유사도가 높은 이미지(98% 이상 유사)중 실제 육안 확인 시에도 중복에 가까운 이미지는 배제
 - 이미지 데이터 정제는 수행기관의 저작도구를 활용한 일관성과 균질성 보장(메타데이터, 해상도 등)

1.6.2 제약조건

- ◆ 일부 제약(semi-constraned)
- ◆ 데이터별 슈퍼헤더 12종 구성
 - 데이터별로 슈퍼헤더 12종을 구성하고 각 슈퍼헤더는 최소 3% 이상(7,500장) 구축하여 편향성을 방지한다. (MS-COCO의 슈퍼헤더를 기반, 과제에 맞게 항목을 커스터마이징한다.)

1.6.3 속성

- ◆ 수집된 이미지 데이터셋은 정제단계에서 전처리하여 640*480 픽셀을 기본 크기로 조정한다.
- ◆ 이미지의 밝기, 선명도, Denoise, jpeg Format 변환 등 전처리 작업을 수행한다.

촬영방식	촬영도구	원본 규격
클라우드소싱	휴대폰	1920*1080 픽셀, jpeg

1.7 기타정보

1.7.1 포괄성

- ◆ 이미지 수집자는 전국의 다양한 청년, 경단녀, 노인을 대상으로 선발하며, 해외 거주중인 자는 본 과제의 이미지 수집 작업자군에서 배제한다.
- ◆ 어린이, 노인, 개인생활, 실내 가구 배치로 데이터를 나눠 범용성을 확보하면서도 각 데이터 별로

특색있는 안전관련 데이터를 포함할 수 있도록 한다. 특히 어린이와 노인 관련 데이터에서는 안전 관련 어노테이션 비중을 일정량 포함하도록 한다.

1.7.2 독립성

- ◆ 클라우드 소싱 방식으로 클라우드 워커가 직접 VQA 이미지를 수집
 - 한국인의 생활 속 밀접하고 친숙한 객체, 생활환경, 상황이해 등을 고려한 학습용 데이터를 전국의 클라우드 워커를 통해 직접 수집
 - 각 분류별 가장 적합한 작업자군을 확보하여 이미지 수집
 - 사전에 NIA와 수행기관에서 권리를 소유하는데 동의하는 내용의 서류를 수집하여 저작권 문제가 발생하지 않도록 함

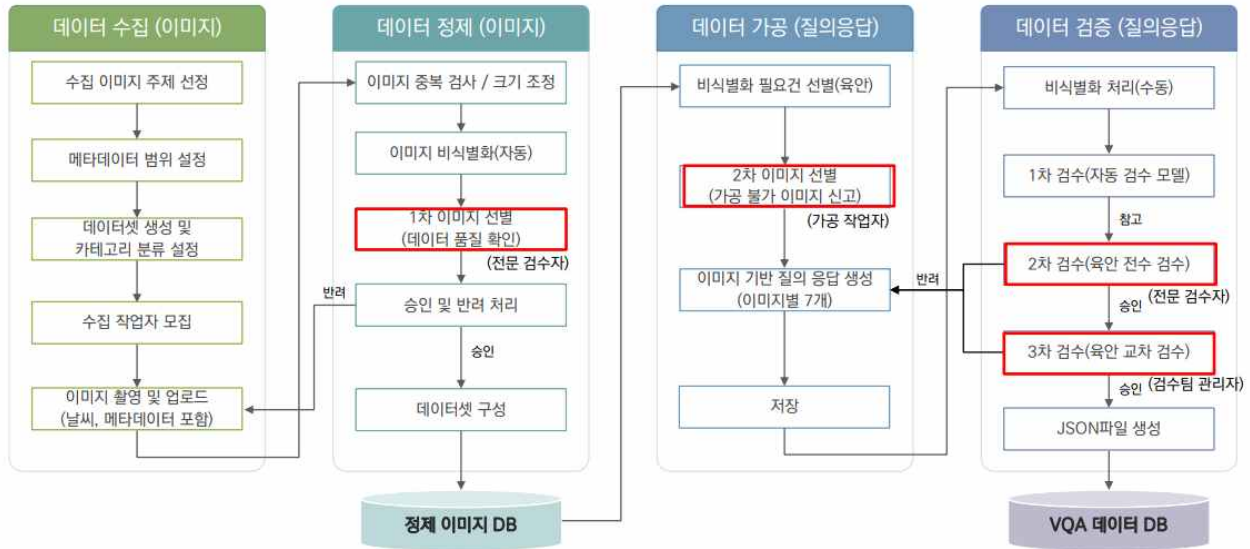
1.7.3 유의사항

- ◆ VQA AI학습데이터를 통한 파급효과
 - 한국 환경에 맞는 KVQA 학습데이터를 이용한 AI 서비스로 경쟁력 강화
: 이미지 중심의 AI와 텍스트 중심의 AI가 최근에 VQA 또는 GQA를 통해 융합하여, 시각지능을 높여 줌으로써, 실감대화 가능한 서비스가 출시됨
 - XR, 로봇 등이 상황을 이미지로 입력 받고, 이에 대하여 사물과 관련된 의미나 관계적 질의응답을 제공하는 기술을 통해서 사람과 기계 간, 기계와 기계 간 관계성을 찾을 수 있음

2. 데이터 구축 가이드

2.1 데이터 구축 개요

💡 데이터 구축 프로세스



◆ 시각정보기반 질의답변 학습용 데이터 구축은 원천 이미지를 수집하고 이에 대해 정제를 마친 이미지에 대한 질문과 답변을 생성, 검수자가 검증을 통해 최종 VQA 학습용 데이터를 구축하는 과정으로 진행된다.

◆ 시각정보기반 데이터구축은 수집단계, 정제단계, 가공단계, 검수단계 및 배포단계의 총 네 단계로 구분한다.

- 수집단계 : 수집할 원천 데이터의 주제와 범위를 설정, 수집 작업자를 통해 원천 데이터를 수집하는 단계.
- 정제단계 : 수집된 이미지를 1차로 비식별화 처리(자동)하고 전문 검수자를 활용해 선별하고 승인, 반려하여 원천 데이터를 정제하는 단계. (정제가 완료된 이미지에 한해 수집 작업자에게 보수가 지급됨)
- 가공단계 : 가공 작업자를 통해 정제된 이미지와 관련된 질의응답을 생성하는 단계
- 검증단계 : 전문 검수자를 통해 최종 비식별화를 완료하고, 1차 자동 검수, 2차 육안 전수 검수, 3차 육안 전수 교차 검수를 진행하여 데이터를 확정 짓는 단계

2.2 문제정의

2.2.1 임무 정의

◆ 어린이와 노인의 안전에 기여할 수 있는 시각지능 데이터를 구축

◆ 한국형 이미지 및 한국어 기반 VQA를 구축하여 다양한 활용 유도

- 해외에서 활발하게 발전되고 있는 영어기반의 VQA 학습용 데이터 상황과 달리, 공개된 한국어 VQA 학습용 데이터는 몹시 부족한 상황
- VQA 데이터를 이용해 개발한 KVQA 모델을 개발자 소스 코드 공유사이트인 GITHUB와 한국정보화진흥원 AI-HUB에 알고리즘 소스 코드를 공개
- 국내·외 개발자들이 공개한 KVQA 알고리즘을 계속적으로 개선하고, 업그레이드한 알고리즘을 공유하도록 라이선스 정책을 제시함
- 모바일 기반의 4종 시범 서비스 개발로 어린이와 노인의 생활 안전 확보에 기여, 범용성이 높은 민원 상담 서비스, 기물 가상 배치 서비스를 통해 VQA에 대한 대중의 관심도 및 활용성 향상

KVQA데이터 구축 대상	실내 가구 및 가전기기 배치 환경 VQA	어린이 생활 및 거주 환경 VQA	노인 생활 및 거주환경 VQA	개인 일상 및 거주환경 VQA
KVQA서비스 개발	가구 및 가전기기 배치 상담서비스	어린이 돌봄 및 안전지원서비스	노인 돌봄 및 긴급지원서비스	개인 일상생활 이미지 기반 대화 서비스
목표	이미지 기반 공간 내 가구 및 가전기기의 배치 관계를 고려하여, 공간이미지에 대한 적절한 배치를 제시해주는 서비스 개발	이미지 기반 어린이 생활환경에 대한 정황을 파악하여, 안전 또는 위험 상황 파악 후 알람을 울리는 기술개발	이미지 기반 노인 정황을 파악하여, 실감 대화할 수 있고, 돌발상황이나, 긴급상황을 인식할 수 있는 기술개발	개인의 일상생활 이미지 기반 내 정황(Context)을 이해한 실감 대화 가능한 기술개발
기대효과	<ul style="list-style-type: none"> - 한국인 가정 거실, 방, 화장실 등 일상 거주 환경에 대한 사물의 배치 VQA 데이터를 이용한 상담 AI를 위한 데이터로 제공 가능 - 자가 인테리어 컨설팅이 필요한 경우 상담 AI를 위한 데이터로 제공 가능 	<ul style="list-style-type: none"> - 한국 어린이의 생활 환경인 교육기관, 길, 교차로, 골목 등 거주 환경 VQA를 이용한 안전 점검 AI를 위한 데이터로 제공 가능 	<ul style="list-style-type: none"> - 국내 노인들의 생활 환경인 요양원, 거주지, 거리, 노인정, 복덕방, 마을회관 등 거주 환경 VQA를 이용한 상담 AI 및 위험상황 자동 알림 AI를 위한 데이터로 제공 가능 	<ul style="list-style-type: none"> - 국내 개인 일상생활인 직장, 도서관, 놀이공원, 음식, 풍경 등 개인 또는 가족 단위 일상 VQA를 이용한 AI 서비스 데이터로 제공 가능 - 한국인에 맞는 이미지 기반 맞춤형 상담 AI 데이터로 제공 가능

2.2.2 데이터 구축 유의사항

◆ 개인정보 보호조치

- 1) 이미지상 비식별화 : 이미지상 사람의 초상, 차량 번호판 등 개인을 특정할 수 있는 요소(지로 용지, 명함, 계좌 정보 등)는 Blur처리하여 비식별화 한다. 단, 광고로 노출되는 연예인의 초상 등은 비식별화 필수 대상으로 보지 않는다.
- 2) 질의응답 텍스트 : 질의응답 텍스트 내 개인정보(이름, 연락처, 주소, 계좌 등)나 그 외 민감정보가 포함되지 않도록 해당 데이터를 포함하여 가공하지 않도록 하고, 발견시 비식별화(*일부 별표처리) 또는 배제 조치한다.

◆ 민감정보 배제

- 1) 정치 성향, 성차별, 종교 관련, 유해 요소가 포함되는 이미지, 텍스트는 데이터에서 배제한다.
 - 작업 매뉴얼을 통해 강력 제한, 전수 검수와 교차 검수시 육안으로 확인하여 배제

◆ 저작권 및 소유권

- 1) 원천 이미지 데이터, 가공 단계에서 생성되는 질의응답 데이터는 저작권 문제가 발생하지 않도록 NIA와 수행기관에서 권리를 소유한다는 내용의 작업 동의서를 사전에 확보한다.

2.3 획득·정제

2.3.1 원시데이터 선정 및 수집 절차

◆ 데이터 선정 지표

이미지 대상	선정지표	비고
이미지	<ul style="list-style-type: none"> • 과제 주제 부합 (4종 데이터-어린이, 노인 생활 등 / 하위 12종 카테고리-실내, 야외, 차량, 가전 등) • 이미지 여부 • 이미지 품질 적정성(음란, 혐오, 민감, 폭력 배제) • 이미지 종류(광각, 밝기, jpeg) 	수집자에게 위임하며, 검수자 승인 시 적합으로 판정
객체	<ul style="list-style-type: none"> • 이미지에 최소 5개 이상의 식별 가능한 객체 존재 • 객체의 치우침 • 객체의 밀집도가 과하지 않은 환경에서 수집 	
환경	<ul style="list-style-type: none"> • 장면 중복, 적절성 • 카테고리별 치우침(다양성) 	상동
장면	<ul style="list-style-type: none"> • 중복성, 객체 구성 치우침 	상동

◆ 데이터 수집 기준

※ 일반 이미지 수집 기준
<ol style="list-style-type: none"> 1. "올바른 촬영 이미지"란, 7가지의 질문과 답변을 생성할 수 있는 이미지를 기준으로 한다. 2. 최소 5개 이상의 *객체가 들어가도록 이미지 수집 (10 객체 내외로 권장) 3. 종류가 명확히 구분되는 객체 최소 5개 이상, 같은 종류의 객체는 1개만 인정 4. 하늘, 벽, 구름, 바닥 등은 객체가 아닌 배경에 해당함. 따라서 객체로 인정하지 않음 5. 이전에 촬영된 이미지와 각도 변화 없이 80% 이상 객체가 겹칠 경우 중복으로 간주. 6. 객체가 변경되어도 동일한 배경에서 같은 각도로 촬영하는 경우 중복으로 간주 7. 동일 객체의 경우 30도 이상 각도 변화가 없는 경우 중복으로 간주 8. 객체란 단순히 물체만을 의미하는 것은 아니며, 식별 기호 등 이미지상 의미를 부여할 수 있는 경우도 객체로 인정함. ex) 수집시 바닥은 객체로 인정되지 않으나, 바닥에 그려진 어린이 보호구역 표시는 객체로 인정 9. 객체의 특성이 보이는 부위가 잘려있거나, 왜곡된 경우 온전한 객체로 인정하지 않음 10. 필터가 없는 기본 카메라로 촬영하며, GPS좌표와 시간 정보 등의 메타 정보(EXIF)를 포함할 것

※ 3D 기반 이미지 수집 기준

1. Azure kinect 등 공간 스캔이 가능한 장비로 공공기관, 공공장소 등의 실내를 촬영한 3D 오브젝트 내에서 평면 사진을 캡처하여 추출.
2. 방문할 기관의 동의를 얻고 코로나 방역과 관련된 수칙을 철저히 지키며 수집을 진행.
2. 최소 4개 이상의 객체가 들어가도록 이미지 수집
3. 특수 장비를 활용하고, 공공기관의 방문 일정을 조율해야 하는 수집 작업인만큼 일반 이미지 수집과 동일한 중복 기준을 적용하는 것은 과도함. 따라서 기존 사진과 10도 이상 각도 변화 시 별도의 이미지로 인정
4. 객체란 단순히 물체만을 의미하는 것은 아니며, 식별 기호 등 이미지상 의미를 부여할 수 있는 경우도 객체로 인정함. ex) 바닥에 그려진 어린이 보호구역 표시도 객체로 인정
5. 촬영 장소에 대한 위치, 시간 정보는 메타데이터가 아닌 별도 문서로 관리
6. 수집된 이미지는 실내 가전, 가구 데이터에서 적합한 카테고리를 찾아 업로드

◆ 데이터 분류 기준

- 데이터별로 슈퍼헤더 12종을 구성하고 각 슈퍼헤더는 최소 7,500장 이상을 구축하여 편향성을 방지한다. (MS-COCO의 슈퍼헤더를 기반, 과제에 맞게 항목을 커스터마이징한다.)
- 각 슈퍼헤더는 최대 20% 이하의 분포(최대 50,000장)를 유지한다.

데이터	슈퍼헤더	예시 장소	예시 개체
어린이 생활 환경 /노인 생활 환경 /개인 생활 환경	가전(가구)	가구 매장, 가전 매장, 부엌, 안방, 거실 등	TV, PC, 스피커, 전자레인지, 냉장고, 세탁기, 싱크대, 의자, 책상, 식탁, 소파, 침대, 장롱, 전등, 실외기 등
	매장	마트, 편의점, 판매점 등	매장 외관, 과자, 생활용품, 매대, 카운터, 소화전, 창고, 음식 등
	골목	골목길	골목길, 상점 입구, 간판, 전봇대, 맨홀, 표지판, 쓰레기봉투, 주택 입구
	공공장소	지하철, 놀이터, 회관, 공원 등	공원, 벤치, 지하철 입구, 지하철 계단, 공중 화장실(내부 금지), 장애인 화장실, 운동기구, 에스컬레이터, 엘리베이터 등
	물건	안방, 거실, 매장 등	사무용품, 장식품, 전자기기, 악세서리, 가방, 침구류 등
	생활	생활 활동, 출퇴근길, 운동 시설 등	업무현장, 교육현장, 재활용 분류장, 거실, 보도, 요양시설 등
	실내	집, 사무실, 병원, 전시장 등	창문, 전등, 의자, 책상, 통행로, 문, 소방시설, 청소도구, 실내 체육시설 등
	야외	공원, 체육시설, 관광지 등	등산로, 공원, 벤치, 수목원, 자전거, 가게 외관, 건물외관, 야외 체육시설 등
	음식	식당, 카페, 부엌, 포차, 노점 등	밥, 반찬, 패스트푸드, 음료 등
	여가공간	극장, 미술관, 침실, PC방, 카페, 찜질방 등	카페, 테이블, 의자, 침대, TV, 책, 매점, 오락도구(보드게임 등)
	변화가	사거리, 도심지, 변화거리, 유흥가 야외 등	간판, 전단지, 사람, 차량, 네온사인, 전광판, 노점, 주점, 백화점 등
	차량	주차장, 도로, 자전거 도로, 교통수단 등	오토바이, 자동차, 자전거, 전동킥보드, 트럭, 버스 등
실내 가전 및 가구 배치	가구	가구 매장, 부엌, 안방, 거실 등	의자, 책상, 식탁, 소파, 침대, 장롱, 장식장, 책장 등
	가전	가전 매장, 부엌, 안방, 거실 등	TV, PC, 스피커, 전자레인지, 냉장고, 세탁기, 식기세척기, 가스레인지, 에어컨, 청정기 등
	공공장소	지하철, 회관, 다중 이용 업소 실내 등	장애인 보조시설, 지하철역 내부, 소방시설, 자판기, 승강기, 스크린도어, 요양시설내부 등
	매장	마트, 편의점, 판매점 등	매장 인테리어, 구조물, 판매상품, 에스컬레이터, 매대, 카운터, 소화전, 창고 등
	물건	안방, 거실, 매장 등	사무용품, 장식품, 전자기기, 악세서리, 가방, 침구류 등
	사무공간	사무실, 업무공간 등	사무실 인테리어, 사무실 집기, 필기구, PC, 의자, 책상, 회의실, 칸막이, 커피포트, 실내화 등
	생활	다중 이용 시설, 출퇴근길, 운동 시설 등	업무현장, 교육현장, 생활 쓰레기, 거실, 보도, 요양시설, 주점, 식당, 카페 등

휴게공간	휴게실, 침실, 카페 등	의자, 책상, 소파, 시설 내 휴게실, 커피포트, 급식실 등
실내	집, 사무실, 병원, 전시장 등	창문, 전등, 의자, 책상, 통행로, 문, 소방시설, 청소도구, 실내 체육시설 등
음식	식당, 카페, 부엌 등	밥, 반찬, 패스트푸드, 음료 등
개인공간	독서실, 안방, 개인 차량 내부, 욕실 등	침대, 책상, PC, 개인 운동기구, 개인 오락용품, 거실, 침실, 욕조 등
여가공간	극장, 미술관, 침실, PC방, 카페, 찜질방 등	카페, 테이블, 의자, 침대, TV, 책, 매점, 오락도구(보드게임 등)

◆ 저작도구를 통한 수집 데이터 업로드

2.3.2 정제 절차

◆ 중복 및 유사 이미지 검사

◆ 자동모델을 통한 비식별화 (1차 비식별화)

○ 개인정보 포함 이미지의 경우 비식별화 처리



<비식별화 처리 전, 후 비교>

- 수집된 이미지에서 개인식별이 가능한 정보(개인 휴대폰 번호, 상세 주소 등)에 대하여 비식별화 한다.
- 육안으로 식별 가능한(100픽셀 이상) 사람의 안면은 대하여 blur 기법으로 처리하여 비식별화 한다.
- 육안으로 식별 가능한(100픽셀 이상) 차량번호판 등에 대하여 blur 기법으로 처리하여 비식별화 한다.

◆ 이미지 크기 변경

- 원천 이미지 데이터는 jpeg 형태로 640 x 480 픽셀 크기로 크기를 조정한다.
640 x 480 픽셀로 변경이 불가능한 비율(세로, 정사각형, 16:9)의 이미지는 가로를 기준으로 640을 맞추고 비율은 원본대로 고정한다.
- 640 x 480 픽셀을 벗어나는 비율의 이미지가 전체 이미지의 20%를 넘지 않도록 한다.
- 원본 내 메타데이터는 변경된 이미지 내에 유지하거나 별도의 JSON파일로 이미지 파일과 대응되게 관리한다.

2.3.3 정제 기준

◆ 이미지 선별 기준(정제단계)

- 선별은 전수 육안 검사를 통해 미준수 또는 금지사항 위반 해당시 반려처리

준수사항	금지사항
1. EXIF 정보에 GPS 좌표, 촬영 시간이 반드시 포함되어야 함 2. 스마트폰 무필터 기본 카메라를 이용해 촬영 (아이폰은 JPG로 저장설정 필수) 3. 4:3 비율로 촬영설정을 변경한 후 가로 모드로 촬영(세로 모드 NO, 세로로 촬영한 사진을 가로 사진으로 수정 NO) 4. 객체가 손가락으로 가려지거나 잘리지 않게 촬영되어야 함 5. USB, 클라우드, 카카오톡 원본 화질 전송을 통하여 사진파일을 PC로 전송 (일반화질, 고화질 전송 시 EXIF 손실 우려) 6. 각 객체가 육안으로 식별이 가능할 것 7. 촬영 프로젝트에 맞게 분류해서 업로드 할 것 8. 이미지 내 5개 이상의 객체가 포함될 것	1. 본인 동의 없는 인물, 인적 사항 정보 촬영 가급적 배제, 수집된 사진에서도 비식별화 처리 2. 촬영 사진을 제외한 수집된 사진 등록금지 3. 유해물, 음란물 금지 4. 보안시설 촬영금지 (군 시설, 공공기관 등) 5. 노이즈, 빛 번짐 과 같은 식별 불가한 사진 사용금지 6. 촬영 이미지는 재사용 및 공유금지 7. 카카오톡 원본 화질 전송을 제외한 메신저를 통한 전송 금지 (EXIF 정보손실 우려) 8. 사회 통념상 부적절한 연출이나 유해성 요소, 개인을 특정할 수 있는 민감한 정보가 포함되지 않을 것

◆ 2.3.1의 데이터 수집기준을 준수하여 원시데이터 선별(승인, 반려 처리)

2.4 어노테이션(가공 단계)

2.4.1 어노테이션 절차

- 정제 이미지별로 저작도구를 사용하여 이미지에 연관된 질문·답변 데이터를 생성하는 절차이다.
- 크라우드 소싱 가공 작업자를 활용하여 웹 기반 저작도구로 진행한다.
- 이미 정제된 이미지 중 부적합 이미지가 있는 경우 검수자에게 역으로 반려를 보낼 수 있으며 이를 원천 데이터에 대한 교차 검증으로 본다. 적합 이미지는 어노테이션을 진행함으로 교차 검증의 승인을 확정한다.
- 정제 이미지의 비식별화 필요성 여부를 재차 판단하여 검수자에게 비식별화를 요청하는 작업을 포함한다.

2.4.2 어노테이션 기준

- 폐쇄형 질문(Closed-ended Question) : 응답자가 세부 설명 없이 명료하게 사실적인 정보를 답할 수 있도록 질문하는 것이다. '예', '아니오'로 대답할 수 있는 질문, '또는' 열거한 특정 범주에서 대답을 고를 수 있는 질문의 형태이다. 이미지당 평균 5개 내외로 생성되도록 관리한다.
- 개방형 질문(Open-ended Question) : 응답자가 원하는 형식으로 대답할 수 있도록 의문사 형태로 질문하는 것이다. 이는 '누가', '언제', '어디서', '무엇을', '어떻게', '왜' 등을 이용한 질문으로 응답자로 하여금 다양한 답변을 얻을 수 있다. 이미지당 평균 2개 내외로 생성되도록 관리한다.
- 이미지 학습을 통해 생성한 한국어 질문과 답변의 오류 교정, 문맥상 부적절한 문장 또는 단어는 제외하거나 변경 처리한다.
- 불완전한 질문으로 구성되거나 한국어 문법을 따르지 않는 심각한 문장 오류 사례는 반려한다.
- 생성된 질문은 이미지와 연관된 질문과 답변이어야 하며, 장면과 상황으로부터 유추할 수 없는 정도로 비약적인 추론 질문은 제외한다.

```

[
  {
    "info": { --> 데이터셋 정보
      "year": "2021", --> 년도
      "version": "0.1", --> 버전
      "description": "VQA 학습 데이터", --> 데이터셋 설명
      "contributor": "유클리드소프트", --> 관련자
      "url": "https://www.labelon.kr", --> 데이터셋 url
      "date_created": "2021-01-15 18:38:58" --> 데이터 생성일자
    },
    "data_type": "TTA 10% 데이터세트", --> 데이터셋 구분
    "data_subtype": "train", --> 데이터셋 하위구분
    "images": [ --> 이미지 정보

    {
      "image_id": 000, --> 이미지 번호
      "image": "파일명.jpg",
      "category": "생활" --> 슈퍼헤더 카테고리
      "weather": "맑음" --> 날씨 메타데이터
    }

    "question": [ --> 질의응답 정보

    {
      "question_id": 000000, --> 질문 식별번호
      "image_id": 000, --> 이미지 번호
      "question": "질문",
      "answer": "답변",
      "answer_type": "판별 값"
    }

    ]
  }
]

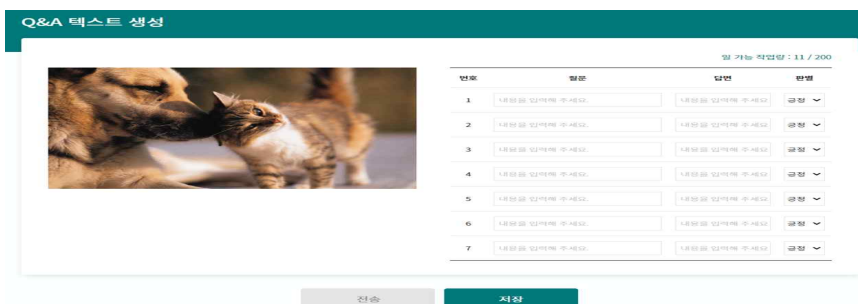
```

VQA 데이터셋 질문의 JSON 파일형태 예시

- 어미를 구분하여 경어나 평어를 만드는 것에 대해서는 데이터 활용의 다양성을 위해 제한하지 않으나, 질문은 반드시 의문문으로 작성되어야 한다.

◆ 질문/답변 어노테이션

- 작업자 한 사람은 이미지를 확인 후 질문과 답변을 생성하고 판별박스(Confidence)를 드롭다운으로 노출, 수정하여 저장한다.



저작도구의 질의응답 어노테이션 화면 예시

◆ 판별 박스

- 긍정: 질문에 대한 답변이 참이거나 참에 거의 가깝다고 볼 수 있는 경우
- 부정: 질문에 대한 답변이 거짓이거나 거짓에 거의 가깝다고 볼 수 있는 경우
- 아마도: 이미지를 통해 유추하기 어렵거나 명확한 참, 거짓으로 나뉘지 않는 경우

2.4.3 저작도구

- VQA 학습 데이터 저작도구는 어노테이션/라벨링 작업과 데이터 구축을 전체 프로세스를 지원하는 크

라우드 소싱 플랫폼이다.

- 저작도구는 이미지 데이터의 저장, 도구 관리자 기능, 작업자관리, 질문생성기, 작업관리(작업배정, 결과제출, 반려 등)와 검수 기능을 가진 프로세스와 서버 및 DB로 구성한다.
- 저작도구에는 작업자의 작업 현황과 결과 검색 및 작업 이력이 관리되며, 통계자료와 작업보수의 확인이 가능하도록 기능을 구성한다.

작업자관리

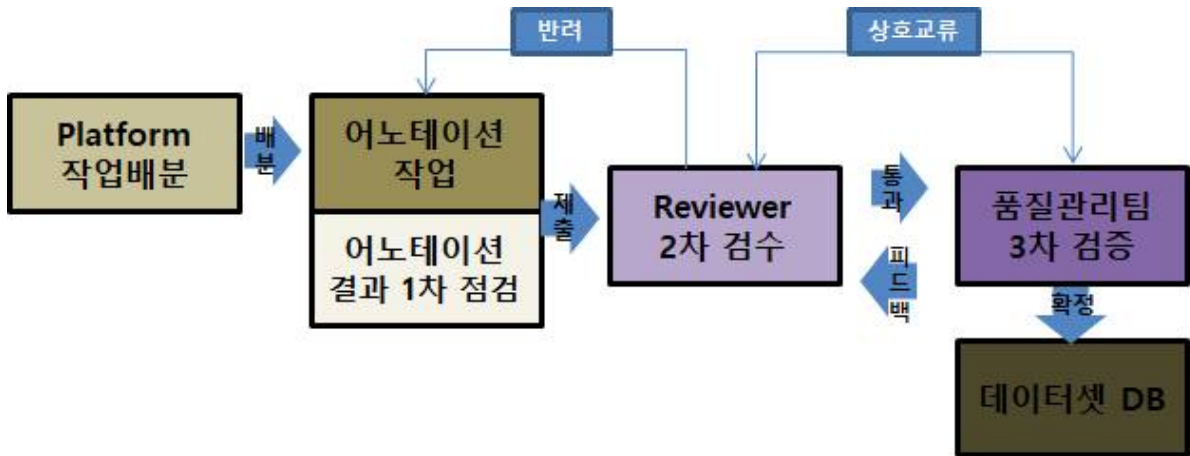
작업	작업기간	작업자정보	합계
6-QUESTION	08.01 - 08.31	2_037_김민교	0.49
15-ANNOTATION	08.06 - 08.10	2_039_유은영	0.49
14-QUESTION	08.06 - 08.10	2_040_양윤영	0.49
3-ANNOTATION	08.01 - 08.31	2_043_김소정	0.49
2-QUESTION	08.01 - 08.31	2_044_박재민	0.49
11-ANNOTATION	07.20 - 07.31	2_045_최원미	0.49
10-QUESTION	07.06 - 07.13	2_046_김유빈	0.49
7-ANNOTATION	07.06 - 07.13	2_047_경정우	0.49
6-QUESTION	07.06 - 07.13	2_048_권해연	0.49
3-ANNOTATION	07.06 - 07.13	2_049_김형병	0.49
2-QUESTION	07.13 - 07.20	2_050_최세롬	0.49
		2_051_박정은	0.49
		자동배분	합계 100.00

회원정보	29명
2102-26_정성현()	
2_038_김지훈()	
2_041_윤채호()	
2_042_홍영국()	
2_056_이지연()	
2_057_이지연()	
2_070_전철우()	
2_077_나현영()	
2_094_정강현()	
314-15_한재엽()	
314-21_김미순()	
314-32_송가도()	
314-53_백종욱()	
314-61_김근아()	
314-73_차선규()	

<작업자 관리화면 예시>

2.5 검수

2.5.1 검수 절차



- ◆ 1차 자동 검수 : 자동 검수 모델(KAERI 용역 개발)을 통해 모델이 판단한 답변과의 일치여부를 참고하여 2차 육안 검수에 참고용으로 활용한다.
- ◆ 2차 전수 검수 : 전문 검수자가 어노테이션된 가공 작업물에 대해서 전수 검수하여 정상 작업물은 통과 처리하고, 부적절하거나 오류가 있는 작업물은 반려 처리한다.
- ◆ 3차 교차 검수 : 저장된 VQA 데이터셋에 대하여 기존 검수자 외의 또 다른 전문 검수자(검수 매니저)가 2차 검수시 승인건에 대해 Cross Check를 진행한다.
- 3차 검수의 Cross Check 또한 2차와 동일하게 육안 전수 검사로 진행한다.
- ◆ 가공 작업에서 비식별화 필요 판단으로 처리된 건에 대하여 수동으로 비식별화 처리한다.
- ◆ 가공 작업 또는 자동 비식별화에서 걸러지지 못한 비식별화 필요 대상도 육안 확인하여 수동으로 비식별화 처리한다.

2.5.2 검수 기준

[질문 검수 기준1] 자연스러운 한국어 문장인가
<ul style="list-style-type: none"> • 자연스러운 어순, 적절한 단어를 사용하고 있는지 확인. • 문장이 잘리거나 미완성이거나 3 ~ 12 개 어절 범위를 벗어난 과도한 질문 • 단어가 잘못되었거나 문장이 어색한 부분이 있는지 확인 • 어순이 부자연스럽거나 비문, 단어 내에서 불필요한 띄어쓰기를 했거나 눈에 띄는 맞춤법 오류시 반려
[질문 검수 기준2] 생성된 질문이 이미지와 연관이 있는가.
<ul style="list-style-type: none"> • 질문이 이미지와 관련된 것을 묻고 있는지 확인. • 이미지 내 실제 객체가 아니어도, 분위기, 정황, 맥락 등을 질문할 수 있음. • 장면과 상황으로부터 유추할 수 없는 정도로 비약적인 추론 질문은 반려처리.
[답변기준]
<ul style="list-style-type: none"> • 답변은 되도록 명사형으로 작성, 객체 특정을 위해 수식어를 넣는 경우라도 4어절 이상은 불가 •

질문과 답변의 검수 기준

2.5.3 품질 검증 계획

- 생활 및 거주 환경 데이터셋은 다양성, 정확성, 유효성을 기준으로 품질검증을 실시하고자 한다.

품질 검증 항목은 다음과 같으며, 상세한 품질 검증은 TTA와 협의하여 확정된 품질 검증 계획서를 기준으로 한다.

<품질 검증 계획 요약표>

품질특성	항목명	측정 지표	정량 목표	지표 및 목표 설정 근거
다양성	카테고리 분류	수량	12종 / 종별 7,500장(목표치 100% 달성 기준 3%에 해당하는 수량) 이상	MS-COCO data의 Super header 12종을 과제에 맞게 커스터마이징한 카테고리(MS COCO data의 최소, 최대값 비율을 기준으로 분포 기준 설정)
	질문문장 길이 편향성	비율	3~12어절 80% 이상	객체 특정이 가능한 최소 어절 수 이상, 과다한 어절 수 이하로 구성
	정답 유형	비율	yes/no 유형 60~70% number 유형 8~20% other 유형 20~30%	데이터, 카테고리 구분 없이 전체 구축량을 기준으로 Yes/no의 비율이 과다하지 않도록 조절 (단, 총 구축량*700만건*을 상회하여 구축한 분량은 제외하고 비율 확인)
구문정확성	구조 정확도	오류율(%)	1% 이하	
	형식 정확도	오류율(%)	1% 이하	
의미정확성	Image-질문-정답 매칭 정확도	정확도(%)	70% 이상	일정 수량을 샘플링하여 육안 검사
유효성	질의응답 (YSE/NO 유형)	EM	51.13	본 과제와 가장 유사한 난이도로 판단되는 해외 VQA 챌린지 사례를 탐색, VizWiz challenge 2018 에서 기록된 7개의 팀중 중간 순위에 해당하는 4위 hdhs팀의 3가지 유형(YES/NO, Number, Other)을 기반으로 목표를 설정하고 유형별로 일부 상하향 조정하여 목표 설정(종합 상황)
	질의응답 (Number 유형)	EM	28.26	
	질의응답 (Other 유형)	EM	37.69	

2.5.4 검수 도구

작업결과물에 대한 검토와 확인관리를 체계적으로 협업하고 오류 발생을 최소화하기 위한 크라우드 기반 저작도구를 이용하여 검수 작업을 관리한다.

☐ 검수자관리

작업목록		검수자정보		
작업	작업기간	2)02-13_최승연()	0.49% (전체 0건 중 약 0건)	검수자060
6-QUESTION	08.01 ~ 08.31	2)02-14_최경민()	0.49% (전체 0건 중 약 0건)	검수자070
15-ANNOTATION	08.06 ~ 08.10	2)02-18_김영민()	0.49% (전체 0건 중 약 0건)	검수자080
14-QUESTION	08.06 ~ 08.10	2)02-19_최정형()	0.49% (전체 0건 중 약 0건)	검수자090
3-ANNOTATION	08.01 ~ 08.31	2)02-20_최기은()	0.49% (전체 0건 중 약 0건)	검수자100
2-QUESTION	08.01 ~ 08.31	2)02-22_이성민()	0.49% (전체 0건 중 약 0건)	검수자110
11-ANNOTATION	07.20 ~ 07.31	2)02-24_송선희()	0.49% (전체 0건 중 약 0건)	검수자020
10-QUESTION	07.06 ~ 07.13	2)02-28_최거원()	0.49% (전체 0건 중 약 0건)	검수자030
7-ANNOTATION	07.06 ~ 07.13	2)02-29_서민영()	0.49% (전체 0건 중 약 0건)	검수자040
6-QUESTION	07.06 ~ 07.13	2)02-30_이변주()	0.49% (전체 0건 중 약 0건)	검수자050
3-ANNOTATION	07.06 ~ 07.13	2)02-32_정선우()	0.49% (전체 0건 중 약 0건)	검수자060
2-QUESTION	07.13 ~ 07.20	2)02-33_소지현()	0.49% (전체 0건 중 약 0건)	검수자070
		2)02-34_전승구()	0.49% (전체 0건 중 약 0건)	검수자070

201명

[저장]

검수자관리 화면 예시

☐ 작업목록 확인 3 건 | 기간 3 건

No	이미지	질문(변경전)	질문(변경후)	처리	질문상태	결재상태
1		남자는 무엇을 들고 있습니까?	남자는 무엇을 들고 있습니까?	<input type="radio"/> 승인 <input type="radio"/> 반려	확인	기안
2		이 사람은 모자를 쓰고 있습니까?	이 사람은 모자를 쓰고 있습니까?	<input type="radio"/> 승인 <input type="radio"/> 반려	확인	기안
3		저 남자는 무엇을 하는거야?	저 남자는 무엇을 하는거야?	<input type="radio"/> 승인 <input checked="" type="radio"/> 반려 자연스러운 문장으로 변경요청함. [저장]	확인	기안

검수/검증 작업관리 화면 예시

2.5.5 기타 품질관리 활동

구축한 VQA 데이터셋을 활용한 해커톤 경진대회를 개최하여 데이터셋의 적정성과 활용도를 검증하고 행사결과는 통계 처리와 분석을 통해 데이터셋의 유용성 평가에 활용하고 도출된 의견을 데이터셋 구축에 반영하여 데이터셋의 품질 개선에 활용한다.

2.6 활용

2.6.1 활용 모델

2.6.1.1 모델 학습

- 서비스 모델 개발에 적용하는 모델은 Co-attention 기반의 한국어 VQA를 활용하며, 파라미터 관련 사항은 개발단계에 추가한다.

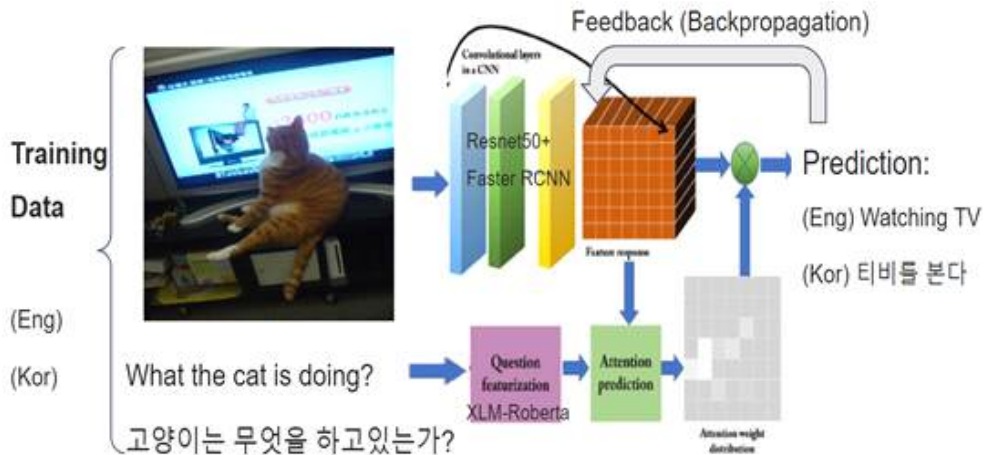


그림 16 Co-attention기반의 한국어 VQA 모델

- 학습데이터의 입력으로 이미지와 질의(영어, 한국어), 응답(영어, 한국어)을 입력받는다.
- 입력으로 받은 이미지는 ResNet 50을 백본으로 사용하고 Faster RNN을 이용하여 이미지의 지역(Region)정보를 표현하는 하나의 매트릭스로 자질을 추출한 결과를 담는다
- 이미지와 질문의 단어 자질을 양방향 attention 기법과 Multi-layered Perception(MLP) 기반의 분류기를 활용하여 정답을 예측한다.

2.6.1.2 서비스 활용 시나리오

- 어린이, 노인, 개인 일상, 실내 가구 4가지의 영역으로 일상에 밀접한 응용 서비스 개발
- 어린이 생활 VQA 데이터를 활용한 '어린이 안전 지킴이 AI' 서비스 개발
- 노인 생활 VQA 데이터를 활용한 '노인 생활 돌봄 AI' 서비스 개발
- 개인 일상 생활 VQA 데이터를 활용한 '개인 생활 추천 상담 AI' 서비스 개발
- 실내 가구 및 가전 배치 VQA 데이터를 활용한 '3D 환경 기반 실내 공간 가구 및 가전기기 배치 상담 AI' 서비스 개발
- 대상시설 및 주변환경에서 수집한 장면 이미지를 학습하고 VQA 모델기반 상황인식 서비스용 서버를 활용한 한국어 VQA 서비스제공 API 프로그램 구성

2.6.2 데이터 제공

◆ 시각정보기반 VQA 서비스

- 시각정보기반 학습용 데이터셋은 NIA에서 운영하는 AI Hub사이트(www.aihub.or.kr)에 공개하며, 데이터셋과 함께 데이터 구조와 설명, 데이터 사용환경, 사용방법 등을 제공한다.
- 데이터셋 산출물은 AI Hub 공개에 앞서 데이터에 대한 검증과 평가를 하고 공인기관의 검증을 받은 데이터셋을 공개 포털에 등록한다.
- 학습 데이터셋 저작도구에 대한 기능과 사용방법 등 매뉴얼을 AI허브에 공개하고 용이하게 접근할 수 있도록 지원한다.