

개요: 위해물품 엑스레이 이미지 데이터셋이란?

위해물품 엑스레이 이미지 데이터 셋은 공항 및 항만의 위해물품 분야 90%, 저장매체 10% 구성으로 (주)엠펙시스템에서 구축하였으며, 위해물품의 경우 27개 품목, 570개 샘플, 이미지는 445,798건으로 구성되며 저장매체의 경우 7개 품목, 85개 샘플, 이미지는 50,158 건으로 구성되어 있다.

X-Ray 위해물품 자동 인식 분야의 특수한 이미지로 표준 데이터로 활용될 수 있다. 항만, 항공 등 테러 또는 은닉성 위험 물품에 대한 감시가 필요한 기관, 기업, 공공기관 등 산업재산 및 정보 유출 방지를 위한 목적으로 활용도가 높을 것으로 예상된다.

기내반입금지 적발 현황

구분	반입금지 물품 적발건수								
	모사총기류	가위,도검류	공구,흥기류	인화성	화학물질	라이터	비활성탄	기타	소계
2013년	590	212,448	27,049	41,580	178	136,351	3,052	1,546,952	1,968,200
2014년	3,603	171,384	20,867	42,892	55	117,873	5,570	1,730,693	2,092,937
2015년	5,238	179,721	21,590	48,258	19	146,973	2,548	1,643,689	2,048,036
2016년	8,237	206,361	29,058	68,315	60	188,320	517	2,570,953	3,071,821
2017년	9,413	208,039	31,197	74,143	84	145,023	79	2,100,786	2,568,764
2018년8월	8,294	154,032	22,541	60,784	82	98,796	1,139	1,758,352	2,104,020

*자료: 김영진 의원실(출처: 인천국제공항)

*가위류 날 길이 6cm이상(기타는 액체류, 젤류,스프레이류, 음식류)

그림 1 인천공항 기내 반입 금지 물품 적발 현황(출처 인천국제공항)

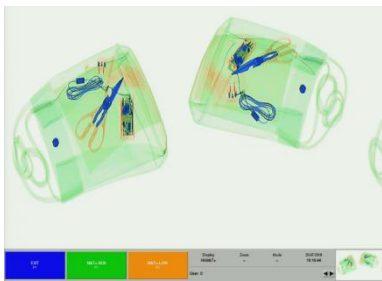

데이터셋의 구성

표준 데이터 셋 411,926건중 실제 X-Ray 검출 데이터의 물체와의 겹침을 고려하여 품목별, 단일 과 복합으로 데이터 방식을 정의하고, X-Ray 제조사별 (Rapiscan, Smith, Astrophysics) 이미지 표출의 차이가 있으나, 데이터셋은 동일(1:1:1) 수량으로 생성하였다.

데이터 종류	구분	내용	제공 방식
단일	단일기본	샘플1개 생성	JSON, XML 파일

데이터셋	단일+비품목	샘플1개에 일반 품목을 동시에 넣어 생성	
복합 데이터셋	복합 품목	샘플 1개이상, 타 위해물품과 동시에 생성	JSON, XML 파일
	복합+비품목	샘플1개이상과 일반 품목을 동시에 생성	

- 엑스레이 제조사별 생성 이미지 비교

Rapiscan	Smith Detection	Astrophysics
		
1회 생성시 3개 표출	1회 생성시 이미지 2개 표출	1회 생성시 1개 표출

데이터셋의 설계 기준과 분포

데이터셋 설계시 국토부 고시 위해 물품 중 공항, 항만 등의 위해물품과 물리보안 분야의 32품목 570종을 선정 하였고, 저장매체 7개 품목 35종의 학습 데이터셋을 구축하였다.

위해물품			저장매체		
품목	샘플	구축데이터	품명	샘플	구축데이터
32개	570종	445,798	7개	85종	50,158

그림 2 데이터셋 구성

위해물품						
 12,000(15)	 20,000(25)	 20,000(25)	 14,000(20)	 14,000(20)	 7,500(15)	 7,500(15)
 7,500(15)	 5,000(10)	 20,000(25)	 7,500(15)	 15,000(30)	 10,000(20)	 24,500(35)
 24,500(35)	 24,500(35)	 17,500(25)	 15,000(25)	 15,000(25)	 12,500(25)	 5,000(10)
 16,000(20)	 16,000(20)	 5,000(10)	 16,000(20)	 12,500(25)	 20,000(25)	
저장매체						
 7,500(15)	 10,000(20)	 10,000(20)	 5,000(10)	 5,000(10)	 2,500(5)	 2,500(5)

위해물품							
알코올	5종	총	20종	플라이어	20종	라이터 오일	15종
송곳	20종	총기 부품	25종	휴대용 가스	10종	저장매체	
도끼	20종	망치	20종	톱	35종	휴대폰	15종
야구배트	10종	수갑	5종	가위	35종	USB	20종
건전지	5종	칼	35종	드라이버	25종	노트북	20종
탄약	15종	라이터	25종	고체연료	10종	테블릿PC	10종
끌	15종	액체(100ml)	25종	스패너	25종	SSD	10종
전자담배(궐련형)	10종	성냥	10종	휴대용 배터리	30종	하드디스크	5종
전자담배(액상형)	10종	금속파이프	25종	신나	5종	외장하드	5종

그림 3 품목별 분류 분포

데이터 구조

- 호환이 가능한 구조의 파일 형태 이미지 구조

File Format	BMP	JPG	PNG
압축여부	무손실	손실	무손실
사이즈 (Full HD기준)	4MB	900KB	1.5MB
표현 색상수	24Bit	24Bit	24Bit
생성속도	느림	빠름	빠름
적합성	2nd	3rd	1st

- Bounding Box(사각형)와 Segmentation(폴리곤) 라벨링 구조

	Bounding Box	Segmentation
목적	데이터를 학습 시키기 위한 정보	
형태	Rectangle	Polyline
활용	Object Detection, Classification	Object Detection, Classification
주요 사용	Yolo, SSD, Faster-RCNN	Mask RCNN

- 장비별 JSON 파일 하나에 Annotation 정보가 기입되는 형태와 이미지 당 1개의 XML
- 파일별 Annotation 정보가 기입되는 형태로 분리.

기본 템플릿 : json

```
{
  "images" : [],
  "categories" :[],
  "annotations":[]
}
```

<Image 정보 포맷>

```
{
  "id": (정수)[이미지 ID Key값]
  "dataset_id": (정수)[해당 이미지가 참조 되어 있는 dataset ID key],
```

```

"path": "(문자열)[이미지 참조 경로],
"width": (정수) [이미지 가로 pixel 길이],
"height": (정수) [이미지 세로 pixel 길이],
"file_name": "(문자열)[이미지 파일명] "
}
<categories 정보 포맷>
{
"id": (정수)[물체 Key],
"name": "(문자열)[물체명]",
"supercategory": "(문자열)[물체의 항목]",
"color": "(문자열 RGBCode)[Annotation시의 색상]",
"metadata": "(배열)[기타 필요 메타 데이터 입력 필드]"
}
<annotation 정보 포맷>
{
"id": (정수)[Annotation ID],
"image_id": (정수)[참조 이미지 ID],
"category_id": (정수)[Category 즉 물체 정보],
"segmentation": (2차원 실수)[polyline의 좌표 정보
( 각 점 에 대해 x,y를 반복하여 입력)],
"area": [segmentation 보조정보],
"bbox": (정수 or 실수)[x,y,width,height],
"iscrowd": (bool)[복잡도 표시],
"color": "[annotation color 정보]",
"metadata": "(배열)[기타 필요 메타 데이터 입력 필드]"
}

```

데이터 구축 과정



그림5 학습용 데이터 구축 과정

구분	업무내용	산출물
물품선정	<ul style="list-style-type: none"> • 일정수립 • 데이터 생성 물품 선정 • 샘플 구매 및 제작 물품 설계 및 구매 • 과제 진행 장비 구매 • 조직구성 	과제준비보고서 - 과제 수행계획서 - 샘플 리스트 - 제작 사양 도면 - 시스템 구성도

물품분류	<ul style="list-style-type: none"> • 데이터 생성 방안 상세 기획 • 소프트웨어 툴 개발 (커스터마이징) • 시스템 시험 테스트 	과제 사용 툴 SW 테스트 리포트
이미지생성	<ul style="list-style-type: none"> • 물품별 작업 계획서에 따른 이미지 생성 • 이미지 데이터 저장 및 분류 	작업 내역 보고서
이미지검증	<ul style="list-style-type: none"> • 오류 이미지 검사 • 오류 이미지 레포팅(관리대장) 	오류이미지 리스트
메타생성	<ul style="list-style-type: none"> • 물품별 작업 계획서에 대한 라벨링 작업 • 라벨링 데이터 저장 및 분류 	작업 내역 보고서
메타검증	<ul style="list-style-type: none"> • 오류 라벨링 검사 • 오류 라벨링 레포트(관리대장) 	오류 라벨링 리스트
AI학습테스트	<ul style="list-style-type: none"> • 품목 학습 및 결과 확인 • 교차 품목 학습 및 결과 확인 • 전체 학습 및 결과 확인 • 각 학습별 로그 분석, 결과 저장 	학습 계획서 학습 결과 보고서
시범서비스	<ul style="list-style-type: none"> • DB서버 설치 및 구축 • 최종 데이터베이스 구축 • SQL Query 제작(입력, 수정, 삭제) • 검색서비스 (외부 서비스 방안) • 사용자 매뉴얼 	검수확인서 인터페이스 정의서 사용자 매뉴얼 Software Tool 소스

그림6 학습용 데이터 구축 상세 과정

검수와 품질 확보

데이터셋 구성 및 이미지 생성에 대한 이미지생성 가이드 라인을 작성하고, 검수를 위한 인력의 전문성을 고려하여 품질을 확보하고, 다양한 샘플 데이터를 추가하여 생성하였다.

구성인원	검증업무	기타
데이터셋	3명(CBT 교육이수 및 공항 업무 경험자)	
이미지	3명(대기업 보안 검색 업무 경력자)	
엑스레이 스캐너	3개사 제품(라피스캔, 스미스, 아스로피직스)	



그림 7 품질 확보를 위한 2단계 품질 검수 체계

데이터 구축 담당자

수행기관(주관) : ㈜엠폴시스템 (전화: 02-6929-0735), 이메일: rndcenter@mpole.co.kr