

# 유해화학물질 누출영상의 오토 라벨링 오인식 유형에 관한 연구

김무빈, 김경배\*

서원대학교, \*서원대학교

rlaanqls124@naver.com, \*gbkim@seowon.ac.kr

## A Study on the Types of Autolabeling Misrecognition of Hazardous Chemical Spill Images

Kim Mu Bin, \*Kim Gyoung Bae

Seowon Univ, \*Seowon Univ.

### 요약

화학산업의 규모가 커짐에 따라 화학사고 또한 증가하고 있다. 서원대학교와 바이브 컴퍼니, 우경정보기술은 빅데이터를 활용한 인공지능기반의 유해화학물질 사고 대응 시스템을 구축 중이다. 시스템 구축을 위해서는 AI학습을 위한 라벨링 된 데이터가 필요하다. 많은 양의 자료가 필요한만큼 오토 라벨링 방식을 활용한다면 학습데이터 구축의 효율성을 높일 수 있을 것이다. Anaconda의 labelme를 통해 영상을 이미지로 분할하여 유해화학물질 사고 대응시스템의 학습데이터 생성을 위한 오토 라벨링 기법의 인식율과 오인식 유형을 파악하였다. 오토 라벨링으로 생성된 태깅 영역이 smoke 중 일부만 인식하는 유형이 가장 많은 비율을 차지하였고, flame이 아닌 부분을 flame으로 인식하는 유형이 그 다음 비율을 차지하였다. 이외에 smoke를 전혀 인식하지 못하거나 사물을 flame으로 인식하는 유형이 적은 부분을 차지하였다. 이를 통해 향후 유해화학물질 사고 대응 시스템을 위한 라벨링 데이터의 문제점을 개선하고 학습하여, 이를 토대로 인공지능기반의 유해화학물질 사고 대응 시스템을 구축할 수 있을 것이다.

### I. 서론

최근 국내 화학산업의 규모는 '19년 168조원으로 세계시장 점유율 6위를 차지하며 국가 제조업 총 생산의 9.3%를 차지한다. 국내 화학산업의 생산은 석유화학업종 및 고무·플라스틱 업종의 지속적 증가 속에 정밀화학업종이 성장을 주도하고 있다. 화학산업은 2019년 168조원에서 2030년에는 202조원까지 늘어날 전망이다[1].

화학 산업이 커지는만큼 화학사고 또한 증가하고 있으며 화학물질안전원의 화학물질 사고현황에 따르면 2014.01.08.부터 22.08.31까지의 사고 중 누출 사고가 가장 많은 것으로 나타났다[2].



그림 1. 화학물질 사고 현황

서원대학교와, 바이브컴퍼니, 우경정보기술은 빅데이터를 활용한 인공지능기반의 유해화학물질 사고 대응 시스템을 구축 중이다.[3] 해당 시스템은 유해화학물질 사고 발생 시 화재나 누출 영상을 보고 어떤 물질인지 미리 사전에 파악하여 사후피해를 줄이는 시스템이다. 시스템 구축을 위해서는 AI학습을 위한 라벨링 된 데이터가 필요하다. 많은 양의 자료가 필요한만큼 오토 라벨링 방식을 활용한다면 학습데이터 구축의 효율성을

높일 수 있을 것이다.

따라서 본 논문에서는 유해화학물질 사고 대응 시스템 구축을 위한 데이터를 생성하기 위한 오토라벨링 기법의 인식율과 오인식 유형을 파악하였다.

### II. 본론

#### 1. 데이터 전처리(이미지 분할)

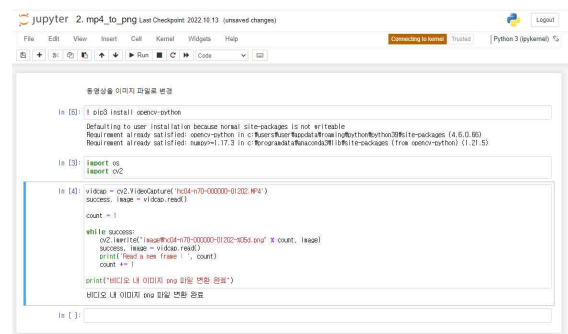


그림 2. 이미지 분할

5개의 유해화학물질의 누출 영상을 5,782개의 이미지 파일로 분할하여, 라벨링에 사용할 수 있도록 데이터 전처리 과정을 거쳤다.

## 2. labelme 태깅



그림 3. labelme 태깅

Anaconda의 labelme를 통해 flame과 smoke 두 개 태그를 설정하여 연기와 불꽃을 구분하는 오토 라벨링을 진행하였다.

## 3. 오토 라벨링의 오인식율

영상 구분	이미지 개수	오인식 이미지수	오인식율
A	1,027	863	84%
B	369	137	37%
C	421	229	54%
D	2,688	1,745	65%
E	1,277	469	37%

표 1. 오토 라벨링 오인식율

위 표의 수치처럼 영상의 환경이나 조건에 따라 flame과 smoke를 아예 인식 못하거나 바운딩 박스안에 flame과 smoke가 완벽히 들어가지 못하는 등 낮은 인식율을 보였다.

## 4. 오토라벨링의 오인식 유형

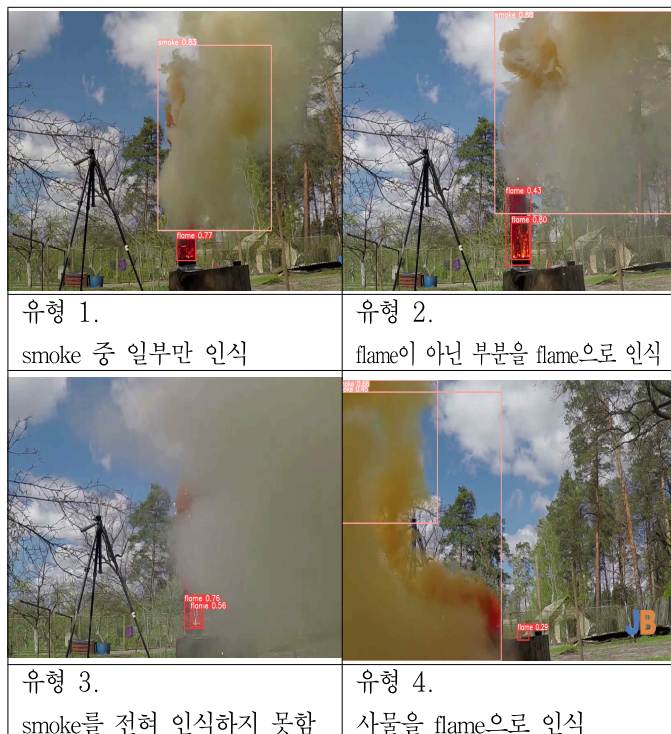


그림 4. 오인식 유형

위의 그림처럼 오토 라벨링에서의 오류 발생 유형은 총 4가지로 나타났다.

유형 1.	유형 2.	유형 3.	유형 4.
83%	4%	0.6%	0.1%

표 2. 오인식 유형 비율

오인식 전체 비율 중 가장 많이 오류가 많이 발생한 유형은 유형 1로 83%, 대부분의 비율을 차지하였다. 유형 2가 4%로 그 다음 비율을 차지하였고 다른 2개의 유형은 발생 비율이 상대적으로 적었다.

## III. 결론

본 연구에서는 유해 화학물질 영상을 프레임단위로 이미지화하여 오토 라벨링 방식으로 라벨링하였다. 오토 라벨링 방식은 효율적이지만 특정 상황이나 조건에서 인식하지 못하는 등 낮은 인식율의 성능을 보였다. 오토 라벨링으로 생성된 태깅 영역이 smoke 중 일부만 인식하는 유형이 83%로 가장 많은 비율을 차지하였고 flame이 아닌 부분을 flame으로 인식하는 유형이 4%로 그 다음 비율을 차지하였다. 이외에 smoke를 전혀 인식하지 못하거나 사물을 flame으로 인식하는 유형이 적은 부분을 차지하였다. 오인식된 라벨링 데이터의 유형을 파악하여 향후 생성한 데이터의 문제점을 개선하고 학습하여 이를 토대로 인공지능기술 기반의 유해화학물질 사고 대응 시스템을 구축할 수 있을 것이다.

## ACKNOWLEDGMENT

본 연구는 과학기술정보통신부와 정보통신산업진흥원 주관으로 소방청 컨소시엄에서 수행하는 “AI융합 유해화학물질 관독시스템 사업”(2022~2024)의 지원을 받았음

## 참 고 문 헌

- [1] 한국화학산업연합회, “한국 화학산업의 경제 및 사회적 영향 분석”, 2020.
- [2] 화학물질 안전원 화학물질종합정보시스템 통계자료 (<https://nics.me.go.kr>)
- [3] 김연진, 박봉섭, 김경배 “인공지능기술 기반의 유해화학물질 사고 대응에 관한 연구”, 한국통신학회, 2022