

유해화학물질 판독을 위한 인공지능 학습 데이터 라벨링에 관한 연구

김연진, 조숙경, 박봉섭*, 김경배

서원대학교, *소방청

anne6497@naver.com, skyindb@naver.com *leeraksa@korea.kr, gkim@seowon.ac.kr

A Study on Labeling of Artificial Intelligence Learning Data for the Identification of Hazardous Chemicals

Kim Yeon Jin, Jo Sook Kyung, Park Bong Seop*, Kim Gyoung Bae

Seowon Univ., *National Fire Agency

요약

소방청 화재통계연감에 따르면 화재건수는 감소하는 추세이지만 화학적 요인의 화학사고 및 화재발생 건수는 증가하고 있다. 증가하고 있는 국내 화학사고의 발생 및 피해를 최소화하고자 서원대학교, 바이브컴퍼니, 우경정보기술은 컴소시업을 구성하여 AI융합 유해 화학물질 판독시스템 연구를 진행하고 있다. 본 논문에서는 인공지능 기반의 유해화학물질 판독을 위한 학습용 데이터 구축 단계에서 사용되는 바운딩박스 라벨링 방법의 문제점에 대해 고찰하고, 유해화학물질의 누출·화재·폭발사고 영상 및 이미지에 대한 적절한 바운딩박스 라벨링 방법과 기준을 제시하였다. 유해화학물질에서의 바운딩박스 라벨링 방법에 대한 기준을 제시하여 화학물질 특성에 따른 유해화학물질 사고 예측 및 사고대응 시나리오를 마련하여 인적·물적인 사회적 손실을 최소화하고자 한다.

I. 서론

2021년 소방청의 화재통계연감[1]에 따르면 2020년 화재건수 38,659건은 전년도에 비해 1,444건이 줄어, 3.6% 감소한 수치로 2017년(44,178건)을 정점으로 감소세로 돌아서고 있다. 화재건수는 감소하는 추세인 반면, 화학적 요인의 화학사고 및 화재발생 건수는 증가하고 있다. 화학물질안전원의 자료에 의하면 2019년까지 감소추세이던 화학사고는 2020년부터 다시 증가[2]하고 있으며, 이는 화학물질 생산 및 취급시설의 노후화 등으로 인해 발생함을 알 수 있다.

국내 화학사고의 발생 및 피해를 최소화하고자, 과학기술정보통신부와 정보통신산업진흥원의 지원을 받아 서원대학교, 바이브컴퍼니, 우경정보기술은 컴소시업을 구성하여 AI융합 유해 화학물질 판독시스템 연구를 진행하고 있다. 해당 연구에서는 실제 발생한 화재 동영상과 화학물질 종류 등에 따른 영상 데이터를 수집하여 데이터 라벨링 과정을 거쳐 데이터를 가공한다. 가공 데이터를 이용한 인공지능 학습을 통해 화학물질의 종류, 사고 유형 등을 판독한다. 판독 결과는 현장에 출동하는 소방관들에게 신속하고 올바른 초기대응을 지원한다.

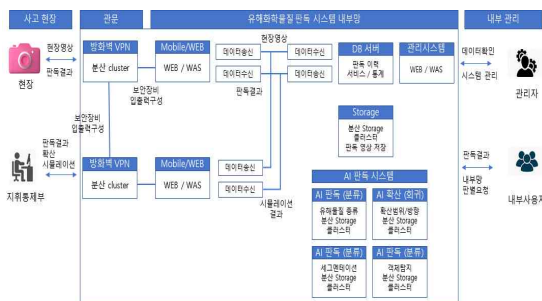


그림 1 AI 유해화학물질 판독 시스템 구성도

본 시스템의 개발을 위해서는 화학물질 사고 중 화재·폭발·누출 등의 영상 데이터에 대한 학습 데이터 구축과 학습 데이터로 만들기 위한 데이터 전처리 과정이 매우 중요하다. 유해화학물질 화재 속성에 대한 정보를 구축하기 위해서는 바운딩박스 라벨링을 통해 연기나 화염, 불꽃을 구분해야 한다. 하지만 그림 2와 같이, 바운딩박스 기법은 직사각형 안에 개체 이외의 빈공간으로 인해 오류가 발생하는 문제를 가진다.

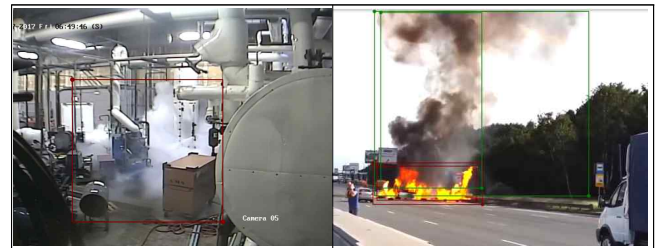


그림 2 바운딩박스 라벨링의 문제점

본 논문에서는 인공지능 기반의 유해화학물질 판독을 위한 학습 데이터 구축의 첫 번째 단계인 라벨링 과정 중 유해화학물질 영상의 바운딩박스 (bounding box) 방법과 기준을 제안한다. 제안된 방법은 화재·누출 등 화학사고의 다양한 상황에서 학습용 데이터의 훈련 시 발생하는 오류를 줄일 수 있다.

II. 본론

1. 유해화학물질 사고 유형

그림 3과 같이 유해화학물질 사고는 크게 누출 화재 폭발의 유형을 나타낸다.

- (누출) 연기가 희미한 경우
- (화재·폭발) 화염과 연기가 겹쳐져 있는 경우
- (화재·폭발) 화염 및 연기에 사물이 겹쳐져 있는 경우

○ (화재·폭발) 화염 및 연기가 분리되어 여러 군데 나타나는 경우



그림 3 유해화학물질의 누출·화재·폭발

2. 유해화학물질 영상의 학습 데이터 라벨링

유해화학물질 영상 데이터를 수집 후 라벨링을 위한 데이터 선별 및 이미지를 추출한다. 인공지능 학습용 데이터 구축을 위해서는 식별 대상의 바운딩박스 및 세그멘테이션 작업이 필요하다. 라벨링 기법 중 하나인 바운딩박스는 물체를 직사각형 모양의 박스 안에 포함되도록 그리는 라벨링 방법으로 데이터 라벨링 작업에서 가장 일반적으로 사용된다. 다양한 유해화학물질 사고 영상의 학습용 데이터에 대한 라벨링은 바운딩박스 범위의 기준에 대한 모호성을 해결해야한다. 이러한 모호성을 해결하고자 다양한 상황에서 유해화학물질 사고영상의 바운딩박스 라벨링 방법에 대한 기준을 그림 4와 같이 제시한다.



그림 4 다양한 유해화학물질 사고 상황에서의 적절한 라벨링 방법 제시

첫째, 누출 상황에서 연기가 희미하게 발생한 경우이다. 연기가 희미하지만 넓은 범위로 연기가 퍼져 있을 때 모든 연기를 바운딩박스 안에 포함시키게 되면 연기 외의 빈 공간이 많아지게 된다. 연기가 희미하면서 동시에 빈 공간도 많아지면, 연기를 제대로 인식하지 못할 확률이 높아지게 된다. 이 경우에는 모든 연기를 다 담지 못하더라도 가장 명확하게 나온 연기의 일부분을 라벨링 하는 것이 학습 데이터를 구축하기에 적절한 방법이다.

둘째, 화염과 연기가 겹쳐져 있는 경우이다. 이 경우도 위와 마찬가지로 화염과 연기가 겹쳐 있다라도 화염·연기가 중복되지 않은 영역을 선택하여 가장 명확한 화염과 연기만을 라벨링 하는 것이 가장 좋다.

셋째, 화염 및 연기에 사물이 겹쳐져 있는 경우이다. 연기의 경우, 넓게 확산되는 특성을 가지기 때문에 대부분 화염이나 연기가 사물과 겹쳐져 있는 경우가 많다. 사물을 피해 바운딩박스로 라벨링을 해도 되지만, 이 경우에는 바운딩박스보다는 여러 개의 점을 가진 선을 활용하여 특정 영역을 라벨링 하는 폴리곤 기법을 사용하는 것이 더욱 적절한 방법이 될 수 있다.

넷째, 화염 및 연기가 분리되어 여러 군데 나타나는 경우이다. 이러한 경우에는 여러 개의 바운딩박스를 이용하여 라벨링 할 수 있다. 다만, 빈 공간과 다른 사물과 겹쳐지는 부분이 최대한 없도록 하는 것이 좋다.

III. 결론

본 논문에서는 인공지능 기반의 유해화학물질 판독을 위한 학습용 데이터 구축 단계에서 사용되는 바운딩박스 라벨링 방법의 문제점에 대해 고찰하였다. 또한, 유해화학물질의 누출·화재·폭발사고 영상 및 이미지에 대한 적절한 바운딩박스 라벨링 방법과 기준을 제시하였다.

연기가 넓고 희미하게 퍼져있는 누출의 경우와 화염·연기가 겹쳐있는 화재나 폭발의 경우에는 모든 식별 대상을 박스 안에 포함시키기 보다는 가장 명확한 대상의 일부분만을 라벨링한다. 또한 화염 및 연기에 사물이 겹쳐져있는 경우에는 특정 영역을 라벨링 하는데 더 효율적인 폴리곤 기법을 사용하여 라벨링 한다. 폭발을 하며 화염 및 연기가 분리되어 여러 군데 나타나는 경우에는 한 박스 안에 포함시키기 보다는 다수의 바운딩박스로 라벨링한다.

본 논문에서 유해화학물질에서의 바운딩박스 라벨링 방법에 대한 기준을 제시함으로써, 향후 가공된 학습용 데이터를 이용한 인공지능 기반의 유해화학물질 판독에서의 인식율을 높여 화학물질 특성에 따른 유해화학물질 사고 예측 및 사고대응 시나리오를 마련하여 인적·물적인 사회적 손실을 최소화하고자 한다.

ACKNOWLEDGMENT

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

참 고 문 헌

- [1] 소방청, 2021 화재통계연감
- [2] 화학물질안전원 (<https://nics.me.go.kr/>)
- [3] 김연진, 박봉섭, 김경배, “인공지능기술 기반의 유해화학물질 사고 대응에 관한 연구”, 한국통신학회 학술대회논문집, pp. 359-360, 2021.