

임베디드 환경에서 딥러닝 기반 전염병 확산 방지 시스템

(Epidemic Spread Prevention System Based Deep Learning in Embedded Environment)

김강희*, 김준현*, 황연우**, Nguyen Minh Nguyen**, 임종태**

(Kanghee Kim, Joonhyun Kim, Yeonwoo Hwang, Nguyen Minh Nguyen, Jongtae Lim)

요약

최근 코로나 바이러스 질병의 확산으로 일상생활, 사회 및 글로벌 경제에 많은 영향을 미쳤다. 본 논문에서는 코로나 바이러스와 같은 전염성 강한 질병의 확산을 예방하기 위해 열화상 카메라와 마이크를 이용한 컴퓨터 비전 기반 알고리즘을 제안한다. 이 알고리즘은 열화상 이미지를 이용해 학습한 YOLOv6n 모델로 사람의 마스크 착용 여부를 검출하며 검출된 바운딩 박스를 이용해 사람의 체온을 측정하고 사람을 추적하여 재실자 수를 계산한다. 또 마이크를 이용해 음향 데이터를 입력 받아 CNN 모델로 기침 소리를 인식한다. 최종적으로 마스크 착용 여부와 재실자 수, 기침 소리 여부, 높은 체온 객체 존재 여부, 밀폐 공간의 면적을 이용해 밀폐 공간의 오염 위험도를 계산한다. 제안하는 알고리즘은 전염성이 강한 질병의 확산 예방의 효과도 볼 수 있지만 열화상 카메라를 사용하기 때문에 사람을 식별할 수 없어 이미지 데이터 저장 과정에서 카메라를 이용한 컴퓨터 비전의 문제점인 개인 정보 보호 문제를 해결할 수 있다.

■ 중심어 : 피플 카운터 ; 인공지능 ; 딥러닝 ; 임베디드 보드 ; 객체 추적

Abstract

The recent spread of the COVID-19 virus has had a significant impact on daily life, society, and the global economy. In this paper, we propose a computer vision-based algorithm using thermal cameras and microphones to prevent the spread of contagious diseases like the COVID-19 virus. This algorithm detects whether a person is wearing a mask using the YOLOv6n model trained on thermal images, measures the person's temperature using the detected bounding boxes, and tracks the person to calculate the number of people present. Additionally, a microphone is used to input a data, and a CNN model is used to recognize coughing sounds. Finally, the algorithm calculates the pollution risk of a confined space by considering factors such as mask usage, the number of people present, coughing sounds, the presence of high-temperature people, and the area of the confined space. The proposed algorithm not only helps prevent the spread of highly contagious diseases but also addresses privacy concerns by not identifying individuals, as thermal cameras are used, thus mitigating the issue of personal data protection in computer vision systems that rely on cameras.

■ keywords : People Counter ; Artificial Intelligence ; Deep Learning ; Embedded Board ; Object Tracking

I. 서론

현대 사회에서 발생하는 전염병의 예방 및 대응은 더욱 중요해졌다. 특히, 최근 코로나 바이러스

(COVID-19)의 급속한 확산은 전 세계적으로 대규모의 사회적, 경제적 영향을 초래하였다[1, 2]. 이로 인해 전염병 대응을 강화하고, 예방에 중점을 둔 효과적인 방안을 모색하는 것이 필수적으로 대두되고 있다 [3, 4].

* 준회원, 홍익대학교 전자전기공학과

** 정회원, 홍익대학교 전자전기공학과

This research was supported by Ministry of SMEs and Startups(RS-2024-00448168, RS-2024-00434571), Ministry of interior and Safety(RS-2024-00461780).

접수일자 : 2025년 12월 05일

게재확정일 : 2026년 02월 12일

수정일자 : 1차 2026년 01월 15일, 2차 2026년 02월 05일

교신저자 : 임종태 e-mail : jlim@hongik.ac.kr

전염병 확산 방지를 위한 다양한 노력 중 하나는 컴퓨터 비전 기술을 활용하여 전염병에 대한 대응을 강화하는 것이다[5]. 특히, 환경 모니터링부터 시작하여, 안전한 환경 조성 및 질병 확산 방지를 위한 혁신적인 시스템 구축이 필요하다[6].

이미지 및 비디오 분석에 관한 컴퓨터 비전 기술은 이미 다양한 분야에서 적용되어 왔다[7]. 전염병 대응과 관련된 컴퓨터 비전 연구도 활발하게 진행되고 있지만[8], 일반 카메라로 환경 모니터링을 할 경우 개인 정보 보호 문제를 피할 수 없다[9].

본 논문에서는 개인 정보 보호 문제를 고려하면서 코로나와 같은 전염병에 대한 대응을 위해 열화상 카메라에 컴퓨터 비전 기술을 적용하고자 한다. 열화상 이미지에 YOLOv6n을 적용하여 마스크 착용 여부를 검출하고 체온을 측정한다. 또한, 객체를 추적하여 밀폐된 공간에서의 재실자 수, 높은 체온을 가진 객체, 마스크를 착용하지 않은 객체 수 및 CNN 모델 기반 음향 인식 알고리즘으로 인식한 기침 소리 발생 여부, 밀폐 공간 면적을 이용하여 오염 위험도를 계산했다.

II. 관련 연구

1. CNN

CNN은 주로 이미지나 영상 처리에 사용되는 딥러닝 알고리즘이다. CNN은 이미지 내의 중요한 특징을 자동으로 추출하고, 그 특징들을 기반으로 분류나 예측할 수 있다.

CNN은 $2D\ n \times n$ 픽셀 이미지를 입력한다. 각 계층은 필터 또는 커널이라고 불리는 2D 뉴런 그룹으로 구성된다. CNN의 특징 추출 계층의 뉴런은 인접한 계층의 모든 뉴런과 연결되지 않는다. 하지만 이전 계층의 입력 이미지 또는 특징 맵에서 일정 크기의 일부 겹치는 뉴런에만 연결된다. 입력의 이 영역을 지역 수용영역(Local Receptive Field)라고 한다. 계층 간의 뉴런 연결

수의 감소는 학습 시간과 오버피팅(Overfitting)의 가능성을 줄인다. 필터의 모든 뉴런은 이전 입력 계층(또는 특징 맵)의 동일한 수의 뉴런에 연결되며 동일한 일련의 가중치와 편향을 가진다. 이러한 요소는 학습 속도를 높이고 신경망의 메모리 사용을 줄인다. 따라서 특정 필터의 각 뉴런은 입력 이미지의 다른 부분에서 동일한 패턴을 찾는다. 서브샘플링 계층은 신경망의 크기를 줄이고 지역 수용영역 및 동일 필터 내에서 공유되는 가중치와 함께 이미지의 이동, 크기 및 왜곡에 대한 신경망의 오류를 줄인다. 서브샘플링을 달성하기 위해 종종 최대/평균 풀링 또는 지역 평균화 필터를 사용한다. CNN의 최종 계층은 출력값을 분류하며 계층 간의 모든 뉴런이 완전히 연결된다[10]. CNN 모델에는 AlexNet[11], Inception[12], ResNet[13], VGG[14] 등이 있다.

2. 객체 검출(Object Detection)

객체 검출은 어떤 이미지에 존재하는 객체를 찾아내고, 해당 객체들을 사각형 바운딩 박스(Bounding Box)로 레이블링하여 그 존재 여부를 표시하는 것을 목표로 한다. 객체 검출 방법은 크게 Region Proposal 기반 방법과 Regression/Classification 기반 방법으로 두 가지 유형으로 나뉜다. Region Proposal 기반 방법에는 주로 R-CNN [15], Fast R-CNN [16], Faster R-CNN [17], Mask R-CNN [18]이 포함된다. Regression/Classification 기반 방법에는 주로 AttentionNet [19], YOLO [20], Single Shot MultiBox Detector (SSD) [21], YOLOv2 [22]가 포함된다.

Region proposal 기반 방법은 2 stage 프로세스, 인간 두뇌의 집중 메커니즘과 유사하다. 먼저 전체 이미지를 스캔한 후 RoI(Regions of Interest)에 집중한다. 관련 작업 중 가장 대표적인 것은 Overfeat [23]이다. 이 모델은 CNN을 sliding window 방법에 적용하여, 주요 특징맵으로부터 바로 바운딩 박스를 예측하고 이로부터

객체를 분류한다. 이미지 전체를 스캔한다는 점에서 큰 메모리와 긴 시간이 걸린다는 단점이 있다.

Regression/Classification 기반 방법은 1 stage 프로세스이다. 입력 이미지를 SxS 그리드로 나누고, 이미지 픽셀에서 직접적으로 바운딩 박스 좌표와 클래스 확률로 매핑한다. 따라서 Region proposal 기반 방법보다 시간 비용이 적다는 장점이 있다.

3. 객체 추적(Object Tracking)

객체 추적은 연속된 이미지에서 객체를 검출하고 추적하는 컴퓨터 비전의 중요한 작업 중 하나이다. 객체 추적 대상은 첫 프레임에서 지정되며 다음 프레임에서 검출 및 추적되어야 한다. 객체 추적 알고리즘을 설계할 때 고려해야 할 점들은 다음과 같다. 먼저 배경의 변화 및 조명 변화와 같은 복잡한 상황에서도 객체를 추적할 수 있어야 한다. 다음으로 환경 변화뿐만 아니라 객체 자체의 변화에도 대응할 수 있어야 한다. 예를 들어 객체의 움직임과 같은 변화에도 대응해야 하므로 객체 추적 알고리즘은 객체의 외관적 특성을 검출하고 추적할 수 있어야 한다. 마지막으로 실시간으로 객체를 추적하기 위해 객체 추적 알고리즘은 개발 환경에 맞는 높은 처리 속도를 가져야 한다[24].

객체 추적은 단일 객체 추적(Single Object Tracking, SOT)과 다중 객체 추적(Multiple Object Tracking, MOT)으로 나눌 수 있다. 단일 객체 추적은 하나의 객체를 추적하는 작업으로 이미지나 비디오에서 특정 대상을 추적하는 것이 목표이다. 다중 객체 추적은 하나 이상의 객체를 추적하는 작업으로 동시에 여러 객체의 위치와 움직임을 추적하는 것이 목표이다. 본 논문에서는 실시간으로 비디오에서 여러 객체를 추적하는 것이 목표이기 때문에 다중 객체 추적에 중점을 둔다.

다중 객체 추적 알고리즘은 검출을 통해 객체

를 추적한다. 비디오 프레임에서 객체를 검출해 바운딩 박스를 추출하고, 이러한 검출을 사용하여 추적 프로세스를 수행한다. 이때, 동일한 대상의 바운딩 박스에 동일한 ID를 할당하기 위해 특정 알고리즘을 사용한다.

Deep SORT[25]는 다중 객체 추적 알고리즘으로 딥러닝을 이용한 고유 특징 추출 및 유사도 측정, GPS 및 항공 등에서 객체의 다음 위치를 예측하는 데 이용되는 칼만 필터, IoU 측정 방식을 결합한 알고리즘이다. Deep SORT에서는 Re-id를 위해 헝가리안 알고리즘[26]을 이용하는데, 이에 필요한 정보로 고유 특징과 IoU 값을 이용한다. 하지만, 이 값들만을 이용하면 미감지된 객체로 인한 정보 소실과 감지된 객체 영역의 측정 노이즈 때문에 Re-id에 어려움이 있다. 그래서 칼만 필터를 채용하여 객체 감지의 노이즈를 줄이고, 미감지 객체가 있을 경우에는 필터의 예측값을 Re-id 알고리즘에 이용한다.

III. 제안하는 시스템

1. 객체 검출 모델

본 논문에서는 시스템에 적합한 검출 모델을 선정하는 데 두 가지 주요 기준을 고려하였다. 첫 번째로는 객체 추적이 실시간으로 작동할 수 있는지 여부로, 이를 위해서는 객체 검출 모델의 속도가 빠르기에 대한 여부가 모델 선정에 중요한 부분이 되었다. 두 번째로는 실험 환경에서 충분한 성능을 발휘할 수 있는지 여부이다. 속도가 빠른 모델이 정확도 면에서 부족할 수 있기 때문에, 적절한 성능을 유지하면서도 실시간으로 작동할 수 있는 모델을 선택하는 데 주안점을 두었다.

객체 검출 모델로는 YOLOv6n을 선정하였다. 그림1에서 YOLO의 버전별 성능을 확인할 수 있다. 임베디드 보드에서 실시간으로 동작해야 하기 때문에 각 버전별 속도가 빠른 모델들의 성능을 비교했다.

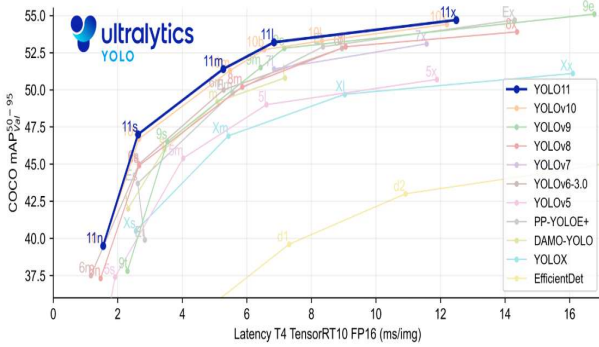


그림 1. YOLO 버전별 성능

속도(ms)는 낮을수록, mAP는 높을수록 좋은 모델이라고 판단하여 mAP/ms를 계산하였다. 그 중 제일 큰 값을 가진 YOLOv6n을 객체 검출 모델로 선정하였다. 본 논문에서는 마스크 착용 여부를 객체를 검출하였다.

표 1. YOLO 모델별 성능 비교

Model	Speed (ms)	COCO mAP	mAP/ms
YOLOv6n	1.1	37.5	34.1
YOLOv8n	1.3	37	28.5
YOLOv10n	1.4	39	27.9
YOLOv11n	1.4	39	27.9

2. 객체 추적 모델

객체 추적 모델은 검출한 객체의 바운딩박스를 이용해 알고리즘을 구현하였다. 구현한 객체 추적 알고리즘은 그림 2와 같다. 먼저 마스크 착용 여부를 검출한 바운딩 박스의 인덱스와 중심 좌표를 3 프레임 동안 저장한다. 3 프레임 중 처음 프레임에서의 바운딩 박스들과 마지막 프레임에서의 바운딩 박스들의 중심 좌표의 거리를 계산한다. 마지막 프레임에서 검출된 바운딩 박스와 처음 프레임에서의 검출된 모든 바운딩 박스 간 거리들 중에서 최솟값을 찾아, 마지막 프레임에서의 검출된 바운딩 박스의 인덱스를 중심 좌표 간 거리가 최소인 처음 프레임에서의 검출된 바운딩 박스 인덱스로 바꾼다. 인덱스가 바뀌면 최솟값에 해당하는 처음 프레임에서의 인덱스와 마지막 프레임에서의 인덱스에 해당하는 행과 열을 삭제한다. 이 과정을 반복하여 마지막 프레

임에서의 검출된 바운딩 박스의 인덱스가 모두 바뀌거나 처음 프레임에서의 검출된 바운딩 박스의 인덱스가 없으면 다음 단계를 진행한다. 마지막 프레임에서의 검출된 바운딩 박스의 인덱스가 모두 바뀌지 않았는데, 인덱스 간 거리의 배열에서 처음 프레임에서의 검출된 바운딩 박스의 인덱스가 없으면, 바뀌지 않은 바운딩 박스의 인덱스를 바꾼 바운딩 박스의 인덱스의 최대에 1을 더한 값으로 바꾼 후 새로 나타난 객체로 저장한다. 마지막 프레임에서의 검출된 바운딩 박스의 인덱스가 모두 바뀌었지만, 처음 프레임에서의 검출된 바운딩 박스의 인덱스가 남게 되면, 사라진 객체로 저장하고 객체 추적 알고리즘이 종료된다. 객체 추적 알고리즘이 종료된 후 마스크 착용 여부를 검출한 바운딩 박스에 인덱스를 출력하도록 하였다.

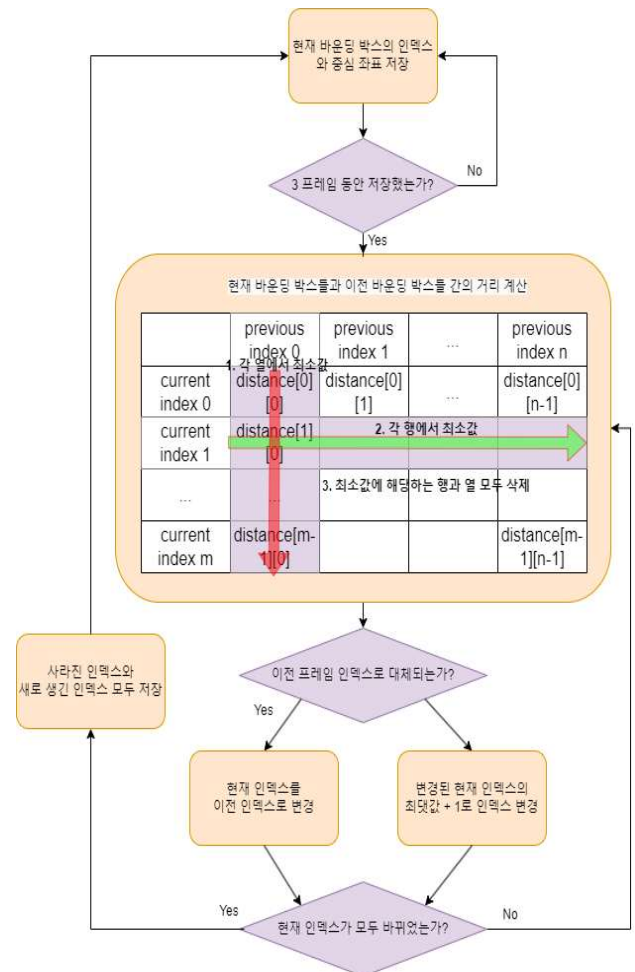


그림 2. 객체 추적 알고리즘

3. 피플카운터

객체 추적을 이용하여 피플카운터를 구현하였다. 객체 검출 모델로만 피플 카운터를 구현하려면 화면 내에 객체를 모두 포함해야한다. 그래서 밀폐된 공간 내에서 모든 객체가 포함되지 않더라도 충분히 피플 카운터를 사용할 수 있도록, 객체 추적을 이용하였다. 피플 카운터 알고리즘은 다음과 같다.

객체 추적 알고리즘을 이용한 피플 카운터는 그림 3과 같이 동작한다. 처음에 실행되면 열화상 카메라로 보이는 사람의 수를 출력한다. 다음으로 출입구에 해당하는 문을 바운딩 박스로 만든다. 문에 해당하는 바운딩 박스에서 추적하고 있는 바운딩 박스가 사라지면 사람이 나갔다고 판단하여 재실자 수를 1을 빼주고(a) 반대로 문에 해당하는 바운딩 박스에서 추적하고 있는 바운딩 박스가 새로 생기면 재실자 수 1을 더해준다.(b)

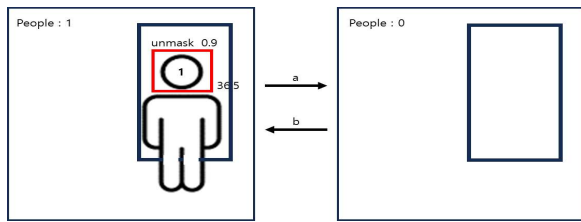


그림 3. 피플카운터

4. 음향 인식 알고리즘

CNN 모델을 이용하여 배경 소리와 구분하여 특정 음향을 인식하기 위한 음향 인식 알고리즘을 구현하였다. 이미지 처리에서 많이 사용되는 CNN 모델은 1차원 음향 데이터를 2차원 MFCC 데이터로 변환하여 입력함으로써 음향인식이 가능하다. 이때, MFCC 이미지 데이터는 음향의 특징을 추출한 데이터이다. CNN 모델에서 특징 추출 부분은 Convolution, BatchNormalization, ReLU, MaxPooling을 3번 반복하였고, 분류 부분에서는 Flatten, Dense, ReLU, Softmax를 사용하였다. 구현한 CNN 모델의 경우 파라미터수가 적어 임베디드 보드 상에서 1초 이내의 음향

분류가 가능하므로, 실시간 처리를 고려해 해당 모델을 채택하였다. 본 논문에서는 음향 인식 알고리즘을 이용해 배경소리, 기침소리를 인식하였다.

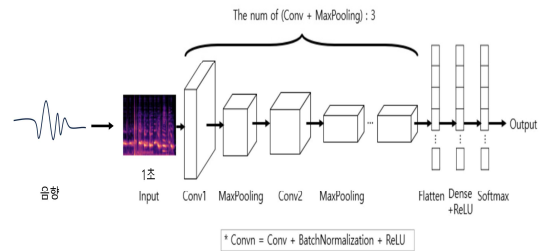


그림 4. 음향 인식 알고리즘

5. 오염위험도

밀폐된 공간에서 검출한 객체의 마스크 착용 여부, 재실자 수, 높은 열 감지 여부, 기침 소리 인식 여부 및 밀폐 공간의 면적을 이용해 오염위험도를 계산하였다. 오염위험도는 밀폐된 공간 내의 전염병 감염 가능성이 있는 사람에 대한 위험도와 감염병 전염 가능성을 높이는 원인에 대해 종합적으로 계산하였다.

밀폐된 공간에서의 감염자의 여부가 해당 공간에서의 전염 가능성 여부를 결정하는 가장 큰 요인이기에 전염병 감염 가능성이 있다고 판단되는 높은 체온 객체 수, 기침 소리 여부에 대해 전염 가능성을 높이는 행동인 밀폐된 공간 대비 높은 재실자 수, 마스크 미착용 객체 수보다 높은 가중치를 부여한다.

9개국 성인 24,410명을 대상으로 연구된 코로나 바이러스(COVID-19) 전염병에 대한 증상은 발열 78% 기침 57%로 나타남을 알 수 있다[27]. 해당 데이터를 통해 발열 증상이 감염자 판별에 높은 신뢰도를 주는 것을 확인할 수 있다. 또한, 기침의 경우 전염병 증상과 감염 증상이 아닌 기침의 구분이 정확하지 않기 때문에 발열에 대해 더 높은 가중치를 주는 것이 합리적이라고 판단하였다.

따라서, 오염 위험도의 수식에서 판단 우선순위에 대한 직관적 표현을 위해 정수의 가중치를 사용하여 밀폐된 공간 대비 높은 재실자 수, 마

스크 미착용 객체 수의 가중치는 1로 기침 소리 여부는 2로 높은 체온의 객체 수의 가중치를 3으로 책정하였다.

$$\text{오염위험도} = \frac{\text{재실자수}}{\text{밀폐된 공간의 넓이}(m^2)} + \text{마스크 미착용 객체수} + 2 \times \text{기침 소리 여부} + 3 \times \text{높은 체온 객체수}$$

(1)

IV. 실험 및 결과

실험은 Raspberry Pi 4B 보드에서 진행하였으며, Ubuntu 20.04 및 Python 3, Tensorflow 2 라이브러리를 이용하였다. CPU는 ARM Cortex-A72 1.5GHz이고 GPU는 Broadcom VideoCore VI이다. 또한 딥러닝 계산에 용이한 GPU 대신 구글 사의 Coral USB Accelerator를 이용하였다.

표 2는 제안하는 시스템과 다른 시스템과의 성능 비교를 보여준다.

표 2. 제안하는 시스템의 성능 비교

Model	Object Tracking (fps)
YOLOv6n+DeepSORT	~1
YOLOv6n+ 제안하는 객체 추적 알고리즘	~8

객체 검출 모델은 YOLOv6n 모델을 사용하였으며, 객체 추적 모델은 제안하는 객체 추적 알고리즘을 사용하였다. 널리 사용되는 객체 추적 알고리즘인 DeepSORT을 사용했을 때는 1fps로 실시간으로 객체 추적을 할 수 없었지만, 제안하는 객체 추적 알고리즘을 사용하였을 때 8fps로 실시간으로 객체를 추적하였다. 그림 5는 제안하는 객체 추적 알고리즘으로 밀폐된 공간에서 피플카운터를 하는 결과이다.

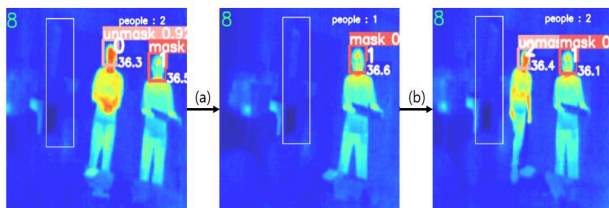


그림 5. 객체 추적 결과

배경음의 경우 다양한 음향이 포함되기 때문에 수집에 어려움이 없었지만, 기침소리의 경우에는 데이터 수집에 어려움이 있었다. 그래서 기침 소리는 음향의 길이나 크기를 조절하면서 기존보다 약 30배만큼 데이터 증강을 하였다. 음향 인식 알고리즘을 학습할 때 데이터 증강 전보다 데이터 증강 후 학습한 경우 F1 score가 0.75에서 0.98로 향상한 걸 확인하였다.

표 3. 음향 데이터 수

	데이터 증강 전 데이터 수	데이터 증강 후 데이터 수
배경음	16,641	16,641
기침 소리	374	9,724

표 4. 음향 인식 알고리즘 성능 비교

	Accuracy	F1score	FPR	FNR
데이터 증강 전	0.87	0.75	0.17	0.1
데이터 증강 후	0.98	0.98	0.02	0.02

밀폐된 공간에서 검출한 객체의 마스크 착용 여부, 재실자 수, 높은 열 감지 여부, 기침 소리 인식 여부 및 밀폐 공간의 면적을 이용해 계산한 오염위험도를 그림 6처럼 이미지에 출력하였다.



그림 6. 오염위험도 출력

V. 결론

본 논문에서는 전염병 확산 방지를 위한 시스템을 연구하였다. 성능이 낮은 임베디드 보드에서 실시간 객체 추적에 가벼운 모델을 사용하였

기 때문에 정확도는 다소 떨어지지만, 8프레임으로 채실자수를 준수하게 계산하는 걸 확인하였다. 출입구에 대해 객체를 추적하여 피플카운터를 구현하였으므로, 제안된 시스템은 출입구가 특정되는 밀폐된 공간에 적합하다. 하지만 밀폐된 공간 안에 있는 모든 객체를 포함하지 않아도 되고 임베디드 기기의 저렴한 가격으로 다양한 환경에 적용 가능하다. 또한 열화상 카메라를 사용하였기 때문에 일반 카메라의 개인 정보 문제도 해결할 수 있다. CNN 기반 음향 인식 알고리즘을 이용해 기침 소리를 감지하였다. 음향 데이터 수집에 어려움이 있어 데이터 증강을 하였고, 0.98의 정확도로 기침 소리를 감지하였다. 밀폐된 공간에서 오염위험도를 계산하였기 때문에 살균 UVC LED와 연동한다면 전염 예방을 위한 공기 정화가 가능할 것이다.

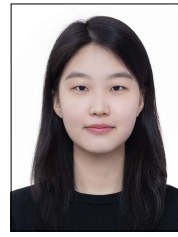
REFERENCES

- [1] Maital, Shlomo, and Ella Barzani, "The global economic impact of COVID-19: A summary of research," *Samuel Neaman Institute for National Policy Research*, pp. 1-12, 2020.
- [2] Yeyati, Eduardo Levy, and Federico Filippini, "Social and economic impact of COVID-19," *Brookings Institution*, 2021.
- [3] Güner, Hatice Rahmet, İmran Hasanoglu, and Firdevs Aktaş, "COVID-19: Prevention and control measures in community," *Turkish Journal of Medical Sciences*, vol. 50, no. 9, pp. 571-577, 2020.
- [4] Zhou, Wang, *The coronavirus prevention handbook: 101 science-based tips that could save your life*, Simon and Schuster, 2020.
- [5] Ulhaq, Anwaar, et al., "Computer vision for COVID-19 control: a survey," arXiv preprint, arXiv:2004.09420, 2020.
- [6] Khandelwal, Prateek, et al., "Using computer vision to enhance safety of workforce in manufacturing in a post covid world," arXiv preprint, arXiv:2005.05287, 2020.
- [7] Szeliski, Richard, *Computer vision: algorithms and applications*, Springer Nature, 2022.
- [8] Suganthalakshmi, R., et al., "COVID-19 facemask detection with deep learning and computer vision," *International Journal of Engineering Research and Technology (IJERT)*, ICRADL, 2021.
- [9] Schiff, Jeremy, et al., "Respectful cameras: Detecting visual markers in real-time to address privacy concerns," in *Protecting Privacy in Video Surveillance*, pp. 65-89, 2009.
- [10] Shrestha, A., and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040 - 53065, 2019, doi:10.1109/ACCESS.2019.2912200.
- [11] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, 2012.
- [12] Szegedy, Christian, et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [13] He, Kaiming, et al., "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [14] Simonyan, Karen, and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint, arXiv:1409.1556, 2014.
- [15] Girshick, R., et al., "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 580 - 587, 2014.
- [16] Girshick, R., "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440 - 1448, 2015.
- [17] Ren, S., et al., "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 91 - 99, 2015.
- [18] He, K., et al., "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2980 - 2988, 2017.
- [19] Yoo, D., et al., "AttentionNet: Aggregating weak directions for accurate object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2659 - 2667, 2015.
- [20] Redmon, J., et al., "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779 - 788, 2016.
- [21] Liu, W., et al., "SSD: Single shot multibox detector," in *Proceedings of the European*

Conference on Computer Vision (ECCV), pp. 21 - 37, 2016.

- [22] Redmon, J., and A. Farhadi, "YOLO9000: Better, faster, stronger," *arXiv*, arXiv:1612.08242, 2016.
- [23] Sermanet, P., et al., "OverFeat: Integrated recognition, localization and detection using convolutional networks," *arXiv*, arXiv:1312.6229, 2013.
- [24] Soleimanitaleb, Z., M. A. Keyvanrad, and A. Jafari, "Object tracking methods: A review," in *Proceedings of the 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, Mashhad, Iran, pp. 282 - 288, 2019.
- [25] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," *arXiv preprint arXiv:1703.07402*, 2017.
- [26] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1 - 2, pp. 83 - 97, 1955.
- [27] Grant, M. C., et al. "The Prevalence of Symptoms in 24,410 Adults Infected by the Novel Coronavirus (SARS-CoV-2; COVID-19): A Systematic Review and Meta-Analysis of 148 Studies from 9 Countries." *PLOS ONE*, vol. 15, no. 6, 2020.

 저자 소개



김강희(준회원)

2025년 홍익대학교 전자전기공학 공
학사

2025년~ 현재 홍익대학교 전자전기공
학전공 석사과정 재학

<주관심분야 : 인공지능, 딥러닝>



김준현(준회원)

2025년 홍익대학교 전자전기공학 공
학사

2025년~ 현재 홍익대학교 전자전기공
학전공 석사과정 재학

<주관심분야 : 인공지능, 딥러닝>



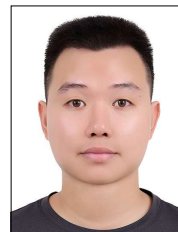
황연우(정회원)

2021년 홍익대학교 전자전기공학 공
학사

2024년 홍익대학교 전자전기공학과
석사

2024년~ 현재 홍익대학교 전자전기공
학과 박사과정 재학

<주관심분야 : 인공지능, 딥러닝>



Nguyen Minh Nguyen(정회원)

2019년 Ton Duc Thang University
학사

2022년 세종대학교 대학원 석사

2025년 홍익대학교 전자전기공학과
공학박사

<주관심분야 : 분산 학습, 연합 학습, 객체 감지>



임종태(정회원)

1989년 서울대학교 전자공학과 공학
사

1991년 서울대학교 전자공학과 공학
석사

2001년 The University of Michigan
at Ann Arbor 공학박사

2004년 9월 ~ 2008년 2월 한국항공대
학교 항공전자 및 정보통신공학부 조교수

2008년 3월 ~ 현재 홍익대학교 전자전기공학부 교수

<주관심분야 : 디지털 통신 및 방송 시스템, 디지털 신
호 처리>