YOLO를 이용한 마스크 착용 상태 검출 연구

정우성[1], 신선우[2], 서상현[3*]
중앙대학교 컴퓨터예술학부[1-2]
중앙대학교 예술공학부[3*]
e-mail : yesung111@naver.com[1], dkqjwlwoo@naver.com[2], sanghyun@cau.ac.kr[3*]

Mask wearing state detection technology research using YOLO

Woosung Chung[1], Sunwoo Shin[2], Sanghyun Seo[3*]
School of Computer Art, Chung-Ang University[1-2]
College of Art and Technology, Chung-Ang University[3*]

요 약

본 논문에서는 딥러닝을 기반으로 한 마스크 착용 상태를 검출하는 기술을 제안한다. 마스크 착용 상태를 착용, 오착용, 미착용으로 나누어서 RGB 카메라를 통해 실시간으로 입력받은 영상 채널에서 마스크 착용 상태를 검출하는 연구이다. 합성곱 신경망(Convolutional Neural Network)을 기반으로 한 딥러닝 모델인 You Only Look Once(YOLO) 모델을 버전 별로 비교하여 최적의 알고리즘을 찾고, 다양한 포즈의 마스크 착용 이미지로 이루어진 학습 데이터를 구축하였다. 본 논문의 내용은 이후 인구가 밀집되는 실내 환경에서 방역 시스템을 구축하는 데에 사용할 수 있다.

1. 서 론

코로나19 사태가 지속되고 있는 가운데, 바이러스의 확산을 예방하는 방역 시스템의 중요성이 증가하고 있다. 특히 마스크를 착용하게 되면 바이러스의 확산을 크게 줄일 수 있기에, 마스크 착용은 중요한 방역 지침 중 하나이다. 공연장 등 인구가 밀집되는 실내 환경에서는 마스크 착용의 중요성이 더 높아진다. 하지만, 관리원이 있다고 하더라도 사람의 눈으로 많은 사람의 마스크 착용 상태를 동시에 확인하기는 어렵다. 이로 인해, 실시간으로 마스크 착용 상태를 확인할 수 있는 모니터링 시스템에 대한 요구가 증가하고 있다.

따라서, 본 논문에서는 딥러닝 기술을 기반으로 하여 마스크 착용 상태를 검출할 수 있는 기술을 제안한다. 마스크 착용 상태를 착용, 오착용, 미착용으로 분류하고, RGB 카메라를 통해 실시간으로 입력받은 영상 채널에서 합성곱 신경망(Convolutional Neural Network, CNN) 기반 알고리즘을 통해 마스크 착용 상태를 검출하는 기술이다. 합성곱 신경망을 기반으로 한 You Only Look Once (YOLO) 모델들을 Mean Average Precision(mAP)과 F1-Score를 통해서 비교하여 가장 높은 정확도를 가진 최적의 알고리즘을 찾았다. 또한 여러 가지 비정상적 마스크 착용 상태를 검출하기 위해 다양한 마스크 포즈 이미지로 이루어진 학습 데이터를 구축했다. 더 나아가 인구 밀집된 지역에서 얻은 이미지 데이터를 수집해서 다중 객체 인식을 목표로 했다.

본 논문에서 진행된 연구는 환기가 안 되는 폐쇄 공간에서 중요한 방역 프로세스를 활용될 것으로 예상한다. 특히, 문화시설에서 이용자들이 안심하고 즐길 수 있는 디지털 방역 시스템을 마련함으로써 이용자의 건강 증진을 활성화할 수 있을 것으로 예상한다.

2. 개 념

합성곱 신경망(Convolutional Neural Network, CNN)[1]은 이미지, 영상 처리에 주로 사용되는 신경망이다. 다차원 배열의 데이터를 처리하기 때문에 컬러 이미지를 처리하는 데에 특화되어 있다. 합성곱 신경망은 합성곱층(Convolution Layer)이라는 층에서 입력 데이터의 특징을 추출하여 학습을 진행한다. 필터(Filter)를 스트라이트(stride)라는 간격을 통해서 훑으면서 입력 데이터에서 특징을 추출하게 된다. 이러한 연산을 컨볼루션(Convolution)이라고 한다. 컨볼루션 연산을 거치게 되면 데이터의 크기가 줄어들게 된다. 이 과정에서 발생하는 데이터의 손실 예방하기 위해서 해당 층을 특정 값으로 채우는 패딩(Padding) 과정을 거친다. 합성곱층에서 출력되는 데이터는 활성화 함수(Activation Funtion)를 거쳐서 출력 신호로 변환된다.

You Only Look Once(YOLO)[2] 모델은 합성곱 신경망 모델 중에 하나로, 이미지, 영상에서 객체를 검출하는 데에 사용된다. YOLO는 바운딩 박스 좌표(Bounding Box Coordinate)와 분류를 따로 진행하던 기존의 객체 검출 알고리즘과는 달리 이 두 작업을 한 번에 처리한다. 따라서 이미지를 분할할 필요없이 이미지를 한 번만 인식시켜도 객체 검출이 가능하고, 객체 검출 속도가 다른 객체 검출 모델들에 비해 빠르다.

mean Average Precision(mAP)은 객체 검출(Object Detection) 분야에서 사용하는 성능평가지표이다. 1개의 객체의 평균 정밀도(Average Precision)를 구하고, 여러 객체 검출 알고리즘에 대하여 평균(mean) 값을 구한 것이다. F1-Score 라는 지표 또한 사용되는데, 정밀도(Precision)과 재현율(Recall)의 조화 평균을 뜻한다. 식으로 나타내면 아래 수식과 같다.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

mAP와 F1-Score 모두 높을수록 객체가 높은 정확도로 검출된 것이다.

Frame Per Second(FPS)는 속도를 평가하는 지표로 1초당 몇 프레임(Frame)의 이미지를 처리할 수 있는지를 뜻한다. FPS가 높을수록 이미지 처리 속도도 빠르다.

3. 본 론

3.1 학습 데이터 구축

본 논문에서는 그림 1과 같이 마스크를 제대로 착용했을 때는 “착용”, 입이나 턱까지만 썼을 때는 “오착용”, 착용하지 않았을 때는 “미착용”으로 마스크 착용 상태를 분류했다.

(그림 1) 본 논문에서 분류한 마스크 착용 상태, 왼쪽 위부터 순서대로 착용, 미착용, 오착용(입, 혹은 턱까지만 착용)으로 분류

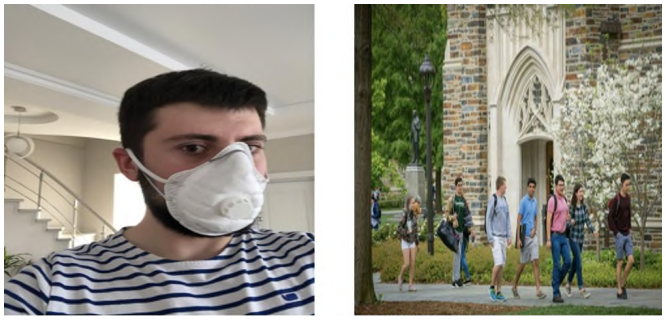
또한, 그림 2와 같이 마스크가 아닌 물체로 코와 입을 가리는 이미지 데이터를 추가하고 미착용이라고 분류함으로써, 단순히 얼굴을 가리는 것이 아닌 마스크의 모양과 위치에 따라서 객체 검출 모델이 마스크 착용 상태를 검출할 수 있도록 하였다.

(그림 2) 마스크가 아닌 물체로 얼굴을 가린 이미지 데이터, 본 논문에서는 미착용으로 분류

수집한 이미지 데이터들은 Yolo Mark라는 어노테이션(Annotation) 툴을 통해서 레이블링(Labeling)하였다. 또한, 이미 학습된 모델을 기반으로 자동으로 레이블링하는 수도 레이블링(Pseudo Labeling)[4]] 을 사용했다. 본 논문에서는 총 1,848개의 이미지를 학습 데이터로 사용했다.

3.2 모델 비교

본 논문에서는 Yolo v3, Yolo v4, Yolo v4 tiny의 세 가지 모델을 비교하였다. Tiny 모델의 경우 Yolo v3, v4보다 단순한 네트워크를 가지고 있어서 연산량이 적다. 위 세 가지 모델을 500개의 평가 데이터(Validation Data)를 통해 비교했다. 평가 데이터는 그림 3과 같이 가까이서 촬영한 사진과 인구가 밀집된 지역에서 촬영한 사진 등 다양한 상황에서 구축한 이미지 데이터를 활용했다.



(그림 3) 본 논문에서 사용한 평가 데이터의 예시

평가 지표로는 Mean Average Precision(mAP), F1-Score, Frame Per Second(FPS)를 사용했다. 사용한 개발 환경은 표1과 같다.

(표 1) 본 논문의 개발 및 실험 환경

운영 체제	Windows 10
그래픽 카드	NVIDIA GeForce 2060 Super
CUDA Version	10.1
CuDNN Version	8.0.5
OpenCV Version	4.5.0

3.3 결과 및 분석

세 개의 모델을 비교한 결과는 표 2와 같다. 여기서 FPS 값은 실시간으로 구동할 때의 값이다.

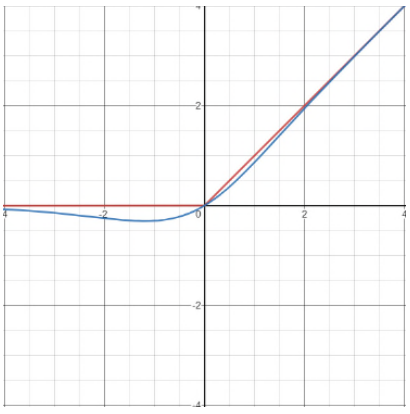
(표 2)Yolo v4, Yolo v3, Yolo v4 tiny 모델의 비교 결과

	Yolo v4	Yolo v3	Yolo v4 tiny
mAP (%)	98.62	83.70	61.64
F1-Score	0.98	0.87	0.75
FPS(Real-Time)	15	15	15

표 2와 같이 Yolo v4, Yolo v3, Yolo v4 tiny의 순서대로 정확도가 높게 나왔다. Yolo v4 tiny는 다른 두 모델들보다 네트워크 구조가 간단하다. 따라서 정확도가 떨어지는 대신 검출 속도가 빠르다. 하지만 실시간으로 표 1의 환경에서 테스트를 진행했을 때는, 세 모델의 FPS 값이 모두 똑같이 나왔기 때문에 Yolo v4 tiny의 간단한 구조가 가지는 장점이 무의미하다.

Yolo v4와 Yolo v3는 둘다 낮지 않은 정확도가 나왔지만 Yolo v4가 더 높게 나왔다. 두 모델의 네트워크를 비교해봤을 때, Yolo v3는 다운샘플링(Downsampling) 과정에서 Leaky ReLU를 활성화 함수로 쓴 반면, Yolo v4는 Mish를 사용했다. Leaky ReLU와 Mish는 아래 수식과 같고, 두 함수를 그래프로 표현하면 그림 3과 같다.

$$LeakyReLU = \max(0.01x, x)$$
$$Mish = x\tanh(\ln(1+e^{x}))$$



(그림 3) Leaky ReLU와 Mish 함수를 표현한 그래프, 빨간색이 Leaky ReLU, 파란색이 Mish

그림 3처럼 Mish 함수는 Leaky ReLU 함수와 달리 음수를 허용하는 방식을 사용하기 때문에 그라디언트가 더 잘 흐르게 된다.[5] Leaky ReLU를 사용하는 Yolo v3의 합성곱층의 아웃풋보다 Yolo v4의 합성곱층의 아웃풋이 더 해상도가 높은 이미지이다. 따라서 최종적인 모델의 정확도가 더 높다.

또한, Yolo v4에서는 데이터 증강을 위해 한 이미지에 여러 개의 클래스를 넣는 방법을 택했다.[6] 따라서 모델이 다양한 상황에서 마스크 착용 상태를 검출할 수 있게 된다.

총 세 개의 모델을 실시간으로 작동시켜 비교했을 때, FPS 값은 모든 모델이 같고, Yolo v4가 가장 높은 정확도를 보인다.

본 논문은 3.1에서 구축한 데이터를 Yolo v4를 통해 훈련한 최종 모델을 제작했다. Yolo v4는 네트워크 안에서 데이터 증강을 하는 방식을 채택했기에, 다양한 환경에서의 데이터 이미지가 필요한 본 논문의 연구에 가장 적합한 모델로 판단할 수 있다.

3. 결 론

본 논문은 대표적인 합성곱 신경망 중에 하나인 Yolo 모델 세 가지를 비교하여 마스크 객체 검출에 가장 적합한 모델을 찾고, 그 이유에 대해 분석했다. 위 결과를 통해 최종적인 알고리즘을 제작했고, 이 알고리즘은 폐쇄 공간에서 방역 시스템을 구축할 만큼 높은 정확도를 가진다. 본 논문의 연구 내용은 실내 환경에서 방역 시스템을 구축하는 데에 사용될 수 있고, 코로나19로 피해를 입은 문화 시설에서 중요한 방역 시스템을 활용할 것으로 예상된다.

다만, 실시간으로 구동했을 때의 속도가 실시간(24FPS 이상)으로 구동하지 않는다는 한계점이 있다. 본 논문의 저자는 이후 특징 맵의 크기를 다양하게 설정하여 탐지 속도가 일반 Yolo v4 모델보다 높은 정확도와 속도가 모두 높은 알고리즘을 제작할 예정이다.

감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2021년 문화기술연구개발 지원사업으로 수행되었음 (과제번호: R2021040028)

참 고 문 헌

[1] Albawi, Saad & Abed Mohammed, Tareq & ALZAWI, Saad. (2017). Understanding of a Convolutional Neural Network. 10.1109/ICEngTechnol.2017.8308186.

[2] Joseph Redom, Santosh Divvala, Ross Cirshick & Ali Farhadi, (2015), You Only Look Once: Unified, Real-Time Object Detection, arXiv:1506.02640

[3] Maxwell, A.E.; Warner,T.A.; Guillén, L.A. Accuracy

Assessment in Convolutional NeuralNetwork-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review. Remote Sens. 2021, 13, 2450. https://doi.org/10.3390/rs13132450
[4] Lee, Dong-Hyun. (2013). Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. ICML 2013 Workshop : Challenges in Representation Learning (WREPL).
[5] Diganta Misra, Mish: A Self Regularized Non-Monotic Nerual Activation Function
[6] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection