JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 22, No. 11, pp. 1931-1939, Nov. 2021

Check for updates

딥러닝 기반 앙상블을 이용한 마스크 착용 상태 검출 기술 연구

신 선 우[1] · 정 우 성[1] · 이 태 민[2] · 서 상 현[3*]
[1]중앙대학교 예술공학대학 컴퓨터예술학부 학사과정
[2]중앙대학교 다빈치SW교육원 특임교수
[3*]중앙대학교 예술공학대학 예술공학부 교수

A Study on Detecting Mask Wearing Status using Ensemble based on Deep Learning

Sunwoo Shin[1] · Woosung Chung[1] · Taemin Lee[2] · Sanghyun Seo[3*]

[1]Bachelor's Course, School Of Computer Art, Chung-Ang University, Anseong-Si, Gyeonggi-Do 17546, Korea
[2]Special Affair Professor, Davinci SoftWare Education Institue, Chung-Ang University, Seoul 06974, Korea
[3*]Professor, College of Art & Technology, Chung-Ang University, Anseong-Si, Gyeonggi-Do 17546, Korea

[요 약]

코로나19 사태로 인하여 바이러스 확산을 방지할 수 있는 방역 시스템에 대한 요구가 증가했다. 본 논문에서는 딥러닝 기반의 앙상블을 통해 마스크 착용 상태를 검출하는 기술을 제안한다. 콘볼루션 신경망을 기반으로 한 딥러닝 알고리즘을 사용하여 마스크 착용 상태를 검출할 수 있는 모델을 제작했다. 마스크 착용 상태를 착용, 오착용, 미착용으로 분류한 뒤, 다양한 자세의 학습 데이터를 구축하여 정확도를 높였다. 또한, 앙상블 기법을 통해 딥러닝 모델을 결합하여 정확도를 향상하였고, 특징 맵의 크기를 다양화하는 방법들 중 FPN 기법을 사용하여 객체 탐지 성능을 높이면서도 연산량을 줄여 탐지 속도를 높였다. 향후 본 연구는 환기가 어려운 실내 환경에서 효율적인 방역 시스템으로 사용될 것이고, 특히 공연장 등 코로나19 사태로 침체된 문화시설을 사용자들이 안심하고 이용할 수 있도록 방역 환경을 구축하는 데에 기여할 것이다.

[Abstract]

Due to the COVID-19 incident, demand for a quarantine system that can prevent the spread of the virus has increased. In this paper, a mask wearing state detection technology based on a deep learning-based ensemble is proposed. A model that can detect mask wearing status was created using a deep learning algorithm based on a convolutional neural network. Afclassifying wearing a mask into mask, improper mask, and no mask, the accuracy was improved by constructing learning data for various poses. In addition, the combination of deep learning models through ensemble techniques increased accuracy, and among the methods of diversifying the size of feature maps, FPN techniques were used to increase object detection performance while reducing computation volume. In the future, this study will be used as an efficient quarantine system in indoor environments where ventilation is difficult, and in particular, it will be contributed to establish a quarantine environment so that users can safely use cultural facilities such as performance halls.

색인어 : 콘볼루션 신경망, YOLO, 마스크 검출, 앙상블, FPN
Keyword : Convolutional Neural Network, YOLO, Mask Detection, Ensemble, FPN

http://dx.doi.org/10.9728/dcs.2021.22.11.1931

Received 15 October 2021; **Revised** 24 November 2021
Accepted 24 November 2021

***Corresponding Author; Sanghyun Seo**
Tel: +※ 개인정보 표시제한
E-mail: sanghyun@cau.ackr

Ⅰ. 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 딥러닝을 기반으로 한 마스크 착용 상태 검출 기술을 제안한다. 착용, 오착용, 미착용의 세 가지 상태를 레이블로 분류하여, RGB 카메라를 통해 입력받은 영상 채널을 딥러닝 기술을 통해 분석하여 실시간으로 마스크 착용 상태를 검출하는 기술이다.

본 연구는 마스크 착용 상태를 높은 정확도로 검출하는 것을 목표로 하였다. 이를 위해 다양한 환경과 포즈의 원본 데이터를 수집하면서, 특수한 상황이나 조건에서도 마스크 착용 상태를 검출할 수 있도록 알고리즘을 개발했다. 특히 본 연구는 다중 객체 인식, 즉, 한 사람에 대해서가 아닌, 여러 사람을 대상으로 마스크 착용 상태를 검출하기 위해서 인구가 밀집되는 지역에서 영상 데이터를 확보하였다. 더 나아가 여러 가지 딥러닝 모델들을 앙상블 기법을 통해 조합해서, 단일 모델을 사용했을 때보다 정확한 예측 결과를 도출해낼 수 있는 방향으로 알고리즘을 제작했다.

또한, 본 연구는 다중 객체 인식, 즉, 한 사람에 대해서가 아닌, 여러 사람을 대상으로 마스크 착용 상태를 검출하기 위해서 인구가 밀집되는 지역에서 영상 데이터를 확보하였다. 실시간으로 마스크 착용 상태를 검출하기 위해서, 특징 맵(Feature Map)의 크기를 다양화하는 방식으로 객체를 감지하는 속도를 높였다.

2장에서는 기존에 진행된 마스크 착용 상태 검출과 객체 탐지 기술 연구를 검토한다. 3장에서는 딥러닝 기반 마스크 착용 상태 검출 기술의 개요를 살펴본다. 4장에서는 본 연구를 통해 제작한 검출 시스템을 기술한다. 5장에서는 검출 시스템의 실험 결과를 기술한다.

Ⅱ. 관련 연구

딥 러닝(Deep Learning)을 기반으로 하여 마스크 착용 여부를 검출하는 기술들은 연구되어 왔다. 일반적으로 객체를 탐지하는 연구를 응용하여 마스크 착용 여부를 검출한다. 객체를 탐지할 수 있는 연구들 중에서 대표적으로 You Only Look Once(YOLO)[1]가 존한다. YOLO는 Region Based Convolutional Neural Networks (R-CNN)[2]와는 달리 이미지를 분할하지 않고, 통합된 네트워크로 구성되어 있다. 또한 물체의 위치를 추정하는 Localization과 객체의 분류 작업인 Classification을 동시에 수행하는 원 스테이지 디텍터(One Stage Detector)로 빠르고 간단하게 객체를 탐지할 수 있다. YOLO는 객체의 위치를 예측한 영역인 Anchor를 한 객체당 하나씩만 설정함으로써, 다른 객체 탐지 알고리즘보다 영역의 중첩 현상인 Overlap으로부터 자유로운 편이다[4].

RetinaNet[3] 또한 YOLO와 같은 원 스테이지 디렉터 알고리즘이다. RetinaNet은 ResNet-FPN을 백본 네트워크로 사용하는데, FPN은 Feature Pyramid Network의 줄임말로 다양한 스케일의 특징 맵을 생성하는 네트워크이다. 일반적인 원 스테이지 디텍터 모델들은 객체와 배경 클래스의 불균형으로 인해 정확도가 낮게 나오는데, RetinaNet은 이를 오분류되는 객체에 대해 더 큰 가중치를 부여하는 Focal Loss 손실 함수를 사용함으로써 이 문제를 해결하였다. YOLO가 크기가 작은 객체를 검출하지 못하는 반면 RetinaNet은 FPN을 활용하여 작은 사이즈의 객체를 잘 검출할 수 있다는 장점이 있다.

본 논문은 실시간 객체 탐지를 목표로 하고 있으므로, 원 스테이지 디텍터 알고리즘인 YOLO를 사용했고, YOLO의 단점을 보완하기 위해서 RetinaNet을 앙상블하여 사용했다.

마스크 착용 여부를 검출하는 연구로 대표적으로 J.Yu[5]의 연구가 존재한다. [5]는 YOLO v4 모델을 이용한 마스크 착용 여부 검출 연구에서 가상의 배경과 기존의 이미지를 합성함으로써 스케일링(Scaling) 과정에서 이미지의 비율을 유지하고, 다른 크기를 가진 3개의 피처 맵을 객체의 크기에 따라서 적용하는 어댑티드 스케일링(Adaptive Scaling)을 사용했다. 이 방법을 사용함으로써 기존의 모델보다 정확도와 속도 측면에서 향상된 성능을 가진 모델을 만들어낼 수 있다. 해당 연구는 특징 맵의 크기를 다양화함으로써 객체 검출의 속도는 높지만, 데이터의 유형이 다양하지 않아서 제한된 상황에만 사용할 수 있다.

또한, S.Sethi[6]는 YOLO v3 모델을 이용하여 마스크 착용 여부 검출 연구를 진행하였다. 해당 연구에서는 다양한 포즈의 마스크 착용 상태를 데이터로 활용함으로써 특수한 경우의 마스크 착용 상태 또한 검출할 수 있다. 다만, 딥러닝 모델을 개선하지는 않아, 객체 검출 속도는 일반적인 YOLO v3 모델과 같다.

본 논문은 [5]가 제안한 어댑티드 스케일링과 같이 다양한 크기의 특징 맵을 사용하여 객체 검출의 성능을 높였고, [6]와 같이 다양한 포즈의 마스크 착용 상태 데이터를 확보함으로써 다양한 상황에서 마스크 착용 상태를 탐지할 수 있도록 했다.

[5]와 [6]은 단일 모델을 활용하여 마스크 상태를 검출하는 연구를 진행했다. 단일 모델은 연산량이 적어서 속도는 빠르지만, 정확도가 떨어지는 단점이 있다. 따라서 본 연구는 앙상블 기법(Ensembles)을 활용하여 모델의 정확도를 높인다. 앙상블 기법이란 기계학습(Machine Learning)에서 두 개 이상의 학습 모델의 예측을 결합하는 기법이다. 앙상블 기법은 일반적으로 부스팅(Boosting), 배깅(Bagging) 등과 같이 학습 알고리즘에 앙상블을 적용하는 기법이 있고, 보팅(Voting)

등과 같이 학습이 완료된 예측 결과에 적용하여 더 나은 결론을 내리는 기법이 있다. 앙상블 기법을 사용함으로써, 단일 모델을 사용했을 때보다 높은 정확도를 확보할 수 있다[7].

A.Casado-Garcia[8]는 CNN 기반의 객체 검출 알고리즘을 앙상블 기법을 통해 결합하는 알고리즘을 제안한다. 해당 연구에서는 단일 객체 탐지 모델들을 보팅을 통해서 결합하였다. 해당 연구에서는 일반적으로는 단일 모델을 사용했을 때보다 앙상블을 사용했을 때 더 높은 정확도로 객체를 검출한다. 하지만 간혹 단일 모델에서는 검출되는 객체가 검출이 안 되는 사례가 있다.

M.Loey[9]는 마스크 착용 상태 검출에 앙상블 학습을 적용했다. ResNet-50(Residual Networks)의 말단 레이어를 제거하고 이를 각각 서포트 벡터 머신(Support Vector Machine), 결정트리(Decision Tree), 그리고 이 둘을 앙상블 기법으로 결합한 결과로 대체한 세 가지의 모델의 성능을 비교하는 연구를 진행하였다. 결과적으로 앙상블한 결과를 말단 레이어로 넣은 모델이 전반적으로 높은 정확도를 보였다. 본 논문에서는 위 연구가 머신러닝 모델들을 간접적으로 앙상블 기법을 결합한 것에서 한 걸음 나아가 두 가지의 딥러닝 모델을 직접 앙상블 기법으로 결합한다.

앙상블 기법은 객체 검출의 속도를 향상할 수 있지만, 여러 개의 모델을 동시에 사용하는 것이기 때문에 일반적으로 단일 모델을 사용할 때보다 속도가 저하된다. 본 논문에서는 J.Yu의 연구와 같이 특징 맵의 크기를 다양화하는 방법을 통해서 학습 및 객체 검출의 속도를 향상하는 방향으로 연구를 진행하였다. 본 논문에서는 기존 연구들에서 마스크 상태 검출에 사용한 CNN 계열의 알고리즘인 YOLO 모델을 사용했다.

또한, 다양한 포즈의 데이터를 확보 및 구축하고, 앙상블 기법을 활용하여 두 가지 모델을 결합하여 알고리즘을 구축하여 기존에 진행되었던 연구들보다 높은 정확도와 범용성을 확보하는 쪽으로 연구 방향성을 설정했다. 더 나아가 다양한 크기의 특징 맵을 활용하는 방향으로 기존 객체 검출 모델들을 개선하여 더 높은 속도를 확보했다.

Ⅲ. 검출 시스템 제작

3-1 시스템 개요

[그림 1]은 본 논문에서 마스크 검출 시스템을 제작하는 과정을 전체적으로 보여준다. 필요한 이미지 데이터를 수집하여, 단일 딥러닝 모델들을 학습시키고, 모델들을 앙상블하여 성능을 높였다. 또한, 특징 맵의 크기를 조절할 수 있는 모델들을 사용하여 객체 검출 성능을 높여서 높은 정확도의 마스크 착용 여부 검출 시스템을 제작하였다.

3-2 학습 데이터 구축

본 논문에서는 마스크 착용 상태를 세 가지로 나누어 레이블링하여 학습 데이터를 구축하였다. [그림 2(a)]와 같이 코밑까지 마스크를 쓰는 “코스크”나 턱 밑까지만 쓰는 “턱스크”인 오착용(Improper Mask) 상태, [그림 2(b)]와 같이 제대로 착용한 마스크 착용(Mask) 상태, [그림 2(c)]와 같이 착용하지 않은 미착용(No Mask) 상태이다.

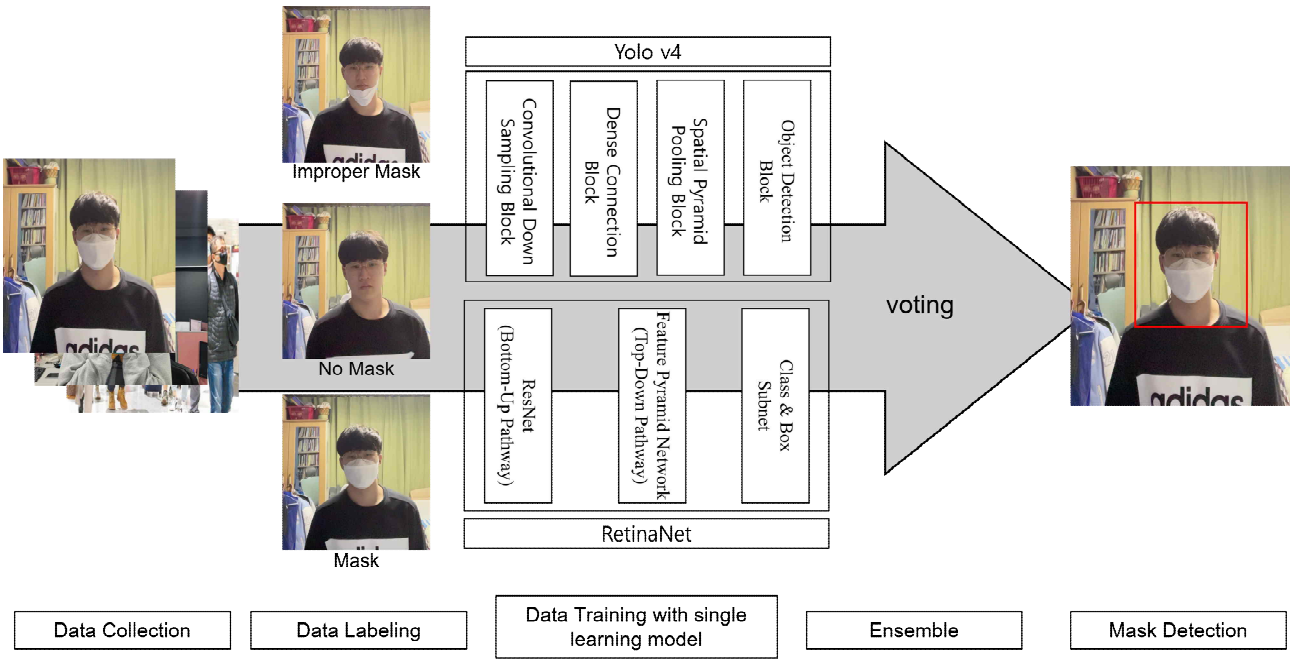

그림 1. 검출 시스템 개요도
Fig. 1. A schematic diagram of the detection system

본 데이터는 정면뿐만 아니라 정측면, 측면, 후측면 등 여러 각도에서의 이미지를 포함하고 있어서, 다양한 상황에서 마스크 착용 상태를 검출할 수 있다.

또한, [그림 3(a)]와 같이 마스크가 아닌 물체가 코와 입을 가리고 있는 데이터를 추가하고 이를 미착용으로 분류하여, 딥러닝 모델이 마스크로 코와 입을 가리고 있을 때만 착용으로 인식할 수 있도록 했다. 더 나아가 본 연구가 실내 환경에서 적용될 수 있도록, [그림 3(b)]와 같이 인구 밀집 지역에서 다양한 각도와 해상도로 촬영한 오픈 소스 데이터를 추가했다. 오픈 소스 데이터는 Kaggle의 Face Mask Detection 데이터셋[10]을 사용했다.

(a) Improper Mask

(b) Mask (c) No Mask

그림 2. 마스크 착용 상태별 레이블 이미지 : (a) 오착용 (턱스크, 코스크), (b) 착용, (c) 미착용

Fig. 2. Label images for each mask wearing state, (a) Improper Mask, (b) Mask, and (c) No Mask

(a) (b)

그림 3. 다양한 포즈의 데이터 : (a) 마스크가 아닌 물체로 코, 입을 가린 이미지 (b) 인구 밀집 지역에서의 이미지

Fig. 3. Image data of various poses (a) images covering the nose and mouth with objects other than masks (b) images in densely populated areas

구축된 학습 데이터는 [그림 2]나 [그림 3(a)]처럼 단일객체만 있는 이미지와 [그림 3(b)]처럼 다중 객체가 포함된 이미지가 합쳐져 있다. 본 데이터는 [10]의 오픈 소스 이미지 1,355장과 성인 남녀 48명에게서 수집한 이미지 데이터 31,789장 중에 3,645장이 포함되어 있다. 이미지 데이터를 더 늘려도 모델의 정확도에 큰 차이가 없었기에 5,000장을 최종 데이터셋으로 구축했다.

3-3 단일 모델을 통한 데이터 학습

구축한 데이터를 단일 객체 탐지 모델을 통해 학습하였다. 학습된 모델은 얼굴의 특징점을 추출하여 마스크 착용 상태를 검출한다. 빠른 속도로 객체를 검출할 수 있는 YOLO 알고리즘을 사용했고, 그 중에서 YOLO v4[11], YOLO v3[12], YOLO v4 tiny를 비교했다. YOLO 알고리즘을 학습/구동하는 데에는 Darknet[13] 패키지를 사용했다.

YOLO v3는 darknet53 네트워크를 기반으로 한 YOLO 모델로 53개의 Convolutional Layer를 가지는 모델이다. YOLO v4는 YOLO v3의 후속 모델로 YOLO v3를 기반으로 한 모델이다. YOLO v3가 다운샘플링 과정에서 Leaky ReLU를 활성화 함수로 사용하는 것과 달리, YOLO v4는 Mish를 사용하여 음수에 대해 약간의 허용이 가능하다. 따라서 이미지 데이터의 픽셀 변화도가 더 완만해진다. YOLO v4 tiny 모델의 경우 YOLO v4 모델의 특징을 가지고 있지만 네트워크가 간략화되어 있어 정확도는 조금 낮지만, 빠른 속도로 객체 탐지를 할 수 있다.

위 세 모델을 [10]의 데이터셋에서 100장을 선별 후, 사용하여 테스트를 진행했다. 테스트 환경으로는 “AMD Ryzen 7 5800x 8-core Processor” 프로세서와 “NVIDIA GeForce RTX 3070 TI” GPU가 사용됐다. 각 모델의 평가를 위해서 F1-Score, mAP, IOU를 사용했다. mAP는 재현율(Recall) 값들에 대응하는 정밀도(Precision) 값들의 평균을 의미하는 AP를 한 개의 객체별로 구하고, 여러 개의 객체 탐지기에 대해서 평균을 구한 것을 뜻한다. F1-Score는 재현율과 정밀도의 조화평균을 뜻하고 수식 (1)로 표현할 수 있다. IOU는 실제 B-Box와 모델이 예측한 B-Box의 면적에 대해 연산한 값이다. 세 가지 평가 지표 모두 높을수록 정확하게 탐지했음을 뜻한다.

$$F1-Score = \frac{2*\mathrm{P}recision*Recall}{\mathrm{P}rrecision+Recall} \quad (1)$$

세 가지 모델의 평가 결과는 [표 1]과 같다. YOLO v4, YOLO v3, YOLO v4 Tiny 순으로 정확도가 높다. 또한, YOLO v4 tiny의 빠른 검출 속도로 실시간 검출을 구현하려고 했으나, Darknet 패키지에서 웹캠을 통해 실시간으로 YOLO 알고리즘을 구동했을 때 Frame Per Second(FPS) 값이 약 15로 고정되기 때문에, 15프레임 이상의 속도가 의미가 없다. 따라서 정확도가 떨어지는 tiny를 사용할 필요가

없고, 세 모델이 모두 15프레임은 유지할 수 있기 때문에, 정확도만을 비교하여 가장 높은 YOLO v4 알고리즘을 마스크 착용 여부 검출에 사용했다.

표 1. 3가지 YOLO 단일 모델 정확도 비교
Table 1. Compare the accuracy of 3 YOLO single models (YOLO v4, YOLO v3, YOLO v4 Tiny)

Model	mAP (@0.50)	F1-score	IOU
YOLO v4	92.71	0.92	74.91
YOLO v3	85.43	0.90	73.19
YOLO v4 Tiny	70.96	0.82	72.98

3-4 앙상블 기법

YOLO 알고리즘은 높은 성능으로 객체를 검출할 수 있다. 하지만 속도가 빠른만큼 작은 객체를 잘 탐지하지 못한다는 단점이 있다. 본 논문에서는 앙상블 기법을 사용해 YOLO와 작은 객체를 검출할 수 있는 RetinaNet을 결합하여 YOLO의 단점을 보완하여 더 나은 성능을 가진 모델을 개발했다.

[8]과 [15]의 연구에서 제안한 객체 탐지 알고리즘을 앙상블하는 방법은 적용 단계에 따라 [그림 4]와 같이 크게 세가지로 나뉜다. 첫 째, [그림 4(a)]와 같이 Localization과 Classification 이후 앙상블하는 방식이다. 이는 대게 서브 모델로 예측한 결과를 바탕으로 메인 모델에서 앙상블을 진행한다.

그림 4. C.Garcia[8]가 제안한 세 가지 객체 탐지 알고리즘의 앙상블 기법 (각기 다른 객체를 다른 색깔의 B-Box로 표현했다.)

Fig. 4. The ensemble techniques of the three object detection algorithms proposed by C. Garcia[8] (Different objects are represented in different colors of B-Box.)

단, 해당 방식은 사용할 수 있는 네트워크의 조합이 한정적이다. 이와 같은 방법으로 우리의 연구에서 사용하는 YOLO와 RetinaNet을 결합할 수 없기 때문에 사용할 수 없다.

둘 째, Localization 이후 Classification stage 이전에 적용을 하는 [그림 4(b)]와 같은 방법이다. 그러나 이는 인식한 객체의 클래스 수나 한 객체에 대해 인식 가능한 모델의 수를 고려하지 않는다는 단점이 있다. 이러한 경우 어떤 모델이 어떤 객체를 분류 및 예측하는지가 명확하게 드러나지 않기 때문에 false positive의 값이 증가하는 경향을 보이므로 상태 검출 시 적합하지 않아 사용할 수 없다.

마지막으로 Localization 단계 이전에 특징들을 앙상블하는 방식이다. 이 방식은 다수의 모델을 쉽게 앙상블 할 수 있다. 또한 test-time augmentation과 같이 앙상블의 기반이 되는 다양한 추가 기술을 구현할 수 있게 한다. 본 논문에서는 Darknet 기반 YOLO v4 모델과 RetinaNet에 대해 이러한 방식의 앙상블을 적용하였다. 이는 [그림 4(c)]와 같이 하나의 학습 데이터에 대해 두 가지 방식으로 데이터를 증강한다. 이렇게 생성된 세 가지 데이터에 대해 각각 Localization을 수행하여 객체를 인식한다. 그 후 각 데이터에 적용된 증강 방식을 반대로 적용해 원본 데이터와 같은 위치에 Anchor Box가 위치하게 한다. 이렇게 생긴 각 객체에 대한 Anchor Box를 보팅하는 방식으로 앙상블하여 알고리즘의 정확도를 높였다.

본 논문에서는 탐지 속도와 정확도에서 높은 성능 보이는 YOLO와 YOLO의 단점인 작은 사이즈의 객체를 탐지하는 데에 특화되어 있는 RetinaNet을 앙상블 기법, 그 중에서도 Localization 이전에 앙상블하는 방식으로 결합함으로써, 다양한 사이즈의 객체를 높은 정확도로 측정할 수 있는 모델을 만들었다.

3-5 특징 맵 크기 다양화

일반적인 모델들은 특징 맵의 크기가 정해져 있기에, 이미지 해상도의 영향을 크게 받는 단점이 있다. 따라서 특징 맵의 크기가 다양할수록 다양한 해상도의 이미지의 특징점을 정확하게 추출할 수 있다. 따라서 본 논문에서는 Feature Pyramid Network (FPN)[16]를 기반으로 한 YOLO v3 이상의 모델들과 RetinaNet을 사용하였다.

FPN은 [그림 5]의 두 가지 과정을 통해서 다양 크기의 특징맵에서 특징점을 추출한다. 먼저 Bottom-up Pathway로, 가장 하위에 있는 레이어 그림부터 올라가며 스케일링을 2배로 하면서 단계별로 피처 맵을 생성하는 방식이다. 이후 가장 상위에 있는 레이어부터 반대로 내려가면서 피처들을 결합하는 Top-Down Pathway를 진행하게 된다. 이 방법을 통해 각 레이어별로 탐지한 결과를 다른 레이어의 탐지 결과의 레퍼런스로 사용하면서 객체를 탐지하게 된다.

FPN은 한 네트워크에서 모든 피처 맵의 정보를 가지고 있지만, 이미지 하나로 계산이 되고 한 번의 프로세서로 해당 연산이 가능하기 때문에, 연산량과 속도가 줄어들면서도 다양한 해상도의 이미지로부터 객체를 탐지할 수 있다.

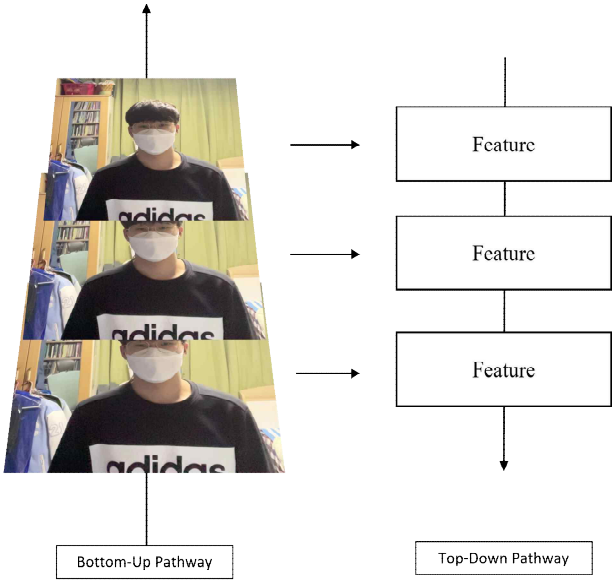

그림 5. Feature Pyramid Network의 원리
Fig. 5. Principle of Feature Pyramid Network

[그림 6]은 FPN을 적용하지 않은 YOLO v2 모델과 RetinaNet을 동일한 이미지로 테스트한 결과다. 인풋 데이터는 416 x 416의 저해상도 이미지다. [그림 6(a)]는 YOLO v2, [그림 6(b)]는 RetinaNet의 결과이다. FPN을 지원하지 않는 YOLO v2가 탐지할 수 없는 객체를 RetinaNet은 탐지한다. 본 논문은 FPN을 기반으로 한 YOLO v3 이상 모델과 RetinaNet을 결합하여 탐지 성능을 높이면서도 연산량을 줄여서 탐지 속도를 높였다.

(a) (b)

그림 6. YOLO v2와 RetinaNet의 저해상도 이미지에 대한 탐지 결과 (빨간색 박스는 마스크를 착용한 객체를 의미한다.)
Fig. 6. Detection results for low-resolution images of YOLO v2 and RetinaNet (The red box means an object wearing a mask.)

Ⅳ. 실험 결과

본 논문은 YOLO 모델중 가장 성능이 높게 나오는 YOLO v4와 RetinaNet을 앙상블 기법으로 결합했다. Localization 과 Classification 이전에 보팅으로 두 모델을 결합하는 방식을 선택했다.

3-3에서 사용한 데이터셋으로 테스트를 진행했고, 실험 환경 역시 같은 환경에서 진행하였다. 이를 통해 나온 성능 결과는 [표 2]와 같다. [표 1]의 YOLO 단일 모델들의 결과들보다 검출 성능이 향상된 것을 확인할 수 있다. 본 논문에서 개발한 최종 모델의 경우 단일 YOLO 모델에서 탐지하지 못한 작은 해상도를 가진 객체를 RetinaNet이 탐지하면서 YOLO 의 단점을 보완하게 된다. [그림 7]은 YOLO 단일 모델들과 본 논문에서 개발한 모델을 동일한 입력 이미지로 테스트한 결과다.

표 2. 앙상블 모델 테스트 결과
Table. 2. Ensemble model test results

Model	Accuracy	Precision	Recall	F1-score
YOLO v4 + RetinaNet	86.68	91.12	94.68	92.86

본 논문에서 개발한 모델은 YOLO 단일 모델보다 작은 사이즈의 객체 검출이나 오검출에서 강점을 보인다. 우선, [그림 7(c)]에서 볼 수 있듯, YOLO 모델들이 검출하지 못한 작은 사이즈의 객체를 검출할 수 있는 것을 확인할 수 있다. 또한, [그림 7(a)]의 YOLO v4나 [그림 7(d)]의 YOLO v3, v4와 같이 얼굴이 아닌 객체를 인식하여 오검출이 발생하는 경우도 단일 모델보다 적게 발생하여 정확도가 단일 모델들보다 높다.

본 논문에서 구축한 데이터는 다양한 각도에서의 이미지를 담고 있어 [그림 7(c)]와 같이 측면에서 촬영한 이미지에서도 마스크 착용 상태를 검출할 수 있다. [그림 3(a)]와 같이 마스크가 아닌 물체로 얼굴을 가린 이미지 데이터를 위와 같은 데이터도 추가함으로써, [그림 7(c)]의 Our Models에서처럼, 마스크를 쓰고 다른 물체가 얼굴을 가린 경우도 착용 상태로 인식할 수 있다. 인구가 밀집될 수 있는 지역에서는 [그림 7(b)]와 같이 얼굴이 다 촬영되는 경우보다는 [그림 7(c)]의 경우처럼 얼굴이 가려져 촬영이 되는 경우가 많은데, 본 논문의 모델은 공연장같이 인구가 밀집되는 지역에서도 효율적으로 마스크 착용 상태를 검출할 수 있다.

[그림 8]은 다양한 마스크 착용상태를 검출한 결과를 보여준다. 빨간 바운딩 박스의 경우 제대로 착용된 얼굴을 탐지해주는 것이고, 초록색 바운딩 박스는 제대로 착용되지 않은 결과를 보여준다. 그림에서 확인할 수 있듯이, 적은 수의 사람, 얼굴의 옆모습이 찍힌 사람, 다 수의 사람등과 같이 다양한 조건에서 모두 마스크 착용이 정상적으로 되어있는지 아닌지를 정확하게 판별할 수 있다.

[그림 9]는 우리의 결과중에 대표적인 실패사례를 보여준다. 딥러닝 기반의 마스크 검출 특성상 마스크를 착용하였는데, 착용하지 않았다고 측정되거나, 마스크를 착용하지 않았는데, 착용되었다고 측정되는 실패사례들은 단순 학습의 문제이기 때문에, 학습데이터의 양에 따라 달라지게 된다.

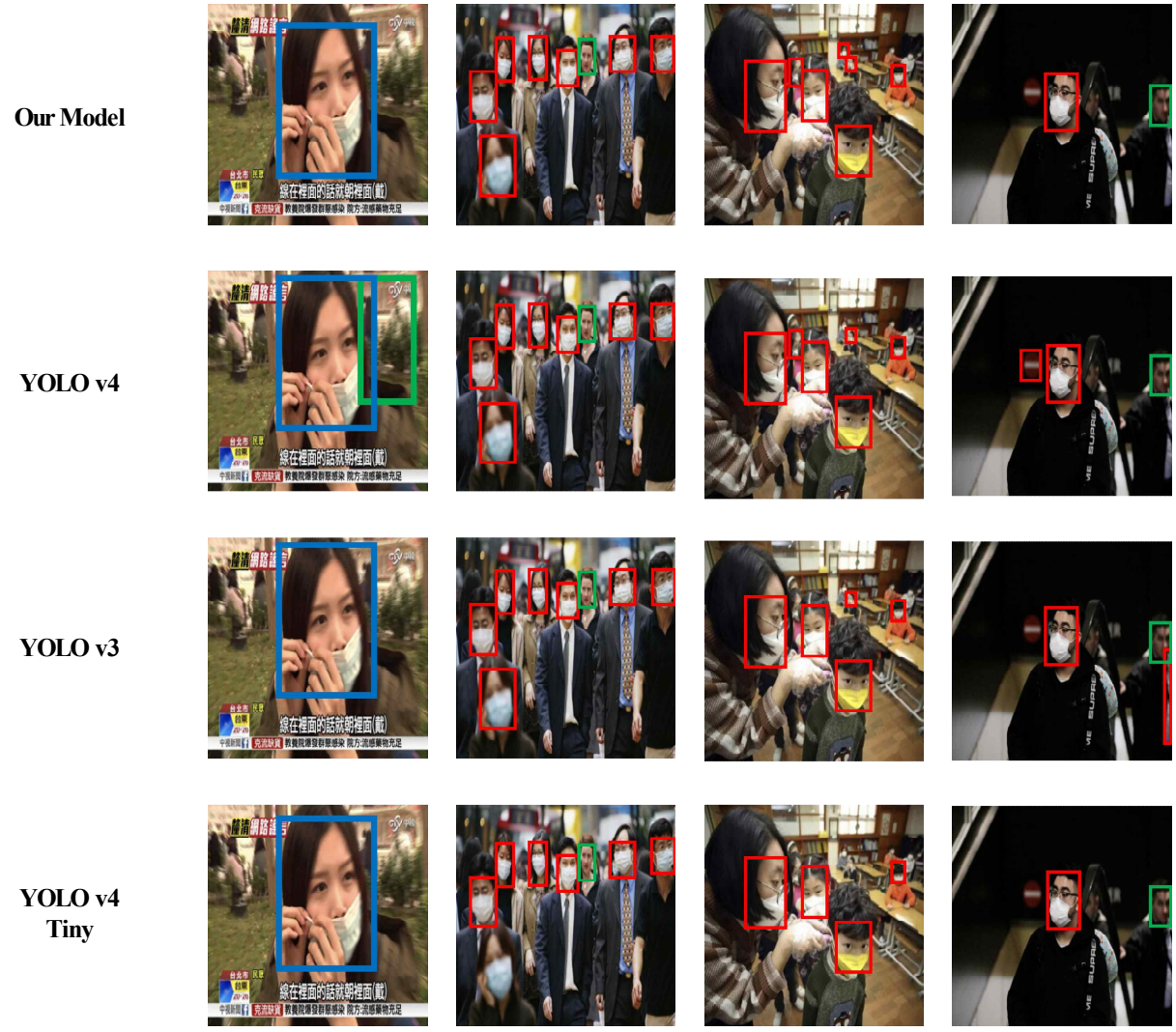

그림 7. 최종 모델과 YOLO 단일 모델들의 마스크 착용 상태 검출 결과
Fig. 7. Final model and YOLO single model's detection of wearing masks.

[그림 9]에서 언급하고 있는 실패 사례들은 단순한 학습데이터의 양에서 오는 문제가 아니라, 학습데이터를 수집하는데 있어서 다양한 정보를 수집해야할 필요성을 보여준다. [그림 9(a)]는 이미지의 여성 객체가 마스크를 쓰고 있다는 것과 쓰지 않고 있다는 바운딩 박스가 두 개 나오게 되는데, 이는 여성 객체의 코와 입을 가리고 있는 마스크의 형태를 잡는 것과, 학습된 마스크의 색상이 분홍색이 없기 때문에 마스크라고 인식을 하지 못하는 부분에서 충돌이 일어나서 생기는 결과이다. 비슷한 이유로 [그림 9(b)]의 경우 이미지의 왼쪽 하단부에 나오는 하얀색 물체가 마스크로 인식되면서 마치 사람이 마스크를 쓰고 있는 것처럼 예측되는 문제점이 생기게 된다. 즉, 데이터를 단순히 많이 모으는 것보다 다양한 케이스를 더 고려하여 수집해야할 필요성을 보인다.

Ⅴ. 결론

본 논문에서는 다양한 포즈의 마스크 착용 상태를 담은 이미지들을 데이터로 사용하면서, 다양한 각도에서 마스크 착용 상태를 검출할 수 있도록 했다. 또한 모델은 YOLO v4와 RetinaNet을 앙상블 기법을 통해 결합한 모델로, 작은 객체를 검출하는 데에 한계를 지닌 YOLO의 단점을 RetinaNet의 FPN 스케일링 기법으로 보완한 모델이다. 따라서, 단일 YOLO 모델보다 저해상도의 인풋 데이터나 작은 크기의 객체를 검출하는 데에 강점을 보인다. 또한, 두 모델을 보팅의 방식으로 결합함으로써, 오검출이 단일 모델에 비해 적게 발생한다. 두 가지 모델을 결합한 모델이고 피처맵의 크기를 다양화했기 때문에 연산량이 증가하여 검출 속도가 느려질 수 있지만, FPN 기법을 사용하여 연산량을 줄여서 빠른 검출 속도

를 확보하였다.

본 연구는 실내 환경 및 인구 밀집 지역에서의 방역 시스템을 구축하는 데에 적극적으로 활용될 것으로 예상된다. 실시간으로 마스크 착용 상태를 검출하면서 환기가 어려운 폐쇄 공간에서 중요한 방역 프로세스가 될 것이다. 특히, 공연장 등의 문화시설을 사용자들이 안심하고 사용할 수 있도록 방역 환경을 구축할 수 있을 것이다. 다만, 본 논문에서 구축한 데이터는 공연장과 같이 저조도 환경에서의 데이터는 부족하고, 광량에 따라 모델의 성능이 어떻게 변하는지에 대한 연구가 필요하다. 이후 저조도 환경에서 사용할 수 있는 조도 센서를 활용하여 광량에 따라서 RGB 카메라 혹은 적외선 카메라를 선택적으로 사용하여 마스크 착용 상태를 검출하는 연구를 진행할 것이다. 또한, 카메라와 객체와의 거리에 따른 모델의 검출 성능 변화를 확인하고 더 다양한 이미지 데이터를 구축하고 모델을 개발할 수 있다.

그림 8. 다양한 디텍션 결과
Fig. 8. Various detection results

(a) (b)

그림 9. 마스크 디텍션 실패 예시
Fig. 9. Example of mask detection failure

감사의 글

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

참고문헌

[1] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, pp. 779-788, June 2016. https://doi.org/10.1109/CVPR.2016.91

[2] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Ohio, pp. 580-587, 2014. https://doi.org/10.1109/CVPR.2014.81

[3] T. Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318-327, 1 Feb. 2020. https://doi.org/10.1109/TPAMI.2018.2858826

[4] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, and V. Pattabiraman, "Comparative analysis of deep learning image detection algorithms," *J Big Data* 8, 66 , pp. 1-27, 2021. https://doi.org/10.1186/s40537-021-00434-w

[5] J. Yu, and W. Zhang, "Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4," *Sensors* 21, no. 9: 3263, May 8, 2021. https://doi.org/10.3390/s21093263

[6] S. Sethi, M. Kathuria, and T. Kaushik, "Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread," *Journal of Biomedical Infor¬matics*, Volume 120, pp. 1-12, Aug 2021, 103848, ISSN 1532-0464, 2021. https://doi.org/10.1016/j.jbi.2021.103 848.

[7] R. Maclin, and D. Opitz, "Popular Ensemble Methods: An Empirical Study," *Journal of Artificial Intelligence Research*, 11, pp 169-198, Aug 1999. https://doi.org/10.1613/ jair.614

[8] Á. C. García, and J. Heras, "Ensemble Meth¬ods for Object Detection." *Frontiers in Artificial Intelligence and Applications*, pp. 2688 – 2695, 2019. https://doi.org/10.3233/FAIA200407

[9] M. Loey, G. Manogaran, M. Hamed, N. Taha, N. Eldeen, and M. Khalifa," A hybrid deep transfer learning model with machine learning methods for face mask detection in the era

of the COVID-19 pandemic". *The journal of the International Measurement Confederation*, Vol. 167, 108288. Jan 2021. https://doi.org/10.1016/j.measurement.202 0.108288

[10] Face Mask Detection, Available: https://www.kaggle.com/andrewmvd/face-mask-detection

[11] A. Bochkovskiy, C. Wang, and H.M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint*, pp. 1-17, 2020, arXiv:2004.10934

[12] J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint*, pp. 1-6, 2018, arXiv:1804.02767

[13] AlexeyAB. "darknet". Available: https://github.com/AlexeyAB/darknet

[14] fizyr. "keras-retinanet". Available: https://github.com/fizyr/keras-retinanet

[15] J. Li, J. Qian, and Y. Zheng, "Ensemble R-FCN for Object Detection," in Proceedings of the *International Conference on Computer Science and its Applications*, Taiwan, Vol. 474 of CSA'17, pp. 400–406, Dec 2017. https://doi.org/10.1007/978-981-10-7605-3_66

[16] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Hawaii, pp. 936-944, July 2017. https://doi.org/10.1109/CVPR.2017.106

신선우(Sunwoo Shin)

2019년 : 중앙대학교 예술공학대학 컴퓨터예술학부

2019년～현 재: 중앙대학교 예술공학대학 컴퓨터예술학부 학부생

※관심분야 : 딥러닝(Deep Learning), 머신러닝(Machine Learning), 엔터테인먼트 테크놀로지 (Entertainment Technology) 등

정우성(Woosung Chung)

2019년 : 중앙대학교 예술공학대학 컴퓨터예술학부

2019년～현 재: 중앙대학교 예술공학대학 컴퓨터예술학부 학부생

※관심분야 : 딥러닝(Deep Learning), 머신러닝(Machine Learnin), 예술공학(Art & Technology)

이태민(Taemin Lee)

2011년 : 중앙대학교 컴퓨터공학과 (공학사)

2013년 : 중앙대학교 일반대학원 (공학석사-컴퓨터그래픽스)

2019년 : 중앙대학교 일반대학원 (공학박사-컴퓨터그래픽스)

2019년 : 중앙대학교 다빈치SW교육원 특임교수

※관심분야 : 비사실적 렌더링(Non-Photorealistic Rendering), 색상 이론(Color Theory), 감성 컴퓨팅(Emotional Computing), 인공지능(Artificial Intelligence)

서상현(Sanghyun Seo)

1998년 : 중앙대학교 컴퓨터공학과 (공학사)

2000년 : 중앙대학교 첨단영상대학원 영상공학과(공학석사-컴퓨터그래픽스)

2010년 : 중앙대학교 첨단영상대학원 영상공학과(공학박사-컴퓨터그래픽스및가상환경)

2011년～2013년: 프랑스 리옹 1대학, LIRIS 연구소, Post-Doc

2013년～2016년: 한국전자통신연구원, 선임연구원

2016년～2019년: 성결대학교 미디어소프트웨어학부 조교수

2019년～현 재: 중앙대학교 예술공학대학 컴퓨터예술학부 부교수

※관심분야 : 컴퓨터그래픽스(Computer Graphics), 비사실적 렌더링(Non-Photorealistic Rendering), 가상/증강현실(Virtucal Reality/Augmented Reality), 게임 기술(Game Technology)