

# 딥러닝 기반 마스크 착용 모니터링 시스템

김규태\*, 박성재\*, 서동현\*, 임채현\*, 오병우\*\*

## Deep Learning-Based Monitoring System for Mask Wearing

Kyu-Tae Kim\*, Sung-Jae Park\*, Dong-Hyun Seo\*, Chae-Hyeon Im\*, and Byoung-Woo Oh\*\*

### 요 약

세계적인 팬데믹 사태인 코로나바이러스-19(COVID-19)로 인해 일상생활에서 마스크 착용은 필수가 되었다. 그러나, 많은 사람이 모여 있는 실내에서 모든 사람의 마스크 착용 여부를 지속적으로 파악하기는 어렵다. 본 논문에서는 실내에서 사람을 대신하여 마스크 착용 여부를 효율적으로 모니터링하는 방법을 제안한다. 제안 시스템은 마스크 착용 식별을 위해 객체 검출 딥러닝 모델 중 하나인 YOLOv5를 사용한다. 모델 학습은 마스크를 정확히 착용한 이미지, 착용하지 않은 이미지, 정확하지 않게 착용한 이미지를 데이터셋으로 사용하여 학습하고, 웹캠을 사용한 실시간 영상을 통해 마스크 착용여부를 식별한다. 사용모델의 가중치는 실험을 통해 실시간 영상에 적합한 가중치를 채택한다.

### Abstract

Because of the global pandemic, Coronavirus-19 (COVID-19), wearing a mask has become a necessity in everyday life. However, it is difficult to constantly determine whether everyone wears a mask indoors where many people are gathered. In this paper, we propose a method to efficiently monitor whether a mask is worn on behalf of a person indoors. The proposed system uses YOLOv5, one of the deep learning models of object detection, for mask wear identification. Model learning is learned using images wearing masks correctly, images not wearing them, and images wearing them incorrectly as datasets, and real-time images using webcams to identify whether to wear masks. Weights of the use model adopt appropriate weights for real-time images through experiments.

### Key words

deep learning, object detection, YOLOv5, mask detection, monitoring

### 1. 서 론

세계적으로 유행중인 코로나바이러스(COVID-19)로 인해 생활 속 거리두기와 마스크 착용 등의 예

방법이 중요시되고 있다. 사람이 많이 모인 실내의 경우 마스크 착용의 중요성이 더욱 강조되지만 여전히 마스크를 착용하지 않는 사람들도 존재한다. 마스크 착용을 강제하기 위하여 대부분의 건물 실

\* 금오공과대학교 컴퓨터공학과 학부생

\*\* 금오공과대학교 컴퓨터공학과 교수 (교신저자)

내에서 마스크 미착용자 식별이 필요하고 이를 위하여 인력 및 비용이 소모되고 있는 실정이다.

실내에 많은 인원이 모여 있는 경우 육안으로 마스크 착용 여부를 지속적으로 파악하기에는 인력의 소모가 불가피하므로 인력 소모를 줄일 수 있는 방법이 필요하다.

인력 소모를 줄이기 위해 이미지나 영상 내의 사물을 파악하는 기술로 딥러닝 분야의 객체 검출[1]을 이용할 수 있다. 객체 검출은 물체의 종류와 위치를 검출하는 기술로 어떤 물체인지 분류하는 classification 문제와 물체의 위치 정보를 나타내는 localization 문제를 다루는 분야이다. 이를 통해 마스크 착용 식별을 자동화하여 인력의 소모를 줄일 수 있다. 이를 활용하는 마스크 착용 식별 시스템이 존재하지만, 대부분 시스템이 검출 대상을 한 명으로 삼고 있으므로 다수의 사람을 대상으로 하는 시스템이 필요하다.

본 논문에서는 객체 검출 기법 중 YOLOv5[2] 모델을 사용하여 카메라를 통해 실내에서 다수의 사람을 대상으로 마스크 착용 여부를 식별함으로써 마스크 착용, 미착용 여부를 확인할 수 있는 마스크 착용 모니터링 시스템을 제안한다.

## II. 관련 연구

최근 딥러닝을 활용한 컴퓨터 비전 분야의 발전으로 객체 검출 알고리즘 모델의 성능 역시 발전되고 있고 다양한 분야에서 활용되고 있다. [3]에서는 SAR-Ship 데이터셋에서 배를 검출하기 위해 모델들의 성능을 비교하였고 결과는 표 1과 같다.

표 1. SAR-Ship 데이터셋 결과 비교

Table 1. Comparison results for the SAR-Ship dataset

구분	FPS	AP <sub>50</sub>
Mask-RCNN	8.85	86.56
YOLOv3	19.23	86.58
YOLOv4-tiny	76.92	79.83
YOLOv4	22.22	87.36
EfficientDet-d0	13.16	86.37
YOLOv5	25.64	88.16

모든 경우에 일반화시킬 수는 없지만 특정 데이터셋에 대한 객체 검출 모델 중 준수한 FPS(Frame Per Second)와 가장 높은 AP<sub>50</sub>(Average Precision)를 나타내는 YOLOv5 모델이 FPS와 AP<sub>50</sub>부분에서 실시간 영상에 대한 처리에 가장 적합하다고 판단하여 본 논문에서 제안하는 시스템의 학습에 사용한다.

YOLOv5의 전체적인 모델 구조는 그림 1과 같다. YOLOv5는 일반적인 YOLO Network와 같이 크게 Backbone, Neck, 그리고 Head로 구성되어 있다. 그 중 Backbone은 특징 추출기로서 입력받은 이미지로부터 특징 맵을 추출하는 부분으로 YOLOv5 모델에서는 YOLOv4[4]와 유사하게 CSP-Darknet을 사용한다. 그리고 Backbone의 BottleneckCSP는 ResNet[5]에서 사용한 것과 같이 layer의 입력을 layer의 출력에 연결하는 shortcut connection 방식을 사용한다.

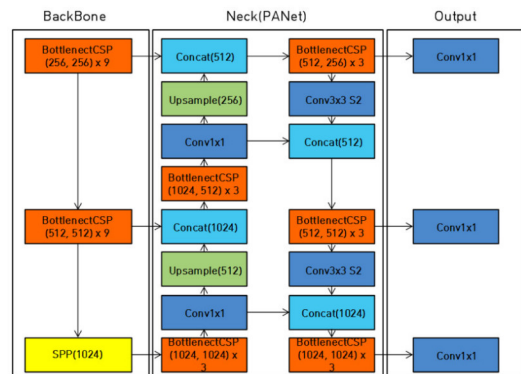


그림 1. YOLOv5 모델 구조

Fig. 1. YOLOv5 model architecture

BottleneckCSP는 배치 정규화를 적용한 convolution layer 2개와 적용하지 않은 convolution layer, 총 4개의 convolution layer를 생성하여 shortcut connection으로 연결한 2개의 convolution layer 연산 값과 1개의 단순 convolution layer 연산 값을 생성하여 합친 후 다른 1개의 convolution layer를 통과시킴으로써 연산한다. 그 후 이 연산을 depth 만큼 반복함으로써 더 깊은 모델을 만든다.

SPP(Spatial Pyramid Pooling layer)[6]는 고정된 크기의 입력을 받아야 하는 fully connected layer로 인하여 신경망을 통과시키기 위해 입력 이미지의 크

기를 고정시킴으로써 발생할 수 있는 정보왜곡을 해결하기 위해 제안되었다. SPP는 입력 이미지의 크기에 관계없이 이미지를 convolution layer에 통과시키고 특징 맵을 추출한다. 그 후 추출된 특징들을 SPP에 적용하여 고정된 크기의 특징 벡터를 추출한 뒤 fully connected layer의 입력으로 사용한다.

Neck으로 사용되는 PANet[7]은 기존의 FPN (Feature Pyramid Network)의 구조가 최하위 layer에서 최상위 layer까지 도달하는데 많은 layer를 지나게 됨으로써 하위 단계의 특징이 결과 예측에 적게 반영된다는 문제점을 해결하기 위해 제안되었다. 최상위 layer까지의 경로를 단축시키는 bottom-up path augmentation을 생성하여 하위 단계의 특징 상위 단계에 도달할 수 있도록 함으로써 원활한 정보 흐름이 가능하도록 한다.

Head는 추출된 특징 맵을 바탕으로 찾는 물체를 검출하는 부분으로 초기에 anchor box를 설정한 뒤 물체의 픽셀 정보에 따라 3가지 scale에서 최종 bounding box를 결정한다.

### III. 마스크 착용 모니터링 시스템

#### 3.1 시스템 구성

본 논문에서 제안하는 모니터링 시스템은 연결된 웹캠을 통해 입력된 실시간 영상의 프레임에서 마스크 착용 여부를 식별한다. 그림 2는 마스크 착용 모니터링 시스템의 전체적인 구성도이다.

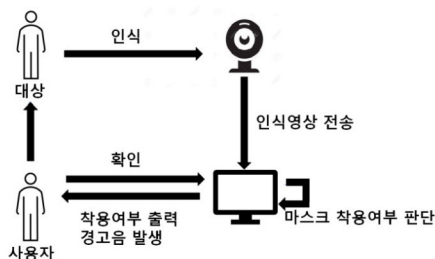


그림 2. 마스크 착용 모니터링 시스템  
Fig. 2. Monitoring system for mask-wearing

시스템은 마스크 착용 여부를 식별하고 결과를

화면에 출력하여 인식된 사람이 마스크를 정확히 착용한 경우 with\_mask, 마스크를 착용하지 않은 경우 without\_mask, 마스크를 정확하지 않게 착용한 경우 mask\_wearred\_incorrect의 텍스트와 함께 검출된 영역을 bounding box로 표시한다. 이 과정을 통해 마스크 착용 여부를 식별하여 실시간으로 화면에 출력한다.

#### 3.2 데이터셋

본 논문의 시스템은 마스크 착용 식별을 위해 Kaggle[8]에서 853개의 이미지를 수집하여 학습, 검증, 테스트 과정에서 with\_mask, without\_mask, mask\_wearred\_incorrect의 세 개의 class를 가지는 데이터셋을 사용한다. 그림 3은 사용한 Kaggle에서 수집하여 활용한 데이터셋의 예시이다.



그림 3. 사용 데이터셋 예시  
Fig. 3. Example of used dataset

학습 시의 학습 데이터, 검증 데이터, 테스트 데이터의 비율은 7:2:1로 한다. 해당 데이터셋은 세 개의 class가 포함된 단일 인물의 이미지와 몇 개의 class가 복합된 이미지로 구성되었으며 416x416 크기의 853개 이미지에서 4058개의 객체를 사용하여 학습한다.

#### 3.3 가중치 설정

YOLOv5는 YOLOv5-s(small), YOLOv5-m(medium), YOLOv5-l(large), YOLOv5-x(xlarge) 총 4가지의 가중치로 구성된다. [9]에서는 YOLOv5에서 제공하는 사전 학습된 가중치의 평가지표에 대한 비교 결과를

공개하였으며 표 2와 같다.

표 2. YOLOv5 사전 학습 비교

Table 2. Comparison of YOLOv5 pretrained checkpoints

구분	FPS	Speed	AP <sup>test</sup>
YOLOv5-s	455	2.2ms	36.7
YOLOv5-m	345	2.9ms	44.5
YOLOv5-l	264	3.8ms	48.2
YOLOv5-x	167	6.0ms	50.4

표 2의 결과를 참조했을 때 FPS는 YOLOv5-s가 가장 좋은 결과를 보여주고, AP<sup>test</sup>는 YOLOv5-x가 가장 좋은 결과를 보여준다. 본 논문에서 제안하는 시스템은 실시간 영상에서의 성능이 중요시되므로 FPS 수치가 가장 높은 YOLOv5-s를 가중치로 채택한다.

### 3.4 실험 결과

실험환경은 Google Colab에서 제공되는 환경을 사용하여 학습을 진행한다. 학습에 사용한 설정은 epoch를 1000으로 설정하였으며 optimizer는 adam을 사용한다.

최종 학습된 모델을 테스트 이미지에 적용하여 정상적으로 동작하는지 검증한다. 그림 4는 학습된 모델이 적용된 테스트 이미지로서 마스크 착용 여부 대해 정상적으로 동작함을 보여준다.

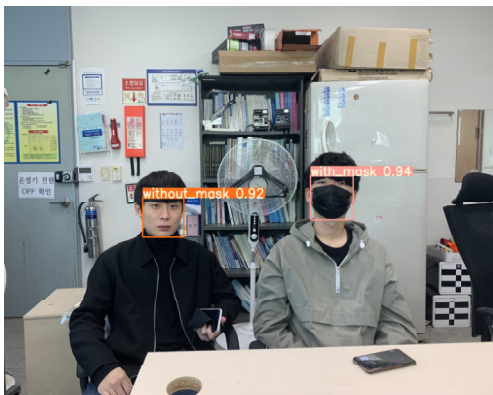


그림 4. 학습된 모델이 적용된 테스트 이미지

Fig. 4. Example of test image

그림 4와 같이 테스트 이미지에서 마스크를 착용한 경우 with\_mask, 마스크를 착용하지 않은 경우 without\_mask의 텍스트와 함께 검출된 영역을 bounding box로 표시한다. 그리고 연결된 웹캠을 사용하여 해당 과정을 실시간 영상에 적용한다. 그림 5는 학습된 모델이 적용된 실시간 영상의 화면이다.



그림 5. 실시간 영상에 적용한 이미지

Fig. 5. Image applied to real-time images

그림 5와 같이 연결된 웹캠을 사용한 실시간 영상에서 마스크를 정확히 착용한 경우 with\_mask, 마스크를 착용하지 않은 경우 without\_mask, 마스크를 정확하지 않게 착용한 경우 mask\_worn\_incorrect 텍스트와 함께 검출된 영역을 bounding box로 표시한다. 이를 통해 실내에서 마스크 미착용자를 식별할 경우 신속한 대처가 가능하다.

## IV. 결 론

본 논문에서는 객체 검출 알고리즘을 활용하여 사람의 실내에서 마스크 착용 여부를 식별하는 시스템을 제안하였다. 제안된 시스템은 마스크 미착용자 식별을 위한 여러 알고리즘 중 FPS와 AP<sub>50</sub>가 높아 실시간 처리에 적합한 YOLOv5 알고리즘을 이용하여 실시간으로 마스크 미착용자 식별이 가능하도록 구현하였다. 이 시스템을 사용함으로써 사람이 많은 실내에서 사람들의 마스크 착용 여부를 쉽게 확인할 수 있어 코로나바이러스(COVID-19) 확산 방지에 도움을 줄 것으로 기대된다.

향후 연구로 적외선 영상에서의 객체를 검출할 수 있도록 학습을 진행하여 극장과 같은 다소 어두

운 실내에서 적외선 영상에 대해 마스크 착용을 식별하는 방법에 대해 연구를 진행할 계획이다.

### 참 고 문 헌

- [1] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikainen, "Deep learning for generic object detection: A survey", International journal of computer vision, vol. 128, no. 2, pp. 261-318, October 2020.
- [2] YOLOv5 Model: <https://github.com/ultralytics/yolov5>. [accessed: May. 04, 2021]
- [3] G. Zhang, Zhi Li, Xuewei Li, Canbin Yin, and Zengkai Shi, "A Novel Salient Feature Fusion Method for Ship Detection in Synthetic Aperture Radar Images", in IEEE Access, vol. 8, pp. 215904-215914, November 2020.
- [4] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao, "Yolov4: Optimal speed and accuracy of object detection", arXiv preprint arXiv:2004.10934, April 2020.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition", Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, June 2016.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition", IEEE transactions on pattern analysis and machine intelligence, vol. 37, no. 9, pp. 1904-1916, September 2015.
- [7] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia, "Path Aggregation Network for Instance Segmentation", Proceedings of the IEEE conference on computer vision and pattern recognition, vol. 4, pp. 8759-8768, September 2018.
- [8] Kaggle Mask Dataset: <https://www.kaggle.com/andrewmvd/face-mask-detection>. [accessed: April. 29, 2021]
- [9] YOLOv5 모델 가중치 비교: <https://github.com/ultralytics/yolov5/releases/tag/v4.0>. [accessed: May. 17, 2021]