

딥러닝 기반 마스크 착용자 및 미착용자 검출

김태현, 우승희, 김정미, 최해철
한밭대학교, 정보통신공학과

bluex08@naver.com, wsh854@daum.net, 11wjdal@naver.com, choihc@hanbat.ac.kr

Mask and Maskless Wearers Detection based on Deep Learning

Taehyeon Kim, Seunghee Woo, Jeongmi Kim, and Haechul Choi
Dept. of Information Community Engineering, Hanbat National University

요약

코로나19 전염병 예방을 위해 공공장소에서의 마스크 착용이 의무화되고 있다. 그러나 사람들이 다양한 이유로 마스크를 제대로 착용하지 않아 감염에 노출되는 위험이 발생하고 있다. 이러한 방역 문제를 해결하고자 본 논문은 영상을 인식하여 마스크를 쓴 얼굴과 쓰지 않은 얼굴을 검출하는 방식을 제안한다. 제안 방법은 마스크 착용자와 미착용자 얼굴 영상을 딥러닝 기반의 YOLO 네트워크로 학습하여, 마스크 착용 유무를 판별한다. 동일 YOLO 네트워크에 대해 여러가지 조건으로 학습을 수행하고, 학습에 사용되지 않은 검증 데이터를 이용해 정확도가 가장 높은 네트워크의 가중치를 선택하였다. 실험결과, 마스크 착용자는 67.2%, 미착용자는 39.8%의 판별 정확도를 보였다. 미착용자에 대해 낮은 정확도를 보인 이유는 학습 데이터의 부족으로 판단되며, 이를 보완하기 위하여 더 많은 학습데이터를 제작하여 성능을 개선시키고자 한다.

1. 작품의 제작 동기

2020년부터 시작된 코로나19로 일상생활에 많은 영향을 받고 있다. 방역당국에서는 국가와 기업의 방역은 물론 개인 방역 역시 중요하게 여기고 있으나 몇몇의 사람들이 다양한 이유로 마스크를 제대로 착용하지 않아 감염에 노출되는 위험이 발생하고 있다. 이러한 이유를 통해 영상 내에 마스크를 쓰지 않은 사람을 검출하여 감염의 위험을 줄이고 방역에 좀 더 기여할 수 있도록 하고자 한다.

2. 작품의 설계 및 구현

얼굴 인식 네트워크에서 실용적이고 간단하며 쉽게 확장 가능한 얼굴 인식 데이터셋을 이용하여 문제를 단순화하여 진행하고자 한다. Faster R-CNN과 YOLO가 대표적인 객체 검출 알고리즘인데, 이들은 CNN을 기반으로 성능 개선을 주도한 대표적인 알고리즘이다. 성능 개선의 모태는 동일하지만, 프레임워크에서 차이점을 갖는데, YOLO의 실시간 얼굴 탐지를 위한 속도는 Faster R-CNN보다 빠르고, 얼굴 정면이 아니어도 쉽게 인식하며, 구현에 있어서 단순하여 이를 채택하였다.

2-1. YOLO

'You Only Look Once'의 약어로 2015년에 Joseph Redmon이 워싱턴 대학교에서 친구들과 함께 YOLOv1을 처음 논문과 함께 발표했다. 발표할 당시에는 Object Detection에서는 대부분 Faster R-CNN이 가장 좋은 성능을 내고 있었다. YOLO는 처음으로 One-shot-detection 방법을 고안하였는데, 이 전까지의 Two-shot-detection으로 object detection의 실시간성을 보완한 방법으로 볼 수 있다.

구분	one-shot-detection	two-shot-detection
장점	높은 실시간성	높은 정확도
단점	two-shot-detection에 비해 낮은 정확도	느린 속도

표 1. detection의 장단점

YOLO는 입력 이미지 또는 Feature Map을 특정 그리드로 나누고 각 Cell 마다 Object Detection을 수행하는 것이다.

2-2. YOLOv3

YOLOv3는 YOLO의 3번째 버전으로 YOLOv2에서 개선된 버전이다. 기본 동작은 YOLOv2와 거의 비슷하고, 큰 변화보다는 최신 기법 적용으로 성능을 끌어올린 것이라고 생각하면 된다.

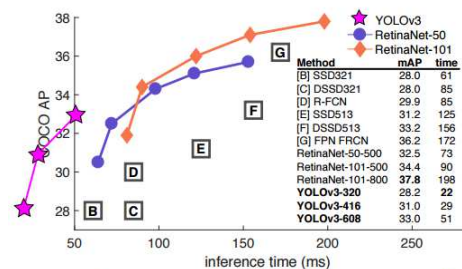


Figure 1. We adapt this figure from the Focal Loss paper [9]. YOLOv3 runs significantly faster than other detection methods with comparable performance. Times from either an M40 or Titan X, they are basically the same GPU.

그림 1. YOLOv3 성능

YOLOv2의 경우 bounding box를 예측할 때 4개의 좌표 (Tx, Ty, Tw, Th)를 구한 후 Bx, By, Bw, Bh로 변환한 후 L2 loss를 통해 학습시켰다. 하지만 YOLOv3는 ground truth 좌표를 위의 공식에서 거꾸로 적용시켜 T*로 변환한다. 예측한 bounding box마다 objectness score를 logistic함수를 적용하여 구한다. 또한 prior box(anchor box)와 ground truth box와의 IOU값이 가장 높은 box만 매칭시킨다. 할당되지 못한 bounding box는 오직 objectness score에 대한 loss 만 발생시킨다.

각각의 bounding box는 multi-label classification을 수행한다. 하나의 box 안에 복수의 객체가 존재하는 경우 softmax 함수를 사용하여 class를 예측하면 적절하게 객체를 포착하지 못한다는 문제가 있다. 따라서 box에서 각 class가 존재하는 여부를 확인하는 binary-cross entropy가 보다 적절하다고 볼 수 있다.

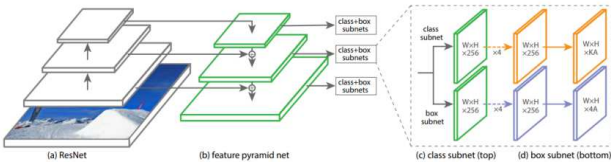


그림 2. YOLOv3 구조

YOLOv3는 위와 같은 일반적인 FPN과 구조가 비슷하다.

왼쪽은 일반적인 SSD와 같은 구조로, feature extractor의 앞쪽에서 나온 feature map은 표현력이 부족하다. 이를 보완하기 위해 오른쪽과 같이 deconvolution으로 feature map의 크기를 확장시켜 high-level feature를 뽑도록 한다. 오른쪽과 왼쪽의 feature map을 concat하여(왼쪽에서 위치정보 등을 가져오는 방식으로) 표현력을 향상시켜 사용한다.

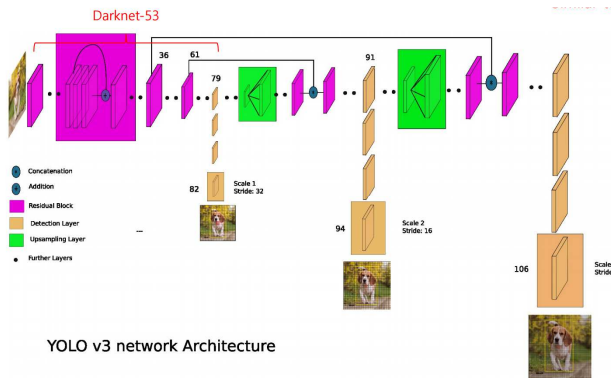


그림 3. YOLOv3 구조 (2)

왼쪽부터 순서대로 가장 큰 객체, 중간 객체, 가장 작은 객체를 찾도록 할당된다. 각각 feature extractor에서 생성된 feature map을(가로 방향) detection layer를 거쳐(세로 방향) 객체를 찾게 된다.

3. 작품의 구현 결과

해당 연구는 YOLO 네트워크를 이용하여 인종과 성별에 구애받지 않고 마스크의 색에 상관없이 다양한 사람들의 얼굴을 인식하여 마스크를 착용한 사람, 착용하지 않은 사람을 구분하여 검출해준다.

학습된 네트워크 모델의 성능을 파악하기 위하여 80%를 train data로 사용하고 20%를 test data로 사용하여 진행하였다.

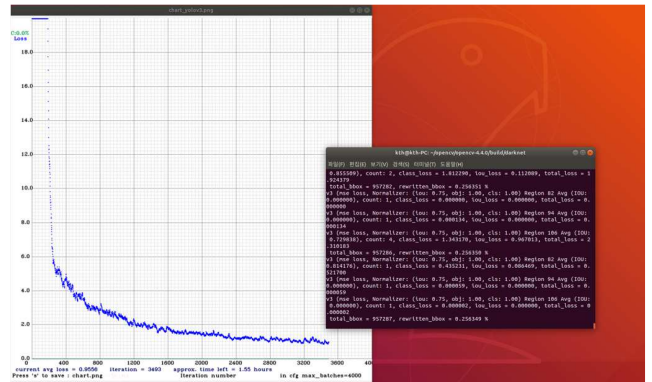


그림 4. 학습 진행 장면

darknet의 train 명령어로 mask : 2,547, no_mask : 823, total : 3,370의 마킹을 바탕으로 학습을 진행하였다.



그림 5. 학습 결과 가중치로 test 명령어 확인 화면

학습 진행 후 얻은 가중치를 이용하여 test 명령어를 통해 결과화면을 얻었고 mask와 no_mask를 검출한 장면을 확인할 수 있었다.

Network's weights	last.weights	final.weights
no_mask	39.88%	32.76%
mask	67.26%	62.25%
mAP50	53.57%	47.51%
Average IOU	47.42%	45.63%

표 2. 학습 가중치의 결과

학습한 가중치의 파일들의 결과를 위의 chart와 같이 나타냈으며 학습의 실험데이터의 평가를 위한 성능으로 IOU(Intersection of Union)의 평균값을 사용하였다. 정확한 결과를 얻고자 Bounding Box가 80% 이상 겹쳤을 때 예측이 성공하도록 설정하였으며, 그에 따라 얻은 가중치를 확인했을 때 위의 chart와 같은 결과를 얻었고, 이에 따라 좀 더 학습의 결과가 좋은 last.weights를 사용하여 결과화면을 얻었다.

Network's weights	최종학습.weights
no_mask	42.78%
mask	71.62%
mAP50	59.18%
Average IOU	53.75%

표 3. 최종학습한 후 가중치의 결과



그림 6. 최종 결과 화면

크롤링 데이터를 (약 3000장) 더 모아서 추가 학습을 시켜 최종적으로 IOU값이 50%가 넘도록 학습을 시킬 수 있었다. 추가 학습을 통해 기존에 오류로 인식되던 트럭에서의 라벨링이 사라진 것을 확인할 수 있었고, 마스크를 착용하지 않은 사람의 얼굴을 추가적으로 검출하여 나타내는 것도 확인할 수 있었다.

4. 작품의 기대효과

현재 사람들이 많이 이용하는 공간이나 시설에서 발열체크 카메라와 온도 측정을 진행하고 있다. 코로나19의 증상이 독감과 비슷한 발열에서부터 시작되었기 때문에 많은 상용화가 되어있는 것이다. 그러나 개인방역으로 마스크를 착용하는 것만큼 중요한 방역은 없다. 본 작품을 토대로 CCTV 또는 카메라에 접목하여 실시간으로 사람들의 얼굴에서 마스크 미착용을 판별할 수 있다면 방역에 좀 더 힘쓰며 개인방역 수칙에 빠른 대처가 가능할 것이고, 나아가 방역수칙을 지키지 않은 사람들에 대한 처벌 역시 빠르게 가능할 것이다.

5. 참고문헌

[1] <https://arxiv.org/abs/1804.02767> - 'YOLOv3: An Incremental Improvement'

[2]<https://seongkyun.github.io/papers/2019/11/20/yolov3/> - 'YOLOv3: An Incremental Improvement'

[3] <https://bestinau.com.au/yolov3-architecture-best-model-in-object-detection/> - 'YOLOv3 Architecture: Best Model in Object Detection'