

Tiny-YOLOv3와 ResNet50을 이용한 실시간 마스크 표정인식

박규리 박나연 김승우 김승혜 김진산 고병철
계명대학교geww171524@gmail.com vivicacaremanse@gmail.com 99n88@naver.com
gapalyt@naver.com kjkjw77@naver.com niceko@kmu.ac.kr

Real-time mask facial expression recognition using Tiny-YOLOv3 and ResNet50

Park, Gyuri Park, Nayeon Kim, Seungwoo Kim, Seunghye Kim, Jinsan Ko,
Byungchul
Keimyung University

요약

최근 휴먼-컴퓨터 인터페이스, 가상현실, 증강현실, 지능형 자동차등에서 얼굴표정 인식에 대한 연구가 활발히 진행되고 있다. 얼굴표정인식 연구는 대부분 맨얼굴을 대상으로 하고 있지만 최근 코로나-19로 인해 마스크 착용한 사람들이 많아지면서, 마스크를 착용했을 때의 표정인식에 대한 필요성이 증가하고 있다. 본 논문은 마스크를 착용했을 때에도 실시간으로 표정 분류가 가능한 시스템개발을 목표로 구동에 필요한 알고리즘을 조사했고, 그 중 Tiny-YOLOv3와 ResNet50 알고리즘을 이용하기로 했다. 얼굴과 표정 데이터셋 등에서 모은 이미지 데이터를 사용하여 실험해 보고 그 적절성 및 성능에 대해 평가해 보았다.

1. 서론

코로나 이전의 표정 인식 기술들은 얼굴의 형태를 인식하여 얼굴의 주요 부위들을 탐색해내는 형식이 대부분이었다. 하지만 공공보건을 위해 모두가 마스크를 쓰는 현재 기존의 딥러닝 표정 인식 기술은 얼굴의 일부가 가려지거나 마스크를 쓰는 경우 사람의 표정을 인식할 수 없다는 한계점이 있다. 따라서 우리는 다양한 마스크 착용 얼굴 표본을 모아 얼굴이 가려진 상태에서도 감정을 파악하고 표정을 분류해 낼 수 있는 기술을 목표로 한다.

마스크를 썼을 경우 가장 불편한 업종을 생각해 봤을 때, 크게 세 가지를 생각해 낼 수 있었다. 수사기관, 판매 직원, 심리 상담사이다. 수사기관의 경우 CCTV 등을 통해 범죄자의 심리적 특성을 분석하여 다음 행동을 예측하는 프로파일링 기법을 사용한다. 얼굴이 가려진 범죄자의 감정을 분석해냄으로써 추가 범죄의 발생 가능성을 크게 낮출 수 있을 것이다. 또한, 매장을 운영하는 직원의 경우 매장의 곳곳에 카메라를 배치하여 마스크 속 고객의 표정을 분석하여 긍정적인 반응의 제품을 유지하고, 부정적인 반응을 보이는 제품을 교체하는 등 매출에 변화를 줄 수 있을 것이다.

마지막으로 심리 상담의 경우 상담사는 상담자의 심리를 파악하며 상담을 이끌어야 한다. 만일 상담자가 마스크를 착용한 경우 상담사는 상담자의 얼굴 전체를 확인하는 데 어려움이 있기 때문에 어떤 심리상태가 표정에 드러나는지 확인하기 힘들 수 있다. 따라서 카메라를 이용해 상담자의 가려진 표정을 확인하고 현재의 대화가 어떤 감정 변화를 일으키는지 확인하여, 상담자의 심리상태에 따라 상담의 진행 방향을

결정하는 데 도움을 줄 수 있을 것이다.

본 논문에서는 딥러닝 알고리즘인 Tiny-YOLOv3와 ResNet50을 사용하여 얼굴이 가려졌을 때 표정 인식을 가능하도록 알고리즘을 개발하고 이를 응용프로그램으로 구현하였다.

2. YOLOv3와 ResNet을 이용한 표정인식

2.1 YOLOv3를 이용한 얼굴 검출

YOLO는 간단한 처리 과정을 통해 영상의 물체를 검출하는 매우 빠른 딥러닝 알고리즘 중의 하나이다 [1]. 이 방법은 기존에 있던 실시간 딥러닝 물체 검출 알고리즘보다 높은 정확도와 속도를 가졌다. YOLO가 개발된 이후로 여러 작은 변화들도 검출해낼 수 있는 YOLO 알고리즘이 여러 버전으로 개선되었다. 그중 tiny-YOLOv3는 작은 객체에 대한 검출 성능을 높인 YOLO v3보다 좀 더 빠른 속도로 객체 검출이 가능하다. 두 방법에 대한 성능 비교를 위해 4개의 객체 class를 두고 학습 데이터 약 20,000장, 검증 데이터 약 2,200장을 사용하여 성능을 비교하여 보았고 그 결과는 표 1과 같다. 정확도는 YOLO v3가 약 9%정도 더 높지만, 속도는 tiny-YOLOv3가 약 4.2 배 빠르다. 따라서 본 논문에서는 실시간으로 얼굴을 검출할 필요가 있으므로 tiny-YOLOv3를 이용해 마스크 착용과 착용하지 않은 얼굴 검출을 하였다.

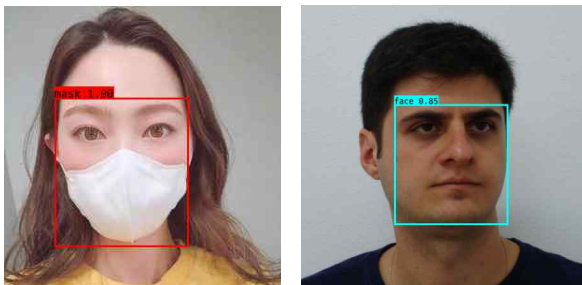
Model	mAP	FPS
YOLOv3	87.79%	30.3
tiny-YOLOv3	73.51%	155.2

<표 1> YOLOv3와 tiny-YOLOv3 성능 비교

본 논문에서는 마스크 착용 여부를 확인할 필요가 있기 때문에 마스크를 착용한 얼굴과 마스크를 착용하지 않은 얼굴 데이터를 수집했다. 약 4,000장의 데이터를 모았으며, 데이터 학습 시간을 줄이기 위해 학습 시킬 영상 크기를 416x416으로 조절해준다. 이후 YOLO_mark라는 라벨링 툴을 사용하여 각 영상에서 검출해야 할 얼굴의 위치가 어디에 위치해 있고, 해당 얼굴의 class가 마스크를 쓴 것(1)인지 안 쓴 것인지(0) 지정해준다. 이 과정을 모든 영상에 반복해 annotation 파일들을 만들었다.

보통 딥러닝 알고리즘에서 학습시키는 데이터는 적어도 10,000장이 필요하지만, 실제 수집한 데이터가 부족함으로 이를 늘이기 위하여 데이터를 학습시킬 때 config 파일 내에서 angle, saturation, exposure의 값을 바꿔가며 학습 데이터를 증가시켰다. 각도의 경우, config 파일을 바꾸는 것만으론 안 되기 때문에 따로 여러 각도로 인위적으로 돌려 데이터를 증가(augmentation) 시킨 다음 라벨링을 했다.

증강된 학습 데이터들을 가지고 tiny-YOLOv3를 학습시켰다. 학습된 가중치 파일을 가지고 객체 탐지를 시작하면 그림 1과 같이 나온다. 웹캠을 연결해 실행했을 때도 같은 결과가 나오음을 확인했다.



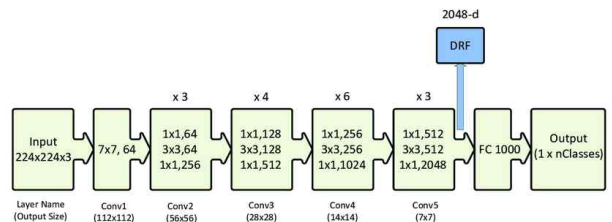
<그림 1> Tiny-YOLOv3를 이용한 객체 탐지

본 논문에서는 마스크 착용 여부 관계없이 표정을 분류할 수 있어야 한다. 따라서 얼굴영역 이외의 부분은 필요 없으므로, 영상을 검출해낼 때 사각형으로 검출 부분을 표시하고 그 좌표값을 저장한다. 그 후, 저장된 영상은 160x160으로 정규화되어 ResNet에 입력되고 표정인식을 위해 학습에 사용된다.

YOLO에서 검출 결과 class가 0이 나오면 마스크를 착용하지 않았을 때 표정 분류를 할 수 있는 ResNet 모델 0을, class가 1이 나오면 마스크를 착용했을 때 표정 분류를 할 수 있는 ResNet 모델 1을 불러오도록 했다. ResNet 모델 0에서의 class는 얼굴 전체를 보고 판단할 수 있기 때문에 표정 분류도 쉬워 angry, fear, neutral, happy, sad, surprise 총 6개의 class로 분류했다. ResNet 모델 1에서의 class는 눈과 눈썹만으로 표정 분류를 해야하기 때문에 angry와 fear, sad의 구분이 어려워서 negative로 통일하였다. 따라서 ResNet 모델 1의 class는 negative, neutral, happy, surprise 총 4개의 class로 분류했다.

2.1 ResNet을 이용한 표정 인식

ResNet 모델은 모델의 층이 많을 때, 학습 과정에서 입력층으로 갈수록 기울기 값이 작아지는 기울기 소실문제(Vanishing Gradient Problem)가 발생한다 [3]. 이 문제를 완화하기 위해 그림 2와 같이 잔여블록(Residual Block)을 사용하였으며, 입력값과 해당 입력값에 대해 하나 이상의 층들의 합쳐져 출력값으로 사용할 수 있도록 이루어져 있다. 이러한 구조를 통해 신경망의 깊이가 깊어질수록 학습 과정에서의 기울기 소실문제에 대한 문제를 완화할 수 있다는 장점을 갖고 있어 기존 신경망보다 높은 정확도를 얻을 수 있다.



<그림 2> ResNet의 구조 [6]

ResNet에는 레이어 개수에 따라 ResNet 18, ResNet 50으로 나뉘어진다. 본 논문에서는 처리속도를 줄이기 위해 ResNet 15모델을 새롭게 만들어 ResNet 50모델과 비교실험을 진행하였다. 두 모델에 사이즈가 160x160인 동일한 영상을 약 3,000장을 가지고 batch size를 16으로 지정해 손실률과 정확도를 비교해보았다. 마스크 착용 시의 표정 분류 성능 결과를 표 2와, 마스크 미착용 시의 표정 분류 성능 결과는 표 3과 같다.

<표 2> ResNet 모델 0(마스크 미착용)의 15와 50 성능 비교

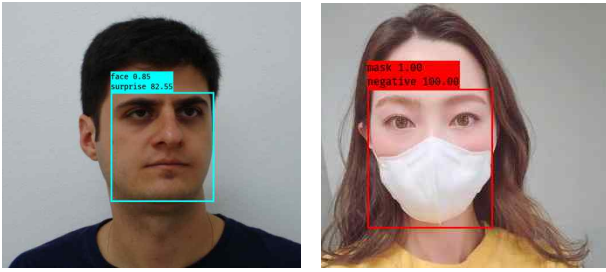
	최종 loss	최종 accuracy
ResNet 15	0.0517	0.9875
ResNet 50	0.0257	0.9891

<표 3> ResNet 모델 1(마스크 착용)의 15와 50 성능 비교

	최종 loss	최종 accuracy
ResNet 15	0.1775	0.9634
ResNet 50	0.1312	0.9573

마스크 미착용에서의 두 표정 분류 모델을 비교해본 결과, ResNet 15는 최종 loss가 0.0517인 반면 ResNet 50의 최종 loss는 0.0257이다. 마스크 착용에서의 두 표정 분류 모델도 ResNet 15의 최종 loss는 0.1775, ResNet 50의 최종 loss는 0.1312이다. 이를 통해 모델 1에서 loss가 0.026, 모델 2에서 loss가 0.0463 더 낮은 ResNet 50이 15보다 더 학습이 잘 된 것을 알 수 있었다. 따라서 본 논문에서는 ResNet 50을 이용하여 표정 분류 모델을 사용했다.

실시간으로 tiny-YOLOv3에서 얼굴을 검출하고 ResNet 50에서 표정 분류를 한 결과를 화면에서 곧바로 확인할 수 있도록 했다. 그림 3과 같이 검출한 부분을 사각형으로 그린 다음 사각형의 왼쪽 위에 마스크 착용 여부와 현재의 표정이 출력되도록 했다.



<그림 3> Tiny-Yolov3와 ResNet50을 이용한 표정 분류

3. 결론

본 논문에서는 마스크를 착용했거나 얼굴이 가려져 있는 상태에서도 얼굴 표정을 알아낼 수 있도록 하기 위해 새로운 얼굴 표정인식 알고리즘들을 제안하고 이를 프로그램으로 개발하였다. 우선 tiny-YOLO v3의 경우 영상 데이터 전처리를 통해 얼굴 위치의 좌표를 구해내고, 마스크 착용 여부에 따라 적절한 클래스로 분류하는 데에 성공했다. 또한 ResNet 50을 이용하여 마스크를 착용했을 때와 착용하지 않았을 때의 표정 분류 영상들을 학습시켜 데이터 분류 학습을 수행하였다. 그 결과 마스크를 착용했음에도 표정 분류를 원활히 해낼 수 있음을 알 수 있었다.

감사의 글

본 논문은 교육부와 한국연구재단의 계명대학교 대학혁신지원사업비를 지원받아 수행된 것입니다.

참고 문헌

- [1] 이경민, 송혁, 김재우, 인치호, 「효율적인 작은 객체 검출을 위한 균형적인 성능의 YOLOv3-tiny」, 한국방송미디어공학회 학술발표대회 논문집, 2018(11), 113-114, 2018
- [2] 남충현, 남은정, 장경식, 「딥러닝을 이용한 마스크 착용 여부 검사 시스템」, 한국정보통신학회논문지, 25(1), 44-49, 2021
- [3] "[3분 딥러닝] 5. 이미지 처리 능력이 탁월한 CNN" Rookie's Programming tistory, 2020년 1월 28일 수정, 2021년 4월 16일 접속, <https://bigdatadiary0819.tistory.com/53>
- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, 「Deep Residual Learning for Image Recognition」, CVPR, 770-778, 2016
- [5] Ammar Mahmood, Ana Giraldo, Mohammed Bennamoun, Senjian An, 「Automatic Hierarchical Classification of Kelps Using Deep Residual Features」, MDPI, 7, 2020