Journal of the Korea Academia-Industrial cooperation Society
Vol. 23, No. 1 pp. 652-659, 2022

https://doi.org/10.5762/KAIS.2022.23.1.652
ISSN 1975-4701 / eISSN 2288-4688

젯슨 나노를 이용한 심층 컨볼루션 신경망 기반 실시간 마스크 착용 검출기 구현

강태운, 김용우*
상명대학교 시스템반도체공학과

An Implementation of CNN-based Real-Time Face Mask Detector using Jetson Nano

Tae-Woon Kang, Yongwoo Kim*
Division of System Semiconductor Engineering, Sangmyung University

요 약 코로나바이러스의 전파를 막기 위한 가장 현실적인 대안은 마스크 착용이나 마스크 착용 의무화가 시행되었음에도 마스크 착용이 지켜지지 않는 경우가 있다. 집회 및 다중이용시설의 경우 마스크 미착용자를 사람이 일일이 확인하는 것은 불가능에 가깝다. 본 논문에서는 인간의 한계를 극복하고, 코로나바이러스의 전파를 방지하고자 딥러닝 기반 실시간 마스크 검출기를 제안한다. 제안 방법은 딥러닝 기반 객체 검출기인 Single Shot Multibox Detector(SSD)를 마스크 검출에 용이하도록 경계 박스의 개수 및 비율을 수정하고 특징 맵의 개수를 최적화하였다. 또한 SSD에서 사용되는 VGG-16 기본 네트워크 구조 대신 모바일 디바이스에 특화된 MobileNetV2 네트워크 구조를 수정 및 최적화하여 기본 네트워크로 적용하였다. 제안한 마스크 검출 기법은 VGG-16 기반 SSD 마스크 검출 기법에 대비하여 젯슨 나노에서 약 3배가량 수행 시간이 감소함을 확인하였다. 또한, 제안한 마스크 검출 기법은 MobileNetV1 기반 SSD 마스크 검출 기법 대비 mAP 성능이 약 7.05% 향상된 마스크 검출 정확도를 보여주었다.

Abstract The most realistic measure to prevent the spread of the coronavirus is to wear a mask, but there are cases where wearing a mask is not strictly followed even though it is mandatory. Furthermore, it is almost impossible for people to check if everyone is wearing a mask in multipurpose facilities or gatherings. Hence, this paper proposes a real-time mask detector based on deep learning to overcome human limitations and prevent coronavirus spread. The proposed method uses the Single Shot Multibox Detector (SSD) to modify the number and ratio of bounding boxes to facilitate the mask detection and optimize the number of feature maps. As a result, the proposed network accurately and quickly detects masks in real-time on Jetson Nano. It was confirmed that the proposed mask detection method reduces the execution time by about three times with the Jetson Nano compared to the VGG-16 based SSD mask detection method. In addition, the proposed mask detection method showed a mask detection accuracy that improved by about 7.05% in mAP performance compared to the MobileNetV1-based SSD mask detection method.

Keywords : Single Shot Multibox Detector, Mask Detection, Embedded Device, Real-Time, Jetson Nano

다음의 성과는 과학기술정보통신부와 연구개발특구진흥재단이 지원하는 과학벨트 지원사업으로 수행된 연구결과입니다.
*Corresponding Author : Yongwoo Kim(Sangmyung Univ.)
email: yongwoo.kim@smu.ac.kr
Received September 2, 2021 Revised October 5, 2021
Accepted January 7, 2022 Published January 31, 2022

1. 서론

최근 코로나바이러스(SARS-CoV2) 사태로 전 세계가 팬데믹에 빠져있다. 코로나바이러스는 현재 다양한 변종으로 변화하고 있으며, 국내 우세종으로 들어온 바이러스는 델타 변이바이러스(lineage B.1.617.2)이다. 델타 변이바이러스는 알파 변이바이러스보다 전파력이 약 1.64배 높고, 감염시 입원 위험이 1.85배 높은 것으로 알려졌다[1]. 코로나바이러스는 비말을 통해 전파되는 것으로 알려져 있다. 비말 확산을 방지하기 위해서는 마스크 착용이 필수적이다. 마스크 착용은 코로나바이러스 전파를 90% 이상 방지한다[2]. 백신 접종을 완료한 사람들 또한 마스크를 착용하지 않으면 언제든지 돌파 감염이 일어날 수 있다. 실제 2021년 7월 미국 매사추세츠주에서 대규모 행사 이후 확진된 469명 중 백신 접종을 완료한 사람 346명(74%)인 통계가 CDC(Centers of Disease Control and Prevention)에 존재하며, CDC 또한 백신 접종 완료자에 대한 마스크 착용을 권장하고 있다[3]. 그러나 일부 사람들은 마스크를 착용하지 않고 집회, 건물 출입 등 다중 이용시설을 방문하여 방역 수칙을 위반하고 있다. 다중 이용시설의 마스크 착용을 확인하기 위해 방역 당국 및 경찰 인력이 투입되고 있다. 이러한 인력이 투입되더라도, 모든 인원의 마스크 착용을 확인하는 것은 불가능에 가깝다. 따라서 본 논문에서는 팬데믹 상황 속 마스크 착용 확인을 위한 딥러닝 기반 실시간 마스크 검출기를 제안하고자 한다.

본 논문에서 제안한 마스크 검출기 특징은 다음과 같다. 첫째, 기존에 연구된 마스크 착용 여부 검출기보다 mAP 및 수행시간이 향상된 Mask-SSD를 제안하였다. 둘째, 얼굴의 특성을 분석하여, 경계 박스의 개수를 약 2배 줄여, 객체 검출 네트워크를 경량화하였다. 마지막으로 제안한 마스크 검출기 Mask-SSD를 임베디드 디바이스인 젯슨 나노에 구현하여, 실시간 검출이 가능함을 보여주었다.

본 논문의 구성은 다음과 같다. 2장에서는 마스크 검출을 위해 사용된 딥러닝 기반 심층 컨볼루션 신경망 네트워크의 관련 연구 및 그에 대한 문제점을 제시한다. 3장에서는 새롭게 제안된 마스크 검출 네트워크인 Mask-SSD의 구조와 제안 방법을 설명한다. 4장에서는 제안된 Mask-SSD의 실험 환경 및 실험 결과를 제시하고 분석한다. 마지막으로 5장에서 결론에 대해 기술한다.

2. 관련 연구

딥러닝 기반 객체 검출 기법은 크게 두 가지 과정으로 구성할 수 있다. 객체의 위치를 찾는 회귀(Regression) 과정과 객체를 분류(Classifcation)하는 과정이다. 위의 두 과정을 한 번에 처리하는 것을 1단계 기반 객체 검출기라고 일컫으며, 2단계 기반 객체 검출기는 순차적으로 두 과정을 처리하여 객체를 검출한다. 1단계 기반 객체 검출기의 대표적인 예로는 SSD(Single Shot Multibox Detector)[4]와 YOLO(You Only Look Once)[5]가 있으며, 2단계 기반 객체 검출기의 대표적인 예로는 Faster R-CNN[6]과 Mask-RCNN[7]이 있다. 1단계 기반 객체 검출기는 일반적으로 2단계 기반 검출기보다 속도가 빠르지만, 정확도는 낮다고 알려져 있다[8].

그중에서도 SSD[4]는 다양한 크기의 특징 맵(feature map)을 추출하여 크기가 다른 여러 객체를 동시에 확인할 수 있는 장점이 있다. SSD는 VGG-16[9]을 기본 네트워크(base network) 또는 백본 네트워크(backbone network)로 사용한다. VGG-16 네트워크는 필터의 크기가 3으로 고정된 컨볼루션 계층을 13번 통과하고, 완전 연결 계층을 3번 통과하여 네트워크 구성하여 분류

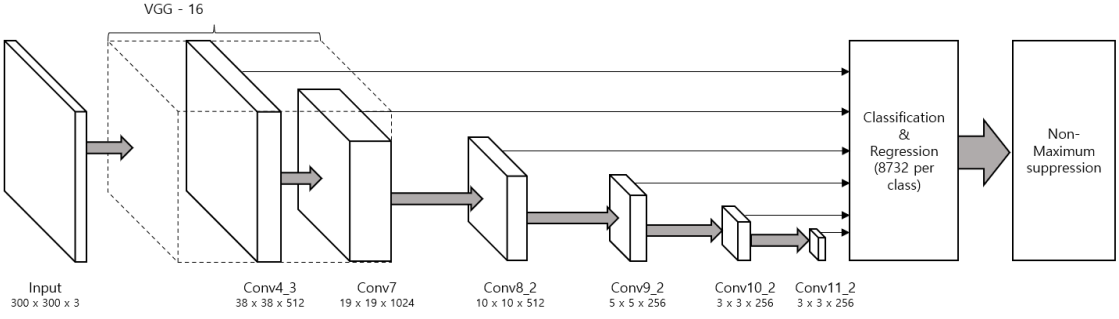


Fig. 1. VGG-16 based SSD architecture.

문제에서 높은 정확도를 달성하였다. VGG-16을 기본 네트워크로 사용한 SSD의 구조는 Fig. 1과 같다. SSD는 총 6개의 특징 맵을 추출하여, 여러 가지 크기의 특징 맵에 대해 각기 다른 개수의 경계 박스(bounding box)를 통해 검출할 객체의 위치 및 정답을 예측한다. 각 특징 맵의 한 픽셀당 4개 또는 6개의 영상 비(aspect ratio)로 경계 박스가 구성되어 있으며 총 8,732개로 구성된다. 이를 토대로 크기가 작은 객체나 크기가 큰 객체를 동시에 검출할 수 있다.

Fig. 1의 마지막 단계에 있는 NMS(non-maximum suppression)는 여러 가지 특징 맵을 통해 추출한 수많은 경계 박스 중 가장 높은 점수를 가지고 있는 경계 박스를 선택한다. 만약 경계 박스에 대한 점수가 IoU(intersection over union) 임곗값 보다 작다면 무효화하고, 다른 경계 박스들과 비교하여 가장 높은 점수를 가지고 있는 경계 박스를 찾는 방식이다. 반복된 과정을 통해 정답과 가장 유사한 경계 박스 하나만을 보여주게 된다.

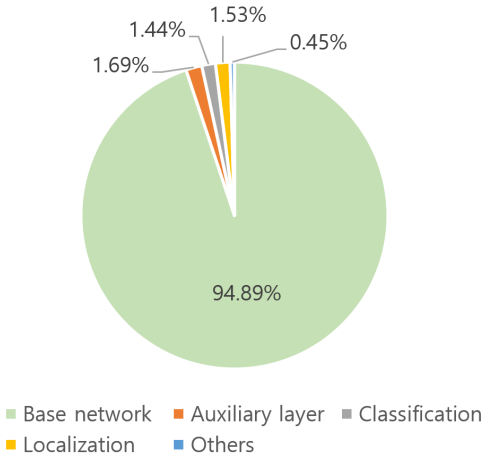


Fig. 2. Inference time analysis of VGG-16 based SSD

Fig. 2는 VGG-16을 기본 네트워크로 사용한 SSD를 구성하는 각각의 네트워크 별 수행 시간을 측정한 결과를 보여준다. 각각의 연산 및 계층마다 실행 시간을 측정한 결과 Fig. 2와 같이 기본 네트워크인 VGG-16을 수행하는 시간 비중이 매우 큰 것을 확인할 수 있는데, 전체 연산시간의 약 94.89%를 차지한다. 따라서, 실시간 객체 검출을 위해서는 SSD에서 적용된 기본 네트워크의 최적화가 반드시 필요하다. 최근 연구에서 모바일 디바이스에 최적화된 심층 컨볼루션 신경망으로 알려진 MobileNetV1[10]과 MobileNetV2[11]가 제안되었다.

MobileNetV1은 Depthwise Seperable 컨볼루션의 개념을 도입하여 만든 네트워크로, 연산량과 모델의 경량화를 통해 모바일 디바이스에서 효율적으로 분류할 수 있도록 만든 네트워크이다[10]. Depthwise Seperable 컨볼루션은 Depthwise 컨볼루션 연산과 Pointwise 컨볼루션 연산으로 나누어진다. Depthwise seperable 컨볼루션은 일반적인 컨볼루션 연산을 진행하였을 때 보다 연산량이 필터 크기의 제곱만큼 줄어든다.

MobileNetV2는 MobilnetV1에 Bottleneck Residual block이 도입된 네트워크이다. Fig. 3은 Bottleneck Residual block을 나타낸다. Depthwise Seperable 컨볼루션 사용 및 스트라이드(stride)값에 따라 입력 데이터를 출력 데이터와 더하는 특징을 갖고 있다[11]. 또한 Fig. 3의 활성화 함수로는 ReLU6가 사용되었다. ReLU6는 기존 ReLU 함수에서 상한선을 6으로 두어 대부분이 0으로 차 있는 희소 특징(sparse feature)들을 통해 조금 더 빠르게 학습할 수 있다[12]. COCO 데이터셋[13]으로 훈련된 MobileNetV1이 기본 네트워크로 사용된 SSD와 VGG-16이 기본 네트워크로 사용된 SSD의 성능을 비교한 결과 성능지표인 mAP(mean average precision)값이 약 1.8% 하락하였다[10].

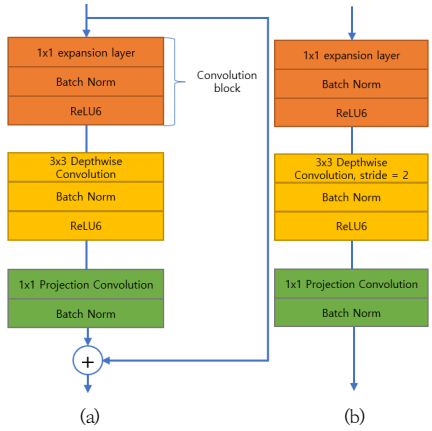


Fig. 3. Bottleneck residual block
(a) stride=1, (b) stride=2

MobileNetV2이 기본 네트워크로 사용된 SSD는 SSD-lite가 제안되었다. SSD-lite는 조금 더 모바일 디바이스에 최적화되게 제안된 네트워크이며, 기본 네트워크를 포함한 모든 컨볼루션 연산을 Depthwise Seperable 컨볼루션 연산으로 변경하여 파라미터의 크기를 약 7배 줄인다. COCO 데이터셋으로 훈련된 MobileNetV2가 기본 네트워크로 사용된 SSD와 VGG-16이 기본 네트워

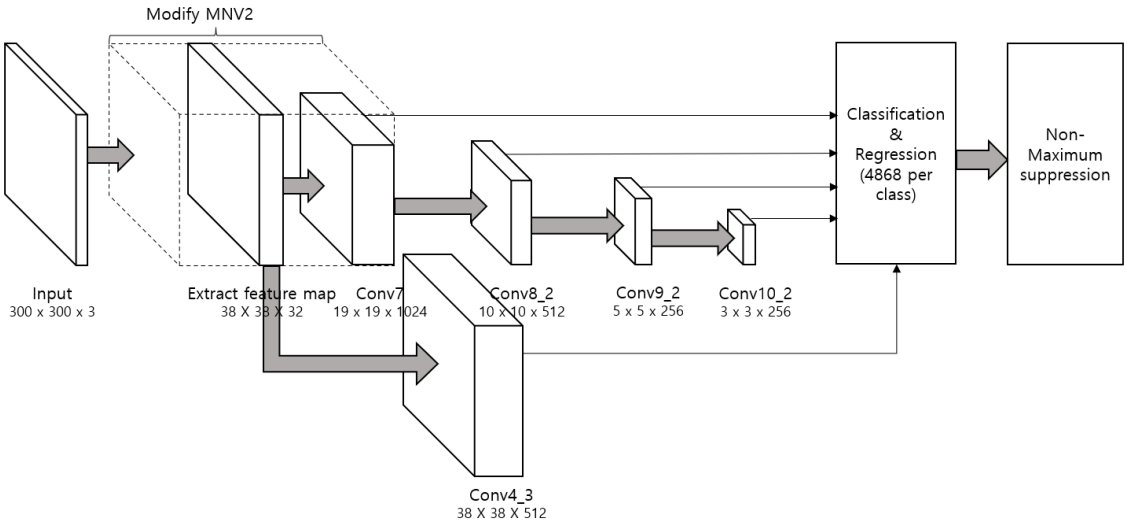


Fig. 4. Proposed Mask-SSD architecture

크로 사용된 SSD의 성능을 비교한 결과 성능지표인 mAP 값이 약 1.1% 하락하였다[11].

SSD에서 사용된 기본 네트워크의 변경으로 파라미터의 수를 줄여 속도 향상을 확인할 수 있으나, mAP 값이 하락하는 문제를 확인할 수 있다. 본 논문에서는 VGG-16을 기본 네트워크로 사용한 SSD와 유사한 정확도를 가지며, MobileNetV1과 MobilNetV2를 기본 네트워크로 사용한 SSD의 수행 시간과 유사한 성능을 가지는 Mask-SSD를 제안하고자 한다.

3. 제안 방법

Fig. 4는 본 논문에서 제안하는 Mask-SSD의 전반적인 구조를 나타낸다. Fig. 1과 대비하여 Conv11_2에 해당하는 컨볼루션 레이어가 삭제되고, 기본 네트워크가 수정된 것을 확인할 수 있다. Conv11_2 레이어의 작은 특징 맵은 큰 객체를 검출하는 특징을 가지는데 이러한 특징은 CCTV와 IP 카메라를 통해 실시간 검출을 수행하는 환경에는 적합하지 않다. 다시 말하면, 영상에 화면 전체에 한 명의 얼굴이 가득 차게 되는 시나리오는 없으므로, 작은 특징 맵을 배제하였다. Table 1은 기본 네트워크의 연산 과정을 보여주고 있는데, 이때 t는 팽창 계수(expansion factor), n은 bottleneck 연산 반복횟수, s는 스트라이드, c는 출력되는 채널의 수를 의미한다.

Fig. 4에 5번째 특징 맵을 추출하여 Conv4_3으로 변

Table 1. Modified MobileNetV2 process

Input	Operator	t	n	s	c
300^2 x 3	Conv2d 3×3	-	1	2	32
150^2 x 32	bottleneck	1	1	1	16
150^2 x 16	bottleneck	6	2	2	24
75^2 x 24	bottleneck	6	3	2	32
38^2 x 32	bottleneck	6	4	2	64
19^2 x 64	bottleneck	6	3	1	96
19^2 x 96	bottleneck	6	1	2	160
10^2 x 160	bottleneck	6	1	1	320
10^2 x 320	Conv2d 1×1	-	1	1	672
14^2 x 672	AvgPool 2×2	-	1	1	672
13^2 x 672	Conv2d 1×1	-	1	1	1024
Last output		19^2 x 1024			

형하는 과정이 있는데, 이때 추출한 특징 맵은 Table 1의 5번째 과정의 입력으로 들어가는 특징 맵을 의미한다. 추출된 특징 맵은 Fig. 3의 Convolution block과 같은 연산을 거쳐 변환한다. 제안 방법을 통해 SSD의 예측 레이어를 수정하지 않고 필터의 크기가 1인 컨볼루션 연산을 통해 간단히 채널의 출력값만을 수정하여 특징 맵을 깊이 있게 만들어 연결해 준다.

Fig. 5는 Mask-SSD의 기본 네트워크로 사용된 변형된 MobileNetV2의 구조를 보여주고 있다. Fig. 5의 Residual Block은 Fig. 3과 동일한 구조이다. MobileNetV2와 다르게 분류기는 삭제하였고, 평균 풀링 레이어와 Fig. 3의 convolution block을 사용하여 Fig. 4의 Conv7과 같은 구조로 만들었다.

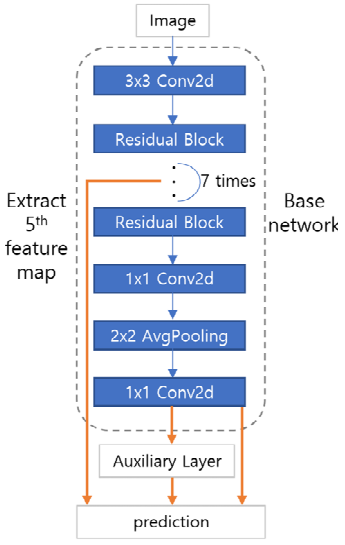


Fig. 5. Network Structure of Modified MobileNetV2

Table 2. The number of prior box by feature maps with aspect ratio

Feature map	Feature map form	Scale of bounding box	Aspect ratio	The number of bounding box
Conv4_3	38,38	0.1	1:1 + extra	2,888
Conv7	19,19	0.2	1:1, 1:2, 2:1 + extra	1,444
Conv8_2	10,10	0.375	1:1, 1:2, 2:1 + extra	400
Conv9_2	5,5	0.55	1:1, 1:2, 2:1 + extra	100
Conv10_2	3,3	0.725	1:1, 1:2, 2:1 + extra	36
TOTAL	-	-	-	4,868

Table 2는 각각 추출된 각 특징 맵의 크기, 이미지 대비 경계 박스의 비율, 영상 비, 총 경계 박스의 개수가 기술되어 있다. 기존 SSD 대비 절반에 가까운 4,868개의 경계 박스를 갖고 있는 것을 확인할 수 있다. 마스크 미착용 및 마스크를 착용한 사람의 검출을 목적으로 하기에, 사람의 얼굴 비율의 기준이 매우 중요하다. 얼굴의 황금비는 1:1.168이며, 대부분의 얼굴은 2:3으로 구성된 사각형으로 표현할 수 있다고 한다[14]. 결과적으로 사람의 얼굴 비율은 1:2를 넘지 않기에 이러한 가정을 통해 얼굴 비율 이외는 모두 삭제하였다. Table 2에 기재된 Aspect ratio의 extra 비율은 특징 맵의 한 픽셀과 크기가 다른 1:1 크기의 경계 박스를 의미한다.

4. 실험결과

이번 장에서는 본 논문에서 제안된 Mask-SSD의 성능을 확인하고자 진행한 실험 환경 및 실험결과를 자세히 설명한다. 네트워크의 전반적인 처리는 Fig. 6와 같다. 훈련 데이터와 실험 데이터를 나누어 훈련하고 실험 데이터를 통해 성능지표들을 계산하였다. 이후 실시간 동작이 가능한지 확인하고자 제안된 Mask-SSD 객체 검출 기법을 실제 환경인 젯슨 나노에 이식하여 수행 시간을 확인하였다.

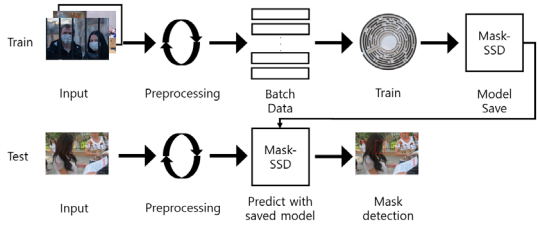


Fig. 6. Overall processing for Mask-SSD

4.1 실험 환경

본 논문에서는 실시간 검출이 가능한지 확인하기 위해 두 가지 환경을 구성하였다. 첫째는 GPU 없이 CPU만 존재하는 PC 환경이며, 둘째는 임베디드 디바이스인 젯슨 나노 보드 환경이다. PC의 경우 CPU는 Intel[R]의 i7-8565U를 사용하였고, RAM은 DDR4 16GB이며 운영체제는 Windows10 64bit에서 실험을 진행하였다. 젯슨 나노 보드의 CPU는 64비트 ARM 프로세서 A57이 사용되었고, RAM은 LPDDR4 4GB이며 운영체제는 Ubuntu 18.04를 사용한 환경에서 실험을 진행하였다. 젯슨 나노 보드의 가장 큰 특징은 128-core NVDIA Maxwell^{TM} GPU가 탑재되어있다. 제안된 Mask-SSD 객체 검출 기법의 구현은 모두 Pytorch 프레임워크를 사용하였고, Pytorch 1.9.0 버전을 사용하였다. 영상처리를 위한 모듈은 OpenCV-python 4.5.3 버전을 사용하였다.

마스크 검출을 위한 데이터셋[15]은 총 7,959개의 마스크 착용 및 마스크 미착용자의 이미지가 섞여 있다. 데이터셋을 육안으로 확인하여 정답 라벨이 없거나, 정답 박스의 좌표가 잘못된 5개의 이미지 데이터는 소거하였다. 데이터를 임의로 섞어 6,115개의 훈련 데이터셋과 1,839개의 실험 데이터셋으로 나누었다. 다양한 데이터는 네트워크 입력 규격에 맞게 크기를 변경하였다. 이와 동시에 객체 검출 네트워크의 성능을 올리기 위해 데이

터 증강 기법을 적용하였다. 훈련 이미지를 확률적으로 회전, 축소, 확대 및 자르기, 좌우 반전, 밝기, 대비, 채도, 색도를 변경하여 정확도를 높인다.

Mask-SSD 훈련은 초기 학습률(learning rate)을 0.001로 설정하고, 옵티마이저로 경사하강법을 사용하여 가중치를 조정하였다. 또한 208 에폭(epoch)이 되었을 경우 학습률에 0.1을 곱하여 학습률을 감소시켜 훈련을 진행하였다. 총 에폭은 418 에폭으로 설정하였고, 배치 크기(batch size)는 8로 설정하여 훈련을 진행하였다.

4.2 성능지표

성능지표로 사용된 정밀도(precision)및 재현율(recall)과 mAP 값을 계산하기 위해 TP(true positive), FP(false positive), FN(false negative) 지표를 사용하였다. TP는 정확하게 정답을 판단한 것을 의미하고, FP는 틀린 것을 정답으로 판단한 것을 의미한다. FN은 정답을 틀렸다고 판단한 것, 즉 탐지하지 못한 것을 의미한다. 성능지표로 사용되는 정밀도와 재현율의 식은 각각 Eq. (1)과 Eq. (2)로 표현할 수 있다. 정밀도와 재현율은 Eq. (3)과 같이 조화평균으로 나타내어 정확도를 나타내는 성능지표인 F1 Score로 표현할 수 있다. 또한, 성능지표로서 정밀도와 재현율 간의 곡선을 단조 적으로 변화시켜 그래프 영역의 넓이를 계산하는 AP(average preciosn)가 있다. 이때 클래스 당 AP의 평균을 구하게 되면 성능지표인 mAP를 구할 수 있다.

$$precision = \frac{TP}{TP + FP} \tag{1}$$

$$recall = \frac{TP}{TP + FN} \tag{2}$$

$$f1\ \ score = \frac{2 * precision * recall}{precision + recall} \tag{3}$$

Where, TP denotes true positive, FP denotes false positive, FN denotes false negative. F1 Score is the mean between rate of change that express the accuracy

4.3 실험결과

Table 3은 4개의 기본 네트워크에 대한 SSD의 결과로 나오는 C(Class), TP, FP, FN, PR(precision, 정밀도), RE(recall, 재현율), F1 score이다. 클래스별 분류에서, M은 마스크 착용자(Masked)에 대한 실험결과이고, U는 마스크 미착용자(UnMasked)에 대한 실험결과이다. 실험은 1,839개의 실험 데이터셋을 이용하여 IoU(Intersection over Union) 값이 임계 값인 0.45보다 큰 경우, 예측한 최소 점수가 임계 점수인 0.5를 넘는 경우, 최대 200개의 정답을 예측하는 환경으로 구성하여 평가를 진행하였다.

Table 3. Comparison of performance between different models

Model	C	TP	FP	FN	PR	RE	F1
VGG-16 SSD	M	974	48	68	0.95	0.93	0.94
	U	1594	75	304	0.96	0.84	0.90
MobileNetV1 SSD	M	818	53	224	0.94	0.79	0.86
	U	825	50	1073	0.94	0.43	0.60
MobileNetV2 SSD-lite	M	823	47	219	0.95	0.79	0.86
	U	793	42	1105	0.95	0.42	0.58
Mask-SSD (This work)	M	911	49	131	0.95	0.87	0.91
	U	1342	91	556	0.94	0.71	0.81

Table 3에서 보는 바와 같이 F1 Score는 MobileNetV1과 MobileNetV2의 점수가 상당히 낮은 것을 확인할 수 있다. 특히 마스크 미착용자에 대한 재현율이 압도적으로 낮은 것을 확인할 수 있다. 이는 정확하게 많은 사람을 판별하지 못했음을 의미한다. Mask-SSD와 MobileNetV1의 F1 Score를 비교하였을 때, 마스크 착용자의 F1 Score는 약 0.05 높고, 마스크 미착용자의 경우 0.21 이상 높은 것을 확인할 수 있다. 또한 MobileNetV2의 F1 Score를 비교하였을 때, 마스크 착용자의 F1 score는 약 0.05 높고, 마스크 미착용자의 경우 약 0.23 높은 것을 알 수 있다.

Table 5는 1,839개의 실험 데이터셋 기반으로 mAP 값과 FPS 값의 성능을 비교하였다. 여기에서는 예측한 최소 점수가 임계 점수인 0.5를 넘는 경우에서 0.01을 넘는 경우로 바꾸어, 200개의 바운딩 박스를 거의 모두 뽑아낼 수 있도록 환경 구성을 변경하였다. FPS 값은 앞서 제시한 PC 환경과 젯슨 나노 임베디드 디바이스 환경에서 초반 30개의 프레임을 제외한 후 200프레임에 대한 평균을 측정한 값이다. 입력 데이터의 크기는 가로세로 모두 300이며 카메라는 HD 화질인 720p로 고정하여 입력 데이터를 받았다. 실험 및 결과를 추출하기 위해 이미지의 크기 조정을 OpenCV의 알고리즘 중 픽셀 영역 사이의 정보를 이용한 보간법인 cv2.INTER_AREA 방법을 사용하였다.

Table 4. Comparison of performance between VGG-based SSD, MobileNetV1-based SSD, MobileNetV2-based SSD-lite and our proposed method.

Model	mAP (%)	PC (FPS)	Jetson Nano (FPS)
VGG-16 SSD	89.4	1.96	2.25
MobileNetV1 SSD	78.15	8.62	9.51
MobileNetV2 SSD-lite	79.04	8.51	11.15
Mask-SSD (This work)	85.2	8.18	6.92

제안한 Mask-SSD 기법의 mAP 값은 VGG-16 기반 SSD 보다 약 4.2% 떨어지나 FPS는 PC 환경에서는 약 4.2배 상승하였고 젯슨 나노 보드에서는 약 3.1배 상승하였다. MobiltnetV1을 기본 네트워크로 구성한 SSD 대비 mAP 값은 제안한 기법이 7.1% 높고, MobileNetV2 SSD-lite 대비하여 6.16% 높은 것을 알 수 있다.

(a)

(b)

Fig. 7. Output images while real time detection. (a) Real time detection result (b) Picutred while real time detect on Jetson Nano

젯슨 나노 보드에 제안한 Mask-SSD를 적용한 결과는 Fig. 7과 같다. 정상적으로 마스크를 착용하지 않고 입장하는 사람(뒤)과 마스크를 착용하고 입장하는 사람(앞)을 정확히 검출하는 것을 확인할 수 있다.

5. 결론

본 논문은 실시간으로 동작하는 딥러닝 기반 마스크 검출기인 Mask-SSD를 제안하였다. 제안된 Mask-SSD는 기존 SSD 대비 추출하는 특징 맵의 수를 적게 구성하고 경계 박스의 개수 또한 절반에 가까운 4,868개를 가지고 객체를 검출하는 특징을 가진다. MobileNetV1 SSD와 MobileNetV2 SSD-lite대비 mAP 값이 높은 것을 확인하였고, VGG-16 기본 네트워크 기반 SSD 대비 약 4배의 FPS 성능 향상을 확인하였다.

향후 연구에서는 CCTV와 같은 저사양 임베디드 장치에서 제안한 마스크 검출기가 실시간으로 동작이 가능하도록 딥러닝 알고리즘을 가속하는 시스템 연구를 진행할 예정이다. 제안한 마스크 검출기는 버스나 지하철 역사와 같이 불특정 다수가 이용하는 공공시설에서 마스크를 정확하게 검출할 수 있다. 또한, 추가 연구로서 실시간으로 교통카드 단말기에 카드를 찍기 전 마스크 착용 여부를 검출하여, 마스크 착용자에 한해 탑승이 가능하도록 하는 방안을 연구할 예정이다.

References

[1] B. Y. Ryu, J. Y. Oh, M. J. Sin, S. H. Kim, I. H. Kim, "Trends and characteristics of SARS-CoV-2 delta mutant virus," Weekly health and disease, Korea Disease Control and Prevention Agency, Republic of Korea, pp. 2354-2365, Aug. 2021.

[2] Q. X. Ma, H. Shan, H. L. Zhang, G. M. Li, R. M. Yang, et al, "Potential utilities of mask-wearing and instant hand hygiene for fighting SARS-CoV-2," Journal of medical virology, Vol.92, No.9, pp.1567-1571, Mar. 2020.
DOI: https://doi.org/10.1002/jmv.25805

[3] C. M. Brown, "Outbreak of SARS-CoV-2 infections, including COVID-19 vaccine breakthrough infections, associated with large public gatherings—Barnstable County, Massachusetts, Jul. 2021, Morbidity and Mortality Weekly Report," Centers for Disease Control and Prevention, pp.1059-1062, Aug. 2021.
DOI: http://dx.doi.org/10.15585/mmwr.mm7031e2

[4] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, et al, "SSD: Single Shot Multibox Detector," European conference on computer vision, Springer Cham, Amsterdam, Netherland, pp 21-37, Oct. 2016.
DOI: https://doi.org/10.1007/978-3-319-46448-0_2

[5] J. Redmon, S. Divvala, R. Girshick, A.Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proceedings of the IEEE conference on computer vision and pattern recognition, IEEE, Las Vegas, USA, pp.779-788, Jun. 2016.
DOI: https://doi.org/10.1109/CVPR.2016.91

[6] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, Jun. 2017.
DOI: https://doi.org/10.1109/TPAMI.2016.2577031

[7] K. He, G. Gkioxari, P. Dollár, R. Girshick, "Mask R-CNN," Proceedings of the IEEE international conference on computer vision, Venice, Italy, pp. 2980-2988, Oct. 2017.
DOI: https://doi.org/10.1109/ICCV.2017.322

[8] P. Soviany and R. T. Ionescu, "Optimizing the Trade-Off between Single-Stage and Two-Stage Deep Object Detectors using Image Difficulty Prediction," 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pp. 209-214, 2018.
DOI: https://doi.org/10.1109/SYNASC.2018.00041

[9] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," International Conference on Learning Representations 2015, San Diego, USA, May. 2015.

[10] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, et al, "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, Apr. 2017.

[11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. C. Chen, "MobileNetv2: Inverted Residuals and Linear Bottlenecks," Proceedings of the IEEE conference on computer vision and pattern recognition, Utah, USA, pp.4510-4520, Dec. 2018.
DOI: https://doi.org/10.1109/CVPR.2018.00474

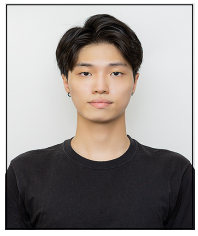
[12] Alex Krizhevsky, "Convolutional deep belief networks on cifar-10," Unpublished manuscript, University of Toronto, pp.1-9, 2010.

[13] T. Y. Lin, M. Maire, S. Belongie, J. Hays, et al, "Microsoft COCO: Common Objects in Context," European conference on computer vision, Springer Cham, Zurich, Switzerland, pp.740-755, Sep. 2014.
DOI: https://doi.org/10.1007/978-3-319-10602-1_48

[14] Valentin Schwind, "The golden ratio in 3D human face modeling," Stuttgart Media University, pp.1-4, 2011

[15] D. Chiang, "Detect faces and determine whether people are wearing mask," Available From: https://github.com/AIZOOTech/FaceMaskDetection (accessed July 11, 2021)

강 태 운(Tae-Woon Kang) [준회원]



- 2019년 3월 ~ 현재 : 상명대학교 시스템반도체공학과 (학사)

〈관심분야〉
딥러닝, FPGA

김 용 우(Yongwoo Kim) [정회원]

- 2009년 2월 : 인하대학교 전자공학과 (석사)
- 2017년 8월 : (주)LX세미콘 선임연구원
- 2019년 2월 : 한국과학기술원 전기및전자공학부 (박사)
- 2020년 3월 : 한국항공우주연구원 선임연구원
- 2020년 3월 ~ 현재 : 상명대학교 시스템반도체공학과 조교수

〈관심분야〉
딥러닝, 영상처리, 디지털 시스템