논문 21-46-03-13 The Journal of Korean Institute of Communications and Information Sciences '21-03 Vol.46 No.03
https://doi.org/10.7840/kics.2021.46.3.526

딥러닝 기반 신호등 인식 시스템에 대한 특수 신호등의 영향 평가

김 형 석◆, 한 영 주*, 박 준 상°

Impacts of Special Traffic Lights on Deep Learning Based Traffic Light Recognition Systems

Hyeong-seok Kim◆, Young-Joo Han*,
Joon-Sang Park°

요 약

신호등 인식 기능은 자율주행 차량 및 첨단 운전자 보조 시스템(ADAS: Advanced Driver Assistance Systems)의 필수 구성요소이다. 교통 신호등은 국제적으로 통용되는 일반적인 체계가 존재하지만, 특정 국가에서만 사용되는 특수 신호등이 존재하므로 이에 대한 고려가 필요하다. 우리나라의 경우 버스 전용 차로를 위한 버스 전용 신호등을 사용하고 있다. 본 연구에서는 YOLOv3 계열의 딥러닝 모델을 이용해 신호등 인식 기능을 구현할 때 버스 전용 신호등이 신호등 인식률에 미치는 영향을 분석한다. 즉, 버스 전용 신호등을 포함하는 학습 데이터셋을 만들어 이를 이용하여 학습을 수행한 신호등 인식 시스템의 인식률을 살펴보고, 버스 전용 신호등이 상대적으로 낮은 인식률을 보이는 요인을 분석한다.

Key Words : Traffic light recognition, Deep learning, Object detection, Image classification, Autonomous driving

ABSTRACT

Traffic light recognition is an essential task in autonomous driving and ADAS(ADAS: Advanced Driver Assistance Systems). Traffic light systems are similar in most countries but there exist special traffic signals adopted to specific countries which must be taken care of. In case of Korea, bus priority traffic lights are being for bus-only lanes. In this study, we analyze the impact of bus priority traffic lights on the performance of a traffic light detection system based on YOLOv3 deep learning model. That is, we evaluate the traffic light detection system trained with a dataset containing bus priority traffic lights which we have collected and discuss reasons behind comparatively low performances in detecting bus priority traffic lights.

Ⅰ. 서 론

최근 들어 첨단 운전자 보조 시스템(ADAS)과 자율주행 차량의 사용이 일반화되고 있는데 이는 다양한 교통 상황, 기후, 주변 환경에서도 신뢰성 있는 성능을 담보할 수 있는 딥러닝 기반 알고리즘의 적용[1-5]에 기인함이 크다. ADAS 및 자율주행 시스템을 구성하고 있는 요소기술들 중 차선, 보행자, 및 신호등 등의 주변 환경을 인식(perception)하는 인식 기술은 안전 운행의 핵심이라 할 수 있고 딥러닝 기반 알고리즘들이 가장 많이 활용되고 있는 요소기술이라 할 수 있다. 본 연구에서는 이러한 인식 기술들 중 신호등 인식 기술에 초점을 맞춘다.

교통 신호등은 다수의 국가에서 통용되는 국제 표준[6]이 존재하지만, 특정 국가에서만 사용되는 특수 신호등 또한 존재한다. 우리나라의 경우, 중앙 버스 전용 차로의 존재로 인해 버스 전용 신호등이 일부 교차로에서 운영되고 있는데, 버스 전용 신호등은 원형이 아닌 버스 형상을 사용하기에, 일반 신호등에 최적화된 신호등 인식 시스템의 인식률이 영향을 미칠 수 있고 그 영향이 어느 정도 인지 살펴볼 필요가 있다.

이를 위해, 본 연구에서는 버스 전용 신호등과 일반 차량용 신호등이 모두 존재하는 학습 데이터셋을 제작하고, 이 학습 데이터셋을 이용하여 실시간성

※ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2019R1F1A1060357).
◆ First Author : Hongik University Computer Engineering Department, gud2169@naver.com, 학생, 정회원
° Corresponding Author : (ORCID:0000-0002-6459-1060)Hongik University Computer Engineering Department, jsp@hongik.ac.kr, 정교수, 종신회원
* Vieworks Co., Ltd, orora71@gmail.com, 연구원
논문번호 : 202012-313-C-LU, Received December 9, 2020; Revised January 4, 2021; Accepted January 7, 2021

(real-time) 딥러닝 모델인 YOLOv3[7] 계열 모델 기반의 신호등 인식 시스템을 학습시켜 버스 전용 신호등이 신호등 인식 시스템의 성능에 미치는 영향을 분석해 본다. 먼저, YOLOv3-tiny 신경망 모델을 이용하여 일반 신호등과 버스 전용 신호등의 인식률을 비교하고, 버스 전용 신호등이 낮은 인식률을 보이는 요인을 분석한 한 후, 이 요인이 다른 YOLOv3 계열 신경망 모델에서도 성능 저하의 동일한 요인으로 작용하는지를 살펴본다.

Ⅱ. 관련 연구

딥러닝을 이용해 신호등과 같은 객체를 인식하는 연구는 매우 활발히 이루어져 왔다. 딥러닝을 사용한 객체 인식 연구는 크게 영역 제안 기반 방법과 회귀/분류 기반 방법으로 나눌 수 있다[8].

영역 제안 기반 방법은 객체가 있을 것으로 예상하는 영역을 영역 제안 네트워크(region proposal network)를 통해 추출한 후, 추출한 영역에 대해 분류 및 지역화를 수행하여 원하는 객체를 검출하는 방법이다. 영역 제안 기반 방법을 사용한 연구로는 R-CNN[9], SPP-NET[10], FRCN[11], Faster R-CNN[12], Mask R-CNN[13] 등이 있다.

회귀/분류 기반 방법은 영상을 격자(grid) 모양의 영역으로 분할한 후, 각 영역별로 객체를 예측한다. 영역별로 분류를 수행하여 객체 인식 문제를 단일 회귀 식으로 단순화함으로써, 영역 제안 기반 방법에 비해 빠른 수행속도를 갖는다. 회귀/분류 기반 알고리즘에는 AttentionNet[14], YOLO[15], SSD[16], YOLOv2[17], YOLOv3 등이 있다.

신호등 인식의 경우, 높은 신뢰도와 실시간성을 위한 빠른 수행속도를 만족해야 한다. 이러한 이유로 본 연구에서는 빠른 수행속도를 갖는 회귀/분류 기반 알고리즘 중 높은 성능을 보이는 YOLOv3 모델을 이용하여 실험을 수행하였다.

Ⅲ. YOLOv3

YOLOv3는 앞서 설명한 것과 같이, 영상을 격자 모양의 영역으로 분할한 후, 각 격자의 중심으로부터 추정한 경계 영역(bounding box, anchor box)에 대한 클래스별 신뢰도를 산출함으로써 원하는 물체를 인식한다. 이러한 원리를 통해, 기존 물체 인식을 위한 네트워크에 비해 획기적으로 빠른 속도를 보였다. YOLO는 2016년 초기 버전을 발표한 이후, 지속적으로 성능이 개선된 버전을 발표하고 있다. YOLOv2는 기존 YOLO와 달리 배치 정규화(batch normalization) 방법과 앵커박스(anchor box)를 사용하여 학습의 속도와 안정성을 증가시켰다. YOLOv3는 ResNet[18]에서 제안한 스킵 연결(skip connection)구조를 사용해 이전보다 더 깊은 네트워크를 훈련 시켰으며, 기존 모델들과 비슷한 성능을 보였지만, 기존 모델들에 비해 빠른 수행속도를 갖는다는 장점이 있다.

본 연구에서는 YOLOv3 모델뿐 아니라, 함께 제안된 YOLOv3-tiny, YOLOv3-spp 모델[19]을 신호등 인식 문제에 적용하였다. 영상을 입력으로 받아서, 각 클래스에 대한 예상 위치와 신뢰도 등을 출력해야 하는 영상 내 객체 인식 네트워크는 네트워크 내에서 다양한 방법을 통해 차원을 축소하며, 유의미한 정보를 추출한다. YOLOv3 모델은 특정 컨볼루션 레이어들의 보폭(stride)을 2로 설정하여 차원을 축소하는 과정을 거치지만, YOLOv3-tiny 모델은 파라미터의 수를 줄여서, 빠른 수행 속도를 갖기 위해 최대 풀링(max pooling)레이어를 사용한다는 점에서 차이를 보인다. YOLOv3-spp 모델은 SPP(Spatial Pyramid Pooling) 기법[20]을 사용하여 다양한 상황에서 객체 검출이 용이하게 하였다. SPP 기법은 최대 풀링 레이어를 다양한 크기로 적용함으로써 다양한 크기의 객체를 효율적으로 검출[21]할 수 있도록 돕는다.

Ⅳ. 실험 및 결과

본 장에서는 버스 전용 신호등이 딥러닝 기반 신호 인식 시스템에 미치는 영향을 분석하기 위해 YOLOv3, YOLOv3-tiny, YOLOv3-spp 모델을 이용해 수행한 실험에 관해 설명한다.

4.1 데이터셋 및 학습

버스 전용 신호등이 딥러닝 기반 알고리즘에 미치는 영향을 분석하기 위해서는 일반 신호등과 버스 전용 신호등이 포함되고, 구분되어서 레이블링 된 데이

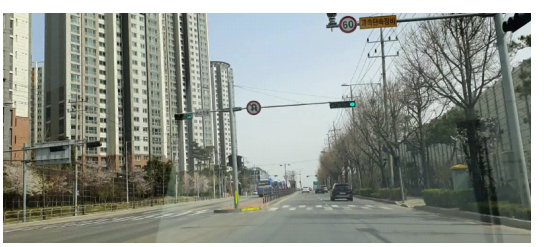

그림 1. 학습용 데이터셋의 이미지 샘플
Fig. 1. Sample image in training dataset

터셋이 필요하다. 이를 위해 그림 1과 같은 차량의 블랙박스를 통해 촬영된 영상을 직접 획득한 후, 레이블링 과정을 거쳐 실험을 위한 데이터셋으로 사용하였다. 레이블링은 YOLO_mark[22]를 이용하였다. 데이터셋은 2288×1080의 크기를 갖는 영상 1024장으로 이루어지며, 학습을 위한 데이터셋은 714장, 성능 평가를 위한 데이터셋은 310장의 영상을 사용하였다. 데이터셋은 일반 신호등과 버스 전용 신호등 두 가지 클래스로 이루어진다. 학습에는 0.001의 학습률(learning rate), 가중치 감쇠(weight decay)는 0.0005, 배치 크기(batch size)는 64를 사용하였다. 인식률 평가를 위해 사용한 지표인 AP(average precision)는 정밀도-재현율 곡선(precision-recall curve)을 통해 산출하는 클래스별 검출 성능을 나타내며, mAP(mean average precision)는 클래스별 AP의 평균을 나타낸다. F1 score는 정밀도와 재현율의 조화평균이다.

4.2 신호등 유형 별 인식 성능 비교

먼저 버스 전용 신호등이 신호등 인식 시스템의 성능에 미치는 영향과 그 요인을 살펴보기 위해 YOLOv3-tiny 모델을 사용하여 각각 416×416, 1024×1024 두 가지 크기의 입력 레이어(input layer)를 갖도록 설정하여 학습한 후 인식 성능을 비교하였다. 대부분의 딥러닝 기반 객체 인식 시스템에서는 제한된 컴퓨팅 자원의 효율적 활용을 위해 원본 영상이 크기 재조정(resizing) 과정을 거친 후 입력 레이어에 전달된다. 입력 레이어 크기를 증가시키면 신경망 모델 전체의 파라미터 수 또한 증가하기에 필요로 하는 시스템 메모리의 크기가 증가한다. 표 1에서 보이는 바와 같이 입력 레이어의 크기가 작은 경우, 버스 전용 신호등(Bus priority traffic signal)의 AP는 일반 신호등(Normal traffic signal)에 비하여 매우 낮은 수치를 보인다. 이에 반하여 입력 레이어의 크기가 작은 큰 경우 그 차이가 미미하다. 이를 통해 입력 영상의 크기 재조정이 버스 전용 신호등의 인식률에 큰 영향을 미치는 것을 알 수 있고 영상의 크기 재조정 과정에서 버스 전용 신호등이 영상 고유의 특성을 잃어버리는 현상이 일어나고 있음을 추정할 수 있다. 입력 영상의 낮은 해상도에 의한 딥러닝 기반 인식 알고리즘의 성능 저하 문제는 이전 연구들[23,24]에서 관찰되어왔다. 그림 2는 재조정을 거친 신호 인식 알고리즘의 버스 전용 신호등 입력 영상 예시이다. 2280×1080 크기의 원본 영상이 416×416으로 크기가 재조정된 경우, 1024×1024로 크기가 재조정된 경우에 비해 영상의 많은 부분이 소실되어서 버스 전용 신호등 영상 고유의 특성을 잃어버리는 것을 관찰할 수 있다. 마지막으로 표 1에 제시된 결과를 보면, 입력 레이어 크기를 증가시키면 mAP뿐 아니라 F1 score도 증가함을 알 수 있다.

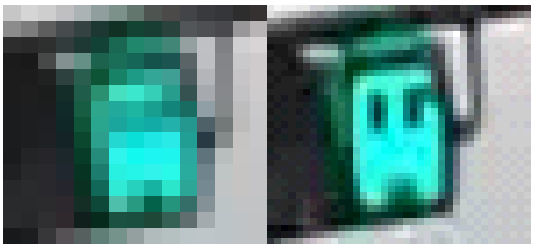

그림 2. 크기 재조정을 거친 버스 전용 신호등 입력 이미지 (좌: 416×416, 우: 1024×1024)
Fig. 2. Sample input image of bus priority signal after resizing (Left : 416×416, Right: 1024×1024)

4.3 신경망 모델별 인식 성능 비교

입력 레이어의 크기를 증가시키는 방법은 인식 성능을 향상시킬 수 있으나 신경망 모델 전체의 파라미터 수 또한 증가하기에, 시스템 메모리가 제한된 환경에서는 신경망 모델에 따라 그 적용이 제한될 수 있다. 즉, 상대적으로 단순한 YOLOv3-tiny와 달리 파라미터가 많은 YOLOv3와 YOLOv3-spp 모델의 경우, 1024×1024와 같이 큰 크기의 입력 레이어를 사용하는 시스템은 하드웨어적인 제약으로 인해 그 구축이 어려운 경우가 존재한다. 이에, 본 연구에서는 YOLOv3-tiny뿐 아니라 YOLOv3, YOLOv3-spp 모델 기반 신호등 인식 시스템의 성능에 대한 버스 전용 신호의 영향 평가와 성능 저하 요인 파악을 위해 영상

표 1. YOLOv3-tiny의 입력 레이어 크기 별 성능
Table 1. Performance of YOLOv3-tiny with different input layer sizes

Model	Original Image Size	Input Layer Size	AP		mAP	F1 score
			Normal traffic signal	Bus priority traffic signal		
YOLOv3 tiny	2280×1080	416×416	70.83%	18.53%	44.68%	0.70
	2280×1080	1024×1024	86.66%	85.30%	85.98%	0.78

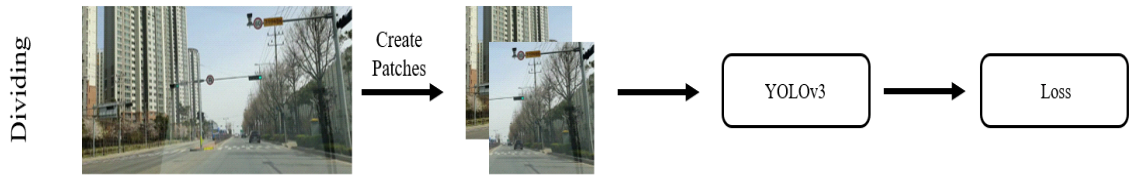

그림 3. YOLOv3로의 분할된 영상의 입력
Fig. 3. Input of partitioned image to YOLOv3

분할 기법을 사용하였다. 여기서 영상 분할 기법은 그림 3과 같이 원본 하나의 영상 프레임(frame)을 서로 겹치지 않는 동일한 크기의 2개 영상 프레임으로 분할한 후, (학습 시와 평가 시에 모두) 각각의 영상을 신호 인식 시스템의 입력 영상으로 사용하는 방법을 의미한다. 즉, 2280×1080 크기의 영상 프레임을 2개의 1240×1080 영상 프레임으로 분할한 후 각각을 신경망 모델에 입력한다. 영상 분할 기법을 통해 신경망 모델의 파라미터 수의 변화 없이 입력 영상 해상도를 약 2배 증가시키는 효과를 얻을 수 있고 이를 통해 입력 영상 크기 재조정의 신호등 인식률에 대한 영향을 살펴볼 수 있다.

표 2에서는 3가지 YOLOv3 계열 신경망 모델에 대해 서로 다른 크기의 입력 영상을 사용할 때의 인식 성능을 비교하였다. 이 때, YOLOv3-spp와 YOLOv3의 경우, 416×416의 한가지 입력 레이어 크기를 갖도록 하였고, YOLOv3-tiny의 경우, 416×416와 1024×1024의 2가지 입력 레이어 크기를 갖는 경우에 대해 살펴보았다. 표에서 Original Image Size는 신경망 모델에 입력되는 원본 영상의 크기를 나타내는데, 영상 분할 기법을 적용하지 않은 경우, 2280×1080의 해상도를 갖는 원본 영상이 416×416로 크기가 재조정되고, 영상 분할 기법을 적용한 경우 (표에서 partitioned image로 표시된 행에 해당) 1240×1080의 영상이 416×416로 크기가 재조정된다. 표에서 보이는 바와 같이 영상 분할 기법을 적용하여 높은 해상도의 입력 영상을 사용한 효과를 추구한 경우, 모든 신경망 모델에 대하여 일반 신호등의 AP가 평균적으로 약 1.18배 증가하였고, 버스 전용 신호등의 AP는 평균 약 1.74배 증가하였다. 특히, YOLOv3-spp의 경우 버스 전용 신호등의 AP가 약 2.5배 증가하였다. 즉, 입력 영상 해상도의 변화에 (또는 크기 재조정 비율에)

표 2. 뉴럴 네트워크의 영상 분할 기법 사용 여부 별 성능 비교
Table 2. Performance comparison of the neural networks with using image dividing method

Model	Original Image Size	Input Layer Size	AP	AP	mAP	F1 score
			Normal traffic signal	Bus priority traffic signal		
YOLOv3	2280×1080	416×416	74.06%	35.52%	54.79%	0.76
	1240×1080 (partitioned image)	416×416	92.12%	76.67%	84.39%	0.83
YOLOv3 spp	2280×1080	416×416	78.88%	21.68%	50.28%	0.74
	1240×1080 (partitioned image)	416×416	91.37%	75.93%	83.65%	0.81
YOLOv3 tiny	2280×1080	416×416	70.83%	18.53%	44.68%	0.70
	1240×1080 (partitioned image)	416×416	87.36%	37.09%	62.22%	0.73
YOLOv3 tiny	2280×1080	1024×1024	86.66%	85.30%	85.98%	0.78
	1240×1080 (partitioned image)	1024×1024	94.76%	91.04%	92.90%	0.90

따른 인식 성능 향상이 일반 신호등에 비하여 버스 전용 신호등의 경우에서 두드러지게 나타났고, 이를 통해 입력 영상의 해상도가 YOLOv3 모든 모델에 대해 버스 전용 신호등 인식 성능에 많은 영향을 미치고 있음을 알 수 있다. 신호등 유형별 구분 없이 인식 성능을 살펴보면, mAP는 평균 약 1.37배 증가하였고 F1 score 역시 평균 약 1.09배 증가하였다. 가장 좋은 인식 성능을 보인 경우는 YOLOv3-tiny 모델을 사용하여 입력 이미지 크기 재조정 비율을 최소화 한 경우, 즉, 화면 분할 기법을 사용하여 입력 영상의 크기를 1240×1080으로 하고 입력 레이어의 크기를 1024×1024로 한 경우로, 91.04%의 mAP와 0.90의 F1 score를 보였다.

V. 결 론

본 연구에서는 버스 전용 신호등이 YOLOv3 기반 신호등 인식 시스템의 성능에 미치는 영향을 분석하고, 성능 저하를 일으키는 요인을 살펴보았다. 이를 통해 YOLOv3 계열의 딥러닝 모델을 사용하는 신호등 인식 시스템에서는 입력 영상의 해상도 및 입력 레이어 크기에 따라 인식 성능이 큰 영향을 받고 높은 해상도를 갖는 입력 영상을 사용하도록 설정함으로써 버스 신호등의 인식 성능을 높일 수 있음을 알 수 있었다.

추후 연구는 다양한 기상 및 조도 환경에서 신호 인식 시스템이 올바르게 동작하는지 확인하고, 이에 대한 해결책을 제시하는 것이다.

References

[1] H. Xu, Y. Gao, F. Yu, and T. Darrell, "End-to-end learning of driving models from large-scale video datasets," in *Proc. IEEE Conf. CVPR*, pp. 2174-2182, Honolulu, Hawaii, Jul. 2017.

[2] H. M. Eraqi, M. N. Moustafa, and J. Honer. "End to-end deep learning for steering autonomous vehicles considering temporal dependencies," *arXiv preprint, arXiv:1710.03804*, Oct. 2017.

[3] S. Hecker, D. Dai, and L. Van Gool, "End-to-end learning of driving models with surround-view cameras and route planners," in *Proc. ECCV*, pp. 435-453, Munich, Germany, Sep. 2018.

[4] V. Rausch, A. Hansen, E. Solowjow, C. Liu, E. Kreuzer, and J. K. Hedrick. "Learning a deep neural net policy for end-to-end control of autonomous vehicles," in *Proc. ACC*, pp. 4914-4919, Seattle, USA, May 2017.

[5] C. Park and G. Cheol, "Implementation of autonomous driving system in the intersection area equipped with traffic lights," *J. Korean Soc. Automotive Eng.*, vol. 27, no. 5, pp. 379-387, May 2019.

[6] Wikipedia, *Vienna Convention on Road Traffic*(2018), Retrieved Dec. 01, 2020, from https://en.wikipedia.org/wiki/Vienna_Convention_on_Road_Traffic

[7] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement" *arXiv preprint, arXiv:1804.02767*, Apr. 2018.

[8] Z.-Q. Zhao, et al., "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. and Learn. Syst.*, vol. 30, no. 11, pp. 3212-3232, Jan. 2019.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. CVPR*, pp. 580-587, Columbus, Ohio, Jun. 2014.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904-1916, Jan. 2015.

[11] R. Girshick, "Fast R-CNN," in *Proc. IEEE ICCV*, pp. 1440-1448, Santiago, Chile, Dec. 2015.

[12] S. Ren, et al., "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149, Jun. 2016.

[13] K. He, et al., "Mask R-CNN," in *Proc. IEEE ICCV*, pp. 2961-2969, Venice, Italy, Oct. 2017.

[14] D. Yoo, S. Park, J.-Y. Lee, A. S. Paek, and I. S. Kweon, "Attentionnet: Aggregating weak

directions for accurate object detection," in *Proc. IEEE Conf. CVPR*, pp. 2659-2667, Boston, Massachusetts, Jun. 2015.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. CVPR*, pp. 779-788, Las Vegas, Nevada, Jun. 2016.

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Proc. ECCV*, pp. 21-37, Amsterdam, Netherlands, Oct. 2016.

[17] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proc. IEEE Conf. CVPR*, pp. 7263-7271, Honolulu, Hawaii, Jul. 2017.

[18] K. He, et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. CVPR*, pp. 770-778, Las Vegas, Nevada, Jun. 2016.

[19] GitHub, *AlexyAB/darknet*, Retrieved Dec. 01, 2020, from https://github.com/AlexeyAB/darknet.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904-1916, Jan. 2015.

[21] Z. Huang, et al., "DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection," *Inf. Sci.*, vol. 522, pp. 241-258, Jun. 2020.

[22] GitHub, *AlexyAB/YOLO_mark*, Retrieved Dec. 01, 2020, from https://github.com/AlexeyAB/Yolo_mark

[23] C. F. Sabottke and B. M. Spieler, "The effect of image resolution on deep learning in radiography," *Radiology: Artificial Intell.*, vol. 2, no. 1, Jan. 2020.

[24] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *arXiv preprint, arXiv:1905.11946*, Sep. 2020.