# TensorRT 엔진과 SSD를 이용한 Face detection

유혜빈*, 김상훈*
*한경대학교 전기전자제어공학과
e-mail: kimsh@hknu.ac.kr

# Objedet detection using TensorRT engine and SSD

Hye-Bin Yoo*, Sang-Hoon Kim*
*Dept of Electrical, Electronic and Control, Hankyong National University

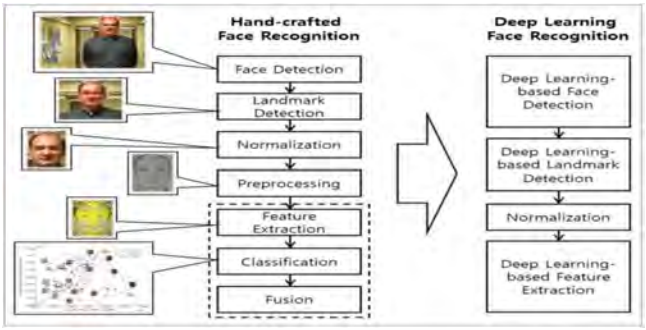
요 약

최근에는 딥러닝 기술의 발달로 물체 인식 및 검출에 관한 기술들 또한 발달하고 있다. 검출에 관한 여러 기법(Faster R-CNN, R-CNN, YOLO, SSD 등) 중 SSD는 다른 기법들과는 다르게 높은 정확도와 빠른 속도가 특징이다. 동시에 여러 detection network들도 쉽게 이용이 가능하다.

본 논문에서는 detection netowork중 Mobilenet V2 network를 이용하여 SSD와 결합해 모델을 훈련하고, TensorRT engine을 이용하여 더 빠른 속도로 검출할 수 있는 방법에 대해 논의한다. 이 방법을 통해 face detector를 만들어 여러 상황에서 쓰일 수 있도록 한다.

## 1. 서론

딥러닝 기술의 발달 [1]로 인해 Computer vision 방식들에도 많은 변화가 생기고 있다. 특히 얼굴 인식 분야에서 기존의 방식인 Hand-crafted Featrue인 HOG, LBP, GABOR등의 특징들이 모두 딥러닝 기반의 특징으로 바뀌었고 얼굴 검출에서도 Viola-hones의 Haar-like Featrue를 Boosting하는 방식에서 딥러닝 기반의 방식으로 바뀌며 계속해서 성능을 개선하고 있다.



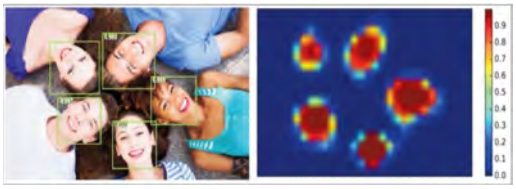

(그림 1) 얼굴 인식 기술의 변화

## 2.본론

### 2.1. 현재 기술의 한계

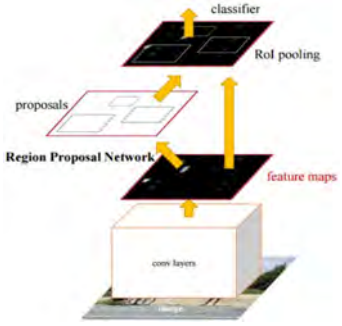
얼굴 검출 관련 딥러닝 기술은 ICMR에서 2015년도에 기본적인 AlexNet [4]을 기반으로 한 얼굴 검출기가 발표되었다. 이 당시에는 AlexNet을 얼굴 영상으로 Tine-tuning하여 [그림 2]와 같은 결과를 산출하였지만, 이때까지는 검출 성능이 높지 않아 최종 단에 SVM을 이용하여 얼굴 유무를 최종 판단하였다.

AlexNet을 이용한 얼굴 검출 방식을 제외하고도 고속 물체 검출기인 YOLO를 fine-tuning한 얼굴 검출기 및 Faster R-CNN을 fine-tuning한 얼굴 검출 알고리즘들이 속속 소개되고 있다.



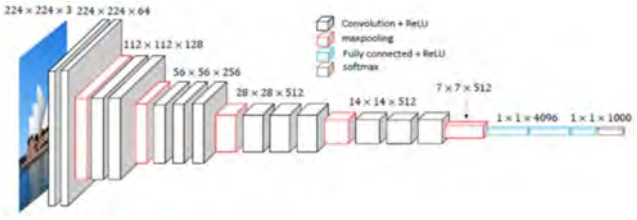
(그림 2) AlexNet

이 중에서도 가장 많이 사용되었던 대표적인 detector는 Faster R-CNN이다. 이 Faster R-CNN은 Detecting을 위해 들어온 image의 후보 영역을 뽑아 뽑은 부분의 특징이나 픽셀을 resampling하고 높은 성능의 classifier를 이용한다.[그림 3] 하지만 Faster R-CNN은 전작인 R-CNN을 열심히 개선했음에도 불구, 속도가 느려(7 FPS with maP 73.2%) 실시간 영상 분석에 사용할 수는 없었다. Faster R-CNN에 비해 YOLO의 속도는 빠른 편이었지만, 그만큼 성능이 낮았다.(45FPS with mAP 63.4%)




(그림 3) Faster R-CNN

보편적으로 사용되고 있는 다른 Deep-learning Detector들은 처음 훈련 시킨 크기만을 입력으로 받을 수 있다. 대표적으로 VGG-16, AlexNet에서 줄 224x224 크기의 이미지를 입력으로 받는다.[그림 4] 즉 원하는 사진에서 객체가 있는지 없는지 확인하기 위해서는 그림을 224x224로 자르거나 변형해야 한다.




(그림 4) VGG-16 architecture

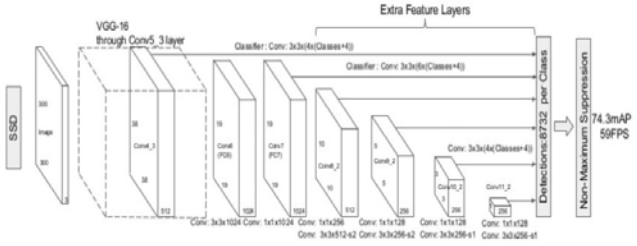
이러한 과정을 위해 Image Pyramid & Sliding Window라던가 Region Proposal Network(Faster R-CNN)등으로 입력 이미지를 변형시켜 네트워크에 집어넣게 된다. 문제는 위의 처리과정을 통해 얻은 여러가지 sample들을 하나씩 network에 넣어 하나씩 검출해야 한다는 것이다. 여러 장의 정보를 처리해야 하기 때문에 그만큼 네트워크를 많이 돌게 되고, 횟수 차이에 의한 속도 저하가 일어나게 된다.

### 2.2. 관련연구

Single shot multibox detector(SSD) [2]는 Single-shot detector라는 말 그대로 사진의 변형 없이 그 한 장으로 훈련 및 검출을 하는 detector를 의미한다. SSD는 후보 영역 추출 과정과 resampling 과정을 제거한 방식을 이용하여 높은 정확성과 빠른 속도를 모두 얻어냈다. (59FPS with mAP 74.3%) (위의 모든 성능 측정은 VOC2007 test 기반)

SSD는 전부 새로 만든 구조가 아니다. 원래 잘 만들어졌단 feed-forward convolutional network에서 feature map을 뽑아내는 과정까지를 하나의 기본 구조로 가지고, 여러 보조적인 몇 가지 구조만을 추가한 것이다.

하지만 single-shot learning을 위해서는 한 가지 큰 문제를 해결해야 한다. 단 한 장의 사진만을 가지고 여러 가지 크기의 물체를 검출해야 한다는 것이다. SSD는 이러한 문제를 기본 구조 뒤에 보조 구조를 붙여 얻은 Multi-scale feature maps를 이용하여 해결하였다. [그림 5]가 SSD의 architecture이다. 기본 구조나 보조 구조에서 얻은 feature map들은 각각 다른 convoltuional filter에 의해 결과값을 얻게 된다.

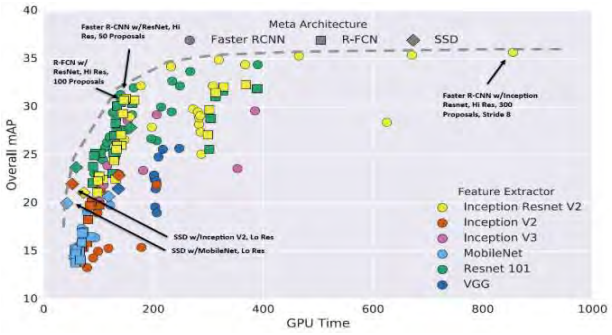



(그림 5) SSD architecture

속도를 제외한 SSD의 가장 큰 장점은 다른 어떠한 detection network이던 single-shot learning에 이용할 수 있다는 것이기에 real-time detector 연구에 용이하다고 볼 수 있다.

## 3. SSD Mobilenet V2

single-shot learning은 detection network 종류에 상관없이 이용이 가능하다고 앞에서 언급하였다. SSD의 장점은 높은 속도와 정확도인데 이를 활용하기 위해 detection network 또한 높은 속도와 정확도를 가지고 있다면 그 효율은 뛰어날 것이다. 이와 같은 이유로 detection netowrk는 Mobilenet v2를 선정하였다.




(그림 6) Accuarcy vs Time

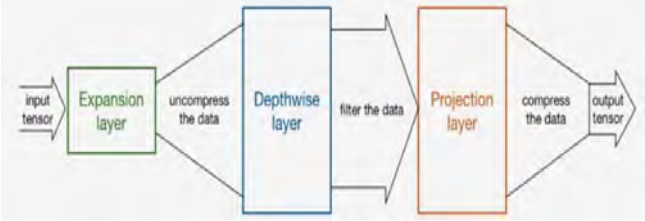
[그림 6]에서 볼 수 있듯이 SSD를 이용한 모델 중 가장 속도가 빠른 것은 Mobilent network임을 확인할 수 있다.


(a)


(b)

(그림 7) ssd의 mobilenet_v2(a)와 inception_v2(b) 속도

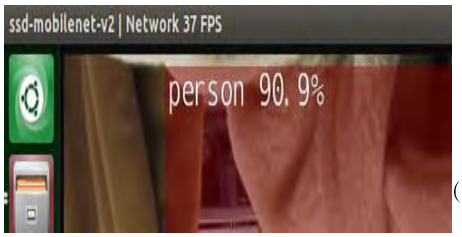
Mobilenet v2 [3]는 Mobilenet V1의 속도 저하 원인이었던 PW의 부담을 인식하고 DW 연산 비중을 올리는 테크닉을 사용한다. Expansion Layer, Projection Layer가 추가되었고, 그 중간에 DW가 존재한다. 즉 ,Expansion Layer PW에서 Channel을 늘려준 상태에서 DW를 한다. Projection Layer에서는 원래의 Channel 개수로 줄여주는 역할을 해가 된다. 즉, Channel을 기준으로 Expansion하고 Projection한다는 것이다.




(그림 8) Mobilenet V2 기술

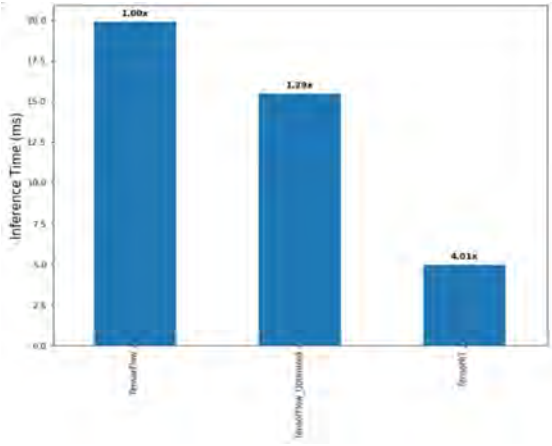
## 4. 실험 및 고찰

coco dataset으로 훈련된 ssd_mobilenet_v2를 Jetson TX2 임베디드 보드에서 실시간 영상으로 실험해본 결과 [그림 9]와 같이 높은 정확도와 빠른 속도를 보여주었다.



(그림 8)

실시간 검출 속도가 평균 35~37 FPS정도인데, 이는 TensorRT 최적화 과정을 거친 것으로 이 과정을 거치지 않으면 10FPS 안팎의 속도를 보이게 된다. [그림 9]를 보면 알수 있듯이 ssd_mobilenet_v2를 tensorflow로만 이용하는 것을 기준으로 보았을 때 TensorRT 엔진을 사용하게 되면 최대 4배에 가까운 성능을 내는 것을 알 수 있다.



(그림 9) TensorRT 최적화 결과

## 5. 결론 및 향후 연구계획

본 논문은 더 효과적인 detection을 하기 위한 연구이다. 기존 Tensorflow만을 이용한 detection은 속도가 느려 제한된 상황에서만 쓰일 수 있었다. TensorRT engine을 이용하여 detection을 하게 된다면 제한된 상황에서도 보다 뛰어난 성능으로 detection이 가능해지게 된다.

향후 연구는 SSD를 이용하여 mobilenet v2 네트워크를 TensorRT 엔진을 사용하여 사람의 얼굴에 관한 detection을 진행할 예정이다. 현재 사람에 관한 것은 'person'이라는 class만 기존 훈련된 네트워크에 있고, 이를 얼굴에 관한 detection으로 더 구체화하여 성별, 인종, 연령대 등으로 검출할 수 있게 하는 것이 최종 목표이다.

## 참고문헌

[1] 황원준 - 딥러닝 기반 얼굴 검출, 랜드마크 검출 및 얼굴 인식 기술 연구 동향

[2] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg1/UNC Chapel Hill Zoox Inc. Google Inc. University of Michigan, Ann-Arbor - SSD: Single Shot MultiBox Detector

[3] Mark Sandler, Andrew Howard Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen/Google Inc. - MobileNetV2 : Inverted Residuals and Linear Bottlenecks

[4] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton - ImageNet Classification with Deep Convolutional Neural Networks