시각장애인을 위한 편의점 제품 인식 애플리케이션

한상혁, 박다수, 임채민, 정지운
한국산업기술대학교 컴퓨터공학부
contea95@kpu.ac.kr, 2016140016@kpu.ac.kr, limkim4120@kpu.ac.kr,
2016150036@kpu.ac.kr

Convenience Store Product Recognition Application for the Blind

Han Sang Hyeok, Park Da Soo, Lim Chae Min, Jeong Ji Woon
Department of Computer Engineering, Korea Polytechnic University

요 약

본 논문은 딥러닝 학습을 통한 객체(편의점제품) 인식 시스템을 소개한다. 편의점 내에서 시각장애인의 접근성인 매우 떨어지고 있다. 그나마 점자가 있는 제품은 음료수 제품이지만 제품 이름이 아닌 범주로 표현하고 있어 원하는 제품 구매를 어렵게 한다. 본 논문에서는 YOLOv5를 통한 딥러닝 학습을 사용하여 정확한 제품을 시각장애인에게 제공할 수 있는 애플리케이션을 개발했다. 사용한 학습데이터 세트는 제품을 직접 찍어 확보했으며, 국내 11개 제품을 포함한다. 학습데이터 세트는 총 23,814장을 사용했으며, 결과 정확도를 나타내는 mAP_0.5:0.95 는 약 0.9790의 성능을 보였다.

1. 서론

최근 객체 인식에 대한 딥러닝 분야가 활성화되면서 전 세계에서 많은 연구가 이뤄지고 있다. 예전에는 딥러닝용 GPU를 구매해야지만 학습할 수 있었지만, 최근에는 AWS같은 클라우드 플랫폼 서비스에서도 딥러닝을 연구할 수 있는 인스턴스를 제공해 누구나 쉽게 딥러닝에 접해볼 수 있게 바뀌었다. 또 딥러닝에 대한 알고리즘 성능은 나날이 발전하고 있으며 객체에 대한 분류를 비롯하여 객체탐지, 객체추적, 객체추론 등 다양한 방식으로 구분되고 있다. 현재는 이러한 객체 인식에 관한 기술과 융합하여 일상생활에서의 불편함을 해소해주는 연구가 진행되고 있으며, 장애인에게 편의성을 제공해주고 있다. 객체 인식 기술은 특히 시각장애인에게 유용한데, 시각장애인은 사물을 인지하는 데 어려움이 있고 같은 형태의 물건을 구별하기 힘들다[1]. 이러한 불편은 편의점이나 가게에서 물건을 구매할 때 자주 나타나는데 대부분 제품은 같은 형태의 다른 맛 제품들이 존재하기 때문이다. 따라서 시각장애인은 자신이 원하는 제품을 구매하기가 쉽지 않다. 2017년 장애인 실태조사에 따르면 시각장애인 정보통신기기 사용 현황에 스마트폰이 절반 이상을 차지하고 있는데 그 이유는 현재 스마트폰에서 제공하는 보이스 어시스턴스가 시각장애인이 스마트폰을 사용하는 데 도움을 주기 때문이다.

본 논문은 딥러닝 기술을 사용해 편의점 내 제품을 학습하여 스마트폰의 장애인 접근성 기술들과 융합해 시각장애인에게 편의점 제품을 인식해 음성으로 알려주는 애플리케이션을 설계하고 개발한다. 딥러닝학습을 수행하기 위해서 편의점 제품에 대한 데이터 세트를 확보하는 것이 우선시되어야 한다.

본 논문의 애플리케이션을 시각장애인이 사용한다면 원하는 제품을 구별하는데 기존에 걸리던 시간보다 빠르고 정확하게 찾도록 도와주고 이를 바탕으로 시각장애인의 독립적인 경제활동을 기대할 수 있을 것이다.

2. 본론

2.1 데이터 구성

본 연구에서 사용한 데이터는 편의점에서 판매하고 있는 제품을 직접 촬영한 데이터 세트이다. 데이터 세트는 총 11개의 제품으로 상자형 과자 5개, 봉지형 과자 3개와 캔 음료수 3개로 구성되어 있다. 제품은 자동으로 돌아가는 원판을 통해 회전하여 다양한 각도로 학습을 할 수 있도록 구성했고 상자형

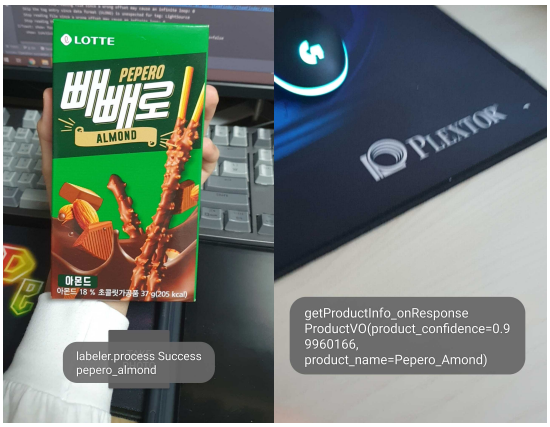
과자와 봉지형 과자의 경우 완전한 옆면이나 뒷면은 같은 제품의 다른 맛과 유사한 경우가 많아 특징을 잡기 어려우므로 제외했다.

데이터 세트는 전체 23,814장을 사용되었으며, 각 제품은 평균적으로 1500장에서 2000장의 데이터를 가진다. 각 제품에 대해 80%는 Train에 사용되었고, 나머지 20%는 Validation과 Test에 사용하였으며 Image Augmentation 작업을 통해 다양한 변화에도 인식할 수 있도록 하였다.

<그림 1> 데이터 세트 예시(롯데 빼빼로 오리지널)

2.2 CNN 방식의 학습

CNN[2]은 수동으로 특징을 추출할 필요 없이 데이터로부터 직접 특징을 학습하는 신경망 아키텍처이다. 주로 이미지 인식에 사용되며 특징을 추출하는 Filter와 Activation 함수를 통해 Conv Filter와 ReLU 함수, 풀링 레이어를 반복적으로 조합해 특징을 추출하고 학습한다. 데이터 세트에 CNN을 사용하여 학습한 결과 사진에 제품이 있으면 <그림 2>의 왼쪽과 같이 제품을 인식하는 데 성공한 것을 볼 수 있다.

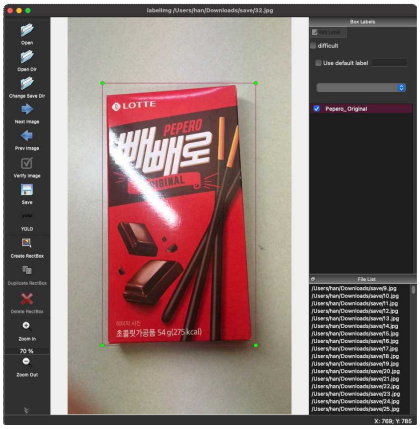


<그림 2> (왼쪽) CNN 인식 결과
(오른쪽) CNN 문제점

하지만 CNN 모델은 사진 내에서 특징을 무조건 잡아 결과를 도출하기 때문에 <그림 2>의 오른쪽과 같이 제품이 없는 사진에서도 제품을 인식하기 때문에 CNN 모델을 통한 학습은 한계가 있다.

2.3 YOLO 방식의 학습

CNN의 한계를 개선하고자 객체의 유무 판별을 위한 검출 기능과 객체가 무엇을 의미하는지 분류하는 기능이 있는 YOLO[3][4]를 사용했다. YOLO 모델은 물체 검출 및 인식이 동시에 가능한 모델이며 Open Source로 공개되어 다양한 backbone과 버전이 존재한다. 본 논문에서는 가장 빠르고 최신인 YOLO v5를 사용하여 객체 검출을 학습하였다. YOLO v5의 특징은 ResNet에서 동일하게 사용한 Bottleneck 기술과 CSP 기술을 합친 BottleneckCSP을 사용하고 있는데 먼저 4개의 Convolution Layer가 생성되는데 각 Layer에서 1번, 4번 Layer는 Convolution과 Batch_norm Layer를 사용하고, 2번, 3번 Layer는 Convolution Layer만 사용해 연산한다. 그 후 CSP구조로 y값 두 개를 생성하는데 1번, 3번 Layer를 통과해 연산하는 y1과 2번 Layer만 통과해 연산하는 y2를 생성한 후 y1과 y2를 합친 결과를 4번 Layer를 통과하여 최종적인 결과를 연산한다. 또한 공간 피라미드 풀링(SPP) 기술을 사용하여 최종적인 고정된 1차원 형태의 배열을 만들어내 검출 정확도를 개선하였다. 또한 YOLO v5에서 기본적으로 제공되는 4가지 BackBone중에서 레이어 수가 적당한 yolov5l.yaml을 사용했다. 전체 데이터 세트 이미지에서 특징을 스스로 학습하는 CNN 모델과 달리 YOLO 모델은 학습하기 위해선 데이터 세트 내 객체가 어디에 있는지 알려주는 위치 정보 (bounding box)가 필요하다.



<그림 3> 학습이미지 및 위치 정보(Bounding Box)

학습데이터 세트에 위치 정보는 .txt 파일로 제공되며 해당 이미지 내의 객체의 좌표로 제공된다. <그림 3>은 학습데이터 세트에 제공하는 이미지 및 위치 정보(Bounding Box)의 예시이다.

본 논문에서의 딥러닝 학습은 Amazon에서 제공하는 AWS를 사용하였고, 학습에 사용된 GPU는 NVIDIA K80 (12GB 메모리)를 사용했고 나머지 학습 모델 배포 서버도 동일한 AWS 인스턴스를 사용했고, Flask 방식으로 서버를 구현하였다.

2.4 학습 결과

본 논문에서의 딥러닝 모델 학습 성능 평가를 위해 물체 검증 알고리즘 성능 평가방법인 AP(Average Precision)을 사용하였다. 여기서 사용되는 TP(True Positive), FP(Flase Positive), FN(False Negative)를 다음과 같이 정의한다.

TP: 모델의 검출한 Bounding Box와 정답 Bounding Box 사이의 IoU(Intersection over Union) Threshold 가 0.5 이상인 경우

FP: 모델의 검출한 Bounding Box와 정답 Bounding Box 사이의 IoU(Intersection over Union) Threshold 가 0.5 미만인 경우

FN: 이미지 프레임 내에 정답 Bounding Box가 존재할 때, 모델의 검출 결과가 없는 경우

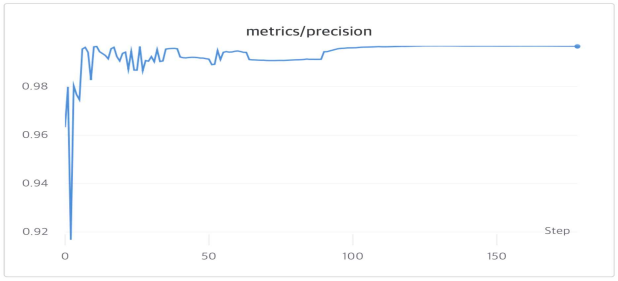
여기서 정의한 TP, FP, FN을 사용하여 검출 성능 지표로 사용되는 Presision과 Recall을 구하고 [식 1,2] Precision-Recall Graph의 아래 면적을 구하는 AP(Average Precision)을 구해 모든 클래스에 각각 AP 연산[식 3]을 반복한 후 그 값들의 평균을 구하는 mAP(mean Average Precision)를 최종 검출 성능 지표로 사용하였다. 식 1, 2, 3은 이 과정을 수식으로 나타낸 것이다.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

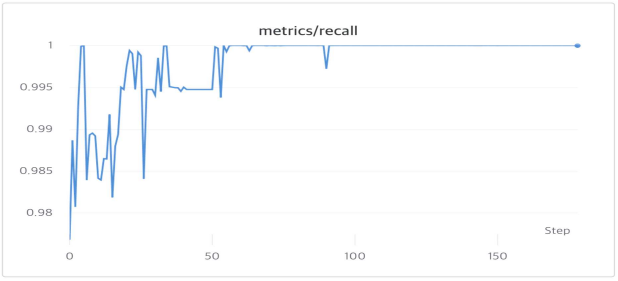
$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$AP = \frac{1}{11} \times (AP_r(0) + AP_r(0.1) + \dots + AP_r(0.9) + AP_r(1.0)) \quad (3)$$

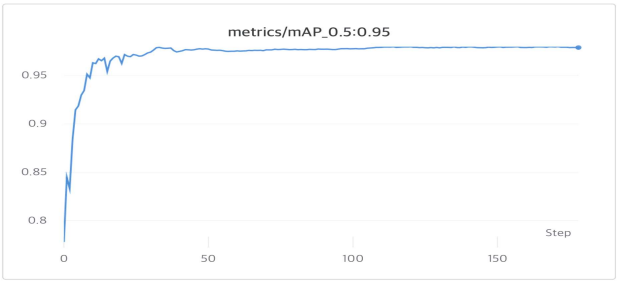
<그림 4, 5, 6>은 약 200번 강화학습을 진행한 YOLO 모델의 Presicion, Recall, mAP을 보여주고 있다.



<그림 4> 학습 모델의 Precision

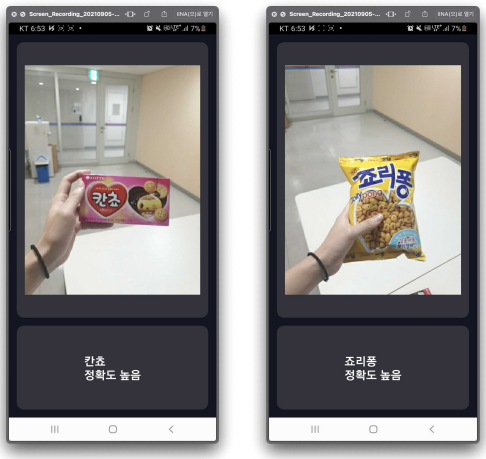


<그림5> 학습 모델의 Recall



<그림 6> 학습 모델의 IoU = 0.5 일 때 mAP

이후 학습된 해당 모델을 Flask로 구현된 서버에 적재하여 배포를 진행했다.



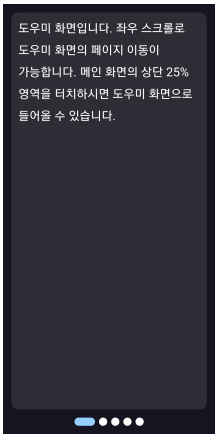
<그림 7> 실제 스마트폰에서 구동한 제품 인식 결과

<그림 7>은 실제 서버에 배포된 모델과 스마트폰이 통신해 편의점 제품을 인식 및 분류하는 사진이다.

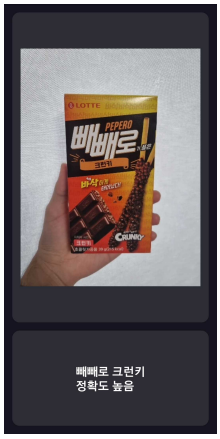
장애인 접근성을 높이기 위해서 확률에 따라 0 - 30%는 '정확도 낮음', 31 - 70%는 '정확도 보통', 71 - 100%는 '정확도 높음'으로 표기하고, 스마트폰 내 시각장애인 모드(Talkback)에 맞춰 읽어주도록 설정했다.

2.5 애플리케이션에서의 장애인 접근성

본 논문에서 개발하는 애플리케이션의 주 사용자는 시각장애인이기 때문에 시각장애인 접근성에 대한 많은 부분을 고려하였다. 튜토리얼 화면이나 도움말 화면의 경우, 버튼을 삽입하는 것 대신 좌우로 스크롤이 가능하도록 구현하였다. 결과 화면의 경우에는 객체 사진의 검증 결과가 제공되기 때문에 해당 콘텐츠를 터치하였을 때 결과를 시각장애인이 알 수 있도록 대체 텍스트[5]를 제공하였다. 또한, 처음 화면으로 되돌아가기 위해 이전 버튼을 삽입하는 대신 화면의 어느 부분을 터치하더라도 화면이 닫히도록 설계했다.



<그림 8> 도우미 화면



<그림 9> 결과 화면

3. 결론

본 논문에서는 시각장애인을 대상으로 스마트폰에서 촬영한 객체를 인식하고 분류하여 사용자에게 알려주는 애플리케이션을 구현하였다. 인식할 객체의 데이터 세트를 직접 촬영해 확보하였고 처음에 진행한 CNN 방식의 딥러닝 학습을 통해 얻은 모델은 좋은 인식률을 보였지만 객체가 없을 때도 인식하는 문제로 인해 객체 검출과 분류를 할 수 있는 YOLO 방식으로 학습 방법을 바꾼 결과 객체의 분류뿐만 아니라 유무 판단까지 할 수 있는 모델을 얻을 수 있게 되었다.

본 논문에서 사용한 데이터 세트는 총 23,814장이며 이 중 Validate를 위해 사용된 데이터 세트는 약 14%인 3,284장의 제품 이미지가 사용되었고, Test를 위해 사용된 데이터 세트는 6%인 1,228장의 제품 이미지가 사용되었다. YOLO v5를 사용해 약 200번 반복 학습으로 산출한 모델의 객체 검출 성능(mAP)는 최종적으로 0.9790으로 가장 높게 나왔다. 이 모델을 배포하여 보이스 어시스턴스가 지원되는 스마트폰 애플리케이션과 연동되면 시각장애인에게 제품을 고르는 데 큰 도움을 줄 것이라고 기대된다.

본 논문에서 학습한 모델의 한계점은 학습 시 연산된 mAP 성능과는 다르게 실제 환경에서 테스트해보면 객체의 각도와 사진의 광량에 따라 정확도가 천차만별로 나온다는 것이다. 따라서 향후에는 다양한 광량과 각도에 맞는 데이터 세트를 구축하여 학습에 사용하고, 특징을 잡는 하이퍼 파라미터에 대한 연구와 함께 각 제품에 대한 데이터 세트 개수도 늘리면 이 문제점은 해결될 것이라 예상된다.

※ 본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다

참고문헌

[1] Dong-Yeon Choi and Min-Jung Kim, "Trends in Alternative Sensory Research Applied to Assistive Technology for the Blind," Disability and Employment, Vol. 25 No. 4 pp. 5-34, 2015.

[2] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neur al information processing systems. 2012.

[3] Young-Hwan Lee and Youngseop Kim, "Camparision of CNN and YOLO for Object Detection", Journal of the Semiconductor & Display Technology, Vol. 19, No. 1. March 2020.

[4] Redmond et al, You Only Look Once: Unified, Real-Time Object Detection, 2015, CVPR

[5] Woochun Jun, A Study on Development of Enhancement Guidelines Application Accessivility for the Disabled, Journal of The Korean Association of Information Education, Vol. 19, No. 1. March 2015, pp.69-76