시각장애인을 위한 YOLOv5 기반 제품 인식 시스템

(YOLOv5-based Product Recognition System for Visually Impaired People)

강 동 훈*, 양 다 영, 최 가 영, 임 창 헌
(Dong-Hun Kang, Da-Yeong Yang, Ga-Young Choi, Chang-Heon Lim)

부경대학교 전자정보통신공학부 전자공학전공

Abstract: When visually impaired people purchase some products without braille labelling in a convenient store, they suffer from identifying some information that directly affects health, such as allergen information. The existing system 'Sullivan Plus' relies on text recognition rather than object recognition, thus it has limitations in providing product names to users in specific situations. To address this difficulty, this paper proposes a system that recognizes objects in real-time in a mobile environment and describes their information in the form of audio. It employs the YOLOv5 model for identifying 65 classes of snacks and beverages and data augmentation such as blurring and noise addition. Experiments show that the recognition accuracies of the proposed system without data augmentation and with augmentation are respectively 86% and 88%.

Keywords : YOLOv5, deep learning, visually impaired, object detection, data augmentation

Ⅰ. 서 론

현대 사회에서 시각장애인들은 많은 불편함과 제약을 겪고 있다. 특히 편의점이나 마트 등에서 제품을 구매할 때, 부실한 점자 표기로 인해 적절한 정보를 확보하는 데 어려움을 겪는다. 매장에서 판매되는 제품에는 대부분 점자 표기가 전혀 없거나, 있더라도 단순 정보만 제공되고 있어 구체적인 제품 식별이 어렵다. 실제로 한국소비자원의 조사에 따르면, 주요 음료 제조업체 191개 제품 중 49.2%만이 점자를 표기하고 있었다[1]. 또한, 점자 표기가 있는 121개 제품 중 85.1%는 '음료' 또는 '탄산'으로만 표기되어 있어 제품 구분에 도움이 되지 않았다. 이처럼 부실한 점자 표기로 인해 시각장애인 소비자들은 큰 불편을 겪고 있다. 마트에서 식품을 구매한 경험이 있는 192명의 시각장애인 중 71.9%가 점자 표기와 관련하여 불편을 경험한 것으로 조사되었다[1]. 일부 제조업체에서는 점자 표기 개선을 위해 노력하고 있으나, 제품 공간 제약과 추가 비용 발생 등의 이유로 브랜드명 점자 표기에는 어려움이 있는 실정이다[2].

기존에 개발되어있는 시각보조 애플리케이션인 '설리번플러스'는 시각장애인들에게 카메라를 통해 인식한 정보를 알려주는 서비스를 제공한다. 그러나 해당 애플리케이션을 사용하여 제품명과 상세 정보를 확인하고자 할 때, 촬영된 이미지 내의 모든 텍스트를 인식하기 때문에 사용자에게 필요 없는 정보를 제공한다는 문제점이 있다. 또한, 여러 상황에서 사용자에게 제품명을 제공하지 못할 가능성이 존재한다. 예를 들어, 제품 측·후면에 대한 이미지와 같이 이미지 내에 텍스트 정보가 없는 경우, 백색광 반사 및 촬영 각도에 의해 제품명이 인식되지 못하는 경우, 과자봉지가 구겨지면서 제품명이 부분적으로 보이지 않게 되는 경우가 있다.

본 논문은 '설리번플러스'에서 사용한 텍스트 인식 방식이 아닌 딥러닝 YOLOv5 기반 객체 인식 방식을 이용해 실시간으로 제품을 인식하여 그에 대한 정보를 제공하는 서비스를 제안한다. 이에 따라 사용자가 불필요한 정보를 얻게 되거나 제품명을 얻지 못하는 문제를 해결할 수 있다. 제품명을 예측한 결과는 제품명과 상세 정보들이 기록된 텍스트 파일로부터 정보를 추출하여 TTS(Text to speech)를 통해 출력된다. 본 논문에서 제공하는

정보는 제품의 제품명과 포함된 알레르기 성분으로 한정한다.

Ⅱ. 제안하는 시스템

1. 시스템 동작 순서

본 논문에서 제안하는 시스템의 동작은 [그림1]과 같다. 신경망 모델에 스마트폰 카메라를 통해 실시간으로 이미지 데이터가 입력되고, 신경망 모델은 입력된 이미지 내의 제품을 인식한다. 다음으로, 스마트폰 화면에 제품 객체에 대한 제품명이 기재된 바운딩 박스가 생성된다. 마지막으로, 스마트폰 화면을 한 번 터치하면 TTS를 통해 제품명이 음성으로 출력되며, 화면을 두 번 터치하면 제품명과 제품에 포함된 알레르기 성분이 음성으로 출력된다. TTS 엔진은 Android Studio에서 제공하는 엔진을 사용하였다.

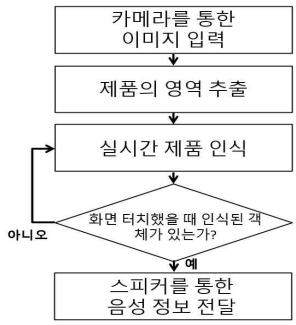

그림 210. 시스템 순서도

2. 학습 데이터셋 구축

본 논문은 시중에 판매되고 있는 과자 총 50종과 음료수 총 15종으로 제품군을 한정하였다. AIHub에서 제공하는 음료 제품 및 Roboflow에서 다른 사용자들이 제작한 프로젝트의 여러 과자 사진 데이터셋을 사용하였다. 동시에, 실제 시스템이 사용될 환경과 유사한 환경의 이미지를 수집하기 위해 다양한 환경에서 직접 촬영하여 총 7693장의 이미지 데이터를 수집하였다. 이후 바운딩 박스를 지정하고 라벨링하는 과정을 거쳤으며, 각각 60%, 20%, 20% 비율로 훈련 데이터(Training set), 검증 데이터(Validation set), 그리고 테스트 데이터(testing date)로 분류하였다.

여러 제품에 대해 다양한 환경에서 촬영된 이미지를 충분히 확보할 필요가 있으므로, Roboflow에서 데이터 증강(Data Augmentation)을 진행하였다. 데이터 증강을 통해 데이터셋을 확장하여 모델의 일반화 성능을 개선하였다. 제품이나 스마트폰 카메라에 이물질이 묻어있는 상황에서도 본 시스템이 작동하도록 데이터 증강 기법 중 Noise 기법 및 Blur 기법을 사용한다[3].

3. 객체 인식 모델

실제 환경에서 해당 애플리케이션을 사용할 때, 한 화면 내에 여러 개의 제품 객체가 존재할 가능성이 있다. 따라서 하나의 클래스를 예측하는 이미지 분류(Image Classification) 모델이 아닌 객체 인식(Object Detection) 모델을 사용하여 여러 개의 객체를 찾아낸 후 개별 정보를 제공할 필요가 있다.

객체 인식 모델은 YOLO, SSD, Faster R-CNN 등이 있는데, 이 중에서 YOLOv5는 높은 정확도를 가지며 작은 모델 크기와 낮은 연산 요구사항을 요구한다. 따라서 본 논문에서는 모바일 환경에서 실시간 객체 인식을 원활하게 진행할 수 있는 YOLOv5를 사용한다.

4. 애플리케이션 동작

[그림 2]는 애플리케이션이 실제로 동작하는 화면을 캡처한 것이다. 편의점에 진열된 과자 및 음료수를 스마트폰 카메라로 촬영하면 제품 객체별로 제품명이 기재된 바운딩 박스가 생성되는 것을 확인할 수 있다. 스마트폰 화면을 한 번 터치하면 바운딩 박스가 생성된 객체의 제품명을 사용자에게 음성으로 전달하며, 화면을 두 번 터치하면 제품명과 함께 알레르기 성분을 전달한다.

그림 211. 애플리케이션 동작 화면

Ⅲ. 성능 평가

서로 다른 네 가지 데이터셋(증강 이전, Noise와 Blur 기법 적용, Noise 기법 적용, Blur 기법 적용)에 대한 신경망 모델 학습을 진행하였다. 학습 시 batchsize는 16과 32로 설정하였고, epoch은 모두 40으로 설정하였다. 평가 방식으로 mAP50(Mean Average Precision at IoU=0.50)을 사용하여 각 모델의 성능을 비교하였다. mAP50은 객체 인식 모델을 평가하는 대표적인 평가지표이며, 모델이 예측한 바운딩 박스와 실제 바운딩 박스 간의 겹치는 정도가 최소 50% 이상일 때의 평균 정밀도를 의미한다. mAP50은 모든 객체 클래스를 고려하기 때문에 여러 클래스에 걸친 모델 성능을 종합적으로 평가하는 데 적절한 평가지표이다.

[그림 3]은 batchsize 16으로 서로 다른 데이터 증강 방식을 적용한 데이터셋을 학습한 모델의 mAP50 값을 그래프로 나타낸 것이다. [그림 4]는 batchsize 32로 동일한 실험을 진행한 결과이다. [표 1]은 [그림 3], [그림 4]의 결과를 표로 나타낸 것이다. batchsize를 32로 설정하고 Blur, Noise 기법을 모두 적용했을 때의 mAP50 값이 가장 높은 것을 확인하였으며, 데이터 증강 적용 전 mAP50 값인 0.855보다 약 3.4%만큼 성능이 개선되었음을 확인하였다. 따라서 애플리케이션 구현 시 사용할 모델로 batchsize 32에 대해 Blur, Noise 기법을 적용한 데이터셋 학습 모델을 선정하였다.

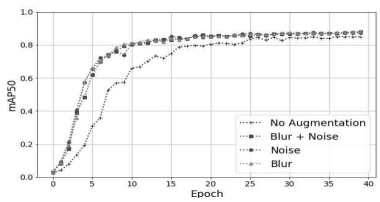

그림 212. 데이터 증강 방식에 따른 mAP50 비교 그래프 (batchsize 16)

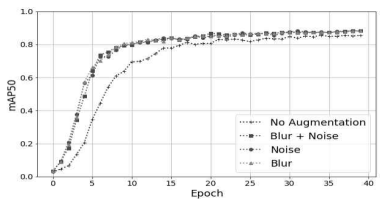

그림 213. 데이터 증강 방식에 따른 mAP50 비교 그래프 (batchsize 32)

표 1. 데이터 증강 적용 여부 및 데이터 증강 방식에 따른 batchsize 별 mAP50 비교

	batchsize 16	batchsize 32
데이터 증강 전	0.864	0.855
Blur + Noise	0.879	0.883
Noise	0.875	0.882
Blur	0.881	0.876

Ⅳ. 결 론

본 논문에서는 시각장애인들이 제품 선택 시 겪는 어려움을 해결하기 위해 딥러닝 기반의 실시간 객체 인식 서비스를 제안하였다. '설리번플러스'에서 사용하는 텍스트 인식 방식과 달리, YOLOv5 모델을 활용하여 제품 이미지를 직접 인식하고 인식된 제품명을 통해 상세 정보를 음성으로 전달하는 새로운 접근법을 채택하였다.

데이터 증강 기법을 적용하여 모델 성능을 개선하였으며, 높은 수준의 인식 정확도를 기대할 수 있게 되었다. 향후 제품군을 확장하고 사용자 인터페이스를 개선한다면, 시각장애인의 제품 선택 편의성을 크게 높일 수 있을 것이다. 본 논문에서 제안하는 시스템은 서론에서 언급한 바와 같이 부실한 점자 표기로 인해 시각장애인들의 제품 선택권이 충분히 보장되지 못하는 문제를 해결하는 데 도움이 될 것으로 기대된다.

참 고 문 헌

[1] 박정태, 시각장애인 식품 점자 표기 소비자문제 실태조사, 한국소비자원, 조사보고서, 2022.

[2] https://news.tf.co.kr/read/economy/2083260.htm

[3] A. Abdulghani, A. M. Abdulghani, W. L. Walters, and K. H. Abed, "Data Augmentation with Noise and Blur to Enhance the Performance of YOLO7 Object Detection Algorithm", 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE), pp. 180-185, July 2023.