

딥러닝 기반의 실시간 영상 처리 기술을 활용한 시각장애인 대상 상품 음성안내 서비스

박영웅¹, 서준혁², 정시현³, 이재훈⁴, 심미경^{*}
^{123*}동국대학교 산업시스템공학과, ⁴LG U+

assayw119@naver.com¹, joon102938@gmail.com², teresa010219@naver.com³, heenayu@lguplus.co.kr⁴, smegyeong@gmail.com^{*},

A Product Voice guidance service for the Visually Impaired Using Real-time Image Processing Technology Based on Deep Learning

Park Young Woong¹, Suh Joon Hyuk², Chung Sihyeon³, Lee Jae Hun, ⁴ Sim Mi Gyeong^{*}
^{123*}Dept. of Industrial & Systems Eng. Dongguk University, ⁴LG U+

요 약

우리나라의 전체 시각장애인 252,046 명[1] 중 점자를 읽을 수 있는 비율은 12.4% 밖에 안되는 상황이기 때문에, 점자 표기의 확대나 관련 교육 외에 시각장애인이 편리하게 필요한 정보를 인식할 수 있는 서비스 제공이 마련되어야 한다. 이에 따라, 본 연구에서는 시각장애인이 편의점이나 마트에서 상품 가격 등 필요한 정보를 확인할 수 있도록 지원하는 서비스를 개발하고자 한다. 구체적으로, 딥러닝 관련 기술을 적용하여, 시각장애인이 정보를 받을 수 있는 음성 채널 확장을 위한 시스템 개발을 목적으로 한다. 이를 위해, 서비스 개발과 관련하여 상품 인식을 위한 Yolo v5(실시간 객체 인식)와 CRAFT(Character Region Awareness for Text Detection, 텍스트 영역 검출), Deep-Text-Recognition-Benchmark(텍스트 인식)을 사용하고, 시각장애인이 이해하고 확인하기 용이한 음성 정보 출력을 위해 TTS(Text to Speech) 알고리즘을 활용한다. 또한, 상품 인식 과정에서 지연 문제를 해결하기 위해 실시간 영상 처리를 통해 상품을 인식한다. 본 논문에서 제시하는 기술과 알고리즘을 통해 실생활에서 시각장애인이 실시간으로 정보를 음성으로 인식하여, 사회적 약자의 사각지대와 정보격차를 줄일 수 있으리라 기대된다.

I. 서 론

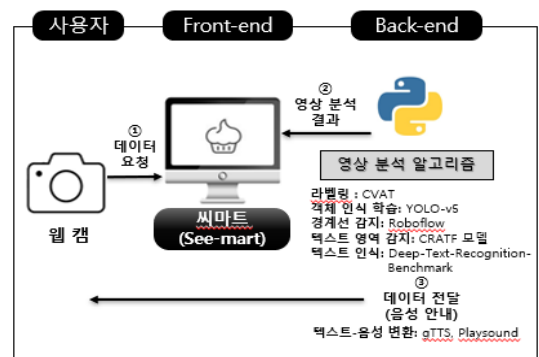
시각장애인이 혼자 편의점이나 마트에서 원하는 물건을 사는 것은 매우 힘든 일이다. 매대에 진열된 상품에는 점자 표기가 없을 뿐만 아니라, 실제로 점자 해독이 가능한 시각장애인은 전체의 12.4%에 불과[1]하기 때문에 점자 표기의 확대만으론 시각장애인의 불편을 해결하는데 한계가 있다. 시각장애인이 마트 이용 시 겪는 불편함을 조사하던 중 인터뷰한 한 대형마트 관계자는, 시각장애인은 특정 제품을 찾는 것과 지나치게 많은 시간이 걸리는 점을 지적하며 개선의 필요성을 언급하였다. 또한 시각장애인의 편의점 이용 영상을 보고 마트나 편의점에 시각장애인을 배려하는 시스템이 얼마나 부족한지 알 수 있었다. 이러한 인터뷰 결과와 실제 시각 장애인의 소비 과정을 본 후 시각장애인에게 제품 정보를 빠르고 간편하게 알려줌으로써 스스로 쇼핑할 수 있도록 돕는 서비스의 필요성을 느꼈다.

본 연구에서 제안하는 상품 인식 서비스는 시각장애인에게 인식한 상품 정보를 음성으로 전달하여 시각장애인의 원활한 소비 생활을 돕는 것을 목적으로 한다. 기존의 시각장애인 대상 객체 인식 서비스는 무분별하게 텍스트를 인식한다는 점과 상품 정보를 얻기 위해 상품 하나하나를 찍어야 한다는 한계점이 존재한다. 또한 서비스 제공 방법에서 시각장애인이 직접 터치로 모드 설정을 하여 원하는 정보를 선택해야 한다는 한계가 있다. 본 연구에서는 이와 같은 문제를 해결하기 위해 객체 인식 & 텍스트 탐지 및 인식과 더불어 실시간 영상 처리 기술의 영역으로 확장한다. 우선 이미지 인식 대신 실시간 영상 처리 기술로 상품을 하나씩 찍어야 하는 번거로움 없이 간편하고 빠른 상품 인식이 가능하게 했고 가격표 식별 후에 인식하고자 하는 텍스트 영역을 찾아 필요한 텍스트만 인식하여 무분별한 글자 인식을 보완했다.

II. 본론

II. I. 서비스 시나리오

본 서비스를 이용하는 시각장애인은 무선 이어폰과 같은 음성 출력 기기를 장착한 후 탈부착이 가능한 웹캠으로 상품 구매 과정을 영상 촬영한다. 본 연구 결과인 서비스 구성도는 (그림 1)과 같으며, 이후 서비스 구성 흐름에 따른 각 단계에 대해 설명한다.

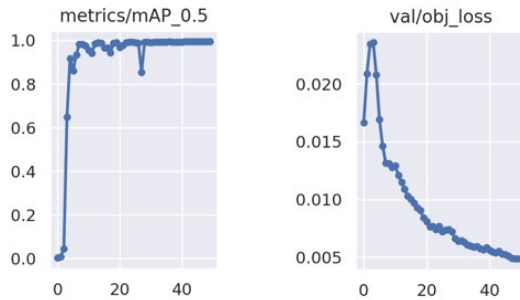


(그림 1) 서비스 구성도

II. II. 가격표 객체 인식

가격표 이미지 데이터는 실제 마트에서 사용되는 가격표를 참고해 직접 제작한 가격표를 찍은 영상의 프레임 이미지로 총 2,477 장이다. 본 연구에서는 이미지 라벨링을 위해 CVAT (Computer Vision Annotation Tool) 프로그램을 사용한다. 데이터 학습은 Object Detection 모델인 YOLO v5(You Only Look Once) [2]을 사용해 진행했다. YOLO v5은 One-Stage Object Detection 이 가능한 모델로, 이전 버전의 모델에 비해

처리 속도가 향상되어 실시간 영상 내에서의 객체 인식 성능이 우수하다. 제작한 가격표를 촬영한 영상에서 가격표 인식을 수행한 결과, 객체 인식 정확도(mAP)와 손실 값(Loss)은 (그림 2)와 같다.



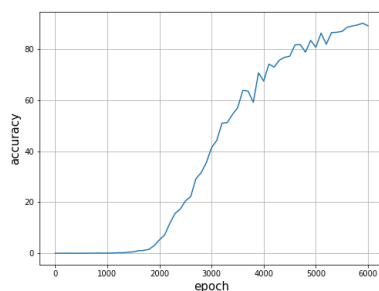
(그림 2) 객체 인식 mAP 및 Loss 그래프

II.III. 텍스트 영역 감지

인식된 가격표에서 상품명과 상품가격 텍스트 영역을 감지한다. 상품명은 경우 주로 한글이지만 영어도 존재하며, 상품가격의 경우 숫자가 존재한다. 따라서 Multi-language Text 에 대한 데이터셋으로 연구하는 모델 중 한국어 설명 자료가 다수 존재하는 CRAFT(Character Region Awareness for Text Detection)알고리즘을[3] 사용하였다. CRAFT 는 네이버 Clova AI 팀에서 발표한 Text Detection 모델로 영어, 한국어, 숫자를 포함한 문자가 있는 영역을 탐지하는 딥러닝 모델이다. CRAFT 는 이미지에 대해 픽셀 단위로 2 가지 정보를 예측하여 텍스트 영역을 추출한다.

II.IV. 텍스트 인식

탐지된 가격표 제품명, 제품 가격 텍스트 영역 속 텍스트 인식을 위해 네이버 Clova AI 팀에서 발표한 Text Recognition 모델인 Deep text recognition benchmark 를 사용했다. 해당 모델은 텍스트 인식 분야에서 인정받는 여러 모델을 조합해 쉽게 학습할 수 있도록 만들어졌으나, 한국어 데이터 셋으로 학습된 사전 학습 모델의 부재로 직접 한국어 학습 과정을 진행하였다. 텍스트 이미지는 AI hub 에서 제공하는 한국어 글자체 중 한글 폰트 50 종에 대한 단어(어절) 인쇄체 이미지를 사용했다. 학습과 검증, 테스트를 위해 데이터를 각각 0.6, 0.2, 0.2 의 비율로 분할했고 분할한 데이터에 대해 이미지 라벨링 텍스트 파일을 lmdb 형태로 변환하여 모델 학습을 진행했다. 이와 같이 Text Recognition 모델을 수행했을 때, 반복 학습이 진행됨에 따라 텍스트 인식 정확도가 90% 이상까지 향상되었으며 정확도 향상 결과는 (그림 3)과 같다.



(그림 3) 텍스트 인식 정확도 그래프

II.V. 음성 안내

인식된 텍스트 형태의 상품 정보가 txt 파일 형태로 저장되면, 저장된 텍스트는 gTTS(google-Text-To-Speech) API 를 통해 음성으로 변환되어 mp3 파일로 저장이 된다. gTTS 는 한글을 지원하므로 한글 텍스트의 음성 변환이 가능하다. 음성안내 mp3 파일을 출력하기 위해 python 언어 기반의 Playsound 모듈을 사용한다.

예를 들어, ‘꼬북칩 초코맛 160g’과 ‘1200 원’ 이라는 정보를 가진 가격표를 인식하면 해당 내용이 담긴 txt 파일이 생성되고, 음성으로 안내하는 mp3 파일의 생성과 함께 실행이 된다.

III. 결론

본 연구에서는 시각장애인들이 편의점이나 마트에서 스스로 상품의 정보를 확인하고 소비활동을 할 수 있도록 돕는 서비스를 제공하고자 딥러닝 기술을 사용하여 상품인식 서비스를 개발하였다. 현재 가격표 객체 인식과 텍스트 탐지 및 인식, 그리고 음성안내 서비스까지 개발 완료하였고, 인식을 향상을 위해 학습용 데이터 셋 추가 확보 및 증강을 통한 반복적인 학습을 시도하고 있다.

해당 서비스의 사용 확대를 위해 추가적인 연구를 진행하고, 관련 연구 내용은 논문을 통해 지속적으로 공유할 예정이다. 또한 앱 개발로의 확장성, 전국의 편의점 혹은 마트와의 협력을 제안하며, 본 연구의 기술이 많은 시각장애인에게 실질적으로 사용되어 사회적 약자들의 정보격차를 줄이는 데 기여하고자 하는 바이다.

ACKNOWLEDGMENT

본 논문은 과학기술정보통신부 정보통신창의인재양성 사업의 지원을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.

참 고 문 헌

- [1] 김성희, et al. 2017 년 장애인 실태조사. 보건복지부, 2017.
- [2] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [3] Baek, Youngmin, et al. "Character region awareness for text detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.