딥러닝 기반 이미지 인식 기술을 활용한 영어 학습 애플리케이션 개발

김유정, 김주연, 이유빈, 이기용
숙명여자대학교 소프트웨어학부
e-mail: {yoojunging, kokoli0528, mosquitoiu, kiyonglee}@sookmyung.ac.kr

Development of an English Study Application using Deep Learning-based Image Recognition techniques

Yoo Jung Kim, Ju Yeon Kim, Yu bin Lee, Ki Yong Lee
Division of Computer Science, Sookmyung Women's University

요 약

본 논문에서는 사용자의 주변사물을 인식하여 영단어로 알려줌으로써 사용자가 실생활에서 영단어를 능동적으로 학습할 수 있도록 돕는 애플리케이션을 개발한다. 본 애플리케이션은 사용자가 카메라로 촬영하거나 사진첩에서 선택한 이미지를 인식하여 사진 속 물체의 영어 단어와 한국어 뜻을 알려주며, 단어의 발음 또한 확인할 수 있고, 직접 단어장에 저장하여 다시 학습할 수 있도록 한다. 이를 위해 TensorFlow를 활용한 딥러닝 기반 이미지 인식 기술을 사용하였으며, 추후 TensorFlow를 통하여 모델을 추가적으로 훈련시킴으로써 이미지 인식의 정확도를 높일 수 있다. 그 외 영어-한국어 번역, 텍스트-음성 변환 등 부가 기능을 통해 사용자가 다양한 방식으로 영단어를 학습할 수 있도록 한다.

1. 서론

현재 영어 공부를 위한 다양한 애플리케이션이 존재하지만, 사용자가 실생활에서 직접 접하는 주변 사물들을 활용하는 영어 학습 애플리케이션은 아직 널리 알려진 바가 없다. 본 논문에서는 사용자가 실생활에서 접하는 주변 사물들을 실시간으로 인식하여, 해당 사물들에 대한 영단어를 학습할 수 있도록 하는 영어 학습 애플리케이션을 개발한다. 본 논문에서 개발한 애플리케이션은 사용자가 카메라로 촬영하거나 사진첩에서 선택한 이미지를 인식하여 이미지 속 물체에 대한 영단어를 텍스트 및 음성으로 학습할 수 있다. 본 논문에서는 이미지 인식 기능 구현을 위해서 TensorFlow[1]를 활용한 딥러닝(deep learning)기술을 사용하였으며, 그 외 영어-한국어 번역, STT(speech to text)와 TTS(text to speech)등 오픈 API를 활용한 부가 기능을 추가하여 사용자가 다양한 방식으로 영단어를 학습할 수 있도록 한다. 본 논문에서 개발한 영단어 학습 애플리케이션은 'Make My Word'라 부르며, 크게 다음과 같은 4가지 기능을 제공한다.

- 카메라 촬영 혹은 사진첩의 사진을 분석하여 물체를 인식하고, 해당 물체에 대한 영단어를 파악하는 기능
- 인식된 물체에 대한 영단어를 음성 또는 텍스트로 입력하여 정답을 확인하는 기능
- 인식된 물체에 대한 영단어를 한국어로 번역하거나 영어로 발음해 주는 기능
- 사용자가 선택한 사진과 함께 사진 속 물체에 대한 영단어와 한국어 뜻을 저장하는 단어장 기능

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 개발한 애플리케이션과 유사한 애플리케이션들을 살펴본다. 3장에서는 본 애플리케이션의 세부 기능 및 실제 수행 예를 보인다. 4장에서는 본 애플리케이션의 개발에 사용된 기술을 설명하며, 5장에서는 결론을 맺는다.

2. 유사 애플리케이션 현황

본 논문에서 개발한 애플리케이션과 유사한 애플리케이션으로는 '찍사'[2]와 '캠딕'[3]이 있다. '찍사'는 카메라와 사전을 결합한 애플리케이션으로 영어로 쓰인 단어나 문장을 촬영하면 그 내용을 번역하여 단어장으로 만들어 준다. '캠딕' 역시 영어 단어

나 문장이 포함된 이미지에서 어려운 단어를 골라 그 뜻을 찾아준다. 하지만 이러한 애플리케이션들은 영어로 이루어진 단어나 문장만을 촬영해야하기 때문에 실생활에서 직접 접하는 사물에 대한 영단어에 대한 호기심을 해결하기가 어렵다. 이에 비해 본 논문에서 개발한 애플리케이션은 별도의 사전 준비를 하지 않아도 실생활에서 접하는 사물에 대한 영단어를 언제 어디서나 학습할 수 있다는 장점이 있다. 이와 같이 이미지 인식 기능을 사용하여 영단어의 학습을 제공하는 애플리케이션은 아직까지는 존재하지 않는 것으로 파악된다.

3. 세부 기능 및 수행 예

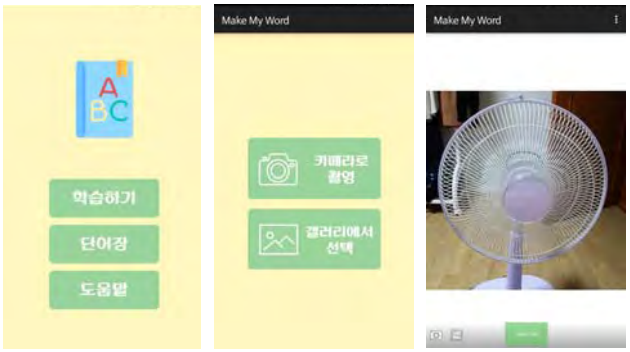
본 장에서는 본 논문에서 개발한 영단어 학습 애플리케이션인 'Make My Word'의 세부 기능과 실제 수행 예를 보인다. 'Make My Word'의 핵심 기능은 이미지 인식 기능으로 인식된 사물에 대한 단어를 사용자가 퀴즈형태로 맞추어 나간다는 것이다. 따라서 영단어 목록이 사전에 내장되어있는 전형적인 영단어 학습 애플리케이션과 달리 사용자의 능동적인 참여를 야기하며, 이를 통해 사용자의 쉽고 빠른 영단어 학습을 돕는다. 'Make My Word'의 세부기능은 다음과 같다.

● **이미지 인식 기능**: 사용자가 실시간으로 촬영하거나, 앨범에서 불러온 사진 속의 물체를 인식하여 그의 단어를 알아낸다. 이후 애플리케이션이 인식한 단어를 사용자가 맞추고 정답을 확인하는 형태로 학습이 진행된다.

● **정답 입력 기능**: 이미지 인식이 끝나면 사용자는 인식된 물체에 대한 영단어를 음성 또는 텍스트로 입력할 수 있다. 사용자가 음성으로 영단어를 입력하는 경우, 사용자의 발음과 가장 유사한 영단어 후보들을 출력하여 사용자가 이 중 하나를 선택하여 입력할 수 있도록 한다. 이후 정답확인을 통하여 사용자가 입력한 영단어의 정답 여부를 출력한다.

● **정답 확인 기능**: 사용자가 영단어를 입력하면 정답 여부를 확인하고, 이미지 속 사물의 명칭을 영어로 출력함과 동시에 해당 단어의 한국어 뜻을 알려준다. 또한 단어의 영어 발음을 들려줌으로써 사용자가 영단어의 발음도 학습할 수 있도록 돕는다.

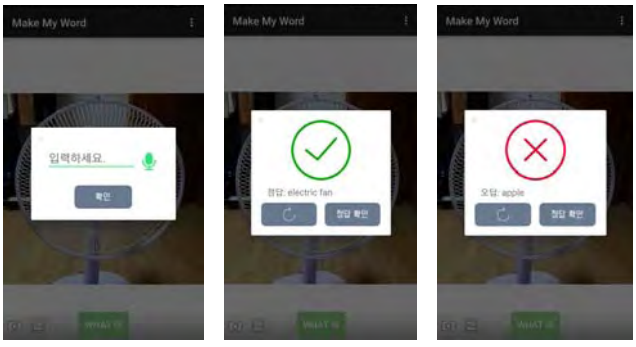
● **단어장 기능** : 본 애플리케이션은 단어장 기능을 제공하여 사용자가 학습한 단어를 따로 저장할 수 있다. 단어장 메뉴에서는 사용자가 이미지 인식 시 사용했던 이미지와 그에 대한 영단어가 단어카드 형태로 보이며, 발음 또한 다시 학습할 수 있다.



(a) (b) (c)

(그림 1) 시작 및 이미지 인식 화면

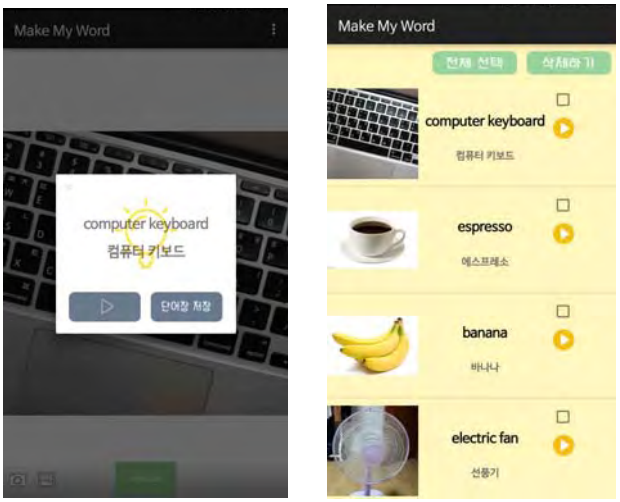
이제 지금까지 설명한 세부 기능의 실제 수행 예를 보인다. (그림 1(a))는 본 애플리케이션의 시작화면이다. 사용자는 이 화면에서 세 가지 주요 기능(학습하기, 단어장, 도움 말) 중 하나를 선택할 수 있다. 사용자가 '학습하기' 버튼을 선택하면 (그림 1(b))와 같은 화면이 나타낸다. 사용자는 카메라로 직접 촬영하거나, 사용자의 사진첩(갤러리)으로부터 사물을 인식할 이미지를 선택할 수 있다. (그림 1(c))는 사용자가 선택한 이미지를 인식하는 화면이다. 화면 하단의 'WHAT IS' 버튼을 클릭하면 이미지 인식이 시작되며, 분석을 마치면 (그림 2)의 화면으로 넘어간다. 현재 본 애플리케이션은 하나의 이미지에 여러 개의 사물이 나타나는 경우 그 중 단 하나의 대표 물체(가장 크고 이미지 중심에 가까운 물체)만을 인식한다.



(a) (b) (c)

(그림 2) 정답 입력 및 확인 화면

(그림 2(a))는 애플리케이션이 인식한 이미지 속 사물의 영단어를 사용자가 맞추어 보도록 유도하는 화면이다. 사용자가 영단어를 입력할 때는 키보드 입력 외에 마이크 버튼을 선택하여 음성 인식 기능으로 영단어를 입력할 수 있다. 이 경우 사용자가 발음한 영단어를 인식하여 사용자의 발음과 가장 유사한 영단어 후보들을 출력, 사용자가 이 중 하나를 선택하여 입력할 수 있도록 한다. 동일한 사물에 대해서 여러 명칭이 있을 수 있기 때문에 (예: computer keyboard, keypad) 사용자가 입력한 영어 단어가 여러 정답 중 하나와 일치한다면 정답으로 처리하여 정답 화면((그림 2(b))을 출력한다. 그렇지 않을 경우 오답으로 처리하여 오답 화면((그림 2(c)))을 출력한다. (그림 2(b)) 또는 (그림 2(c))에서 사용자가 '정답 확인' 버튼을 누르면 (그림 3(a))와 같은 정답 확인 화면으로 넘어간다.



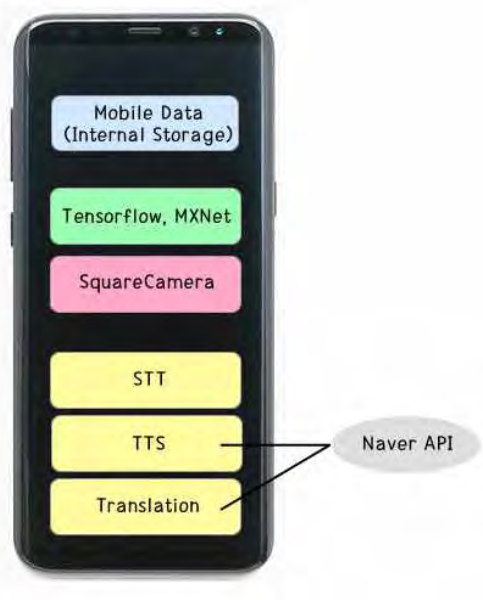
(a) (b)

(그림 3) 정답 확인 및 단어장 화면

(그림 3(a))는 애플리케이션이 인식한 정답을 보여주는 화면으로, 인식한 물체의 명칭을 영어와 한국어 뜻으로 출력하여 보인다. 또한 발음버튼(▷버튼)을 클릭할 경우 영단어가 음성으로 변환되어 출력되므로 사용자가 영단어의 발음을 학습할 수 있다. 사용자가 '단어장 저장'을 클릭할 경우 해당 단어와 이미지가 단어장에 저장된다. 저장된 단어는 (그림 3(b))와 같은 단어장 화면에서 확인할 수 있다. (그림 3(b))는 사용자가 시작 화면에서 '단어장' 버튼을 클릭할 경우 나타나는 화면이다. 좌측에는 사용자가 선택한 이미지가 보이고 우측에는 해당 이미지에서 인식된 물체에 대한 영어 단어와 한국어 뜻이 함께 보인다. 또한 발음버튼을 클릭할 경우 영어 단어의 발음을 다시 학습할 수 있다.

4. 구현 기술

본 장에서는 'Make My Word' 애플리케이션 개발에 사용된 구현 기술들을 설명한다. 본 안드로이드 애플리케이션은 Android Studio에서 Java를 사용하여 구현되었다.



(그림 4) 시스템 아키텍처 구성

(그림 4)는 Make My Word의 시스템 아키텍처를 보인다. 본 애플리케이션은 딥러닝에 기반한 이미지 인식 기능을 구현하기 위해 Google의 오픈소스 라이브러리인 TensorFlow를 사용하였다. 음성 인식(speech-to-text)은 안드로이드에서 제공하는 기능을 이용하였으며, 음성 합성(text-to-speech), 번역(translation) 기능은 네이버 오픈 API를 사용하여 구현하였다. 또한 단어장에 관한 데이터(사진, 영어 텍스트, 한글 텍스트)는 안드로이드 기기 내부 저장소에 저장한다. 본 논문에서 개발한 안드로이드 앱은 다음과 같이 크게 다음 6개의 모듈로 구성되며 각 모듈의 세부 기능 및 구현 방법은 다음과 같다.

● **사진 선택 모듈**: SquareCamera[4]를 사용하여 구현되었으며, 탭 하여 초점 맞추기, 두 손가락 확대/ 축소, 전면 및 후면 카메라, 플래시모드, 인물사진 및 풍경사진모드, 사진저장/보기에 대한 런타임 권한 지원 등이 제공된다.

● **이미지 인식 모듈**: 사용자가 사진첩에서 선택하거나 촬영한 사진을 내부에서 bitmap 형식으로 변환한 뒤, 딥러닝을 위한 Google의 오픈소스 라이브러리인 TensorFlow를 사용하여 가장 유사성이 높은 사물들에 대한 영어 단어 텍스트들을 얻어온다. 이미지 인식에 사용되는 neural network model은 Apache에서 제공하는 MXNet[5]을 이용하였으며, 이 모듈의 구현을 위해 GitHub 소스[6]을 참고하였다. 또한 추후에 이 모듈의 이미지 인식률을 높이기 위해서는, neural network model을 저장하고 있는 기존 파일을 추가적으로 훈련된 새 neural network model을 저장하고 있는 파일과 교체하기만 하면 되므로 유지 보수에 용이하다.

● **음성 인식 모듈**: 사용자가 발화한 음성을 실시간으로 인식하여 텍스트로 전환시켜주는 작업을 수행한다. 이러한 speech-to-text 기능 구현을 위해서 안드로이드 내부에서 제공하는 기능을 사용하였다.

● **음성 합성 모듈**: 텍스트를 입력으로 받아 해당 영어 텍스트의 내용을 음성으로 출력 하는 작업을 수행한다. 또한 출력한 음성을 내부 저장소에 저장하는 기능도 수행한다. 이러한 text-to-speech 기능 구현을 위해서 네이버 오픈 API[7]를 사용하였다.

● **영단어 번역 모듈**: 영어 텍스트를 입력으로 받아 해당 텍스트를 한글로 번역한 것을 출력하는 작업을 수행 한다. 이러한 번역 기능 구현을 위해서 네이버 오픈 API[8]를 사용하였다.

● **단어장 저장 모듈**: 안드로이드 기기 내부의 내부 저장소에 데이터베이스를 생성하고, 데이터를 삽입하거나 삭제하는 작업을 수행한다. 이 모듈은 사용자의 단어장을 저장하는데 사용된다.

본 애플리케이션에서 상기 모듈들은 다음과 같은 순서로 수행된다. **(1) 사진 선택 모듈**을 사용하여 이미지를 촬영하거나 선택하고, **(2) 이미지 인식 모듈**을 이용하여 해당 이미지 속의 물체를 인식한 뒤, **(3) 음성 인식 모듈** 또는 텍스트 입력을 통해 사용자로부터 영단어를 입력받고, (4) 사용자가 '정답 보기'를 선택하면 **영단어 번역 모듈**을 통해 영단어를 한국어로 번역하여 보여준다. (5) 사용자가 영단어의 발음을 듣기 원하는 경우, **음성 합성 모듈**을 통해 음성을 출력하고, (6) '단어장 저장' 버튼을 누르면 **단어장 저장 모듈**을 통해 사용자 기기의 내부 저장소에 단어 및 관련 이미지를 차례대로 저장한다.

5. 결론

본 논문에서는 딥러닝에 기반한 이미지 인식 기능을 사용하여, 사용자가 주변의 사물들에 대한 영단어를 손쉽게 학습할 수 있는 영단어 학습 애플리케이션을 개발하였다. 본 논문에서 개발한 애플리케이션은 사용자가 원하는 사물의 사진을 촬영하거나 사진첩에서 원하는 사진을 선택한 후 딥러닝에 기반한 이미지 인식 과정을 거쳐 물체의 영어 명칭과 한글 뜻을 알려주며, 영단어의 발음을 읽어주기도 하고 나만의 단어장에 저장할 수도 있다. 본 애플리케이션을 통해 사용자는 원하는 사물의 영어 명칭을 쉽고 빠르게 알 수 있다. 따라서 본 애플리케이션은 영어가 필수가 되어버린 현대 사회에 실생활에 적용할 수 있고 능동적인 영어 학습을 돕는 편리한 도구가 될 것으로 기대한다.

Acknowledgment

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.NRF-2015R1C1A1A02037071).

참고문헌

[1] TensorFlow, https://www.tensorflow.org/
[2]찍사,https://play.google.com/store/apps/details?id=com.sejong.zzicsa
[3]캠딕,https://play.google.com/store/apps/details?id=com.limo.hackerdic
[4]SquareCamera, https://github.com/boxme/SquareCamera
[5] MXNet, https://mxnet.incubator.apache.org/
[6] ImageRecognizer-Android, https://github.com/dneprDroid/ImageRecognizer-Android
[7] Naver, "네이버 오픈 API", https://developers.naver.com/products/intro/plan/
[8] Naver, "네이버 오픈 API", https://developers.naver.com/products/intro/plan/