

딥러닝을 이용한 시각장애인 보조도구

김영준, 윤종근, 허재혁, 신재호, 강우철
 인천대학교 임베디드시스템공학과

{youngjun0830,thinkvision,gjwogur71,vlovev1011,wchkang}@inu.ac.kr

An auxiliary mechanism for vision obstruction using DeepLearning

Youngjun Kim, Jonggeun Yun, Jaehyuk Hur, Jaeho Shin, Woochul Kang
 Embedded System Engineering, Incheon University

요 약

우리나라에 장애인 인구의 10% 정도인 약 25만 명의 사람들이 살아가고 있다[3]. 그러한 분들을 위한 여러 복지와 편의시설이 만들어지고 있지만 아직 도로를 안전하게 다니기에는 미흡한 부분이 많다. 시각장애인들이 좀 더 안전하게 생활을 할 수 있도록 하는 보조 장치를 제안한다. 사용자가 필요한 순간의 모습을 촬영한 뒤 딥 러닝으로 축적된 학습데이터를 이용하여 그 장면을 분석한다. 그 결과를 하나의 문장으로 표현하여 이어폰을 통해 사용자에게 서비스를 제공한다. 지원된 음성 서비스를 통해 시각장애인들이 걸어가는 길에 어떠한 장애물이 있는지 알려주어 위험한 상황에 놓이지 않고 안전하게 길을 걸어 다닐 수 있도록 보조해준다.

1. 서 론

전 세계에 많은 시각장애인들을 위해 여러 보조기구들이 마련되고 있지만 여전히 그들에게 불편한 것들이 많다. 그러한 문제점을 개선하기 위한 한 가지 방법으로 그들에게 눈과 같은 역할을 할 수 있는 보조기구가 필요하다고 느꼈다. 시각 장애로부터 안전을 보장받아야 할 뿐만 아니라 사람들과의 원활한 의사소통을 지향할 필요가 있습니다. “딥 러닝을 이용한 시각장애인 보조기구”를 통해 시각 장애인들이 좀 더 안전하고 편안한 생활을 할 수 있게 될 것이다.

시각장애인들에게 눈과 같은 역할을 할 수 있도록 그들이 바라보고 있는 방향에 어떠한 물체가 있는지를 파악하여 그들에게 정보를 주고 생활에 위험한 요소들로부터 좀 더 안전하게 생활할 수 있도록 도움을 주는 것을 목표로 한다. 이러한 기술들을 통해 시각 장애인에게 좀 더 운택한 삶을 제공하고 그들의 삶의 질을 향상시켜주는 것을 이 제품의 최종적인 목표로 둔다.

따라서 본 논문에서는 딥 러닝을 이용하여 학습된 데이터를 기반으로 시각장애인들이 일상생활을 함에 있어서 현재 앞에서 어떠한 상황이 벌어지고 있는지 어떠한 장애물이 있는지와 같은 정보를 제공하는 보조기구를 제안한다. 안경캠을 이용하여 현재 시야에 있는 이미지를 촬영하고 미리 학습시켜놓은 데이터를 이용해서 그 이미지를 분석하여 이미지 캡셔닝(captioning)을 하여 상황이나 물체에 대한 정보를 문장으로 얻어 내고 그 문장을 한글로 번역한 뒤에 그 한글 문장을 블루투스 이어폰을 통해서 음성으로 지원한다.

2. 본 론

딥 러닝을 이용한 시각 장애인 보조기구의 전체적인 흐름은 그림 1과 같이 구성한다.

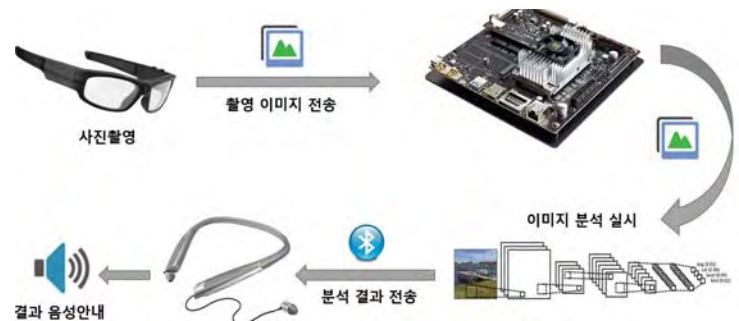


그림 1. 전체적인 흐름

딥 러닝을 이용한 시각 장애인용 보조기구의 시스템을 구성하는 각 컴포넌트들은 다음과 같은 기능을 수행한다.

2.1 딥 러닝 모델 학습

딥 러닝을 이용하여 물체를 분석하고 이미지 캡셔닝(captioning)을 하기 위해서는 많은 양의 DataSet이 필요하고 그 DataSet을 학습시켜서 결론을 도출하는데 사용해야 한다.

우선 DataSet은 Google open dataset과 직접 안경캠을 착용하여 촬영한 Image를 이용하였다. 그리고 이미지 캡셔닝(captioning)을 하기위해서는 이러한 DataSet을 학습시켜주어야 하는데 CNN(Convolutional Neural Network)과 RNN(Recurrent Neural Network) 과정을 거쳐 딥 러닝

모델을 생성하였다. DataSet이 CNN의 Input으로 들어가고 CNN에서 처리되어 나오는 Output이 RNN, 정확히 말하면 LSTM(Long Short-Term Memory)의 Input으로 들어가서 최종적으로 원하는 text 문장을 완성시켜 딥 러닝 모델을 만들어준다. 그리고 만들어진 딥 러닝 모델을 실제 휴대하며 이용할 Jetson Tx2 보드에 Copy하여 사용한다.

2.1.1 CNN

딥 러닝의 주된 기술인 CNN(convolutional neural network) convolution layer와 pooling layer의 구조로 이루어져 있는데 이를 통해 model parameter를 줄여 주어 model complexity가 감소하는 효과를 얻을 수 있다.

convolution layer의 핵심 기술은 sparse weight, parameter sharing, equivariant representations 세 가지로 요약 할 수 있습니다.

sparse weight의 장점은 특징을 찾는 local을 줄여 model complexity를 줄일 수 있다. parameter sharing은 sparse weight와 함께 쓰이면서 최적화된 network를 learning 할 수 있도록 해준다. 아래의 사진을 보면 오른쪽 그림은 하나의 특징을 위해 여러 가지를 고려 하지만 왼쪽 그림은 고려할 특징을 줄여 효율적으로 작동 할 수 있게 한다.



그림 2. Sparse weight

컴퓨터는 이미지가 조금만 돌아가거나 살짝만 바뀌어도 픽셀 단위로 이미지를 판별하기 때문에 다른 결과가 나오게 된다. 하지만 equivariant representations은 약간 바뀌어도 적절한 값을 가질 수 있도록 도와준다.

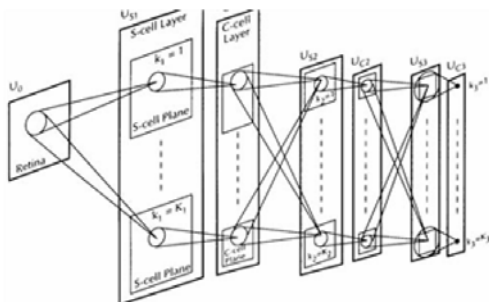


그림 3. Architecture of the convolutional neural network[1]

convolution layer는 그림 3에서 하나의 단계를 표현하지만 실제로는 여러 filter 단계를 걸쳐서 feature map을 표현 해준다[2].

pooling layer는 convolution layer만 사용해서 feature map을 얻을 수 있지만 더 낮은 차원의 feature map을 얻기 위해 사용된다.

2.1.2 RNN

CNN에서 나오는 Output이 RNN, 정확히 말하면 LSTM(Long short-term memory)의 Input으로 들어갑니다. image를 text로 바꾸기 위해서는 sequence를 학습하는 것이 중요한데 sequence를 학습하기 위해서 sequence를 입력으로 다 넣어준다면 입력 사이즈와 Neural Network 사이즈가 너무 커지게 되므로 RNN은 sequence를 모두 입력으로 주지 않고 이전 입력 값들을 계속해서 넣어 주면서 잘 조합되어진 메모리를 사용한다. 그러나 RNN은 계열이 길어지면 계층이 깊어지므로 Vanishing Gradient가 발생하여 Training이 되지 않는다. 그리하여 RNN을 변형시킨 LSTM은 각 데이터의 입출력 및 과거 데이터의 사용/미사용을 제어함으로써, 긴 계열을 취급 가능하게 한다.

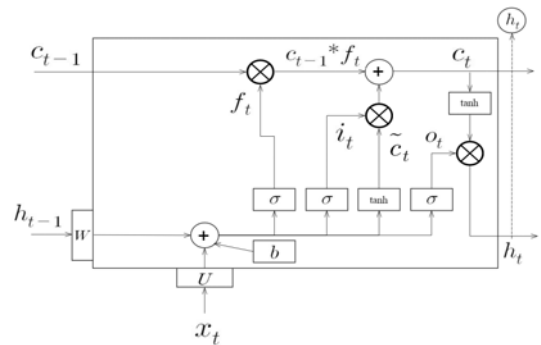


그림 4. Long-Short Term Memory[4]

그림 4에서 f는 과거 계열 데이터의 사용/미사용을 제어하는 Forget Gate이고, i는 입력 데이터의 사용/미사용을 제어하는 Input Gate이고, o는 출력 데이터의 사용/미사용을 제어하는 Output Gate를 의미한다.

LSTM은 결과가 원하는 정보가 아니거나 기대했던 수치와 동떨어져 있는 경우, 그 오차를 출력 쪽에서 역방향으로 분석하게 한 후 각 뉴런의 오류를 정정하여 오차를 줄이는 방법인 오차역전파법(Backpropagation)을 통하여 정확도를 높일 수 있다[2].

최종적으로 원하는 text문장이 완성 될 때까지 위와 같은 과정을 거치면서 이미지 캡셔닝(captioning)을 한다.

2.2 이미지 촬영

시각 장애인이 현재 앞에 상황이나 어떠한 물체가 있는지 알고 싶을 때 분석을 하기 위한 이미지가 필요하다. 안경에 캠을 달아서 현재 바라보고 있는 시야를 촬영하여 이미지를 얻는다. 그리고 계속해서 무분별하게 이미지를

촬영하여 불필요한 분석을 하는 것을 방지하기 위해서 사용자가 분석을 원할 때 버튼을 클릭하여 촬영을 한다. 이렇게 안경캠을 이용하여 얻은 이미지는 Tx2보드에 저장되고 촬영한 이미지를 앞서 Copy 해놓은 딥 러닝 모델을 통하여 분석을 해준다.

2.3 한글지원

분석이 끝나면 이미지 캡셔닝(captioning)을 통해서 이미지에 대한 내용이 text화 되어서 나온다. 이 text화 된 내용은 영어문장으로 나오기 때문에 Naver Translate API를 이용하여 이를 한글로 번역한다.

2.4 음성지원

한글로 번역한 문장을 시각장애인에게 음성으로 지원을 해주기위해서 Naver TTS API를 이용한다. 시각장애인은 블루투스 이어폰을 착용하고 있으면 이어폰을 통해서 문장을 음성으로 지원받는다.

3. 실험 결과

본 논문에서 제안한 딥 러닝을 이용한 시각장애인 보조기구의 목적은 시각 장애로 인한 불편함과 위험성을 줄이고 시각장애인의 삶의 질을 향상시키는데 있다. 그림 5는 딥 러닝을 이용한 시각장애인 보조기구의 흐름도를 나타낸다.

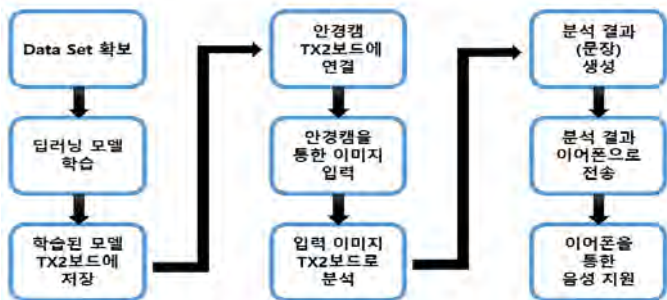


그림 5. 딥 러닝을 이용한 시각장애인 보조기구의 흐름도

그림 6은 사용자가 보조기구를 착용한 모습으로 사용자는 지팡이에 달린 촬영버튼을 눌러 안경에 부착된 카메라로 촬영을 하여 사용자의 시야와 동일한 상황이 분석되어 이어폰을 통하여 상황이 설명된다. 시각장애인들은 보조기구를 사용함으로써 위험하다고 판단이 되는 상황을 판단하고 피할 수 있고 눈이 보이지 않는 답답함을 해소할 수 있다.



그림 6. 착용 사진

그림 7에서는 보조 기구를 사용하여 노트북을 하고 있는 조원을 찍은 모습이다. 실험 결과로는 “젊은 소년은 노트북 컴퓨터 앞에 앉아 있다”라는 사진에서 두드러지는 특징을 잘 설명하는 것을 확인할 수 있다.

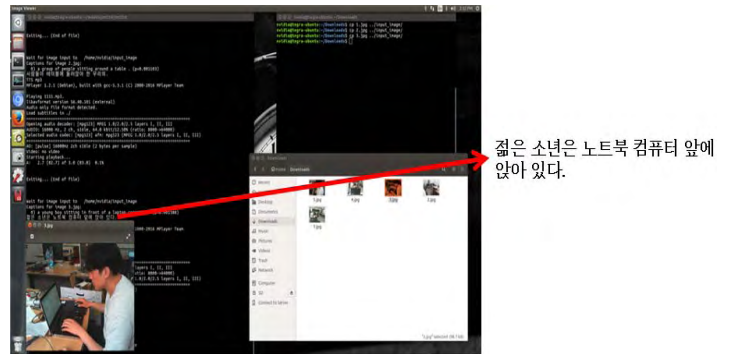


그림 7. 실험 결과

그림 8,9,10에서는 기존에 있던 이미지를 이용하여 이미지 캡셔닝(captioning)을 한 모습이다.



그림 8. 실험 결과



그림 9. 실행 결과

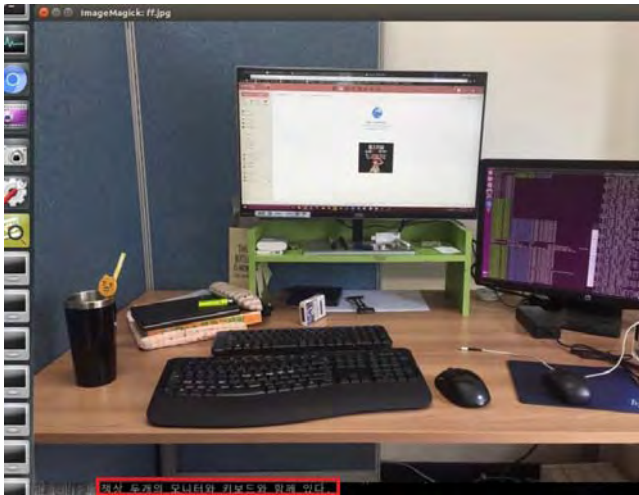


그림 10. 실행 결과

표 1은 위에 그림 8,9,10의 실행 결과가 어떻게 나왔는지 보여주고 있다.

표 1. 그림 8,9,10에 대한 실행 결과

	실행 결과
그림 8	그 곳에는 하얀 접시 파스타와 야채를 얹고 있다.
그림 9	한 남자가 손에 와인 한잔을 잡고 있다.
그림 10	책상 두 개의 모니터와 키보드와 함께 있다.

4. 결 론

시각장애인들이 길에서 마주할 수 있는 문제점들이 점점 증가한다. 예를 들어, 길에서 시각장애인의 길잡이가 되어주는 점자블록이 미관을 해친다는 이유로 점점 사라지고 있다. 그로인해 발생하는 사고가 증가하고 볼라드(차량 진입 억제용 말뚝)와 같은 장애물에 부딪히는 사고도

증가하고 있다. 길에서 시각장애인들의 사고 발생률을 줄일 수 있는 보조기구를 제안함으로써 그들의 삶의 질을 향상시킬 수 있다.

향후 “딥 러닝을 이용한 시각장애인 보조기구”는 GPS 기능을 추가하여 사용자의 위치 또는 주변 가게이름을 알 수 있게 하여 시각장애인이 자신의 위치를 파악할 수 있도록 할 계획이다.

참고 문헌

- [1] Hyeon-Joong Yoo. (2015). Deep Convolution Neural Networks in Computer Vision. IET Transactions on Smart Processing & Computing, 4(1), 35-43.
- [2] 이은주. (2017). CNN과 RNN의 기초 및 응용 연구. 방송과 미디어, 22(1), 87-95.
- [3] 한국장애인고용공단, “한눈에 보는 2016 장애인통계”, 2016.8.31.
- [4] 김양훈, 황용근, 강태관, 정교민. (2016). LSTM 언어 모델 기반 한국어 문장 생성. 한국통신학회논문지, 41(5), 592-601.