시각 장애인을 위한 상황 묘사 어플리케이션

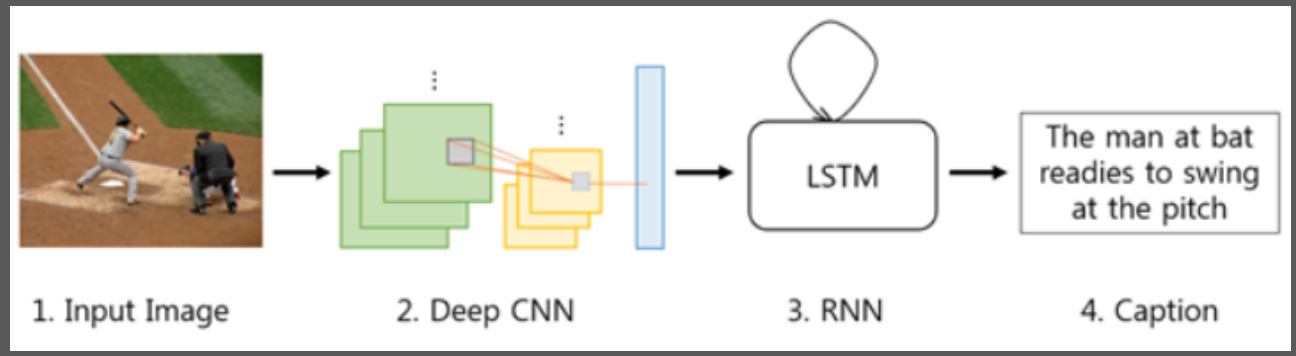
Team 4
정준호(201113275)
전민규(201411802)
김도연(201614157)

‘시각 장애인을 위한 상황 묘사 어플리케이션’ 이란?

- 시각 장애인분들을 위해 카메라로 찍히는 사진의 상황을 글 및 소리로 묘사
- 시각 장애인들이 스마트폰을 달고 걸어가면 30초마다 한 장씩 사진을 찍어 그에 대한 설명을 음성으로 들려줌
- Model
  - Input : 사진 1장
  - output : 사진의 내용을 설명하는 단어(및 형태소)들이 나오고, 그 단어(및 형태소)들을 이어 붙여 문장을 만듦

$$y = y_1, \cdots, y_C, y_i \in \mathbb{R}^K$$

Y : caption, K : vocab의 size, C : caption의 길이



※출처:https://brain.kaist.ac.kr/research.html

새로 만들 SW

- 이미지 캡셔닝 모델 구현
  이미지를 처리하는 CNN과 sequential한 정보를 생성할 LSTM을 seq2seq 구조로 이어 붙여 모델을 구현할 예정
- UI를 보여줄 앱

COTS SW

- TTS API

HW

- 스마트폰(보여줄 화면, 카메라, 스피커)

최종 산출물의 형태 및 기능

- APP
- 사진 한 장을 입력으로 넣었을 때 그 사진에 대한 설명이 텍스트 형태로 나오고 음성 메시지로 들려줌

Related Work

- Show attend and tell (이미지 캡셔닝의 기본 논문들)
- 기존 방법들은 방대한 영어 데이터를 바탕으로 학습된 모델이며 실제 한국어에 적합하게 만들어진 모델은 거의 없음(성능도 매우 안좋음)
- 따라서 우리는 한국어를 위한 데이터를 새로 구축하고 한국어 특성에 맞게 학습을 하도록 함으로써 좋은 성능을 내는 모델을 구현할 예정임. 이는 한국어 자연어 처리 연구에도 큰 의미가 있음
- 아래의 실험은 약 10만개의 데이터를 번역하여 사용하였으며 우리는 더 적은 데이터로도 비슷하거나 더 좋은 성능을 내는 모델을 내는 것이 목표

표 1 학습데이터 유형별 실험결과

Model	B-1	B-2	B-3	B-4
어절 단위	0.289	0.190	0.141	0.111
의미형태소 단위	0.597	0.392	0.286	0.225
형태소 단위	0.615	0.430	0.322	0.251

Risk Analysis

- 데이터를 구축하는 데 어려움이 있음

  존재하는 한국어 데이터가 없고, 방대한 영어 데이터를 직접 번역기로 번역하고 그것이 제대로 된 번역인지 일일이 체크해야 하는 노동이 들어감

- 시간 내에 성능을 올릴 수 있을지의 관건

  현재 한국어를 위한 시도들은 있지만 Related work에서 보인 것 처럼 성능이 매우 좋지 않은 것을 알 수 있음.

- 딥러닝과 pytorch에 대한 이해도

  정말 하고 싶은 주제지만 학부 수업에서 배울 기회가 없었기에 추가적인 공부가 필요

Success Criteria

- BLEU 0.6 이상 (더 적은 데이터(5만?)로 이전 방법과 비슷한 성능 기대)
- 30초 마다 한 번씩 찍고 10초마다 결과 나오도록