



LSTM을 활용한 동작 인식 프로그램

Motion prediction program using LSTM

1. Introduction

- 우리의 주제는 웹캠으로 사용자의 모션을 분석하는 모델을 설계했습니다.
- 모델을 설계하기 위해 LSTM(Long Short-Term Memory)을 활용하였습니다.

2. Background

단순한 이미지분석을 넘어 모션인식에 대한 연구가 활발히 연구가 진행 중입니다.

2015년 BMW에서 물리적으로 버튼을 누르지 않고도 차량을 조작할 수 있는 제스처 컨트롤 기능을 선보였습니다. 뿐만 아니라 애플리케이션을 통한 재활치료, 헬스 등 Motion recognition, HAR(Human Activity recognition)분야에 대한 관심과 발전이 계속되고 있습니다.

이러한 트렌드에 발맞추어 저희는 20BN-jester dataset을 사용하여 8가지의 손 동작과 다른 동작을 구분할 수 있는 모델을 만들었습니다.

3. Enviornment

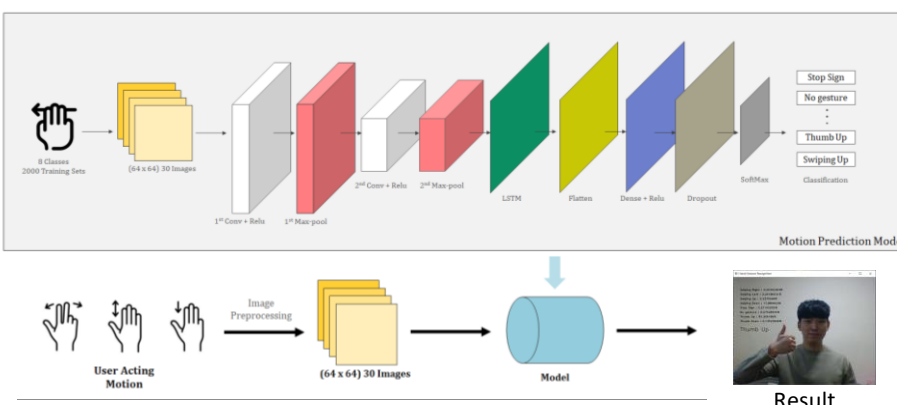


Tensorflow, Keras, StandardScaler, Numpy, Matplotlib, Pandas, CV2 를 사용하였습니다. 신경망 구성을 위하여 Tensorflow에서 지원하는 Conv3D(), Maxpool3D(), ConvLSTM2D(), Flatten(), Dense()을 사용하였습니다.



28개의 동작을 148,092개의 연속적 사진 데이터로 분류한 Jester dataset을 활용하였습니다.

4. Program Structure



위는 학습 데이터를 기반으로 만든 학습 모델입니다. 8개의 모델들의 각 1,500개의 영상을 30장의 이미지로 전처리 하여 학습을 진행하였고, 두 번의 Conv + MaxPool 을 통해 특징을 추출 하였습니다. Flatten Layer은 LSTM -> Flatten 으로 설정하였으며, Dense Layer 와 Dropout Layer을 통해 분류 모델을 설정하였습니다. 위 학습 모델을 바탕으로, 웹 카메라로 들어온 사용자의 이미지를 30장으로 전처리를 진행하고, 학습모델의 입력값으로 설정하여 화면에 학습 모델의 예측값을 출력합니다.

5. Conclusion

Labels	Thumb Down	Thumb Up	Swiping Down	Swiping Up	Swiping Left	Swiping Right	Stop Sign	No gesture	Model Test
Accuracy	79.3%	51.3%	74.6%	70.0%	86%	86%	50.0%	76.6%	87.8%

각 라벨마다 150개의 검증 데이터를 바탕으로 위와 같은 정확도가 집계되었습니다.

기대효과로 더 많은 라벨들을 확대하여 정확도를 높인다면, 다양한 환경에서의 유용한 모델이 될 것입니다.