청각 장애인을 위한 의료 기관에서의 쌍방향 소통 웹페이지 개발

김도하 [1], 김도희 [2], 송여진 [3]
[1,2,3] 이화여자대학교 휴먼기계바이오공학부 학부생

klavday@ewhain.net , kdh3022@ewhain.net , syj1031@ewhain.net

Interactive Communication Web Service in Medical Institutions for the Hearing Impaired

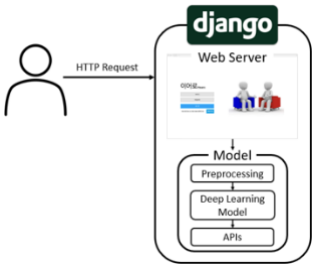
Kim Doha[1], Kim Dohee[2], Song Yeojin[3]
[1,2,3] Dept. of Mechanical and Biomedical Engineering, Ewha Womans University

요 약

청각장애인은 수화 언어, 즉 수어를 통해 의사소통한다. 따라서 본 논문에서는 의료 상황에서 청각 장애인이 겪는 소통의 어려움을 해결하기 위해 의료 상황 중심의 수어 데이터셋을 구축한 뒤, R(2+1)D 딥러닝 모델을 이용해 수어 동작을 영상 단위로 인식하고 분류할 수 있도록 하였다. 그리고 이를 Django를 이용한 웹 사이트로 만들어 사용할 수 있게 하였다. 이 웹페이지는 청각장애인 개인 뿐만 아니라 의료 사회 전반적으로 긍정적인 효과를 줄 것으로 기대한다.

1. 서론

청각장애인은 수화 언어, 즉 수어를 통해 의사소통 한다. 특히 의료 상황에서 자신의 증상에 대해 수어 통역 없이는 전달하기 힘들다. 수어 통역이 우선으로 필요한 영역으로 의료라고 응답한 비율이 38.9%로 가장 높게 나타났다[1]. 본 논문은 의료 상황에서 청각장애인과 의료기관 간의 의사소통 문제점을 해소하는 웹사이트 서비스를 제안한다. 기존의 수어 번역 알고리즘은 ASL(American Sign Language) 데이터를 기반으로 하거나[2] 좌표를 추출하여 학습시킨다.[3] 본 연구에서는 기존 연구의 한계점을 극복하고자 한국 수어 데이터를 바탕으로 3D CNN 모델을 활용한 청각장애인과 의료기관의 쌍방향 의사소통 웹사이트 서비스를 제시한다.



(그림 1) 시스템 구조도.

2. 연구 진행 방향

본 논문에서는 의료 상황에서 청각장애인이 겪는 소통의 어려움을 해결하고자 웹사이트 서비스를 고안하였다. 따라서 의료 상황에서의 데이터를 중심으로 데이터셋을 구축하였다. 기존 연구에서는 주로 ASL 데이터를 기반으로 모델 학습과 수어 인식을 진행해 한국 수어에 대한 연구가 부족하였다. 또한 기존 연구에서는 MediaPipe 등의 활용을 통해 좌표를 추출하는 추가적인 단계를 수행하여 전체적인 진행 시간이 길어진다는 문제점을 지닌다. 수어 번역 프로그램을 개발한 기존 연구에서는 영상을 프레임 단위로 분할하여 이미지 단위로 정보를 처리한다. 이는 영상의 연속적인 정보가 일부 소실된다는 문제점이 있다. 따라서 본 연구는 수어 번역 웹사이트에서 영상의 전체적인 정보를 활용하여 수어 번역을 진행한다. 이를 통해 필요한 정보를 더 효과적으로 보존하고 활용할 수 있다.

3. 데이터셋 및 전처리

AI hub 의 '수어 데이터셋' 영상[4]에서 의료기관에서 자주 사용되는 단어 15 개(가슴, 귀, 너무 아파요, 머리, 목, 무릎, 발, 발가락, 발목, 배, 손가락, 손목, 어깨, 팔꿈치, 허리)를 선별하여 이에 해당하는 영상을 한 단어당 20 개씩 다운로드 받아 모델 학습에 활용하였다. 또한, 모델의 overfitting 을 방지하기 위해 팀원 모두 수어 영상을 개별적으로 촬영하여 한 단어 당 영상의 개수를 41 개로 증가시켰다.

데이터셋 영상의 길이가 모두 달라서 모델에 넣기 위해서는 프레임 수를 통일할 필요가 있었다. 따라서 모든 영상의 프레임 수를 16 으로 통일하는 exponential interpolation 작업을 진행하였다. 여러 interpolation 방법 중 영상의 유의미한 정보인 중앙의 데이터를 강조하는 exponential interpolation 을 선택하였다. 또한 (R(2+1)D) 논문[5]의 데이터셋 크기 기준에 맞추어 128x171 사이즈로 resize 를 진행하고, 112x112 사이즈로 crop 을 진행하였다.

4. 제안 방법

1) YOLOv5

YOLO 는 Backbone 으로는 CSPNet 기반의 CSP-Darknet

을 사용하여 이미지에서 특성 맵을 추출하고, head 에서는 추출된 특성 맵을 바탕으로 객체의 위치를 찾는다. 이러한 기법을 통해 빠른 속도로 객체 탐지 및 분류를 진행할 수 있다.

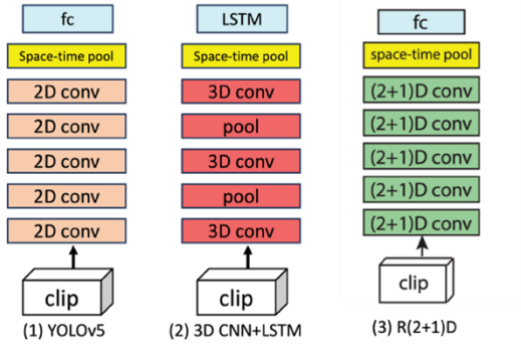
2) 3D CNN+ LSTM

CNN 에 LSTM 을 결합한 모델로, pre trained 된 C3D 모델을 CNN 부분에 사용하고, 모델의 마지막 fully connected 레이어를 제거한 뒤 그 부분에 LSTM 을 결합하였다. CNN 을 이용해 이미지나 영상의 feature 를 추출하여 텐서 형태로 표현하고, 이러한 텐서는 LSTM 레이어로 전달되어 시퀀스 모델링을 수행한다.

3) R(2+ 1)D

ResNet 은 Residual Learning 을 통해 성능을 높이는 데 주력한 모델이다. 주로 사용된 기법은 layer 를 건너뛰어서 학습하는 skip connection 과 bottleneck layer 가 있다. 이를 통해 층을 깊게 쌓을 수 있고, 우수한 성능을 얻을 수 있다. 3D ResNet 은 이러한 ResNet 을 비디오 등의 3D 이미지에 적용할 수 있도록 변형한 모델이다.

(그림 2) YOLO v5[6], 3D CNN+ LSTM, R(2+ 1)D 모델 구조도.



5. 구현

실험의 결과는 (표 1) 과 같이 나타났다. YOLO 는 영상의 연속성을 고려하지 않고 이미지 단위로 학습하기 때문에 연속성이 중요한 단어들에 대한 정확도가 떨어지는 것을 알 수 있었다. 따라서 YOLO 를 채택하지 않았다.

R(2+ 1)D 모델의 Validation accuracy 가 CNN+ LSTM 모델보다 낮은 것을 볼 수 있지만, CNN+ LSTM 모델의 Validation accuracy 가 99%에 달하는 정확도를 보여 학습된 값에만 과한 적합을 보이는 overfitting 이 일어난 것으로 판단해 R(2+ 1)D 를 최종 선택하였다.

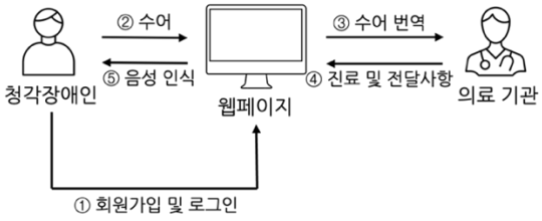
모델명	R(2+1)D	CNN+LSTM	YOLO
Validation Loss	0.5586	0.0343	Confusion Matrix 참조
Validation Accuracy	82.79%	99.18%	Confusion Matrix 참조
* 비고 CNN+LSTM 모델은 오버피팅된 것으로 판단되어 R(2+1)D를 최종 선택			

<표 1> 실험의 결과.

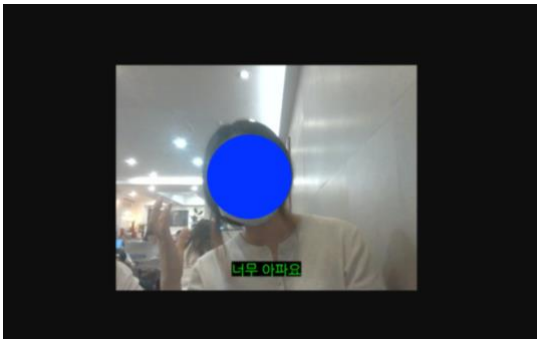
최종 결과물인 웹페이지의 기능 흐름도는 (그림 3) 과 같다. 초기 화면에서 회원가입(①)을 한 후, [진료 시작] 버튼을 눌러 수어 인식을 진행(②)할 수 있다. 웹페이지의 카메라가 환자의 수어를 영상 단위로 인식(③)하고, 인식된 수어는 하나의 완결된 문장으로 번역 및 생성된다. 이는 TTS(Text-To-Speech) 기술을 이용해 음성으로 변환되어 의료 기관에게 전달된다.

또한 청각장애인은 음성 인식 페이지로 넘어가, 의료 기관의 음성 인식(④)을 진행할 수 있다. 인식된 음성은 STT(Speech-To-Text) 기술을 이용해 청각장애인에게 텍스트로 전달(⑤)된다.

웹 페이지의 구현은 Python 기반의 무료 오픈소스 웹 애플리케이션 프레임워크인 Django 를 이용하여 진행하였다. STT 와 TTS 는 Google Cloud 에서 제공하는 API 를 활용하였다.



(그림 3) 웹페이지 기능 흐름도.



(그림 4) 웹페이지 시연 장면.

6. 결론

본 논문에서는 의료 상황에서의 수어 데이터를 사용하여 여러 딥러닝 모델들을 학습하였다. 또한 이를 하나의 웹 페이지에 업로드하여 모델이 수어를 영상 단위로 인식하고 분류할 수 있도록 하였다. 이 웹페이지를 통해 청각장애인이 직접 의료기관과 의사 소통함으로써 청각장애인의 만족도가 증가할 것이다.

※ 본 프로젝트는 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.

참고문헌

[1]이준우 연구 책임자, 한국수어 활용 조사 결과보고서, 국립국어원, 2020 년.

[2]김상우, 김인숙, 정진곤, "CNN 을 활용한 수화번역 시스템," 대한전자공학회 하계학술대회 논문집, 대한전자공학회, 2019 년.

[3]길상현, 이승훈, 오차영, 유승범, 한연회, "딥러닝 기반 자세 및 손 제스처 인식 기술을 활용한 병원 수어 번역 프로그램 설계 및 구현", 한국통신학회 추계종합학술발표회, 한국통신학회, 2021 년.

[4]https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=120&topMenu=100&aihubDataSe=extrldata&dataSetSn=264

[5] TRAN, Du, et al. A closer look at spatiotemporal convolutions for action recognition. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2018. p. 6450-6459.

[6] https://github.com/ultralytics/yolov5/issues/280