출품번호
1426

제69회 전국과학전람회

비수지 신호를 포함한 인공지능 기반 수어 번역 시스템

2023. 08. 14.

출품학생	
지도교원	
구 분	산업 및 에너지
출품부문	학생부

1. 서론

가. 연구 개요

코로나19 바이러스 감염증으로 인한 팬데믹 사태와 4차 산업혁명에 따라 보편화된 비대면 온라인 환경에서 농인들의 원활한 소통을 위하여 수어 인식 시스템을 만드는 것을 목적으로 한다. 수지 신호만을 수어 번역에 이용하는 기존 선행 연구들과는 달리 얼굴 표정, 얼굴 움직임 등으로 이루어진 비수지신호를 포함하여 수어를 단어 단위로 인식하여 해석할 수 있도록 연구를 진행한다. 본 연구에서는 KSL 수어 영상 데이터 세트에 대하여 Mediapipe 프레임워크를 사용하여 얼굴과 손의 특징점을 모두 좌표로 측정한 후 GRU 신경망을 이용한 단어 회귀분석 모델, 그리고 CNN 신경망을 이용한 표정을 통한 의문문 분석, 총 두 종류의 모델을 제작하였다. 단어 회귀분석 모델은 주어진 영상 - 단어 데이터들에서 정상적으로 작동하였으며, 표정으로부터 의문문/평서문을 구별하는 모델은 높은 정확도로 작동하였음을 확인할 수 있었다.

나. 연구 동기

2021년 기준 대한민국의 전체 장애인 수는 약 2,604명, 그 중의 청각장애인은 2번째로 많은 약 32만 명이다. 2020년부터 코로나19로 인해 온라인 화상 회의를 많이 이용하고 있고, 동영상 등을 통한 온라인 수업도 활발하게 진행되고 있다. 하지만 청각 장애인들에게는 이러한 상황이 굉장한 어려움으로 다가온다. 수화 콘텐츠들은 많지 않고, 수화를 사용하는 농인들은 화상 회의에서 속기 통역 등의 서비스를 이용해야만 하는 문제점이 존재한다.

농인은 소리로 언어를 터득하기에 어려움이 존재하므로, ‘수어’를 사용한다. 수어는 청각장애인이 손의 움직임과 표정, 몸짓 등 신체적인 신호를 이용해 의사를 전달하는 시각 언어이다. 흔히 손동작만으로 의미를 전한다고 알려졌으나, 표정 등 여러 가지 방법을 사용해 구사하는 체계적인 언어이다. 일반적인 한국어를 시각적으로 제공하면 의사소통이 편리할 것으로 생각하나, 농인들에게 모국어 즉 제1 언어는 ‘수어’이며, 청력에 문제가 없어 소리를 정상적으로 들을 수 있는 사람 (이하 청인으로 칭함) 들에 모국어인 한국어는 농인들에게 있어서는 제2 언어가 되는 셈이므로 한국어의 자막화는 정답이 될 수 없는 현실이다.

선행 연구 중 한국 수화의 번역에 관한 연구는 대부분 수지 신호만을 이용하였으며, 비수지 신호를 이용한 예도 단어 단위 번역의 정확도를 높이기 위해 사용하였다. 따라서 수지 신호뿐만이 아닌 비수지 신호를 이용해 문장의 전체적인 의미를 결정할 수 있도록 연구를 진행하기로 하였다.

다. 연구내용

본 연구는 청각장애인의 삶의 질 개선과 전체적인 불편함을 덜기 위한 수어 번역 핵심 기술의 구현을 취지로 수어 영상 데이터로부터 그것을 비수지신호와 수지 신호의 합성으로 단어화하는 모델과, 영상에서 표정을 분석하여 문장이 의문문인지, 평서문인지 구분하는 모델을 제작하였다.

2. 이론적 배경 및 선행 연구

가. 이론적 배경

1) 수화

가) 수화

청각장애인들은 소리로 말을 배우는 것이 힘들다. 따라서 이러한 청각장애인들을 위한 언어가 있는데, 그것이 수어이다. 청각장애인 중 수어를 일상어로 사용하는 사람을 농인이라고 하며, 이들에게 수어는 제 1 언어이다. 수어는 수지 신호와 비수지 신호로 구성되어 있다.

나) 수지 신호와 비수지 신호

수어의 두 가지 구성 요소 중 수지 신호는 손가락이나 팔로 그리는 모양, 그 위치나 움직임 등을 의미한다. 한국어, 영어 등의 구어에서의 단어와 비슷한 역할을 한다. 반대로 비수지 신호는 수지 신호를 제외한 얼굴 표정, 얼굴 움직임 등을 의미한다. 구어에서의 세기, 길이, 억양 등과 비슷한 역할을 하며, 의미 전달에서 중요한 역할을 한다. 농인들은 비수지 신호에 수지 신호 못지않게 비중을 두는데, 이는 수화의 형태론과 통사론 등의 문법적 역할을 담당하기 때문이다.

2) 특징점 인식 기술

가) Hand Pose Estimation

Hand Pose Estimation은 손의 모양을 인식하는 인공지능 기술이다. 이 연구에서는 구글의 오픈 소스 프레임워크인 Mediapipe를 이용하였는데 Mediapipe에서의 Hand Pose Estimation의 작동방식은 다 음과 같다. 먼저 BlazePalm이라는 딥러닝 모델을 이용하여 손바닥의 위치를 찾아낸다. 그 후 회귀 알고리즘을 이용한 21개의 손의 특징점 좌표를 찾아내는 모델을 통해 손의 특징점을 잡아낸다.

그림 1 Mediapipe의 Hand Pose Estimation에서의 손 특징점

나) Facial Key points Detection

Facial Mesh는 이미지에서 사람의 얼굴 부분을 추출하고 그 얼굴을 3차원으로 해석하여 얼굴

의 특징점들을 잡는 기술이다. Google에서는 media pipe 얼굴 감지 모델을 이용한다. Mediapipe에서는 얼굴 감지에 BlazeFace라는 다른 모델을 사용하며 이 모델에서는 2D / 3D 얼굴의 특징점 탐지와 영역분할 등의 기능이 존재한다. 연구상에서는 MediaPipe 모듈이 잡을 수 있는 총 얼굴의 468개의 점 중 수어 학습에 핵심적인 점 127개의 점을 추출하고, 이 점의 프레임별 변위를 저장하고 그를 학습시킨다.

다) Word Embedding

자연어를 딥러닝에 사용하기 위해서 단어를 숫자 혹은 벡터로 바꾸는 과정이 필요하다. 이러한 과정을 Embedding이라고 한다. 이런 임베딩 기법으로는 제일 많이 쓰이는 것으로 Word2Vector와 GloVe, FastText가 있는데 본 연구에서는 학습되지 않은 단어에 대해서도 유사도를 찾아내는 FastText를 사용할 것이다.

3) 딥러닝 신경망 종류

가) CNN

Convolution Neural Network(CNN 또는 ConvNet)은 이미지, 비디오, 텍스트, 사운드 등을 다룰 때 많이 사용되는 딥러닝 알고리즘이다. 이미지와 같은 2차원 데이터를 다룰 때 기존 방식대로 하면 이미지를 flatten 하여 1차원으로 만들어 학습시켰는데, 이렇게 하면 데이터의 지역적, 공간적 특징이 소실되기 때문에 학습이 잘 진행되지 않을 수 있다. 이런 문제를 해결하기 위해 고안된 것이 CNN이다.

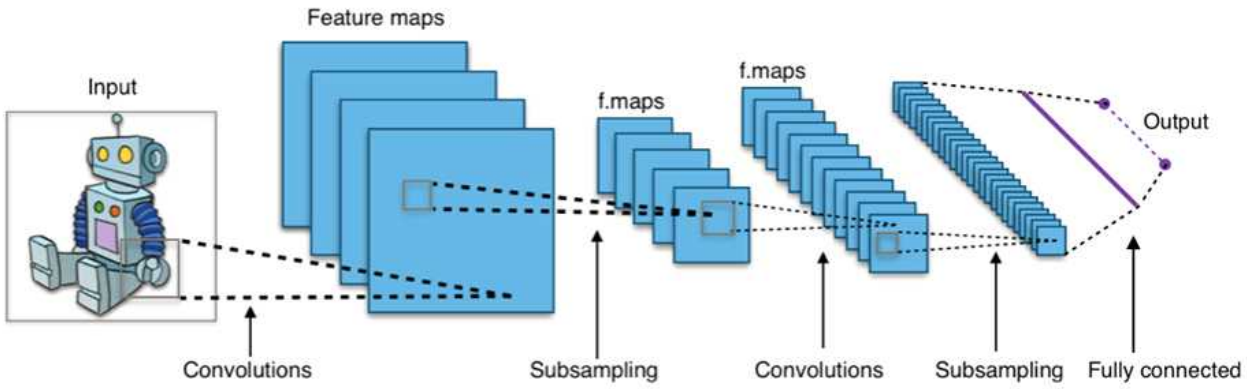

그림 2 CNN

CNN은 <그림 2>와 같은 방식으로 작동한다. 이미지가 들어오면 filter(혹은 kernel)를 이미지에 합성곱하여 feature map을 얻는다. 이러한 feature map은 기존 이미지의 공간적, 지역적 특징을 포함하기 때문에 기존 방식과는 달리 특징을 잘 보존할 수 있다. CNN은 학습을 통해 좀 더 유의미한 특성을 얻는 kernel을 찾아낸다.

이처럼 CNN은 이미지에서 객체, 얼굴, 장면을 인식하기 위해 패턴을 찾는 데 특히 유용하다.

자율 주행 자동차, 얼굴 인식 애플리케이션과 같이 객체 인식과 컴퓨터 비전이 필요한 분야에서 CNN을 많이 사용합니다. 본 연구에서 사람의 표정을 인식하여 문장이 의문문인지, 평서문인지를 판별하기 위한 모델을 개발하는 데 쓰인 가장 핵심적인 신경망의 구조이다.

나) RNN

Recurrent Neural Network(순환신경망)는 음성이나, 문장, 경제 지표, 날씨 데이터 등 순차 데이터나 시계열 데이터를 다룰 때 주로 사용하는 딥러닝 알고리즘이다.

대부분의 신경망은 활성화 함수를 거친 출력은 다음 층의 입력으로 들어가게 된다. 하지만 RNN의 경우 다른 구조를 갖는다. RNN은 아래 그림과 같이 자신의 출력이 다시 자신의 입력으로 돌아오게 되는 구조이다. 이러한 구조를 가진 덕분에 RNN은 연속된 데이터를 다루는 데에 유리하다.

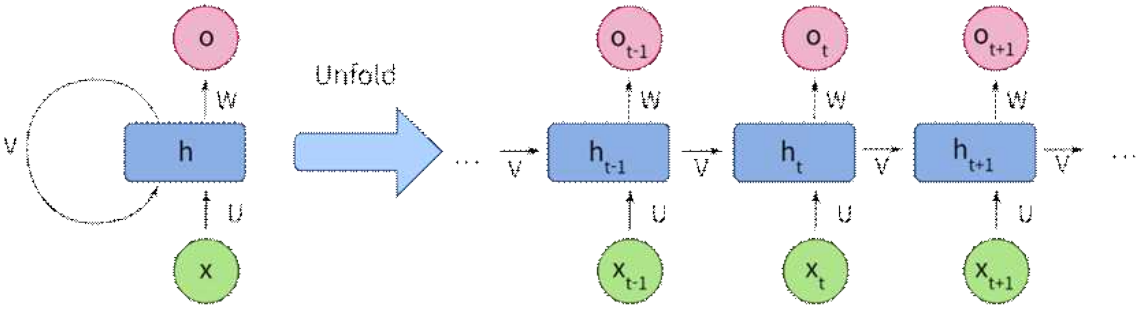

그림 3 RNN

다) GRU

RNN의 경우 input shape가 큰 경우 제일 시계열 시퀀스의 앞부분의 가중치와 뒷부분 시계열 시퀀스 데이터의 가중치가 같아서 데이터양이 커지는 경우 학습에 근본적인 문제가 있다. 그래서 이를 극복하기 위한 모델이 많이 등장했다. 대표적인 두 가지가 LSTM(Long Short Term Memory), 그리고 GRU이다. LSTM은 기존 RNN에 "망각"이라는 기능을 추가한 것으로, 중요치 않은 값을 잊어버리도록 학습되는 것이 포함된다. LSTM은 기존 RNN의 단점을 해결했으나 오히려 더 복잡한 구조를 가진다. 이 구조를 간단하게 개선한 바리에이션이 GRU이다. GRU는 이전 상태를 얼마나 반영할지 정하는 reset gate와 이전 상태와 현 상태를 얼마의 비율로 반영할지 정하는 update gate가 존재한다. 그리고 LSTM보다 더 간단한 연산 과정으로 효율적 학습을 진행할 수 있을 것이다. reset gate에서는 $r=\sigma\left(W_r \bullet [\, h_t-1,\ x_t\ ]\right)$를 거쳐 나온 r로 이전의 정보를 얼마나 잊어버릴지 정한다. update gate에서는 $u=\sigma\left(W_u[\, h_t-1,\ x_t\ ]\right)$를 거쳐 나온 u로 이전의 정보를 얼마나 유지할지 정한다. candidate에서는 hidden layer의 후보 $h_t=\tanh(W \bullet [\, r_t * h_{t-1,} x_t\ ])$를 계산하여 이전 스텝에서 뭘 지울지 선택한다. 그리고 마지막 hidden state에서 $h_t=(\,1-z_t\,)h_{t-1}+z_t h_t$ 를 계산하여 이전에 구한 candidate에 update gate를

적용하여 얼마나 이전 state가 영향을 끼치고 얼마나 candidate가 영향을 많이 끼칠지 정한다.

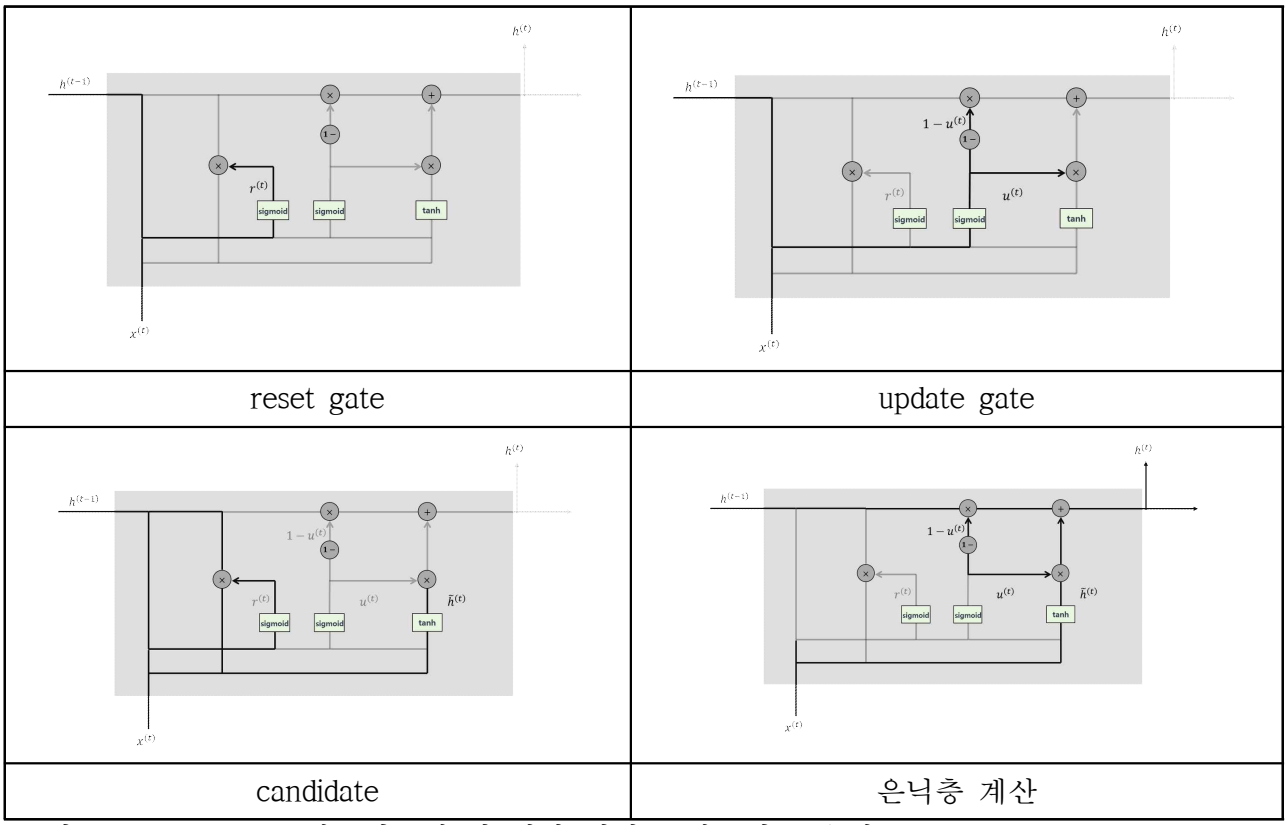

그림 4, 5, 6, 7 GRU 알고리즘의 각 연산 게이트 알고리즘 순서도

나. 선행 연구

1. 수지 신호 기반 수화 번역

가) 딥러닝 기반 OpenPose를 이용한 한국 수화 동작 인식에 관한 연구[1)]

해당 연구에서는 OpenPose를 이용하여 키포인트를 추출하고 RGB 데이터와 함께 Cross Entropy를 사용하여 CNN으로 학습하였다. 비수지 신호를 번역에 사용하기 위하여 얼굴 특징점을 함께 사용하였지만, 문장 전체 의미에 대해 학습한 것이 아니고 단순히 단어 각각에 대하여 학습하였다. 또한 CNN을 사용하여 비슷한 수어가 존재하는 경우 분류가 힘들어지고 서비스화 되기 힘들다는 단점이 존재한다.

나) LSTM을 활용한 수어 단어 인식[2)]

해당 연구에서는 Mediapipe로 추출한 골격 키포인트를 LSTM을 통해 학습하여 수화를 번역하였으나, 비수지 신호를 적용하지 않았다.

2. 수어 발화 시점 탐지

1) 김인혜, 정일홍.(2021).딥러닝 기반 OpenPose를 이용한 한국 수화 동작 인식에 관한 연구.한국디지털콘텐츠학회 논문지,22(4),681-687.

2) 정의손, 조동휘, 박세희, 강현아, 박승보.(2022).LSTM을 활용한 수어 단어 인식.한국컴퓨터정보학회 학술발표논문집 ,30(1),287-288.

가) 키포인트 기반 수어 시작 및 종료 시점 탐지 기법[3)]

해당 연구에서는 Mediapipe로 추출한 수지 신호 키포인트의 좌푯값 변화를 기반으로 수화 발화 시점을 탐지하는 수식들을 설계하였다. 다만 해당 연구는 하나의 수어가 포함된 영상에서 시작점을 탐지하는 기법으로, 여러 개의 수어가 연속된 영상에서의 시작점 탐지에 어려움이 있어 보인다.

나) Watch only once: An end-to-end video action detection framework[4)]

해당 연구에서는 Transformer의 Attention 기법을 활용하여 영상에서의 행동의 시작점과 끝점을 찾는 기법을 연구하였다.

3. 연구 진행

가. 연구 방법 및 절차

1) 연구 환경 구축

연구에서 딥러닝 학습을 위해 사용한 하드웨어와 소프트웨어 사양은 표 1에 정리되어 있다.

CPU	Intel Core i9-10900K Processor
GPU	Nvidia GeFore RTX3080 LHR 10G
Memory	DDR4 64GB
OS	Ubuntu 22.04
Python	3.10
Software	TensorFlow 2.9.2, CUDA 11.8, cuDNN 8.6.0

표 1 연구 환경

2) 데이터 세트 다운로드 및 선별

단어 번역을 위해서는 각 20여 개의 영상이 있는 77여 개 단어로 구성된 수화 데이터 세트인 연세대 Korean Sign Language Dataset) 을 이용하였다. 모든 영상은 1280*720 크기였으며, 영상의 길이는 제각각이었다. 학습을 위해 영상 품질이 낮거나 제대로 전처리 되지 않는 영상들은 제거하였다.

감정 분류를 위해서는 경상북도농아인협회 수화문화원에서 제작한 비수지기호 강의 영상을 단어 단위로 잘라 사용하였다.

---

3) 김근모, 조진성, 김광용, 김봉재, 전기만. (2023). 키포인트 기반 수어 시작 및 종료 시점 탐지 기법. 정보과학회 컴퓨팅의 실제 논문지, 29(4), 184-189.

4) Shoufa Chen, Peize Sun, Enze Xie, Chongjian Ge, Jiannan Wu, Lan Ma, Jiajun Shen, Ping Luo; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 8178-8187

3) 컴퓨터 비전, 자연어 처리 등 딥러닝 및 영상처리 이론 학습

프레임 상에서 얼굴의 표정을 인식하여 의문문과 평서문을 구별하기 위하여 CNN의 기본 구조와 다양한 응용을 해보며 학습하였고, 얼굴을 크롭하고 손의 특징점을 추출하고, 얼굴의 특징점을 추출하는 데에 사용되는 mediapipe 프레임워크의 사용법을 익혔다. 또한 시계열 데이터를 딥러닝에 사용할 것이기 때문에 시계열 데이터 처리에 적합한 RNN계열 신경망인 GRU, LSTM 등의 구조와 작동방식을 공부하였다. 그리고 자연어 처리해야 하므로 자연어를 처리하는데 필수적인 Embedding에 대해 공부하였고, 마지막으로 영상처리를 위해 OpenCV 라이브러리를 이용할 것이므로 OpenCV의 기본적인 사용 법과 응용법 또한 학습하였다.

4) 데이터 전처리

가) 손과 얼굴을 인식하여 단어로 번역하는 신경망

각 동영상을 초당 15프레임씩 이미지로 변환하였다. 수화마다 모두 영상 길이가 달랐기에 모두 일정 프레임씩만 이미지로 변환하였으며, 100프레임보다 적으면 모든 특징점의 변위를 0으로 가 정하였다. 변환된 각 이미지를 구글의 mediapipe의 Hand Pose Estimation과 Face mesh 모델을 사용하여 양손 특징점 21개씩 총 42개와 얼굴 특징점 128개를 추출하였다. 추출한 특징점들의 좌표를 바탕으로 프레임 간의 좌표의 x, y, z 방향 변위를 계산하여 넘파이 배열로 저장하였다. 모델이 예측해야 하는 각각의 단어들은 wordvectors GitHub 레포지토리의 Pre-trained korean FastText모델을 이용하여 임베딩하였다. 각각의 단어들은 모두 200차원의 임베딩 벡터로 변환된다.

나) 얼굴 표정을 인식해 의문문과 평서문으로 구분하는 신경망

각 동영상에서 mediapipe의 Face Detection 모델을 이용하여 얼굴 사진만 따낸 후 모두 50*50의 동일한 사이즈로 resize한다. 그 후 사진이 의문문에서 나온 것인지 평서문에서 나온 것인지 라벨링 한다.

그림 9 전처리된 얼굴 표정 데이터셋

5) 모델 구상 및 시각화, 모델 구현 및 평가

가) 모델 구상

모델은 Python3 와 Tensorflow를 통해 구현된다.

본 연구에서는 두 가지 모델을 사용한다. 첫 번째는 얼굴의 특징점들과 손의 특징점들을 입력으로 받아 하나의 200차원 임베딩 벡터를 출력하는 신경망으로 수어의 수지 신호와 비수지 신호를 이용해 단어의 뜻을 추론하는 모델이다. 두 번째 모델은 얼굴 사진을 입력받아 0과 1 사이의 실수를 출력하는 이진 분류 모델로, 이 문장이 의문문인지 평서문인지 판별한다.

나) 손과 얼굴을 인식하여 단어로 번역하는 신경망 모델

모델의 입력은 손의 특징점과 얼굴의 특징점을 리스트 형태로 묶어 입력받아 두 개의 신경망으로 나누어 입력이 들어가게 된다. 각각은 (10, 3×128)의 형태의 연속된 Facial Key points 정보와 (10, 3×42) 형태의 연속된 Hand Pose Keypoints이다. 두 입력은 별개의 신경망을 따라 진행한 후 더해져 하나의 층으로 합쳐진 후 다시 몇개의 추가적인 층을 거치고 FastText모델을 이용해 임베딩 된 워드벡터와 같은 200차원의 벡터를 예측한다.

다) 표정을 인식해 의문문과 평서문으로 구분하는 신경망 모델

모델의 입력으로 (50, 50, 3) 크기의 사진이 들어오면 CNN 층을 거쳐 이 사진이 의문문인지 평서문인지 판별한다. 5.4. 모델 평가 5.2와 5.3에서 구상한 모델을 구현한 후 loss를 줄이는 방향으로 모델을 수정한다.

4. 연구 결과

가. 모델의 구조

1) 손과 얼굴을 인식하여 단어로 번역하는 신경망

최종적인 모델의 전체 구조는 그림 9와 같다.

먼저 전처리 과정을 거친 얼굴의 특징점 128개의 10프레임 동안의 x, y, z축에 대한 변위 데이터(총 10×384)를 받아 input_face 층으로 입력된다. 그리고 전처리 과정을 거친 양손의 특징점 42개(각각 21개)의 10개 프레임 동안의 x, y, z 축에 대한 변위 데이터(총 10×126)를 받아 input_hand 층으로 입력된다.

각각의 입력은 64개의 unit을 가진 many-to-many GRU layer를 거쳐 64×10차원의 텐서로 바뀐다. 그 후 many-to-one simple RNN을 거쳐 32 크기의 벡터를 출력한다. 그 후 batch normalization을 거쳐 학습의 안정성 보장과 과적합 방지를 한다. 그 후 일반적인 FCN(Fully Connected Layer)를 거치고 한번 더 batch normalization을 해준 후 3개의 FCN을 거쳐 200차원의 벡터를 얻는다. 그리고 손과 얼굴 각각의 신경망의 결과인 200차원 벡터 두 개를 더해 한번 더 FCN을 거쳐 최종적인 예측을 출력한다. GRU와 SimpleRNN층을 제외한 FCN 층층은 활성화 함수로 LeakyReLU를 사용하고 최종 출력층에서는 활성화 함수로 linear를 사용한다. 이 모델은 회귀 모델이므로 최적화 알고리즘으로는 Adam을, 오차함수로는 MSE를 사용했다.

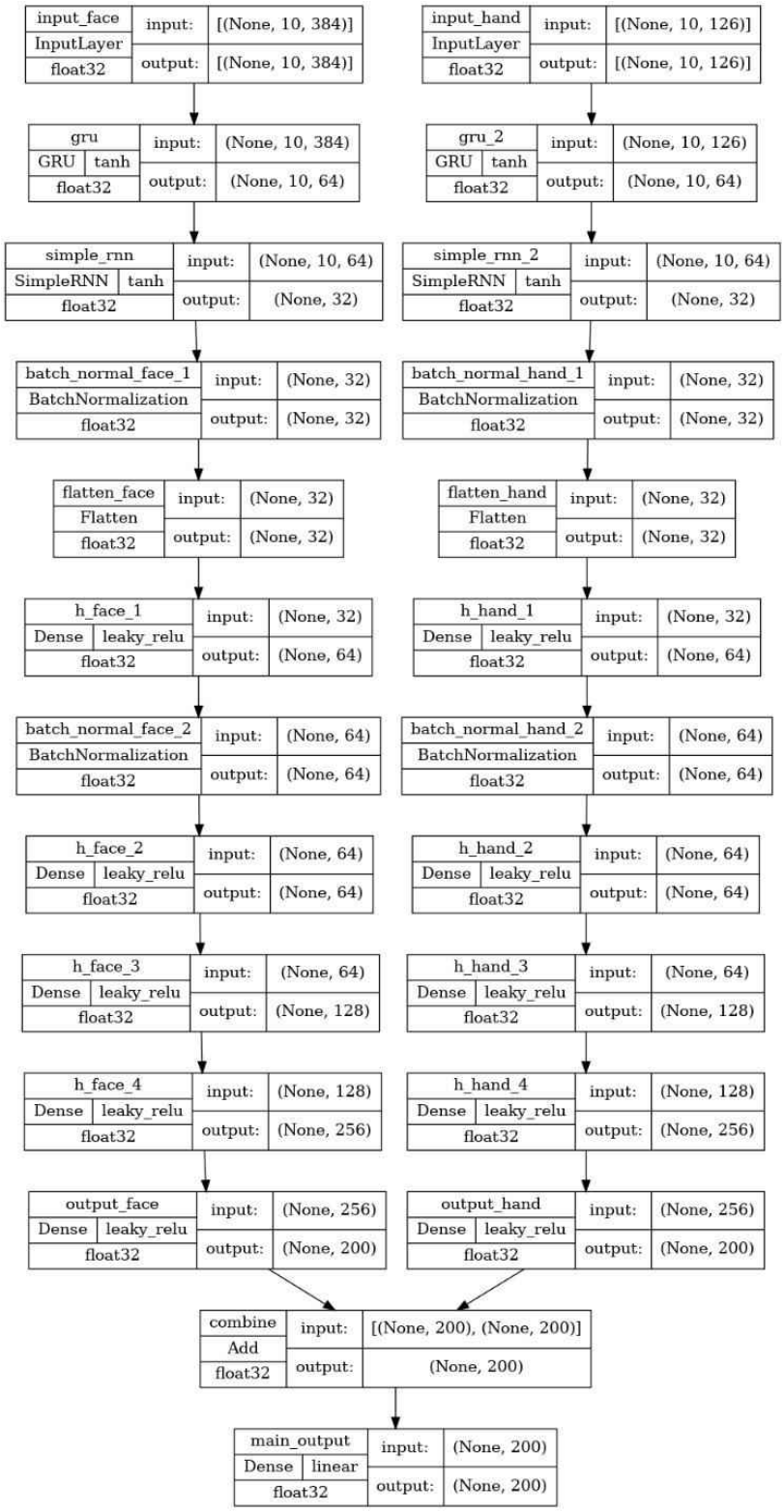

그림 9

2) 의문문 판정 모델

모델의 구조는 그림 10과 같다. 입력으로 50×50으로 resize된 흑백 mono 이미지를 입력으로 받아 8개의 3×3 kernel을 사용하는 Convolution 층을 거쳐 48×48×3 사이즈의 feature map을 얻는다. feature맵이 너무 크기 때문에 MaxPooling을 이용해 subsampling한다. 그 후 flatten 층

을 거쳐 1차원 벡터로 만들고 마지막 FCN layer에서 이 사진이 의문문일 확률을 예측하기 위해 0과 1 사이의 스칼라로 예측하는 sigmoid activation function을 사용해 1 크기의 벡터를 출력한다.

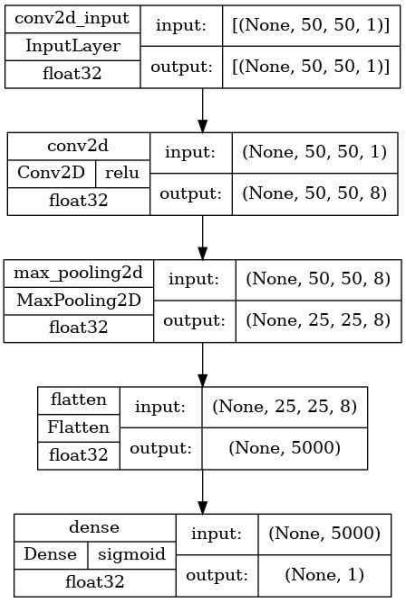

그림 10

나. 학습 결과

1) 손과 얼굴 인식 모델

모델의 학습 결과 그림 11과 같이 train data와 validation data 모두에서 loss가 급격히 감소하고 유지되는 모습을 보였다. 에포크가 길어짐으로 인해 약간의 overfitting이 발생했지만 최종 validation loss가 약 0.0004정도 있었고, 모든 단어에 대해 아주 높은 정확도를 보이지는 않았지만 “기쁨“, “나“ 등의 단어들에 대해서는 새로 만든 수어 영상에 대해서도 높은 정확도를 보였다. 1000 에포크를 설정하여 학습을 진행했을 때는 심각한 과적합이 발생하여 500 에포크까지만 학습을 진행하여 오차를 줄였다.

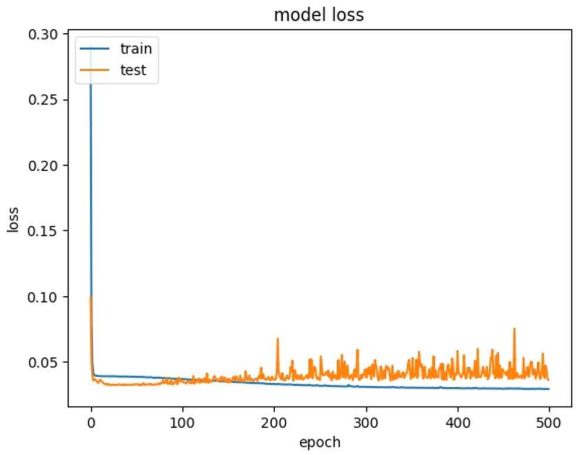

그림 12

2) 얼굴 표정 분류 모델

모델의 학습 결과 그림 12와 같은 그래프를 그렸다. 정확도는 100%, loss는 0에 가까운 결과가 나와 overfitting을 예상했지만, validation data와 test data에서도 같은 결과가 나왔다.

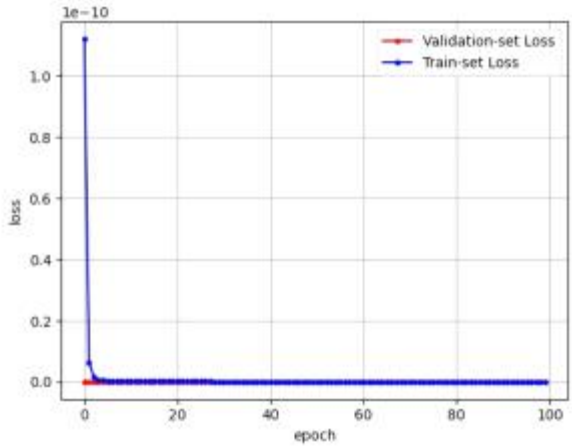

그림 13

5. 고찰

손과 얼굴의 특징점 변위를 기반으로 단어로 번역하는 신경망을 두 가지 방법으로 구축했다. 첫 번째는 전술한 손과 얼굴의 특징점을 처리하는 부분을 각자의 모델로 처리하여 하나의 층으로 합치는 방법의 신경망, 그리고 두 번째는 손과 얼굴 변위벡터를 합쳐 (10, 3*170) 형태의 입력 벡터로 사용, 하나의 모델로 단어 분류를 진행하는 신경망 또한 구현해보았다. 그러나 이 경우에는 학습이 진행될 때 Validation 데이터에 대하여 모델 검증 결과가 잘 출력되지 않았다. 이 이유는, 각각의 데이터셋 영상을 녹화한 사람들 간의 수지 신호에는 확실한 연관성이 있다고 볼 수 있으나 얼굴의 움직임은 조금씩 모두 다르거나 큰 연관성이 보이지 않았다. 카메라와 사람 간의 거리도 모두 달랐고, 조명 환경 등의 이유로 수지 신호보다 얼굴 특징점으로 보여지는 비수지 신호는 확실한 오차 원인이 있었다. 분리하여 각각 학습한 모델을 합치는 구조의 모델은 오차가 있다면 그 쪽을 충분히 무시하는 쪽으로 학습될 수 있으나 한 번에 모든 점을 학습시켜버리면 그들은 연관성이 없어지기 때문에 학습 결과가 원하는 바와 다르게 나왔던 것으로 고찰한다.

얼굴을 인식하여 의문문인지 아닌지를 판별하는 모델의 경우 Loss가 매우 작게 나오는 결과를 보였는데 이 경우는 데이터를 한 사람의 모습만으로 사용했기에 데이터가 편향되어 일반적으로는 완벽히 들어맞지 않을 수 있으며 추후 더욱 많은 데이터를 이용하여 학습시켜 정확도를 개선할 것이다.

6. 결론 및 제언

GRU 모델을 통해 손, 얼굴의 특징점 변위 시계열 데이터를 통해 충분한 정확도로 자연어 단어로 분류할 수 있었으며, CNN 모델을 통해 표정을 인식, 수어로 의사를 전달할 때 비수지 신호 중 하나인 얼굴 표정으로 전달하고자 하는 의문의 의미를 높은 정확도로 판단할 수 있다. 이

연구에 사용된 소스코드는 https://github.com/joon0725/RnE-2nd에서 확인할 수 있다.

7. 추후 연구계획

현재 모델은 전체 수어 영상 중 10프레임만을 따오기 때문에 전체 수어 영상에서 어떤 부분을 모델에 넣어야 정확한 결과가 나오는지는 사람이 직접 해야만 알 수 있다. 하지만 선행 연구 2)에 있는 수어 발화 시점 탐지 연구를 통해 이를 해결할 수 있을 것으로 생각하고 추후 해결할 예정이다. 또한 단어 단위로 번역하는 수어 모델의 출력을 자연스러운 문장으로 만들기 위한 기법 또한 연구하여 추후 적용할 것이다

8. 참고 자료 및 그림 출처

[1] HTTP//keras.io/api/layers/recurrent_layers/gru/

[2] https://www.tensorflow.org/api_docs/python/tf/keras/layers/GRU

[3] 김인혜, 정일홍. (2021). 딥러닝 기반 OpenPose를 이용한 한국 수화 동작 인식에 관한 연구. 한국디지털콘텐츠학회 논문지, 22(4), 681-687.

[4] 석수영 (2016). 수어의 구조와 의미 간의 상관성 고찰. 언어과학연구, 76, 151-173

[5] 윤병천 and 김병하. (2004). 한국수화의 비수지신호에 대한 언어학적 특성 연구. 특수교육저널:이론과 실천, 5(1), 253-277.

[6] 이봉원 외. (2017). 한국수어 교재. 국립국어원.

[7] Aston Zhang 외. Dive into Deep Learning. https://d2l.ai/index.html

[8] 위키미디어 공용