

DNN 기반 수어 번역 모델을 통한 성능 분석

Performance Analysis Using a DNN-Based Sign Language Translation Model

정민재¹, 노승환², 홍준기^{2*}

국립공주대학교 정보통신학과¹, 국립공주대학교 스마트정보기술공학과²

요약

본 연구에서는 수어의 좌표를 압축하여 학습 시간을 획기적으로 단축시킬 수 있는 DNN (Deep Neural Network) 기반 수어 번역 모델을 제안하고 수어 좌표 압축 유무에 따른 정확도와 모델 학습 시간을 비교 분석하였다. 제안한 모델을 사용하여 수어를 번역한 결과, 수어 영상을 압축하기 전과 후의 정확도는 약 5.9% 감소한 반면, 학습 시간은 56.57% 감소하여 수어 번역 정확도 손실 대비 학습 시간에서 많은 이득을 얻는 것을 확인하였다.

■ 중심어 : 딥러닝, DNN, 수어 번역 모델, 성능 분석

Abstract

In this study, we propose a DNN (Deep Neural Network)-based sign language translation model that can significantly reduce training time by compressing sign language coordinates. We compared and analyzed the accuracy and training time of the model with and without sign language coordinate compression. The results of using the proposed model for sign language translation showed that while the accuracy decreased by approximately 5.9% after compressing the sign language video, the training time was reduced by 56.57%, indicating a substantial gain in training efficiency compared to the loss in translation accuracy.

■ Keyword : Deep Learning, DNN, Sign Language Translation Model, Performance Analysis

I. 서론

1.1 연구 배경

2022년 한국장애인고용공단 고용개발원에서 발표한 ‘2022 장애인 통계’에 따르면 2021년 전체 등록 장애인 2,644,700명 중 청각장애인의 수는 41.2만 명으로 약 15.6%를 차지한다. 이는 우리나라 전체 장애인 중 등록된 지체장애인(119.1만 명, 45.1%)에 이어 두 번째로 많은 수치이다 [1]. 농인에게는 글자가 편할 것 같지만, 실제로 농인의 27%가 한글을 거의 이해하지 못하며, 농인의 43%가 한글의 일부만 이해하고 있다. 따라서 대다수의 청인들은 농인이 자신의 의사를 표현할 때 문자를 활용한다고 생각하지만, 사실 한글을 이해하지 못하는 농인은 43%에 이른다[2]. 또한, 한글은 농인에게 외국어와 마찬가지로 낯선 언어이며, 농인들은 불편한 상황에서 직접 시각적으로 표현해야 한다. 이처럼, 농인들은 수어를 모국어로 사용하고 있지만, 통역 서비스의 부족함으로 인해 어려움을 겪고 있다[2].

1.2 기존 수어 번역연구

기존 딥러닝 기술을 통해 한국어 수어를 번역하는 연구가 활발히 진행되고 있다. 딥러닝 기술 중 CNN (Convolutional Neural Network)을 이용한 수어 번역연구[3-7], LSTM (Long Short Term Memory)을 이용한 수어 번역연구[8,9], 그리고 CNN과 LSTM을 함께 사용한 수어 번역연구가 진행되었다[10].

영상 데이터에서 수어 인식을 위해 OpenPose를 사용하여 손의 움직임에 해당하는 양손과 상반신 좌표를 추출하고 Conv1D 모델을 사용하여 레이어 수, 필터 수, Batch Normalization, Dropout 등을 조절하여 성능을 비교 평가하였다. 또한, 실시간 수어 인식을 위해 YOLOv7의 YOLOv7-X 모델을 활용한 수어 번역연구가 진행되었다[3].

또한 수어 생성과 번역이 동시에 이루어질 수 있도록 트랜스포머 기반의 생성자와 조건부 적대 신경망(Conditional GAN) 기반의 감별자를 활용한 양방향 수어 생성 및 번역 통합 모델을 제안한 연구가 진행되었다[6]. 수어 학습을 위해 Conv1D 모델과 LSTM 모델을 결합한 형태로, Conv1D 모델이 특징과 패턴을 추출하고, LSTM 모델이 시계열적 순서를 학습하여 수어를 번역했다[8].

하지만 본 연구에서는 추출된 수어의 좌표를 압축하고 DNN (Deep Neural Network) 기술을 사용하여 수어를 번역하는 모델을 제안하고 수어 좌표 압축 유무에 따른 수어 번역 정확도와 모델 학습 시간을 비교 분석한다. 본 연구에서는 제안한 수어 번역 모델에서 사용되는 좌표 압축 기술의 성능을 학습 시간과 정확도를 기준으로 비교 분석하기 위해 DNN 기술을 사용하였다. 동영상 프레임 단위로 압축하여 입력 데이터의 크기를 줄이고, 그에 따라 모델도 단순한 형태로 선정하여 모델 학습 시간을 단축할 수 있다. 또한, 입력 데이터의 크기가 작다면 새로운 데이터를 입력했을 때 학습된 모델의 연산 시간도 감소한다. 따라서 제안한 수어 번역 모델은 빠른 수어 번역을 통해 수어자가 원하는 의사를 신속하게 전달할 수 있다.

본 논문은 다음과 같이 구성되었다. 2장에서 제안한 DNN 기반 수어 번역 모델을 설명하기 위해 DNN 기술과 손의 좌표를 추출하는 기술을 설명하고, 3장에서 모델에 학습하기 위한 데이터 전처리 과정을 설명하고, 4장에서 제안한 수어 번역 모델을 사용했을 때 수어 좌표 압축 유무에 따른 정확도와 학습 시간과 정확도 변화를 비교 분석한다. 마지막으로 5장에서 본 연구를 요약하고 결론을 맺는다.

II. MediaPipe와 DNN

본 장에서는 제안한 수어 번역 모델에서 수어의

좌표추출을 위해 사용된 MediaPipe와 수어 학습 및 번역을 위해 사용한 DNN 기술을 설명한다.

2.1 MediaPipe

MediaPipe는 구글에서 개발한 실시간 동작 인식, 객체 감지, 얼굴 감지 및 추적, 손의 동작 인식 등의 작업을 수행할 수 있는 오픈소스 라이브러리이다[11].

그림 1은 MediaPipe에서 감지할 수 있는 손의 손목, 엄지, 검지, 중지, 약지, 소지와 관련된 21개 랜드마크 영역을 나타내며, 감지된 손의 21개 랜드마크는 x, y, z의 3차원 좌표로 추출한다.

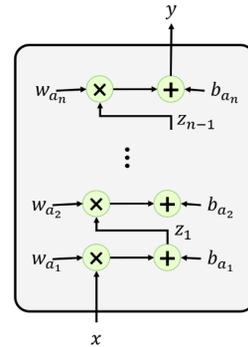


〈그림 1〉 MediaPipe의 21개 랜드마크

2.2 Deep Neural Network (DNN)

DNN은 인공신경망(Artificial Neural Network, ANN)의 한 종류로, 심층신경망을 의미한다. 인공신경망은 생물학적 뉴런에서 영감을 받아 개발되었다. 인공신경망의 뉴런은 다수의 입력값에 가중치를 곱한 후 이를 합산하여 하나의 출력을 생성한다. 일반적으로 ANN은 은닉층이 1개인 신경망을, DNN은 은닉층이 2개 이상인 신경망을 의미한다[12].

그림 2의 w_{an} 과 b_{an} 은 각각 은닉층의 가중치와 편향을 나타내며, n 은 은닉층의 개수를 의미한다. 각 은닉층의 가중치와 편향은 모두 독립적인 변수로, 학습을 통해 업데이트된다. DNN에서 첫 번째 은닉층은 입력 x 를 받아 해당 은닉층의 출력인 z 를 계산하고, 이를 두 번째 은닉층으로 전달한다. 각 은닉층은 이전 은닉층의 출력을 입

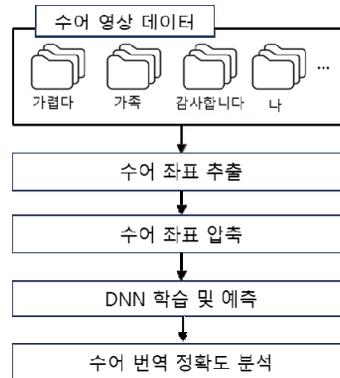


〈그림 2〉 DNN의 구조

력으로 사용하며, 다음 은닉층으로 출력을 전달한다. 다음 은닉층으로 출력을 전달하기 전에 활성화 함수를 적용할 수 있다. 모든 은닉층을 통과한 마지막 은닉층의 출력은 DNN의 최종 예측 결과를 나타낸다.

III. 제안한 DNN 기반 수어 번역 모델

제안한 DNN 기반 수어 번역 모델은 수어의 좌표를 추출하고 프레임을 압축하여 수어 좌표를 학습하여 수어를 실시간으로 번역하는 모델이다. 그림 3은 제안한 DNN 기반 수어 번역 모델의 순서도를 나타낸다.



〈그림 3〉 제안한 수어 번역 모델의 순서도

제안한 DNN 기반 수어 번역 모델의 학습 데

이터는 수어 영상으로부터 MediaPipe를 사용하여 손의 좌표를 추출해 수치 데이터로 저장한다. 저장된 수치 데이터는 수어마다 영상의 길이가 다르기 때문에 바로 모델에 사용하기 적합하지 않으며, 학습 시간을 단축시키기 위해 수어 좌표를 동일한 프레임 단위로 압축한다. 압축 과정을 완료한 수어 좌표 데이터는 DNN 기반 딥러닝 모델을 통해 학습되며, 제안한 모델을 사용하여 시뮬레이션 파라미터별 학습 시간과 수어 번역 정확도를 비교 분석한다.

3.1 수어 영상 데이터

본 연구에서는 한국지능정보사회진흥원의 AI-HUB에서 제공하는 ‘수어 영상’ 데이터를 사용하였다[13]. 이 수어 영상 데이터는 지수자, 지문자의 데이터 영상 생성을 통해 활용 가능한 총 536,000개의 영상 데이터 제공한다. 본 연구에선 효율적인 성능 평가를 위해 선정된 단어는 일상에서 자주 사용하는 대표적인 수어 단어 21개를 선정하였다.

〈표 1〉 연구에 사용된 수어 단어 리스트

가렵다	가족	감사합니다
괜찮다	나	너
맞다	머리	목
못하다	수어	싫다
아니다	아프다	안녕하세요
약속	없다	이름
있다	잘하다	좋다

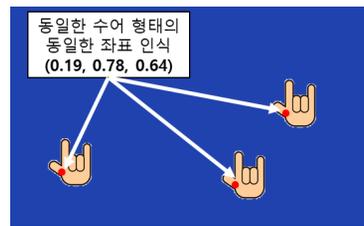
표 1은 제안한 DNN 기반 수어 번역 모델의 성능 평가를 위해 시뮬레이션에 사용된 수어 단어들의 리스트를 나타낸다.

3.2 수어 좌표 추출

제안한 수어 번역 모델은 구글의 MediaPipe 오

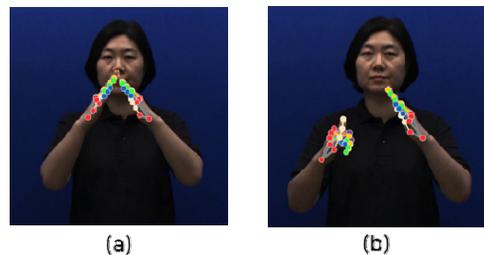
프소스 라이브러리를 활용하여 수어 영상으로부터 손의 위치 좌표를 추출한다. 이 좌표 추출 기술은 수어자의 손의 위치와 형태를 감지하며, 수어 형상으로부터 x, y, z 좌표를 추출한다. 또한, 수어 번역 모델이 좌표를 학습하기 위해 추출된 좌표를 0과 1 사이의 값으로 정규화하는 알고리즘을 제안하여 수어 좌표의 일관성을 보장한다.

예를 들어, 제안한 알고리즘은 그림 4와 같이 수어자의 손 위치에 무관하게 동일한 수어 형태로로부터 모두 동일한 x, y, z 좌표를 추출한다. 따라서, 수어의 의미를 손상하지 않고 좌표의 일관성을 확보할 수 있다.



〈그림 4〉 제안한 수어의 좌표 인식 예시

아래 그림 5는 ‘집’과 ‘사람’으로 이루어진 수어 단어 ‘가족’의 수어를 MediaPipe를 사용하여 추출한 수어의 좌표 랜드마크를 나타낸다.



〈그림 5〉 수어 ‘가족’ 좌표, (a) 집, (b) 사람

MediaPipe를 사용하여 추출한 수어 좌표는 양손의 42개(21×2) 랜드마크와 이에 대응하는 x, y, z 좌표가 모두 포함되어 있으며, 수어의 의미는 라벨 인코딩(label encoding)을 통해 전처리하고

CSV (comma-separated values) 형식으로 변환되어 데이터셋으로 저장한다. 아래 그림 6은 CSV 형식으로 압축된 수어 좌표와 라벨의 예시이다.

	0	1	2	3	...	3777	3778	3779	label
0	0.976963	0.802131	0.569119	0.934849	...	0.949879	0.938902	0.729849	맞다
1	0.385455	0.832083	0.455397	0.473911	...	0.906369	0.929264	0.961651	맞다
2	0.542126	0.812319	0.536373	0.602259	...	0.940129	0.946010	0.958170	맞다
...
2165	0.960791	0.828260	0.744519	0.911962	...	0.954984	0.947970	1.000000	못하다
2166	0.966647	0.813497	0.702841	0.921770	...	0.950624	0.950060	0.974415	못하다
2167	0.550494	0.829729	0.607432	0.642711	...	0.912867	0.931480	0.964304	못하다

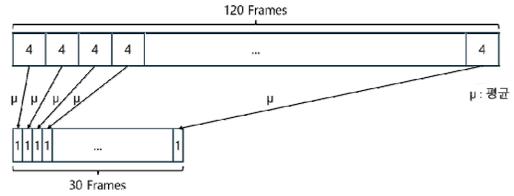
〈그림 6〉 제안한 수어의 좌표 및 라벨 예시

그림 6의 x축 좌표의 개수는 총 3,780개로 이루어져 있으며, 이는 각 수어 영상의 30(프레임 개수) × 42(양손의 랜드마크 개수) × 3(x, y, z 좌표)로 계산된 결과이다. 각 수어 단어 영상의 수어 좌표는 제안한 DNN 기반 수어 번역 모델을 통해 좌표를 학습하고 수어 단어를 번역한다.

3.3 수어 좌표 압축

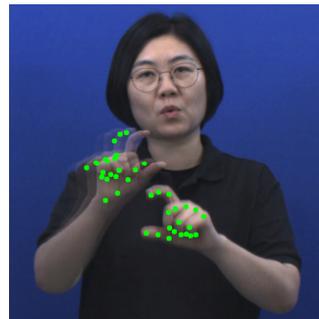
본 연구에서는 수어의 좌표를 학습할 때 소요되는 학습 시간을 획기적으로 단축하기 위해, 추출된 수어의 좌표를 압축하는 기술을 제안하였다. 한국지능정보사회진흥원에서 제공하는 수어 영상의 재생 시간은 3~4초 사이이며, 1초는 30개 프레임으로 구성되어 있으므로 수어 영상의 총 프레임은 90~120개 프레임으로 이루어져 있다. 하지만 빠른 연산을 통한 실시간 수어 번역을 위해선 수어의 의미를 손상시키지 않으면서도 수어의 좌표를 효과적으로 압축하여 영상의 프레임을 효율적으로 압축하는 기술을 제안하였다.

그림 7은 총 120개 프레임의 수어 동영상을 30개 프레임으로 압축한 예시를 나타낸다. 4개 프레임의 좌표 평균값을 계산하여 1개의 프레임으로 압축함으로써, 입력 데이터의 크기를 줄여 수어 번역을 위한 학습 속도를 획기적으로 단축할 수 있다.



〈그림 7〉 수어 좌표 압축 예시

그림 8은 수어 영상 데이터 중 “수어”라는 뜻의 단어에 대해 영상의 일부를 캡처한 후 압축 예시를 시각화하여 보여준다. 그림에서 흐릿하게 보이는 부분은 4개의 프레임이 겹쳐진 화면이며, 초록색 점은 4개의 프레임에서 각 랜드마크 좌표의 평균을 계산하여 1개의 프레임으로 압축한 좌표를 나타낸다.



〈그림 8〉 압축된 수어의 좌표

3.4 DNN을 사용한 수어 번역 정확도 분석

본 연구는 수어 영상 데이터에서 압축된 좌표를 제안한 DNN 기반 수어 번역 모델을 통해 학습하고 수어 좌표 압축 유무에 따른 수어 번역 정확도를 비교 분석하였다. 자세한 성능 분석 결과는 4장에서 설명한다.

IV. 시뮬레이션 결과

4.1 시뮬레이션 파라미터

본 절에서는 DNN 기반 수어 번역 모델이 학

습에서 수행된 모델 학습 환경과 학습을 위해 설정된 학습 모델의 파라미터에 대해 설명한다. 아래 표 2는 제안한 모델의 시뮬레이션 파라미터를 나타낸다.

〈표 2〉 시뮬레이션 파라미터

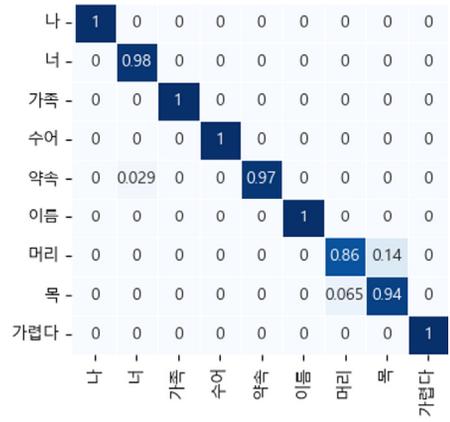
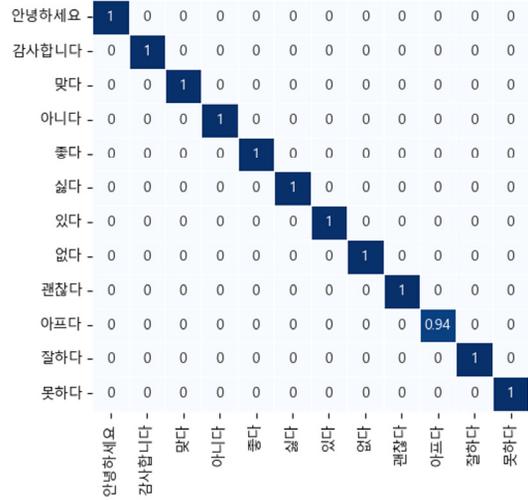
변수	값
No. of input layer	3,780
No. of hidden layer 1	1,890
No. of hidden layer 2	945
No. of output layer	21
Initial learning rate	1e-7
Optimizer	Adam [14]

4.2 수어 번역 정확도 분석

본 절에서는 제안한 DNN 기반 수어 번역 모델의 수어 번역 정확도를 비교 분석한다. 번역 정확도는 번역된 수어 단어와 전체 데이터의 20%인 테스트 데이터셋을 비교한 결과이다. 다음 그림 9와 10은 각각 Epoch 개수가 300일 때 각각 압축 전과 압축 후의 혼동 행렬을 나타낸다.

그림 9 결과에 따르면, 혼동 행렬 중 수어 단어 ‘머리’의 번역률이 가장 낮은 것을 볼 수 있다. ‘머리’를 ‘목’으로 잘못 번역한 경우는 0.14, ‘목’을 ‘머리’로 잘못 번역한 경우는 0.065로 ‘머리’를 ‘목’으로 잘못 번역하는 경우가 많은 것을 볼 수 있다. 그림 10은 압축 후의 혼동 행렬 결과를 나타낸다.

그림 10에 따르면, 수어 단어 ‘목’의 정확도는 0.45로 가장 낮으며, ‘목’을 ‘머리’로 잘못 번역하는 비율이 0.54로 확인되었다. 그림 9와 10을 비교한 결과, 압축 전일 때 ‘머리’를 ‘목’으로 잘못 번역하는 경우가 많고, 압축 후인 모델에서는 ‘목’을 ‘머리’로 잘못 번역하는 경우가 더 많은 것으로 나타났다. 이와 같은 결과는 그림 11과 같이 ‘목’과 ‘머리’는 유사한 수어 좌표 특징을 갖고 있기 때문이다. 이처럼 압축 후 혼동 행렬의 정



〈그림 9〉 수어 좌표 압축 전 혼동 행렬

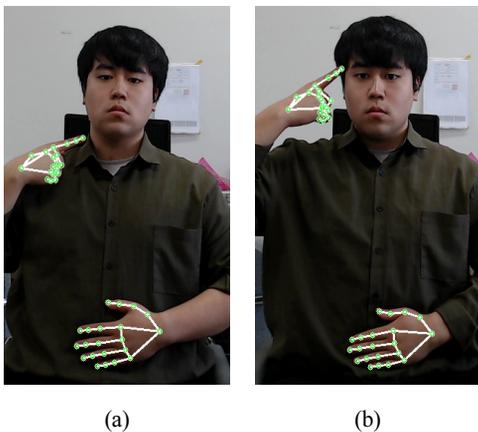
확도가 떨어진 이유는 수어 영상 데이터를 압축하는 과정에서 압축하지 않은 원본 데이터에 비해 특징이 사라지기 때문이다.

그림 11(a)와 11(b)는 각각 ‘목’과 ‘머리’를 나타내는 수어이며, 각 손의 랜드마크를 시각화한 그림이다. 수어에서 신체 부위를 표현할 때는 주로 오른손 검지로 해당 부위를 지칭하며, 제안한 수어 번역 모델은 왼손과 오른손의 거리 차이에 따라 두 단어를 구분한다. 따라서 ‘목’과 ‘머리’의 경우, 오른손은 동일한 자세로 신체의 목과 머리를 지목하므로 왼손의 위치가 중요한 요소로 작용한다. 이로 인해 ‘목’과 ‘머리’는 유사한

안녕하세요	1	0	0	0	0	0	0	0	0	0	0	0	0	0
감사합니다	0	1	0	0	0	0	0	0	0	0	0	0	0	0
맞다	0	0	1	0	0	0	0	0	0	0	0	0	0	0
아니다	0	0	0	1	0	0	0	0	0	0	0	0	0	0
좋다	0	0	0	0	0.87	0	0.1	0	0	0	0	0	0	0
싫다	0	0	0	0	0	1	0	0	0	0	0	0	0	0
있다	0	0	0	0	0	0	1	0	0	0	0	0	0	0
없다	0	0	0	0	0	0	0	1	0	0	0	0	0	0
괜찮다	0	0	0	0	0	0	0	0	1	0	0	0	0	0
아프다	0	0	0	0	0	0	0	0	0	0.98	0	0	0	0
잘하다	0.01	0.02	0	0.01	0	0	0	0	0	0	0	0.92	0	0
못하다	0	0	0	0	0	0	0	0	0	0	0	0	0	1
안녕하세요														
감사합니다														
맞다														
아니다														
좋다														
싫다														
있다														
없다														
괜찮다														
아프다														
잘하다														
못하다														

나	0.99	0	0	0	0	0	0	0	0	0	0	0	0	0.01
너	0	0.95	0	0	0.01	0	0	0	0	0	0	0	0	0
가족	0	0	1	0	0	0	0	0	0	0	0	0	0	0
수어	0	0	0	1	0	0	0	0	0	0	0	0	0	0
약속	0	0	0	0	1	0	0	0	0	0	0	0	0	0
이름	0	0	0	0	0	1	0	0	0	0	0	0	0	0
머리	0	0	0	0	0	0	1	0	0	0	0	0	0	0
목	0	0	0	0	0	0	0	0.54	0.45	0	0	0	0	0
가렵다	0	0	0	0	0	0	0	0	0	0	0	0	0	1
나														
너														
가족														
수어														
약속														
이름														
머리														
목														
가렵다														

<그림 10> 수어 좌표 압축 후 혼동 행렬

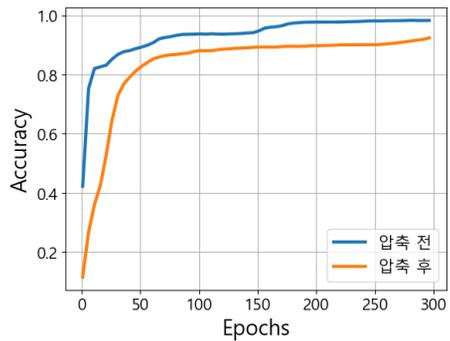


<그림 11> 수어 단어의 좌표: (a) 목, (b) 머리

랜드마크 좌표를 갖게 된다. 만약 그림 11(a)의 왼손이 배꼽 위치가 아닌 더 상단에 위치한다면, 머리를 지칭해도 목으로 번역할 수 있다. 이러한 결과는 제안된 수어 번역 모델이 유사한 수어 좌표 데이터를 번역할 때 정확도가 낮아질 수 있음을 시사한다.

4.3 수어 좌표 압축 유무에 따른 성능 분석

본 절에선 수어 좌표의 압축 전과 후의 학습 시간과 수어 번역 정확도를 비교 분석한다. 그림 12는 수어 압축 유무와 Epoch 개수 변화에 따른 수어 번역 정확도를 나타낸다.

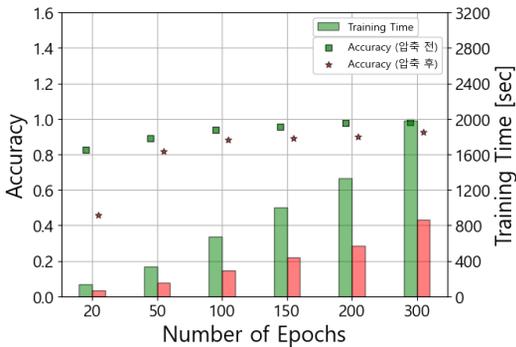


<그림 12> 수어 좌표의 압축 유무에 따른 정확도 비교 분석

그림 12에 따르면, 수어 좌표를 압축할 때 정확도가 감소하는 것을 확인할 수 있다. 이러한 결과는 원본 데이터를 압축하는 과정에서 발생하는 데이터 손실로 인해 정확도가 감소하지만, 입력 데이터의 크기가 감소하면서 학습 시간이 단축되는 장점이 있다. 따라서 다음 절에서는 수어 좌표 데이터 압축 유무에 따른 수어 번역 정확도와 학습 시간을 비교 분석한다.

그림 13은 수어 좌표의 압축 유무에 따른 수어 번역 정확도와 학습 시간을 함께 나타낸 그래프이다. 그림 13의 막대그래프는 학습 시간을 나타내며, 네모와 별은 각각 압축 전과 압축 후 데이

터셋의 Epoch 개수별 정확도를 나타낸다. 그림 13에 따르면, 압축 전과 압축 후 데이터셋 모두 Epoch 개수가 증가할수록 학습 시간이 증가하지만, 압축 후 학습 시간은 압축 전 대비 평균 약 55.27% 감소하는 것을 확인하였다. 아래 표 3은 수어 좌표의 압축 유무에 따른 학습 시간 및 정확도 변화를 비교한 표를 나타낸다.



〈그림 13〉 수어 좌표의 압축 유무에 따른 수어 번역 정확도 및 학습 시간 비교

〈표 3〉 수어 좌표의 압축 유무에 따른 정확도 및 학습 시간 변화를 비교

		Epoch					
		20	50	100	150	200	300
Train ing Time (s)	압축전	136.01	331.18	671.17	1005.21	1330.10	1982.56
	압축후	66.76	152.2	292.61	435.53	571.51	861.11
	변화율 (%)	-50.92	-54.04	-56.40	-56.67	-57.03	-56.57
Accur acy	압축전	0.827	0.889	0.937	0.954	0.978	0.983
	압축후	0.457	0.818	0.88	0.892	0.898	0.925
	변화율 (%)	-44.74	-7.99	-6.08	-6.50	-8.18	-5.90

표 3에 따르면, 수어 영상을 압축한 데이터셋의 학습 시간이 감소하는 것을 확인할 수 있다. 또한 압축 후의 학습 시간은 평균 약 55.27% 감소하는 것을 확인하였다. 이와 같은 결과는 수어 좌표 압축을 통해 입력 데이터의 크기가 감소하여 입력 및 학습 데이터의 크기가 감소하였기 때

문이다.

또한, 학습 초기인 Epoch 20에서는 압축 후 데이터의 학습 시간이 약 50.92% 감소하고, 정확도는 약 44.74% 감소하지만, Epoch가 300까지 증가했을 때 학습 시간은 약 56.57% 감소하고, 수어 번역 정확도는 약 5.90% 손실만이 발생하는 것을 확인하였다. 따라서, 제안한 수어 좌표 압축 기술을 통해 학습이 종료된 시점(Epoch=300)에서 수어 좌표 압축에 따른 손실 대비 짙아진 학습 시간을 통해 더 큰 이득을 얻는 것을 확인할 수 있다.

V. 결론

본 연구에선 수어 영상으로부터 좌표를 추출하고 압축하여 DNN 기술로 수어를 번역하는 모델을 제안하고 수어 좌표의 압축 유무에 따른 수어 번역 정확도와 모델 학습 소요 시간을 비교 분석하였다. 제안한 모델의 시뮬레이션 결과에 따르면 학습이 종료된 시점의 압축 전 데이터와 압축 후 데이터의 정확도는 약 5.9% 감소하는 반면, 수어 모델의 학습 시간은 약 56.57% 감소하는 것을 확인했다. 이와 같은 시뮬레이션 결과를 통해 손실된 수어 번역 정확도 대비 학습 시간에서 더 많은 이득이 발생하는 것을 확인하였다.

또한, 본 연구에선 한정된 수어 단어와 정해진 정수배로 수어의 좌표를 압축하여 평균 좌표값을 사용하였지만, 추후 더 많은 수어 단어들의 프레임 단위로 유사한 수어 좌표들을 함께 묶어 더욱 효과적으로 압축하는 기술로 확장하고 성능을 평가할 예정이다.

아울러 본 연구에선 제안한 압축 기술의 타당성 및 성능 비교분석을 위해 DNN 기술만을 사용하여 성능 분석을 진행하였지만 추후 CNN, LSTM과 같은 다양한 딥러닝 알고리즘과 XGBoost, LGBM (Light Gradient-Boosting Machine)과 같

은 머신러닝 알고리즘을 제안한 수어 번역 모델에 적용하여 성능을 비교 분석할 예정이다.

참 고 문 헌

- [1] 한국장애인고용공단 고용개발원, “2022 장애인 통계,” 2022.
- [2] 이승윤, “고객 상담도 수어 통역되나요?...농인 위한 작은 발걸음,” YTN, 2022.
- [3] 김다연, 이은주, 김유진, 김재호, 이승수, 이홍석, “OpenPose와 CNN 기반 수어 영상 번역 서비스 개발,” 한국콘텐츠학회 종합학술대회 논문집, pp. 393-394, 2021.
- [4] 김인혜, 정일홍, “딥러닝 기반 OpenPose를 이용한 한국 수화 동작 인식에 관한 연구,” 디지털콘텐츠학회논문지, 제22권, 제4호, pp. 681-687, 2021.
- [5] 길상현, 이승훈, 오차영, 유승범, 한연희, “딥러닝 기반 자세 및 손 제스처 인식 기술을 활용한 병원 수어 번역 프로그램 설계 및 구현” 한국통신학회 학술대회논문집, pp. 1015-1016, 2021.
- [6] 김기덕, 이근후, “비디오 캡셔닝을 적용한 수어 번역 및 행동 인식을 적용한 수어 인식” 한국컴퓨터정보학회 학술발표논문집, 제32권, 제1호, pp. 317-319, 2024.
- [7] 이동욱, 김민서, 김남호, 최광미, “실시간 수어 AI 번역 프로그램 구현,” 디지털콘텐츠학회논문지, 제24권, 제10호, pp. 2585-2591, 2023.
- [8] 이재현, 박지수, 성민기, 정세연, 유철중, “딥러닝 기반 양방향 수어 번역 시스템,” Proceedings of KIIT Conference, pp. 801-805, 2023.
- [9] 김신영, 엄서정, 유선영, 김수정, 이경미, “미디어파이프와 장단기기억을 이용한 수화 동작인식 앱 개발,” 디지털콘텐츠학회논문지, 제24권, 제1호, pp. 111-119, 2023.
- [10] 김기찬, 하란, “딥러닝을 통한 실시간 수어 번역

프로그램 개발,” 한국정보과학회 학술발표논문집, pp. 1774-1776, 2022.

- [11] Google AI for Developers, “MediaPipe Solution Guide,” [Online] Available: <https://ai.google.dev/edge/mediapipe/solutions/guide>
- [12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
- [13] 테스트웍스, “수어영상,” [Online] Available: <https://aihub.or.kr/>
- [14] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” 2014. [Online] Available: <https://arxiv.org/abs/1412.6980>

저 자 소 개

정 민 재 (Min-Jae Jeong)



- 2024년 2월 : 국립공주대학교 정보통신공학부 스마트정보기술공학전공 (공학사)
- 2024년 3월~현재 : 국립공주대학교 정보통신공학과 석사과정

<관심분야> 인공지능, DNN 등

E-mail: jeong331440@gmail.com



노 승 환(Soong-Hwan Ro)

- 1987년: 고려대학교 전자공학
학과 (공학사)
- 1989년: 고려대학교 전자공
학과 (공학석사)
- 1993년: 고려대학교 전자공
학과 (공학박사)

• 1994년~현재: 국립 공주대학교 스마트정보기
술공학과 교수

<관심분야> 네트워크, 임베디드시스템, 적외선
영상처리 등

E-mail: rosh@kongju.ac.kr



홍 준 기(Jun-Ki Hong)

- 2010년 11월: Carleton
University 컴퓨터 시스템
공학과 (공학사)
- 2017년 2월: 연세대학교 전기
전자공학과 (공학박사)
- 2016년 8월~2017년 7월: 한

국정보통신기술협회(TTA) 선임연구원

• 2017년 8월~2020년 2월: 영산대학교 전기전
자공학과 조교수

• 2020년 3월~2023년 2월: 배재대학교 컴퓨터
공학과 조교수

• 2023년 3월~현재: 국립공주대학교 스마트정
보기술공학과 조교수

<관심분야> 인공지능, 차세대 통신 등

E-mail: jkhong@kongju.ac.kr