

대규모 언어 모델(LLM)의 **Prompt Engineering**을 통한 한국 수어 한국어 번역 서비스 개발



한지석

허재성

김민혁

지도교수 박진선

목 차

1. 서론	1
1.1. 연구 배경	1
1.2. 기존 문제점	1
1.3. 연구 목표	1
2. 연구배경	3
2.1. 수어 영상 학습을 위한 SlowFastSign 모델	3
2.1.1. 느린 경로(Slow Pathway)	3
2.1.2. 빠른 경로(Fast Pathway)	3
2.1.3 양방향 특징 융합(BI-directional Feature Fusion, BFF)	3
2.1.4 경로 특징 향상(Pathway Feature Enhancement, PFE)	3
2.2. 모델 전체 구조 및 Prompt Engineering 적용	4
2.3. 수어 인식 모듈	4
2.4. Prompt Engineering 모듈	4
2.4.1. 프롬프트 설계	5
2.4.2. 모델 미세 조정	5
2.4.3. 문맥 및 상황 고려	5
2.4.4. 한국어 출력 모듈	5
3. 연구내용	6
3.1. 제공 데이터셋 설명	6
3.1.1. 데이터셋의 규모와 구성	6
3.1.2. 상세 주석 정보	6
3.1.3. 데이터 구축 과정	6
3.1.4. 데이터 전처리	6
3.2. SlowFastSign 모델 학습	9
3.2.1. 환경 설정 및 데이터 준비	10
3.2.2. 모델의 초기화 및 학습	11
3.2.3. 모델 평가	11
3.2.4. 피처 생성 및 저장	12
3.2.5. 설정 파일	12
3.3. 애플리케이션 구현	13
3.3.1. 앱 개발 환경 및 사용 기술	13
3.3.2. 앱 UI 구성	13
3.3.3. 서버 애플리케이션 개발 환경 및 사용 기술	14
3.3.4. 서버 애플리케이션 구성 모듈	15
3.3.5. 프롬프트 엔지니어링	17
3.4. 애플리케이션 동작과정	18
3.4.1. 모바일 애플리케이션 실행	18
3.4.2. 수어 영상 녹화 및 서버로 전송	18

3.4.3. 서버 애플리케이션 동작	19
3.4.4. 프롬프트 엔지니어링	20
3.4.5. 모바일 애플리케이션으로 결과 전송 및 확인	21
4. 연구 결과 분석 및 평가	22
4.1. 개발 일정	22
4.2. 구성원별 역할	23
4.3. WER(Word Error Rate) 분석 및 조기종료	23
4.4. 시행착오 및 한계	25
4.4.1. GPU와 메모리 용량의 제한	25
4.4.2. CTCDecoder와 torch의 버전 충돌	25
4.4.3. 프롬프트 엔지니어링의 모호한 성능 지표	25
4.4.4. 자기 폐색 문제	26
5. 결론 및 향후 연구 방향	26

1. 서론

1.1. 연구 배경

국립국어원(원장 송철의)의 ‘한국수어 사용 실태 조사’의 결과에 따르면 [\[1\]](#), 2016년 8월 「한국수화언어법」이 시행된 이후, 한국수어는 농인의 고유 언어로 공식적으로 인정받았다. 그러나 실제로 농인들이 사회 속에서 한국수어를 얼마나, 어떻게 사용하는지는 명확하게 조사된 바가 없었다. 이에 따라 국립국어원은 2017년 11월, 농인들의 한국수어 사용 실태를 확인하고 향후 정책 수립에 활용하기 위해 대규모 조사를 실시했다. 본 조사는 전국 성인 농인 500명을 대상으로 가구 방문 조사 방식으로 진행되었으며, 조사 결과는 한국수어에 대한 농인의 의사소통 방식, 수어와 농문화에 대한 인식을 파악하는 데 중점을 두었다. 이러한 조사와 연구는 농인의 제1언어인 한국수어가 학습되고 활용되는 환경을 조성하고, 필담에 의존해야 하는 현재의 제약을 극복하여 농인이 차별 없이 소통할 수 있는 사회적 기반을 마련하는 데 기초적인 역할을 할 것이다.

1.2. 기존 문제점

‘부산광역시 농아인 협회’의 자문 결과에 따르면, 상당수의 청각장애인, 특히 고령층은 한글 해독 능력이 부족하여 수어에 의존하는 의사소통 방식을 사용하고 있다. 이는 비장애인과 의사소통에 심각한 장벽으로 작용하며, 길 찾기, 관공서 방문, 의료기관 이용 등 일상적인 활동에 어려움을 초래한다. 수어를 사용하지 않는 사람들과의 소통을 위해 몸짓, 필담, 또는 제한적인 수어 통역 서비스에 의존하는 경우가 많다. 이러한 방식은 정보 전달의 정확성이 떨어지고 시간과 노력이 과다하게 소모될 뿐만 아니라, 오해와 불편을 야기할 수 있다. 수어 자체의 모호성이나 개인적인 표현 방식의 차이로 인해 의사 전달 과정에서 오류가 발생할 가능성이 존재한다. 이는 특히 복잡한 내용을 전달해야 하는 상황에서 심각한 문제가 될 수 있다. 수어 통역사의 지원은 여전히 제한적이며, 특히 관공서, 의료기관 등 공공 기관에서의 접근성이 낮은 경우가 많다. 이로 인해 청각장애인들은 필요한 정보를 제대로 얻거나 서비스를 받는 데 어려움을 겪고 있다. 이러한 문제들은 청각장애인의 사회적 참여를 저해하고 삶의 질을 떨어뜨리는 주요 원인이 되므로, 본 연구는 이러한 문제점들을 해결하기 위한 효과적인 방안을 제시하고자 한다.

1.3. 연구 목표

본 연구는 **AI Hub**에서 제공하는 수어 영상 데이터를 활용하여 청각장애인과 비장애인 간의 원활한 소통을 위한 수어 인식 애플리케이션 개발을 목표로 한다. 기존의 수어 인식 기술의 한계를 극복하고자, **CNN** 기반의 고성능 모델인 **SlowFastSign**을 채택하여 모델 학습을 진행한다. **SlowFastSign**은 공간적 및 시간적 정보를 효과적으로 처리하여 높은 정확도의 수어 인식을 가능하게 한다. 학습된 **SlowFastSign** 모델은 모바일 애플리케이션에 탑재되어 카메라를 통해 실시간으로 수어를 인식하고, **Prompt Engineering** 기법을 통해 인식된 수어 동작을 정확하고 자연스러운 텍스트로 변환한다. 이를 통해 청각장애인은 자신의 수어를 비장애인이 쉽게 이해할 수 있는 언어로 전달할 수 있으며, 비장애인은 수어를 통해 청각장애인과 보다 원활하게 소통할 수 있게 된다. 본 애플리케이션은 단순한 수어-텍스트 변환을 넘어, 실제 소통 환경을 고려한 사용자 친화적인 인터페이스와 정확도 높은 번역 기능을 제공하여, 장애인과 비장애인의 소통 장벽을 낮추고 사회적 포용을 증진하는 데 크게 기여할 것으로 기대된다.

2. 연구배경

2.1. 수어 영상 학습을 위한 **SlowFastSign** 모델

‘SLOWFAST NETWORK FOR CONTINUOUS SIGN LANGUAGE RECOGNITION’ 이 논문^[2]에 따르면, 기존 수어 인식 모델들의 한계를 극복하고 정확도를 향상시키기 위해, 공간적(손 모양, 얼굴 표정) 및 시간적(움직임) 정보를 효과적으로 결합하여 처리하는 **SlowFast** 네트워크를 채택한다. 이 네트워크는 서로 다른 시간적 해상도로 동작하는 두 개의 병렬 경로, 즉 느린 경로(**Slow Pathway**)와 빠른 경로(**Fast Pathway**)로 구성된다.

2.1.1. 느린 경로(**Slow Pathway**)

낮은 프레임 레이트로 동작하여 손 모양, 얼굴 표정 등 공간적으로 의미 있는 특징들을 주로 추출한다. 논문에서 언급된 바와 같이, 긴 영상 데이터 처리의 효율성을 위해 적절한 시간적 간격(**temporal stride**)을 설정하는 것이 중요하며($a=4$), 이는 계산 비용과 정확성 사이의 균형을 고려하여 결정된다. 본 연구에서는 논문에서 제시된 3D 합성곱 연산 및 풀링 연산 등을 통해 공간적 특징을 효과적으로 추출하는 방식을 채택한다.

2.1.2. 빠른 경로(**Fast Pathway**)

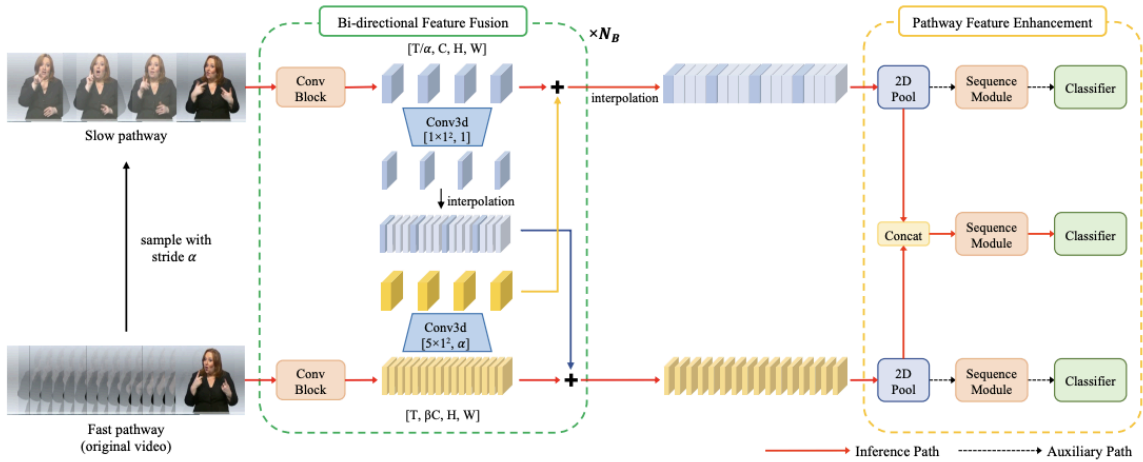
높은 프레임 레이트로 동작하여 팔의 움직임, 몸짓 등 시간적으로 의미 있는 동적 특징들을 추출한다. 빠른 경로는 느린 경로보다 훨씬 더 높은 시간적 해상도를 가지며 (시간적 간격 1), 세밀한 움직임 정보를 포착하는 데 초점을 맞춘다. 하지만 계산 비용을 줄이기 위해 채널 수를 느린 경로의 $1/8$ 로 줄이는 방식을 사용한다. 이를 통해 모델의 경량화를 달성하면서도 동적 특징을 효과적으로 추출할 수 있다.

2.1.3 양방향 특징 융합(**BI-directional Feature Fusion, BFF**)

느린 경로와 빠른 경로에서 추출된 특징들은 단순히 결합하는 것이 아니라, 양방향으로 상호 정보를 교환하는 **BFF** 모듈을 통해 통합된다. 이는 각 경로의 장점을 결합하여 더욱 풍부하고 정확한 정보를 얻는 데 중요한 역할을 한다. 논문에서는 단순한 합산 연산을 통해 효율적인 특징 융합을 구현하며, 계산 비용을 최소화하는 전략을 사용한다.

2.1.4 경로 특징 향상(**Pathway Feature Enhancement, PFE**)

각 경로의 시퀀스 특징을 보강하기 위해 추가적인 서브 네트워크를 활용하는 **PFE** 기법을 도입한다. 이는 각 경로의 특징을 더욱 풍부하게 만들어 수어 인식의 정확도를 높이는 데 기여한다.



[그림 1] 전반적인 Slowfastsign 구조. 프레임워크는 slow pathway와 fast pathway를 사용해 공간적인 특징과 동적 특징을 동시에 추출한다. 양방향 특징 융합(BFF)는 두 경로 간의 풍부한 정보 교환을 용이하게 한다. 경로 기능 향상(PEE)는 서로 다른 샘플 속도에서 얻은 특징을 활용해 각 경로의 시퀀스 특징을 증폭한다. α 와 β 는 각각 시간적 샘플링 보폭과 채널 감소 비율을 나타내며, N_B 는 백본 네트워크의 블록 수를 나타낸다.

2.2. 모델 전체 구조 및 Prompt Engineering 적용

본 애플리케이션은 크게 세 가지 주요 모듈로 구성된다. 수어 인식 모듈, **Prompt Engineering** 모듈, 한국어 출력 모듈. 각 모듈은 서로 유기적으로 연결되어 실시간 수어 인식부터 자연스러운 한국어 문장 생성까지의 과정을 수행한다. 특히, **Prompt Engineering** 기법을 적용하여 기계 번역의 단점을 보완하고, 맥락과 상황을 고려한 자연스러운 한국어 문장 생성에 중점을 둔다.

2.3. 수어 인식 모듈

이 모듈은 사용자의 카메라를 통해 입력되는 수어 영상을 실시간으로 처리한다. 먼저, 영상 전처리 과정을 거쳐 잡음을 제거하고, 필요한 영역을 추출한다. 전처리된 영상은 2.1절에서 설명된 SlowFastSign 모델에 입력된다. SlowFastSign 모델은 영상에서 공간적(손 모양, 얼굴 표정) 및 시간적(움직임) 특징을 효과적으로 추출하여, 수어 동작을 벡터 형태로 표현한 후 학습한 model I. 이 벡터는 다음 단계인 Prompt Engineering 모듈로 전달된다. SlowFastSign 모델의 학습 과정은 대용량의 수어 영상 데이터와 정확한 수어-텍스트 정답

데이터를 사용하여 이루어진다. 모델의 학습 과정에서 다양한 최적화 기법(예: Adam optimizer) 및 손실 함수(예: CTC loss)를 적용하여 모델의 성능을 최대한 향상시킨다.

2.4. Prompt Engineering 모듈

이 모듈은 수어 인식 모듈에서 생성된 벡터를 입력으로 받아, Prompt Engineering 기법을 통해 자연스러운 한국어 문장을 생성한다. Prompt Engineering은 특정한 목표를 달성하기 위해 AI 모델에 입력하는 프롬프트(Prompt)를 전략적으로 설계하는 기술이다. 본 연구에서는 다음과 같은 과정을 통해 Prompt Engineering을 수행한다.

2.4.1. 프롬프트 설계

수어 동작 벡터를 자연스러운 한국어 문장으로 변환하기 위한 적절한 프롬프트를 설계한다. 이때, 수어 동작의 특징(예: 손 모양, 위치, 움직임)을 단순히 나열하는 대신, 주어진 상황(예: 택시 호출, 목적지 도착 등)에 맞춘 문장을 모델에 제공한다. 본 연구에서는 **Few-shot** 학습 방식을 채택하여, 소수의 예시 문장을 통해 상황에 맞는 적절한 문장을 생성할 수 있도록 최적의 프롬프트를 설계한다.

2.4.2. 모델 미세 조정

Few-shot 학습 기반의 프롬프트와 수어 동작 벡터를 결합하여 사전 학습된 **Large Language Model(LLM)**을 직접 활용한다. 대규모 수어-한국어 병렬 데이터 없이도, 소수의 예시와 사전 학습된 모델을 사용해 적절한 문장을 생성하는 방식으로 진행한다. 모델은 몇 개의 예시 문장을 통해 상황과 문맥을 파악하고, 적은 학습 데이터를 바탕으로도 고품질의 한국어 문장을 생성한다.

2.4.3. 문맥 및 상황 고려

생성된 한국어 문장의 자연스러움을 높이기 위해 문맥 및 상황 정보를 적극적으로 반영한다. 수어 동작 벡터를 기반으로 상황에 맞는 프롬프트를 제공하여, 대화의 흐름이나 주변 맥락을 고려한 자연스러운 문장을 생성하도록 한다. 다양한 상황에서 자연스러운 문장이 생성되는지 확인하며, 이를 통해 보다 실용적인 한국어 출력을 기대한다.

2.4.4. 한국어 출력 모듈

Prompt Engineering 모듈에서 생성된 한국어 문장은 사용자에게 출력되며, 인터페이스는 사용자의 편의성을 고려하여 직관적이고 사용하기 쉽게 설계된다. 오류 검출 및 수정 기능을 추가하여, 생성된 문장의 정확성을 높이고 최종적으로 자연스러운 문장을 제공한다.

3. 연구내용

3.1. 제공 데이터셋 설명

우린 데이터셋을 [AIHub\[3\]](#)에 있는 수어 영상 데이터셋을 사용하였다. 이 데이터셋은 총 536,000개의 한국수어 영상 클립으로 구성되어 있으며, 다음과 같은 세 가지 유형의 데이터를 포함한다.

3.1.1. 데이터셋의 규모와 구성

본 연구에 사용된 한국수어 데이터셋은 총 536,000개의 영상 클립으로 구성되며, 실제 상황에서 자주 사용되는 2,000개의 문장과 3,000개의 단어를 20명의 원어민 수어 사용자가 5가지 각도에서 촬영한 50만개의 영상(문장 20만, 단어 30만)과, 52명의 사용자가 촬영한 800개의 손가락 알파벳 단어와 200개의 숫자 영상 21,000개를 포함한다. 문장 데이터는 966개의 의문문을 포함하여 실제 상황을 더욱 정확히 반영하며, 손가락 알파벳 및 숫자 데이터는 클라우드소싱을 통해 수집되어 다양한 발화 스타일을 담고 있다.

3.1.2. 상세 주석 정보

본 데이터셋은 풍부한 주석 정보를 제공하여 인공지능 모델의 정확도와 성능 향상에 기여한다. 각 영상 클립에는 한국수어 동작의 텍스트 표현인 글로스, 정확한 한국어 번역, OpenPose를 이용해 자동 추출한 후 전문가의 검증 및 수동 수정을 거친 2D 및 3D keypoint 정보, 각 글로스의 시작과 끝 시간을 나타내는 타임스탬프, 그리고 얼굴 표정과 머리 동작 등 수어 의미 해석에 중요한 비수어 표현에 대한 주석이 포함되어 있다.

3.1.3. 데이터 구축 과정

데이터 구축은 엄격한 품질 관리 절차를 거쳐 이루어졌다. 다중 카메라 시스템을 사용한 실험실 촬영과 클라우드소싱을 통한 데이터 수집을 병행하여 데이터의 양과 질적 수준 향상에 집중. 전문가 검수 및 농인 연구원의 교차 검토를 통해 데이터의 정확성과 일관성 확보. 특히, 자기 폐색(self-occlusion) 문제를 해결하기 위해 다각도 촬영 기법 활용하였다.

3.1.4. 데이터 전처리

우린 위의 데이터들 중 Slowfastsign 모델에 학습하기 위해 필요한 영상 데이터와 영상 데이터의 이름과 번역되어 있는 한국어가 저장되어 있는 Alhub_train.csv, Alhub_dev.csv 파일을 사용하였다. 영상 데이터 원본은 1960x1080px 인데 모델 학습에 필요한 영상의 크기는 256x256px이기 때문에 resize를 해주었다. 그리고 파일 이름과 한국어 번역, 그리고

영상의 프레임수가 담겨있는 .npy파일을 만들어줬다. .npy은 총 3개가 있는데 하나는 train 관련 파일, 다른 하나는 dev 관련 파일, 마지막으로 gloss가 있는데 이는 한국어 번역 문장을 단어로 나누어 단어를 저장한 파일이다. 그리고 영상은 30fps 기준으로 자르고 256x256px로 resize를 진행하였다. 아래 figure들은 npy 파일을 나타낸다.

```
gloss_dict.npy M X train_info.npy M X
SlowFastSign_KSL > preprocess > KSL > train_info.npy
{
  "KSL":
  ->dataset/KSL/fullframe-196x180bpx,
  "0":
  ({"fileid": "NIA_SL_SEN1932_REAL01", "folder": "train/NIA_SL_SEN1932_REAL01/*.jpg", "signer": "unknown", "Label": "유턴 했다", "num_frames": 169, "original_info": "NIA_SL_SEN1932_REAL01(NIA_SL_SEN1932_REAL01/*.jpg)unknown(유턴 했다)",
    "1":
    ({"fileid": "NIA_SL_SEN1183_REAL01", "folder": "train/NIA_SL_SEN1183_REAL01/*.jpg", "signer": "unknown", "Label": "공방 베스 보다 천 코 차내리다", "num_frames": 234, "original_info": "NIA_SL_SEN1183_REAL01(NIA_SL_SEN1183_REAL01/*.jpg)unknown(공방 베스 보다 천 코 차내리다)",
      "2":
      ({"fileid": "NIA_SL_SEN0034_REAL01", "folder": "train/NIA_SL_SEN0034_REAL01/*.jpg", "signer": "unknown", "Label": "베스 코 내리다 왔다", "num_frames": 201, "original_info": "NIA_SL_SEN0034_REAL01(NIA_SL_SEN0034_REAL01/*.jpg)unknown(베스 코 내리다 왔다)",
        "3":
        ({"fileid": "NIA_SL_SEN1792_REAL01", "folder": "train/NIA_SL_SEN1792_REAL01/*.jpg", "signer": "unknown", "Label": "여기 실종 일어버리다", "num_frames": 188, "original_info": "NIA_SL_SEN1792_REAL01(NIA_SL_SEN1792_REAL01/*.jpg)unknown(여기 실종 일어버리다)",
          "4":
          ({"fileid": "NIA_SL_SEN0783_REAL01", "folder": "train/NIA_SL_SEN0783_REAL01/*.jpg", "signer": "unknown", "Label": "여기 내리다 왔다", "num_frames": 185, "original_info": "NIA_SL_SEN0783_REAL01(NIA_SL_SEN0783_REAL01/*.jpg)unknown(여기 내리다 왔다)",
            "5":
            ({"fileid": "NIA_SL_SEN1808_REAL01", "folder": "train/NIA_SL_SEN1808_REAL01/*.jpg", "signer": "unknown", "Label": "우산 실종 일어버리다", "num_frames": 149, "original_info": "NIA_SL_SEN1808_REAL01(NIA_SL_SEN1808_REAL01/*.jpg)unknown(우산 실종 일어버리다)",
              "6":
              ({"fileid": "NIA_SL_SEN1926_REAL01", "folder": "train/NIA_SL_SEN1926_REAL01/*.jpg", "signer": "unknown", "Label": "차질리다 시간 겹치다", "num_frames": 172, "original_info": "NIA_SL_SEN1926_REAL01(NIA_SL_SEN1926_REAL01/*.jpg)unknown(차질리다 시간 겹치다)",
                "7":
                ({"fileid": "NIA_SL_SEN1362_REAL01", "folder": "train/NIA_SL_SEN1362_REAL01/*.jpg", "signer": "unknown", "Label": "서울대학교 길 왔다 방법", "num_frames": 188, "original_info": "NIA_SL_SEN1362_REAL01(NIA_SL_SEN1362_REAL01/*.jpg)unknown(서울대학교 길 왔다 방법)",
                  "8":
                  ({"fileid": "NIA_SL_SEN1580_REAL01", "folder": "train/NIA_SL_SEN1580_REAL01/*.jpg", "signer": "unknown", "Label": "베스 코 다음 도착 차내리다", "num_frames": 266, "original_info": "NIA_SL_SEN1580_REAL01(NIA_SL_SEN1580_REAL01/*.jpg)unknown(베스 코 다음 도착 차내리다)",
                    "9":
                    ({"fileid": "NIA_SL_SEN0875_REAL01", "folder": "train/NIA_SL_SEN0875_REAL01/*.jpg", "signer": "unknown", "Label": "서울농악연희로 내리다", "num_frames": 198, "original_info": "NIA_SL_SEN0875_REAL01(NIA_SL_SEN0875_REAL01/*.jpg)unknown(서울농악연희로 내리다)",
                      "10":
                      ({"fileid": "NIA_SL_SEN1431_REAL01", "folder": "train/NIA_SL_SEN1431_REAL01/*.jpg", "signer": "unknown", "Label": "카드 고정 된 연희다", "num_frames": 283, "original_info": "NIA_SL_SEN1431_REAL01(NIA_SL_SEN1431_REAL01/*.jpg)unknown(카드 고정 된 연희다)",
                        "11":
                        ({"fileid": "NIA_SL_SEN1896_REAL01", "folder": "train/NIA_SL_SEN1896_REAL01/*.jpg", "signer": "unknown", "Label": "밴드코 실종", "num_frames": 160, "original_info": "NIA_SL_SEN1896_REAL01(NIA_SL_SEN1896_REAL01/*.jpg)unknown(밴드코 실종)",
                          "12":
                          ({"fileid": "NIA_SL_SEN1641_REAL01", "folder": "train/NIA_SL_SEN1641_REAL01/*.jpg", "signer": "unknown", "Label": "군항 코 다음 내리다", "num_frames": 243, "original_info": "NIA_SL_SEN1641_REAL01(NIA_SL_SEN1641_REAL01/*.jpg)unknown(군항 코 다음 내리다)",
                            "13":
                            ({"fileid": "NIA_SL_SEN1672_REAL01", "folder": "train/NIA_SL_SEN1672_REAL01/*.jpg", "signer": "unknown", "Label": "베스 코 다음 내리다", "num_frames": 211, "original_info": "NIA_SL_SEN1672_REAL01(NIA_SL_SEN1672_REAL01/*.jpg)unknown(베스 코 다음 내리다)",
                              "14":
                              ({"fileid": "NIA_SL_SEN0164_REAL01", "folder": "train/NIA_SL_SEN0164_REAL01/*.jpg", "signer": "unknown", "Label": "밴드코 번도 필요", "num_frames": 195, "original_info": "NIA_SL_SEN0164_REAL01(NIA_SL_SEN0164_REAL01/*.jpg)unknown(밴드코 번도 필요)",
                                "15":
                                ({"fileid": "NIA_SL_SEN1431_REAL01", "folder": "train/NIA_SL_SEN1431_REAL01/*.jpg", "signer": "unknown", "Label": "명동로 코 왔다 코 방법", "num_frames": 325, "original_info": "NIA_SL_SEN1431_REAL01(NIA_SL_SEN1431_REAL01/*.jpg)unknown(명동로 코 왔다 코 방법)",
                                  "16":
                                  ({"fileid": "NIA_SL_SEN1883_REAL01", "folder": "train/NIA_SL_SEN1883_REAL01/*.jpg", "signer": "unknown", "Label": "여아 일어버리다 실종", "num_frames": 149, "original_info": "NIA_SL_SEN1883_REAL01(NIA_SL_SEN1883_REAL01/*.jpg)unknown(여아 일어버리다 실종)",
                                    "17":
                                    ({"fileid": "NIA_SL_SEN0454_REAL01", "folder": "train/NIA_SL_SEN0454_REAL01/*.jpg", "signer": "unknown", "Label": "화장실 무명", "num_frames": 144, "original_info": "NIA_SL_SEN0454_REAL01(NIA_SL_SEN0454_REAL01/*.jpg)unknown(화장실 무명)",
                                      "18":
                                      ({"fileid": "NIA_SL_SEN0472_REAL01", "folder": "train/NIA_SL_SEN0472_REAL01/*.jpg", "signer": "unknown", "Label": "지갑 실종", "num_frames": 172, "original_info": "NIA_SL_SEN0472_REAL01(NIA_SL_SEN0472_REAL01/*.jpg)unknown(지갑 실종)",
                                        "19":
                                        ({"fileid": "NIA_SL_SEN0668_REAL01", "folder": "train/NIA_SL_SEN0668_REAL01/*.jpg", "signer": "unknown", "Label": "지갑걸 거고 내리다", "num_frames": 182, "original_info": "NIA_SL_SEN0668_REAL01(NIA_SL_SEN0668_REAL01/*.jpg)unknown(지갑걸 거고 내리다)",
                                          "20":
                                          ({"fileid": "NIA_SL_SEN1538_REAL01", "folder": "train/NIA_SL_SEN1538_REAL01/*.jpg", "signer": "unknown", "Label": "시거로 다음 도착 차내리다", "num_frames": 229, "original_info": "NIA_SL_SEN1538_REAL01(NIA_SL_SEN1538_REAL01/*.jpg)unknown(시거로 다음 도착 차내리다)",
                                            "21":
                                            ({"fileid": "NIA_SL_SEN0873_REAL01", "folder": "train/NIA_SL_SEN0873_REAL01/*.jpg", "signer": "unknown", "Label": "9호", "num_frames": 125, "original_info": "NIA_SL_SEN0873_REAL01(NIA_SL_SEN0873_REAL01/*.jpg)unknown(9호)",
                                              "22":
                                              ({"fileid": "NIA_SL_SEN1932_REAL01", "folder": "train/NIA_SL_SEN1932_REAL01/*.jpg", "signer": "unknown", "Label": "문", "num_frames": 115, "original_info": "NIA_SL_SEN1932_REAL01(NIA_SL_SEN1932_REAL01/*.jpg)unknown(문)",
                                                "23":
                                                ({"fileid": "NIA_SL_SEN0679_REAL01", "folder": "train/NIA_SL_SEN0679_REAL01/*.jpg", "signer": "unknown", "Label": "나사형 거고 차내리다", "num_frames": 182, "original_info": "NIA_SL_SEN0679_REAL01(NIA_SL_SEN0679_REAL01/*.jpg)unknown(나사형 거고 차내리다)",
                                                  "24":
                                                  ({"fileid": "NIA_SL_SEN0558_REAL01", "folder": "train/NIA_SL_SEN0558_REAL01/*.jpg", "signer": "unknown", "Label": "좌회전 왔다", "num_frames": 165, "original_info": "NIA_SL_SEN0558_REAL01(NIA_SL_SEN0558_REAL01/*.jpg)unknown(좌회전 왔다)",
                                                    "25":
                                                    ({"fileid": "NIA_SL_SEN1656_REAL01", "folder": "train/NIA_SL_SEN1656_REAL01/*.jpg", "signer": "unknown", "Label": "병통 찢기 다음", "num_frames": 281, "original_info": "NIA_SL_SEN1656_REAL01(NIA_SL_SEN1656_REAL01/*.jpg)unknown(병통 찢기 다음)",

```

[그림2] train_npy

```
gloss_dict.npy M X dev_info.npy M X
SlowFastSign_KSL > preprocess > KSL > dev_info.npy
{
  "KSL":
  ->dataset/KSL/fullframe-196x180bpx,
  "0":
  ({"fileid": "NIA_SL_SEN1882_REAL02", "folder": "dev/NIA_SL_SEN1882_REAL02/*.jpg", "signer": "unknown", "Label": "약 코 무명", "num_frames": 0, "original_info": "NIA_SL_SEN1882_REAL02(NIA_SL_SEN1882_REAL02/*.jpg)unknown(약 코 무명)",
    "1":
    ({"fileid": "NIA_SL_SEN0043_REAL02", "folder": "dev/NIA_SL_SEN0043_REAL02/*.jpg", "signer": "unknown", "Label": "뽕 12사 후 온 거", "num_frames": 0, "original_info": "NIA_SL_SEN0043_REAL02(NIA_SL_SEN0043_REAL02/*.jpg)unknown(뽕 12사 후 온 거)",
      "2":
      ({"fileid": "NIA_SL_SEN1249_REAL02", "folder": "dev/NIA_SL_SEN1249_REAL02/*.jpg", "signer": "unknown", "Label": "책도넛 보다 다음 차내리다", "num_frames": 0, "original_info": "NIA_SL_SEN1249_REAL02(NIA_SL_SEN1249_REAL02/*.jpg)unknown(책도넛 보다 다음 차내리다)",
        "3":
        ({"fileid": "NIA_SL_SEN0445_REAL02", "folder": "dev/NIA_SL_SEN0445_REAL02/*.jpg", "signer": "unknown", "Label": "신호등 전 거", "num_frames": 0, "original_info": "NIA_SL_SEN0445_REAL02(NIA_SL_SEN0445_REAL02/*.jpg)unknown(신호등 전 거)",
          "4":
          ({"fileid": "NIA_SL_SEN0443_REAL02", "folder": "dev/NIA_SL_SEN0443_REAL02/*.jpg", "signer": "unknown", "Label": "신호등 전 거", "num_frames": 0, "original_info": "NIA_SL_SEN0443_REAL02(NIA_SL_SEN0443_REAL02/*.jpg)unknown(신호등 전 거)",
            "5":
            ({"fileid": "NIA_SL_SEN1275_REAL02", "folder": "dev/NIA_SL_SEN1275_REAL02/*.jpg", "signer": "unknown", "Label": "책시점 다음 코 거고 도착 차내리다", "num_frames": 0, "original_info": "NIA_SL_SEN1275_REAL02(NIA_SL_SEN1275_REAL02/*.jpg)unknown(책시점 다음 코 거고 도착 차내리다)",
              "6":
              ({"fileid": "NIA_SL_SEN0039_REAL02", "folder": "dev/NIA_SL_SEN0039_REAL02/*.jpg", "signer": "unknown", "Label": "책시점 기쁜 환영", "num_frames": 0, "original_info": "NIA_SL_SEN0039_REAL02(NIA_SL_SEN0039_REAL02/*.jpg)unknown(책시점 기쁜 환영)",
                "7":
                ({"fileid": "NIA_SL_SEN0879_REAL02", "folder": "dev/NIA_SL_SEN0879_REAL02/*.jpg", "signer": "unknown", "Label": "난방 고다", "num_frames": 0, "original_info": "NIA_SL_SEN0879_REAL02(NIA_SL_SEN0879_REAL02/*.jpg)unknown(난방 고다)",
                  "8":
                  ({"fileid": "NIA_SL_SEN1929_REAL02", "folder": "dev/NIA_SL_SEN1929_REAL02/*.jpg", "signer": "unknown", "Label": "문", "num_frames": 0, "original_info": "NIA_SL_SEN1929_REAL02(NIA_SL_SEN1929_REAL02/*.jpg)unknown(문)",
                    "9":
                    ({"fileid": "NIA_SL_SEN1740_REAL02", "folder": "dev/NIA_SL_SEN1740_REAL02/*.jpg", "signer": "unknown", "Label": "연어 거고 다음 도착 차내리다", "num_frames": 0, "original_info": "NIA_SL_SEN1740_REAL02(NIA_SL_SEN1740_REAL02/*.jpg)unknown(연어 거고 다음 도착 차내리다)",
                      "10":
                      ({"fileid": "NIA_SL_SEN1227_REAL02", "folder": "dev/NIA_SL_SEN1227_REAL02/*.jpg", "signer": "unknown", "Label": "책시 전 코 도착 차내리다", "num_frames": 0, "original_info": "NIA_SL_SEN1227_REAL02(NIA_SL_SEN1227_REAL02/*.jpg)unknown(책시 전 코 도착 차내리다)",
                        "11":
                        ({"fileid": "NIA_SL_SEN1759_REAL02", "folder": "dev/NIA_SL_SEN1759_REAL02/*.jpg", "signer": "unknown", "Label": "경찰 거고 다음 도착 차내리다", "num_frames": 0, "original_info": "NIA_SL_SEN1759_REAL02(NIA_SL_SEN1759_REAL02/*.jpg)unknown(경찰 거고 다음 도착 차내리다)",
                          "12":
                          ({"fileid": "NIA_SL_SEN0425_REAL02", "folder": "dev/NIA_SL_SEN0425_REAL02/*.jpg", "signer": "unknown", "Label": "여기 왔다", "num_frames": 0, "original_info": "NIA_SL_SEN0425_REAL02(NIA_SL_SEN0425_REAL02/*.jpg)unknown(여기 왔다)",
                            "13":
                            ({"fileid": "NIA_SL_SEN0355_REAL02", "folder": "dev/NIA_SL_SEN0355_REAL02/*.jpg", "signer": "unknown", "Label": "잠시합시다", "num_frames": 0, "original_info": "NIA_SL_SEN0355_REAL02(NIA_SL_SEN0355_REAL02/*.jpg)unknown(잠시합시다)",
                              "14":
                              ({"fileid": "NIA_SL_SEN0357_REAL02", "folder": "dev/NIA_SL_SEN0357_REAL02/*.jpg", "signer": "unknown", "Label": "조심", "num_frames": 0, "original_info": "NIA_SL_SEN0357_REAL02(NIA_SL_SEN0357_REAL02/*.jpg)unknown(조심)",
                                "15":
                                ({"fileid": "NIA_SL_SEN1484_REAL02", "folder": "dev/NIA_SL_SEN1484_REAL02/*.jpg", "signer": "unknown", "Label": "군항 코 다음 차내리다", "num_frames": 0, "original_info": "NIA_SL_SEN1484_REAL02(NIA_SL_SEN1484_REAL02/*.jpg)unknown(군항 코 다음 차내리다)",
                                  "16":
                                  ({"fileid": "NIA_SL_SEN0872_REAL02", "folder": "dev/NIA_SL_SEN0872_REAL02/*.jpg", "signer": "unknown", "Label": "한국농악연희로 차내리다 오케이", "num_frames": 0, "original_info": "NIA_SL_SEN0872_REAL02(NIA_SL_SEN0872_REAL02/*.jpg)unknown(한국농악연희로 차내리다 오케이)",
                                    "17":
                                    ({"fileid": "NIA_SL_SEN1998_REAL02", "folder": "dev/NIA_SL_SEN1998_REAL02/*.jpg", "signer": "unknown", "Label": "보통이 왔다 실종", "num_frames": 0, "original_info": "NIA_SL_SEN1998_REAL02(NIA_SL_SEN1998_REAL02/*.jpg)unknown(보통이 왔다 실종)",
                                      "18":
                                      ({"fileid": "NIA_SL_SEN0590_REAL02", "folder": "dev/NIA_SL_SEN0590_REAL02/*.jpg", "signer": "unknown", "Label": "음급실 거다", "num_frames": 0, "original_info": "NIA_SL_SEN0590_REAL02(NIA_SL_SEN0590_REAL02/*.jpg)unknown(음급실 거다)",
                                        "19":
                                        ({"fileid": "NIA_SL_SEN1357_REAL02", "folder": "dev/NIA_SL_SEN1357_REAL02/*.jpg", "signer": "unknown", "Label": "서울대학교 왔다 거다 걸 거다 방법", "num_frames": 0, "original_info": "NIA_SL_SEN1357_REAL02(NIA_SL_SEN1357_REAL02/*.jpg)unknown(서울대학교 왔다 거다 걸 거다 방법)",
                                          "20":
                                          ({"fileid": "NIA_SL_SEN0296_REAL02", "folder": "dev/NIA_SL_SEN0296_REAL02/*.jpg", "signer": "unknown", "Label": "지갑철 문 미러를 떨어지다", "num_frames": 0, "original_info": "NIA_SL_SEN0296_REAL02(NIA_SL_SEN0296_REAL02/*.jpg)unknown(지갑철 문 미러를 떨어지다)",
                                            "21":
                                            ({"fileid": "NIA_SL_SEN1933_REAL02", "folder": "dev/NIA_SL_SEN1933_REAL02/*.jpg", "signer": "unknown", "Label": "문", "num_frames": 0, "original_info": "NIA_SL_SEN1933_REAL02(NIA_SL_SEN1933_REAL02/*.jpg)unknown(문)",
                                              "22":
                                              ({"fileid": "NIA_SL_SEN1291_REAL02", "folder": "dev/NIA_SL_SEN1291_REAL02/*.jpg", "signer": "unknown", "Label": "이마트 코 거고 다음 차내리다", "num_frames": 0, "original_info": "NIA_SL_SEN1291_REAL02(NIA_SL_SEN1291_REAL02/*.jpg)unknown(이마트 코 거고 다음 차내리다)",
                                                "23":
                                                ({"fileid": "NIA_SL_SEN1952_REAL02", "folder": "dev/NIA_SL_SEN1952_REAL02/*.jpg", "signer": "unknown", "Label": "사건이 일어버리다 실종", "num_frames": 0, "original_info": "NIA_SL_SEN1952_REAL02(NIA_SL_SEN1952_REAL02/*.jpg)unknown(사건이 일어버리다 실종)",
                                                  "24":
                                                  ({"fileid": "NIA_SL_SEN1512_REAL02", "folder": "dev/NIA_SL_SEN1512_REAL02/*.jpg", "signer": "unknown", "Label": "베스 코 다음 내리다", "num_frames": 0, "original_info": "NIA_SL_SEN1512_REAL02(NIA_SL_SEN1512_REAL02/*.jpg)unknown(베스 코 다음 내리다)",
                                                    "25":
                                                    ({"fileid": "NIA_SL_SEN0270_REAL02", "folder": "dev/NIA_SL_SEN0270_REAL02/*.jpg", "signer": "unknown", "Label": "학교 주차", "num_frames": 0, "original_info": "NIA_SL_SEN0270_REAL02(NIA_SL_SEN0270_REAL02/*.jpg)unknown(학교 주차)",

```

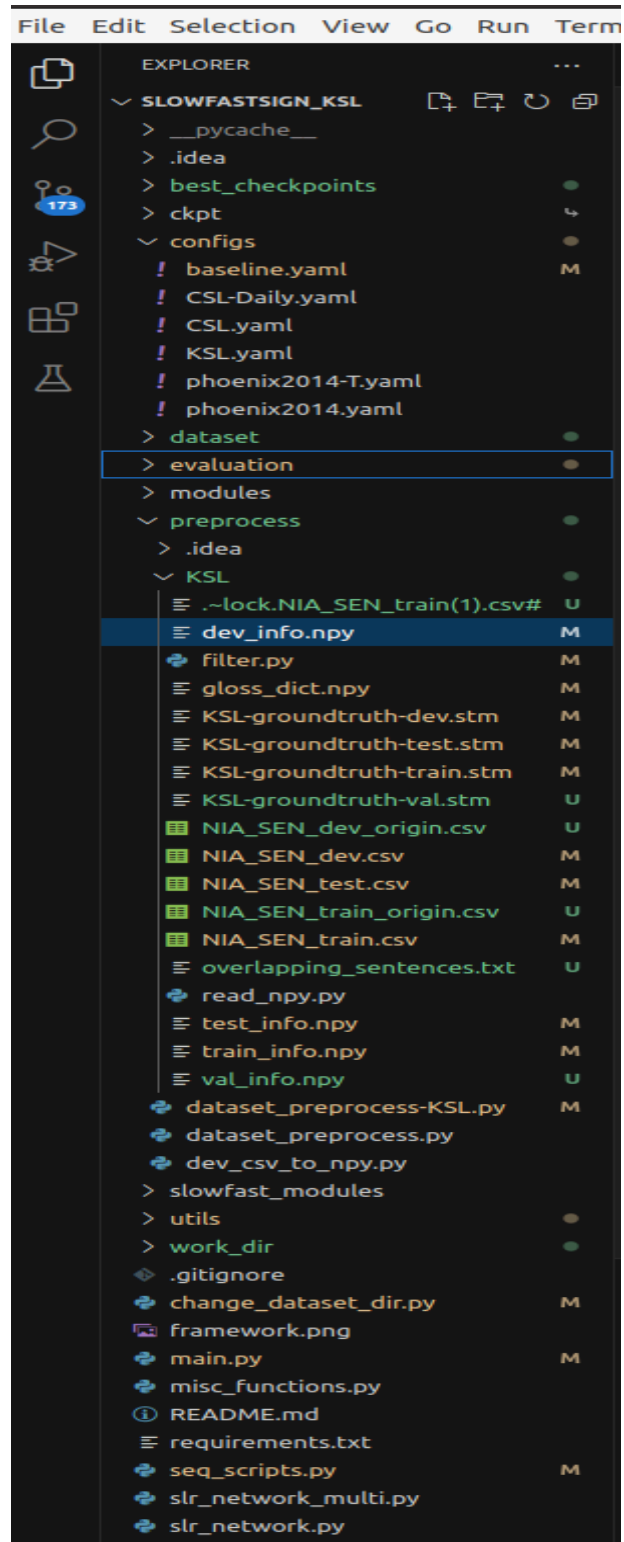
[그림3] dev_npy

```
gloss_dict.npy M  dev_info.npy M
SlowFastSign_KSL > preprocess > KSL > gloss_dict.npy
{
  '1': [1, 2],
  '10분': [2, 7],
  '119': [3, 26],
  '12': [4, 2],
  '12시': [5, 2],
  '1시': [6, 3],
  '1시21분': [7, 2],
  '1호': [8, 20],
  '1화': [9, 2],
  '2호': [10, 2],
  '3': [11, 18],
  '3시': [12, 3],
  '3호': [13, 2],
  '40분': [14, 1],
  '4시': [15, 1],
  '50분': [16, 1],
  '5분': [17, 4],
  '5호': [18, 3],
  '8호': [19, 6],
  '9호': [20, 10],
  '가깝다': [21, 7],
  '가동': [22, 29],
  '가다': [23, 263],
  '가방': [24, 19],
  '같이타다': [25, 28],
  '감사합니다': [26, 2],
  '강남': [27, 2],
}
```

[그림4] gloss_npy

gloss_npy 파일은 train, dev, test의 label 데이터에 포함된 단어를 포함하고 있다. SlowFastSign 모델은 이 gloss_npy 단어 개수만큼의 classifier 역할을 한다.

3.2. SlowFastSign 모델 학습



[그림5] 모델 학습에 필요한 file들

본 연구에서는 한국 수어 영상의 한국어 번역 모델을 개발하기 위해 체계적인 데이터 전처리와 모델 학습 과정을 설계했다, 전체 과정은 크게 환경 설정과 데이터 전처리, 모델의 초기화 학습 및 학습, 평가 단계로 구분된다.

3.2.1. 환경 설정 및 데이터 준비

모델 학습의 첫 단계는 환경 설정과 데이터 준비이다. 이 작업은 `main.py`와 `dataset_preprocess-KSL.py` 파일에서 주로 수행된다.

- a. 환경설정 - `main.py` 파일에서 `CUDA` 장치를 설정하고 필요한 라이브러리를 호출해 예측 성능을 극대화할 수 있는 환경을 구축했다. 이 과정은 `GPU` 사용을 최적화하기 위한 필수적인 전처리 단계이다.

```
import os
import torch
import cv2
import pdb
# 기타 필요 라이브러리...
```

- b. 데이터 전처리 - `dataset_preprocess-KSL.py`에서는 수어 영상 데이터를 `CSV` 파일에서 불러와 각 샘플에 대한 경로 및 라벨 정보를 추출한다. 이 스크립트는 영상을 일정한 크기로 리사이즈하여 모델 학습 전처리를 완벽히 지원한다.

```
def csv2dict(anno_path, dataset_type):
    # CSV 파일 읽기

    df = pd.read_csv(anno_path, encoding='utf-8')

    # 경로 및 라벨 정보 추출

    for index, row in df.iterrows():

        영상의 프레임 경로 설정

        영상의 프레임 수 파악 및 저장
```

3.2.2. 모델의 초기화 및 학습

모델의 초기화와 학습은 'Processor' 클래스가 관리한다. 이 클래스는 모델 전체의 라이프 사이클을 책임진다.

- a. 모델 초기화 - 내부에서는 주어진 구성 파일 **baseline.yaml**을 읽어들이고, 모델, 옵티마이저, 그리고 데이터 로더를 초기화한다. 초기에 설정된 하이퍼파라미터와 경로들은 효율적인 학습을 위한 기반을 제공한다.

```
class Processor():

    def __init__(self, arg):

        # 작업 디렉토리 생성 및 초기 설정

        self.model, self.optimizer = self.loading()

    def loading(self):

        # 모델 클래스 로딩 및 초기화

        model_class = import_class(self.arg.model)

        model = model_class(**self.arg.model_args)

        return model, optimizer
```

- b. 모델 학습 - 'start()' 메소드에서 실제 학습이 이루어지며, 각 에포크마다 데이터 배치를 불러와 훈련시킨다. 옵티마이저를 사용해 손실을 최소화하고, 학습이 진행됨에 따라 점차적으로 모델의 성능을 향상시킨다.

```
def start(self):

    if self.arg.phase == 'train':

        for epoch in range(시작 에폭, 총 에폭 수):

            seq_train(self.data_loader['train'], self.model, self.optimizer, self.device, epoch,
self.recoder)
```

3.2.3. 모델 평가

평가는 'seq_eval' 함수를 통해 수행되며, 특정 간격마다 검증 데이터셋으로 모델의 성능을 측정한다. 이 과정에서 WER(Word Error Rate)를 사용해 모델의 번역 품질을 객관적으로 평가한다.

```
def seq_eval(cfg, loader, model, device, mode, epoch, work_dir, recoder):
```

모델을 평가 모드로 전환

각 배치에 대해:

```
ret_dict = model(vid, vid_lgt, label=label, label_lgt=label_lgt)
```

시스템을 통해 WER 평가

3.2.4. 피쳐 생성 및 저장

학습된 모델 피쳐를 추출해 저장하는 과정도 추가적으로 구현하였다. 이 단계는 추후 어플리케이션에서 피쳐 기반의 전처리나 후처리 과정에 유용하게 활용될 수 있다.

```
def seq_feature_generation(loader, model, device, mode, work_dir, recoder):
```

```
model.eval()
```

비디오 피쳐를 파일로 저장

3.2.5. 설정 파일

설정 파일('baseline.yaml')을 통해 모델의 학습과 관련된 여러 하이퍼파라미터를 관리한다. 이는 실험의 반복성과 설정의 유연성을 제공해 다양한 조건에서 신속하게 모델을 훈련 및 평가를 할 수 있게 해준다.

a. 데이터 및 학습 파라미터 설정

```
feeder: dataset.dataloader_video.BaseFeeder
```

```
phase: train
```

```
dataset: KSL
```

```
num_epoch: 50
```

```
batch_size: 1
```

```
num_worker: 0
```

```
device: 0
```

b. 모델 및 옵티마이저 설정

```
model: slr_network_multi.SLRModel
```

```
model_args:
```

```
num_classes: 1296
c2d_type: slowfast101

optimizer_args:
  optimizer: Adam
  base_lr: 0.00005
  scheduler: multistep
  step: [40, 60]

loss_weights:
  SeqCTC: 1.0
  ConvCTC: 1.0
  Dist: 25.0
```

3.3. 애플리케이션 구현

3.3.1. 앱 개발 환경 및 사용 기술

모바일 애플리케이션 개발은 **Android Studio**를 주요 개발 환경으로 선택했다. 간단한 앱을 빠르게 구현하기 위해 선언형 UI 프레임워크인 **Jetpack Compose**를 사용하였다.

3.3.2. 앱 UI 구성

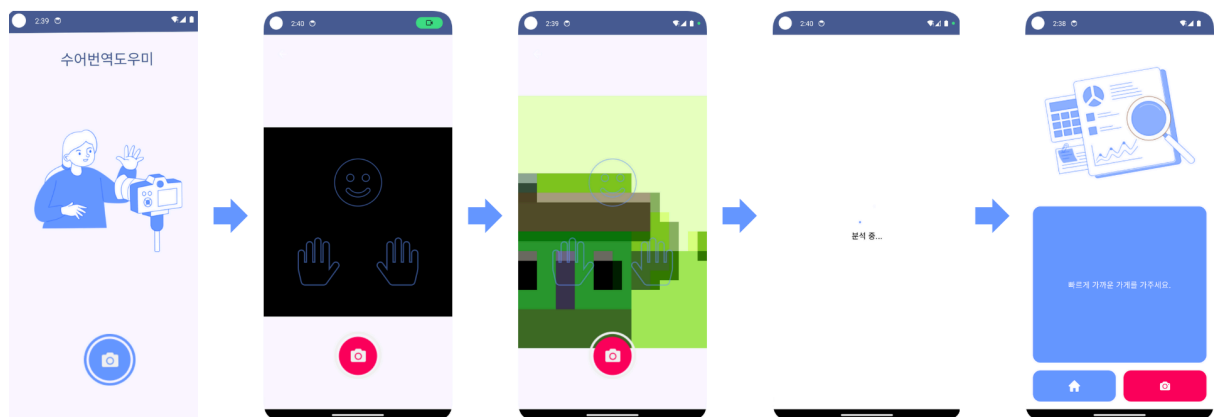
예상 사용자층이 한글 문맹 농인이므로 텍스트 사용을 지양하고, 직관적인 아이콘과 이미지를 활용하여 UI를 설계했다. 화면 UI 구성은 홈, 녹화, 분석, 결과 네 단계로 구성된다.

홈: 앱 실행 후 첫 화면. 카메라 버튼을 누르면 녹화 모드로 전환됨

녹화: 녹화 버튼과 중지 버튼으로 직관적으로 영상을 녹화하는 화면

분석: 녹화된 영상을 서버에 전송하고, 응답이 도착하기 전까지 잠시 대기하는 화면

결과: 서버에서 영상을 분석한 결과 문장이 텍스트 및 음성으로 출력되는 화면



[그림6] 앱 화면

매우 직관적이며, 결과 화면에서 음성을 함께 출력함으로써 택시 상황에서 운전자에게 번역 내용이 잘 전달 될 수 있도록 하였다. 또한 녹화 화면에서 적절한 얼굴, 손 위치에 대한 가이드를 제시하여 자체적으로 영상을 규격화했다.

3.3.3. 서버 애플리케이션 개발 환경 및 사용 기술

서버 애플리케이션을 구현할 때는 평소에 주로 사용해왔던 **VScode**를 개발 환경으로 선택하였다. 웹 프레임워크로는 경량 프로그램을 구축하기 용이하며 **python**으로 작성된 AI 모델 코드를 자연스럽게 통합하기 좋은 **flask**를 사용하였다.

3.3.4. 서버 애플리케이션 구성 모듈

서버 애플리케이션은 다음과 같은 모듈들로 구성된다.

- a. **main.py**: 애플리케이션의 진입점, **Flask** 앱 인스턴스 생성

```
def load_environment_variables():  
    # .env 파일에서 환경 변수 로드  
  
def create_flask_app():  
    # Flask 앱 생성 및 설정  
  
def register_app_routes(app):  
    # 라우트 등록  
  
def run_gunicorn_app(app):  
    # gunicorn을 사용해 앱 실행
```

- b. **config.py**: 애플리케이션의 설정, 환경 변수, API 키 관리

```
def ensure_directories_exist():  
    # 필요한 디렉토리가 존재하는지 확인하고 없으면 생성  
  
def configure_app_folders(app):  
    # 앱의 폴더 경로 설정 및 구성
```

- c. **routes.py**: URL 라우팅 정의, HTTP 요청 처리 로직 구현, 클라이언트의 요청을 적절한 처리 함수로 연결

```
def register_video_processing_route(app):  
    # 비디오 처리 요청을 위한 경로 등록  
  
def handle_video_processing_request():  
    # 요청에서 비디오 파일 확인 및 처리
```

- d. **video_processor.py**: 비디오 처리 관련 로직 구현, 프레임 추출, 이미지 처리 등의 기능 수행

```
def ensure_frames_directory():  
    # 프레임을 저장할 하위 폴더 생성 또는 확인  
  
def split_video_into_frames(video_path):
```

```
# 비디오를 프레임으로 분할하고 저장
def process_video_file(video_file):
    # 업로드된 비디오 파일을 처리하고 프레임을 생성
    # 프레임에서 단어 추출 후 자연스러운 문장 생성
```

- e. openai_helper.py: OpenAI API 연동 로직 구현, 프롬프트 엔지니어링 수행

```
def initialize_openai_client():
    # OpenAI 클라이언트 초기화

def read_examples_from_file(filename):
    # 퓨샷 위한 예시 문장 읽기

def generate_few_shot_examples():
    # 퓨샷 학습 예시 생성(문장 세트)

def ask_chatgpt(words):
    # ChatGPT API 호출 및 응답 처리
```

- f. gunicorn_runner.py: Gunicorn WSGI 서버 실행 설정, 프로덕션 환경에서의 애플리케이션 실행 관리

```
def display_app_configurations(app):
    # 앱의 설정 정보 출력

def run_gunicorn_server(app):
    # Gunicorn 서버 실행
```

3.3.5. 프롬프트 엔지니어링

서버 애플리케이션 중 `openai_helper.py` 파일에서 `few-shot` 프롬프트 엔지니어링을 수행하기 위해 예시 문장과 택시 상황이라는 조건을 제공하였다.

Few-shot 학습을 위해 다음과 같은 예시 문장들이 사용되었다. 이 예시들은 수어에서 사용되는 핵심 단어(**gloss**)들의 조합과 자연스럽게 번역된 문장을 매핑한 자료구조로 구성되었다.

예시 1: {"words": "영수증 카드 현금", "sentence": "영수증 주시고 카드나 현금 중 어떤 걸로 결제할까요?"}

예시 2: {"words": "공항 시간 급해요", "sentence": "공항까지 가야 하는데 시간이 없어서 급해요."}

프롬프트의 성능을 높이기 위해, **gloss**들과 연관이 높은 예시 문장과 다양한 문맥을 제공하려고 노력했다.

각 단어에 대한 문맥 설명 추가

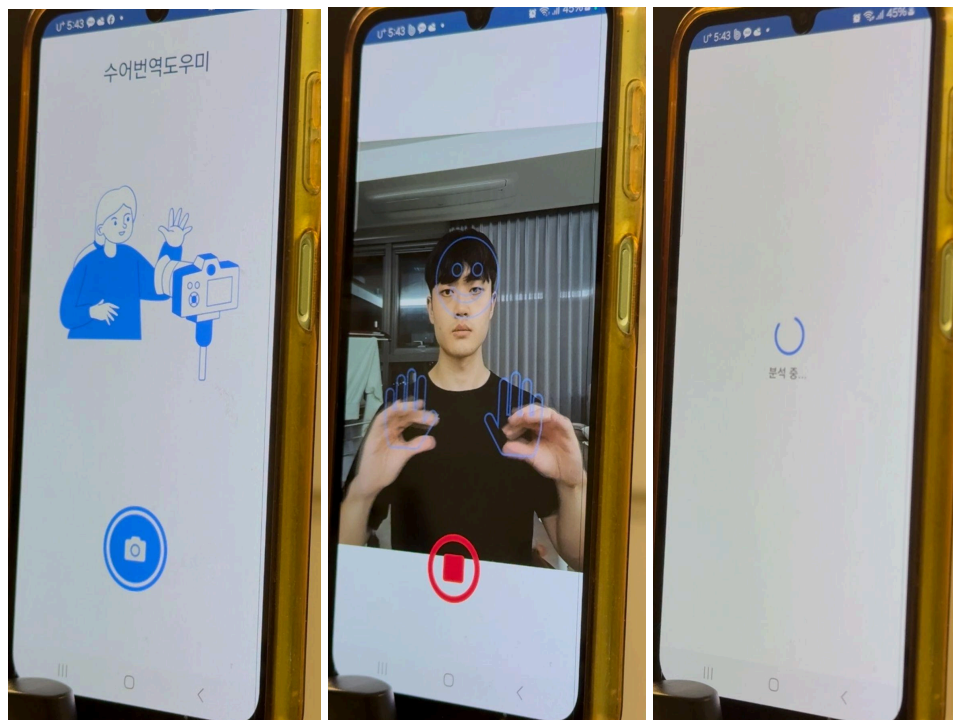
```
context = {
    "영수증": "결제 시 필요하며, 카드나 현금으로 결제할 수 있습니다.",
    "공항": "여행의 출발지나 도착지로 자주 언급됩니다.",
    "시간": "약속이나 일정에 따라 중요합니다.",
    "급해요": "시간이 부족하여 빠른 행동이 필요한 상황을 나타냅니다.",
    "주소": "목적지를 지정하는 중요한 정보입니다.",
    "길안내": "정확한 방향을 알려주는 요청입니다.",
    "차량": "탑승하고 있는 교통수단을 의미합니다.",
    "냉방": "차 안의 온도를 조절하기 위한 요청입니다.",
    "짐": "택시에서 옮기고 싶은 물건을 의미합니다.",
    "정차": "차를 멈추는 행위를 의미하며, 특정 요청을 할 때 사용됩니다."
}

context_description = "\n".join([f"{word}: {context[word]}" for word in words if word in context])
```

3.4. 애플리케이션 동작과정

3.4.1. 모바일 애플리케이션 실행

모바일 애플리케이션이 실행되면, 사용자는 홈 화면의 카메라 버튼을 터치해 녹화 화면으로 진입할 수 있다. 이 녹화 화면은 단말기의 전면 카메라를 사용하며, 수어 영상 녹화 시 얼굴과 손 모양이 화면의 특정 위치에 표시된다. 이는 사용자들이 학습된 모델에 최적화된 영상을 녹화할 수 있도록 유도한다.

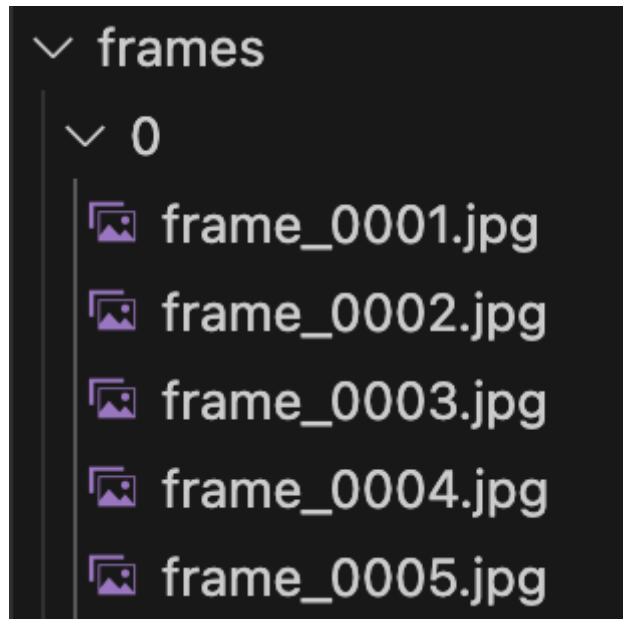


[그림 7] 모바일 애플리케이션 홈, 녹화, 대기

3.4.2. 수어 영상 녹화 및 서버로 전송

수어 영상 녹화가 끝나면, 녹화한 영상이 자동으로 서버에 전송된다. 사용자는 대기 화면에 진입하여 서버로부터 응답이 도착할 때까지 대기한다.

3.4.3. 서버 애플리케이션 동작



[그림8] 모델 입력 프레임

서버 애플리케이션은 업로드된 영상을 사전 정의된 규격에 맞춰 편집한 후, 초당 30 프레임으로 분할하여 학습된 모델이 최적화된 입력을 받을 수 있도록 처리한다. 이렇게 생성된 프레임들은 별도의 폴더에 저장되며, 해당 폴더를 학습된 모델에 전달하면, 모델은 가장 유사한 수어 gloss 조합을 반환한다.

```
Loading data finished.
[ Fri Oct 18 01:53:22 2024 ] Model: slr_network_multi.SLRModel.
[ Fri Oct 18 01:53:22 2024 ] Weights: ./best_checkpoints/dev_75.78_test_76.20_epoch9_model.pt.
<torch.utils.data.dataloader.DataLoader object at 0x7f1306970f10>
  0%|          | 0/1 [00:00<?, ?it/s]----- {'KSL': '../frames', 0: {'fi
leid': '0', 'folder': '0/*.jpg', 'signer': 'unknown', 'label': 'unknown', 'num_frames': 0, 'original_info': 'frames/*.jpg|unknown|unkno
wn'}}
***** 0
***** ../frames/0/*.jpg
100%|          | 1/1 [00:01<00:00, 1.90s/it]
지하철 다음 내리다
[ Fri Oct 18 01:53:24 2024 ] Evaluation Done.
```

[그림9] 모델 출력 gloss 조합

3.4.4. 프롬프트 엔지니어링

모델로부터 얻어낸 **gloss** 조합을 OpenAI의 API를 활용하여 프롬프트 엔지니어링을 진행한다. 이 과정에서는 문맥과 예시 문장을 프롬프트에 포함하여, **gloss** 조합이 자연스러운 문장으로 변환되도록 한다.

프롬프트 설계 시, 택시와 같은 특정 상황에 맞는 맥락을 제공하며, **few-shot** 학습 기법을 통해 미리 정의된 예시 문장과 단어 조합을 활용해 모델의 성능을 최적화한다. 예를 들어, “영수증”, “카드”, “현금”과 같은 **gloss** 조합에 대해 “영수증 주시고 카드나 현금 중 어떤 걸로 결제할까요?”와 같은 문장이 생성될 수 있도록 한다.

이와 같이 프롬프트 엔지니어링은 **gloss**에서 의미 있는 문장을 도출하는 데 최적화되어 있으며, 이를 통해 특정 상황에 적합한 자연스러운 응답을 제공할 수 있다.

```
def ask_chatgpt(words):
    few_shot_examples = generate_few_shot_examples()
    prompt = f"""다음 한글 문장들을 참고해서, 택시 상황에서 쓸 수 있는 문장을 만들어야 해.:
{few_shot_examples}
다음 단어들을 사용하여 택시 상황에 적합한 자연스러운 한국어 문장을 만들어줘.:
Words: {' '.join(words)}
Sentence: """
    try:
        response = client.chat.completions.create(
            model="gpt-3.5-turbo",
            messages=[
                {"role": "system", "content": "You are a helpful assistant that generates natural"},
                {"role": "user", "content": prompt}
            ]
        )
    return response.choices[0].message.content.strip()
```

[그림10] 프롬프트

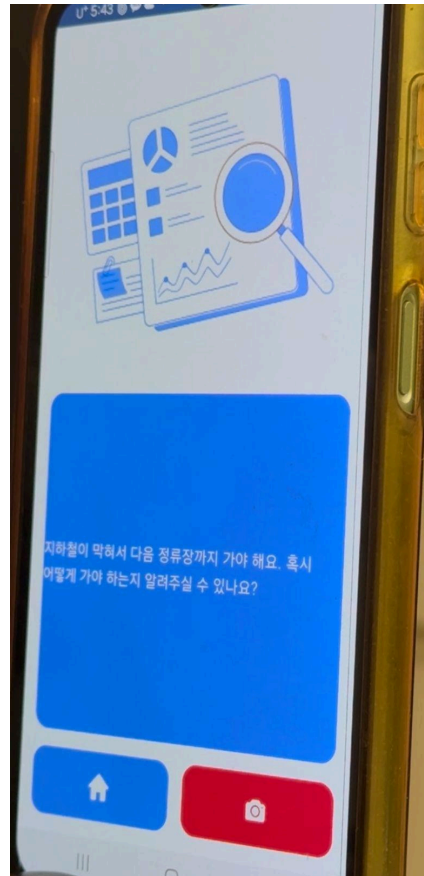
```
단어가 포함된 텍스트 파일의 이름을 입력하세요: /home/viplab/Desktop/sign_language_project/app/AihubModel/output.txt

생성된 택시 상황 문장: 지하철이 막혀서 다음 정류장까지 가야 해요. 혹시 어떻게 가야 하는지 알려주실 수 있나요?
```

[그림11] Prompt Engineering 을 통한 예측 문장 정제

3.4.5. 모바일 애플리케이션으로 결과 전송 및 확인

프롬프트 엔지니어링을 거친 최종 문장은 모바일 애플리케이션으로 전송되어 결과 화면에 표시된다. 이 문장은 사용자가 이해하기 쉽게 텍스트로 나타나며, 동시에 택시 운전자에게 들릴 수 있도록 음성으로 변환되어 출력된다. 이 과정에서 TTS(Text-to-Speech) 기능이 적용되어, 수어 영상을 기반으로 한 문장을 자연스럽게 음성으로 전달함으로써 사용자와 운전자 간의 의사소통을 돕는다.



[그림 12] 모바일 애플리케이션 출력

4. 연구 결과 분석 및 평가

4.1. 개발일정

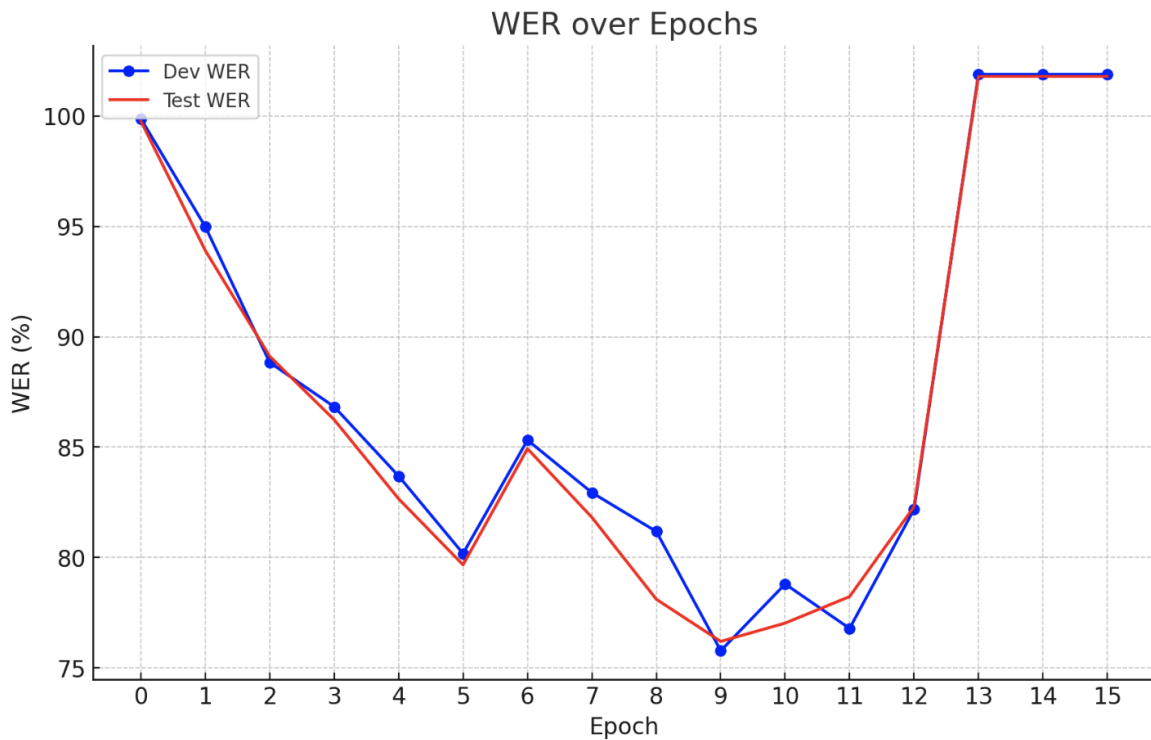
	5월			6월					7월					8월					9월					10월		
주	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3
	데이터셋 선정 및 개발 환경 구축																									
				데이터 전처리																						
							모델 구성																			
											모델 최적화															
														중간 보고서 작성												
																데이터 보충 가능 여부 확인										
																	오류 확인 및 최종 테스트									
																							최종 보고서 작성 및 발표준비			

4.2. 구성원별 역할

이름	역할
김민혁	1. 안드로이드 애플리케이션 개발 2. 웹 서버 애플리케이션 개발 3. 프롬프트 엔지니어링
한지석	1. 데이터 수집 및 전처리 2. Slowfastsing 모델 기반으로 한 아키텍처 설계 3. 모델 학습 및 검증
허재성	1. 데이터 수집 및 전처리 2. Slowfastsign 모델 성능 최적화 3. 모델 학습 및 검증

4.3. WER(Word Error Rate) 분석 및 조기종료

본 연구에서는 조기 종료 (Early Stopping) 알고리즘을 활용하여 최적의 모델 (Optimal Model)을 저장하는 방법을 적용하였다. 조기 종료는 모델이 과적합 (Overfitting)되는 것을 방지하고, 성능이 더 이상 향상되지 않을 때 훈련을 중단하는 기법으로, 모델의 일반화 능력 (Generalization Ability)을 높이는 데 기여한다.



[그림 13] epoch 당 Dev WER과 Test WER

훈련 결과, 검증 WER (Word Error Rate)는 초기 99.87%에서 시작하여 지속적으로 감소하였으며, 최종적으로 75.78%에 도달하였다. 이는 모델이 훈련 초기부터 상당한 성능 개선 (Significant Performance Improvement)을 이룬 것을 나타낸다. 그러나 9 epoch 이후에는 WER이 크게 상승하여 100%가 넘어가는 경향을 보였으며, 이러한 현상은 기울기 폭주 (Gradient Explosion) 문제로 인해 발생하였다. 기울기 폭주는 모델의 파라미터 업데이트가 지나치게 커져 학습이 불안정해지는 상황을 의미한다.

그래프를 분석해보면 최적의 성능이 75.78%에서 나타나며, 이후에는 성능 개선이 이루어지지 않았다. 이에 따라 조기 종료 기준 (Early Stopping Criteria)을 설정하여 모델의 성능이 더 이상 향상되지 않을 때 훈련을 중단하고, 최고의 성능을 기록한 모델을 저장하는 방식을 채택하였다. 이러한 방법은 최적의 모델을 확보하는 데 효과적이었다.

조기 종료 알고리즘을 통해 확보한 최적 모델은 향후 평가 (Evaluation)에서도 높은 성능을 유지할 것으로 기대된다. 본 연구의 결과는 향후 모델 성능 향상을 위한 기초 자료로 활용될 것이다.

4.4. 시행착오 및 한계

4.4.1. GPU와 메모리 용량의 제한

GPU 성능과 메모리 용량의 제한으로 인해 학습 시 **batch size**를 1로 설정해야 했으며, 대량의 데이터를 처리하는 데 어려움이 있었다. 그 결과, 536,000개의 영상 클립 중 2000개의 문장 영상 데이터만을 사용할 수 있었다. 이로 인해 모델의 일반화 성능에 제약이 있었고, 더 많은 데이터를 활용하지 못한 점이 아쉬운 부분으로 남는다.

또한 원래는 **num_workers**를 4로 설정해 데이터를 더 빠르게 로드하고 병렬 처리할 수 있었지만, 하드웨어 사양의 제약으로 인해 **num_workers**를 0으로 설정해 학습을 진행해야 했다. **num_workers**가 0일 경우 데이터 로딩이 메인 프로세스에서 순차적으로 진행되게 된다. 이는 학습 속도가 느려지고, 전체적인 효율성이 저하되는 결과를 초래했다. 이러한 제한이 학습 과정에 추가적인 부담을 주었으며, 더 나은 성능을 위한 데이터 처리 속도를 높일 수 없었던 점이 아쉬웠다.

4.4.2. CTCDecoder와 torch의 버전 충돌

CTCDecoder는 주로 음성 인식(ASR) 및 문자인식(OCR)과 같은 시퀀스 모델링 작업에서 사용되는 CTC(Connectionist Temporal Classification) 알고리즘의 디코딩 방식이다. CTC는 입력 시퀀스와 출력 시퀀스 간의 정렬이 필요하지 않은 문제를 해결하기 위해 고안된 알고리즘으로, 특히 문자나 음성을 인식하는 데 자주 사용된다.

CTCDecoder는 **torch 1.x** 버전과 주로 호환되며, 최신 버전인 **torch 2.x** 버전과는 호환되지 않는 경우가 발생할 수 있다. PyTorch 2.x 버전에서는 일부 API 변경이나 기능 제거가 있었기 때문에, 이러한 업데이트로 인해 CTCDecoder와 같은 라이브러리가 정상적으로 동작하지 않을 수 있다. 따라서 CTCDecoder를 사용해야 하는 환경에서는 **torch 1.x** 버전을 사용해야 호환성 문제를 피할 수 있었다.

4.4.3. 프롬프트 엔지니어링의 모호한 성능 지표

프롬프트 엔지니어링을 진행하면서 다양한 프롬프트를 설계했으나, 각 프롬프트의 성능을 평가하는 명확한 기준을 설정하는 데 어려움이 있었다. 예시 문장과 문맥을 제공하여 성능을 향상시키려는 시도는 있었지만, 그 결과를 객관적으로 측정할 수 있는 지표가 부재했다. 특히, 프롬프트에 따라 생성되는 문장의 자연스러움이나 정확성을 평가하는 것은

주관적일 수밖에 없었으며, 특정 상황에서 유용한 프롬프트가 다른 상황에서는 적합하지 않은 결과를 도출하는 경우도 있었다. 이러한 한계로 인해 프롬프트 엔지니어링의 효과성을 평가하는 작업이 불확실했고, 개선 방향을 구체적으로 도출하는 데 제약이 있었다.

4.4.4. 자기 폐색 문제

모델을 학습시키기 위해 사용한 데이터셋에는 정면뿐만 아니라 상, 하, 좌, 우 5개 방향에서 수어 사용자를 녹화한 영상을 제공하여, 손이 서로 가려지는 자기 폐색 문제를 완화하려는 목적이 있었다. 하지만 실제 개발 과정에서는 기기의 제약과 서비스 제공 공간이 택시라는 환경적 특성 때문에, 정면에서 촬영한 영상만을 사용할 수밖에 없었다. 결과적으로 데이터셋을 충분히 이용할 수 없었고, 자기 폐색 문제를 완벽하게 해결하지 못했다.

5. 결론 및 향후 연구 방향

수어는 손과 신체적 신호를 이용하는 시각적 언어이고, 비수지신호가 수어에서 중요한 역할을 한다는 것을 알게 되었다. 한국 수어 데이터셋은 실생활에서 발생하는 의사소통 문제를 해결하기 위해 구축되었고, 자기 폐색 문제를 해결하기 위해 다양한 각도의 영상을 제공해 수어 인식을 돕는다는 것도 알게 되었다.

Sign Language Detection 의 현재 SOTA 모델인 ‘SlowFast Sign’ 을 파인튜닝 및 학습하며 성능을 향상시켰다. 물론 환경설정 및 데이터 전처리 중 많은 시행착오와 자원의 한계로 어려움이 있었지만, LLM 학습을 위한 환경구축 및 파라미터 조정을 하며 모델을 구축할 수 있었다.

GPU 성능을 높이고 메모리 용량을 확장함으로써, 배치 크기를 늘리고 `num_workers` 값을 조정하여 데이터 로딩 속도를 개선할 필요가 있다. 이러한 개선은 모델 학습 속도를 증가시킬 뿐만 아니라, 다양한 하이퍼파라미터 조정을 가능하게 하여 최적의 성능을 이끌어낼 수 있다. 또한 메모리 용량을 확장함으로써, 더 많은 수어 데이터를 학습할 수 있게 되어 모델의 일반화 성능을 크게 향상시킬 수 있다.

논문 내용에 직접 관련이 있는 문헌에 대해서는 관련이 있는 본문 중에 참고문헌 번호를 쓰고 그 문헌을 참고문헌란에 인용 순서대로 기술한다. 참고문헌은 영문으로만 표기하며 학술지의 경우에는 저자, 제목, 학술지명, 권, 호, 쪽수, 발행년도의 순으로, 단행본은 저자, 도서명, 발행소, 발행년도의 순으로 기술한다.

[1]국생연,18-04 - 24, ‘한국수어 사용 실태 조사’ 결과 발표, 국어생활연구원,
http://malsaem.kr/bbs/board.php?bo_table=news4&wr_id=147

[2] Junseok Ahn, Youngjoon Jang, Joon Son Chung, “SLOWFAST NETWORK FOR CONTINUOUS SIGN LANGUAGE RECOGNITION”

[3] aihub

“<https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=103>”