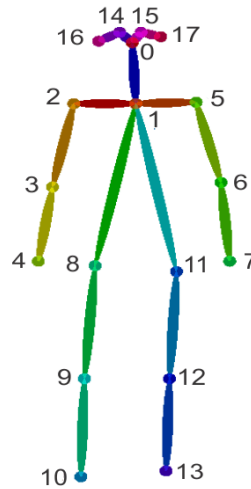


# 핵심 아이디어

수어를 텍스트로,  
AI 기반 실시간 수어 번역 서비스



수어 영상 데이터 수집



데이터 전처리



신경망 학습 모델 기반  
텍스트 생성

## AI HUB 수어 영상 데이터 셋

적용 도메인에 맞추어 선별된 단어 419개에 대한 수어 동영상 데이터

### 데이터 셋 구성

2017년도 : 105개 문장, 419개 단어를 10명의 수어 전문가가 2대의 카메라로 총 10,480개 동영상 구축

2018년도 : 105개 문장, 419개 단어를 20명의 수어 전문가가 3대의 카메라로 총 31,440개 동영상 구축

### 데이터 셋 구조

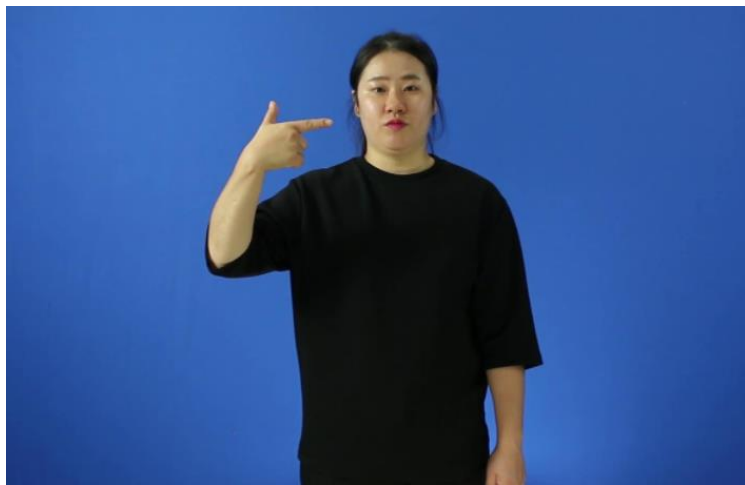
CSV 포맷 파일의 메타 정보로 동영상 연관 정보 전달

번호	언어 제공자 ID	취득연도	방향	타입(단어/문장)	파일명	한국어
8072	10	2017	측면	단어	KETI_SL_0000008072.MTS	갈히다
8380	10	2017	측면	단어	KETI_SL_0000008380.MTS	화재
3759	5	2017	정면	단어	KETI_SL_0000003759.MOV	화장실
3465	5	2017	정면	단어	KETI_SL_0000003465.MOV	감전
6093	8	2017	정면	단어	KETI_SL_0000006093.MOV	배고프다

# 핵심 메커니즘

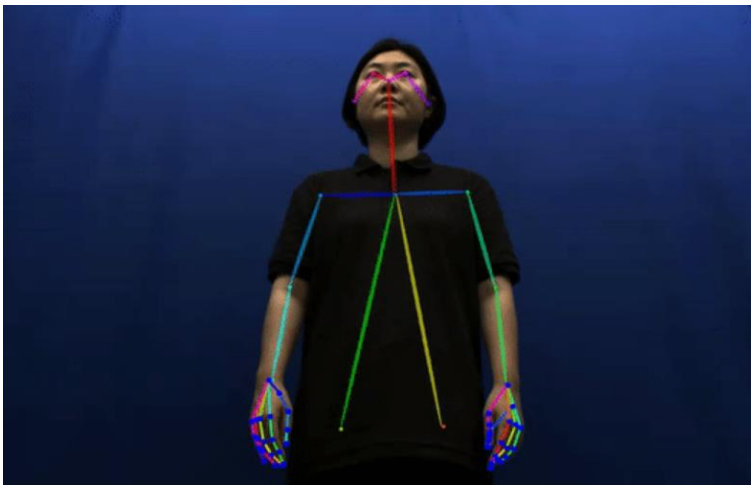
## Data Preprocessing

영상 데이터를 일정 시간으로 분할한 뒤 openpose를 이용해 각 frame에 대한 keypoint값을 추출. 특징 벡터를 정규화함으로써 데이터의 분산을 줄임



### Split frame

약 4초 정도의 수화 영상을  
130-150개 정도의 Frame으로 split 진행



### Extract Key point

Openpose를 이용해 각 frame 별로  
Key point 값을 추출해 json파일로 저장

$$V_x^* = \frac{V_x - \bar{V}_x}{\sigma(V_x)},$$

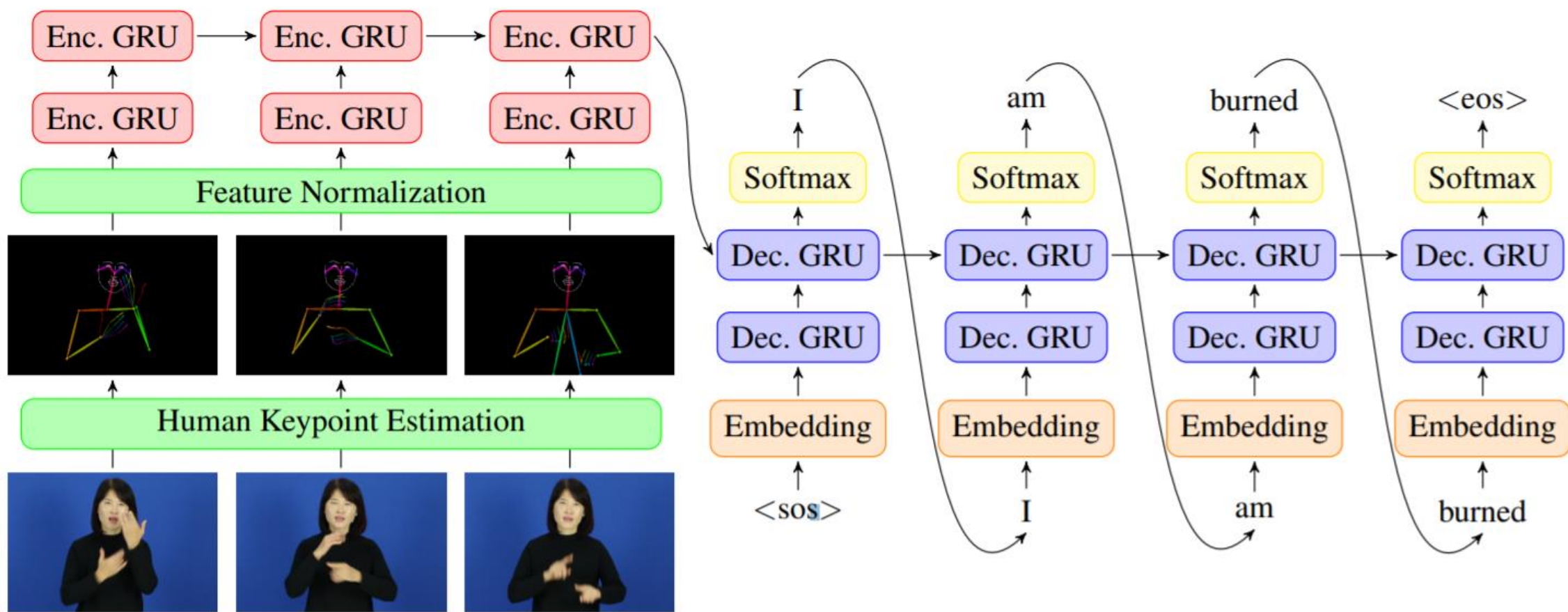
### Feature Vector Normalization

프레임 내 Keypoint의 절대 위치나 신체 부위의 스케일이 다를 수 있기 때문에 데이터의 분산을 줄이는 목적으로 벡터의 평균 및 표준편차를 사용해 특징 벡터 정규화

# 핵심 메커니즘

## Modeling

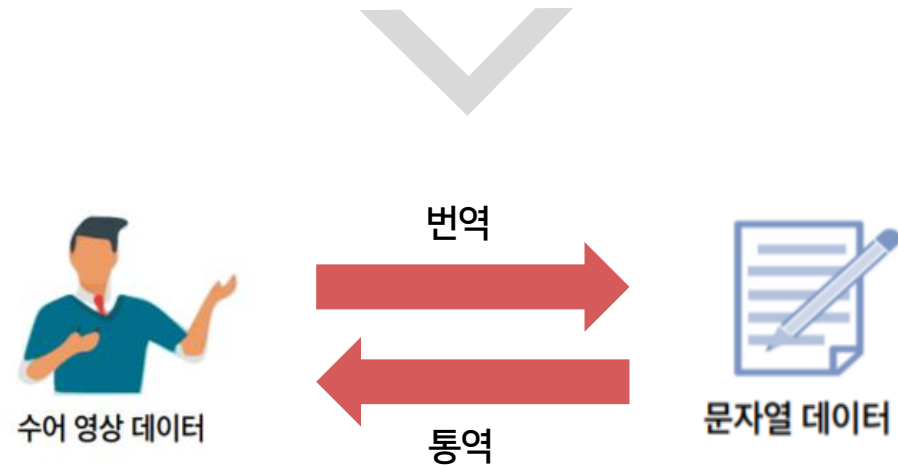
Text 생성을 위해 각 frame 별로 정규화를 진행한 key point를 frame 순서대로 sequence-to-sequence 구조에 기초한 번역 모델에 대한 입력으로 사용



# Vision

## 수어를 텍스트로, AI 기반 실시간 수어 번역 서비스

소통의 한계로 인해 농인이 겪는 사회 전반에 걸친 다양한 고충들과 피해를 줄이고자, 농인의 수어를 인공지능 기술을 통해 청인이 인지할 수 있는 형태(텍스트)로 출력하는 서비스를 제공하고자 함



나아가 텍스트로 입력되는 언어를 수어 체계로 변환한 뒤  
영상으로 매칭하는 서비스를 구상하여 양방향 소통이 가능하도록 하는 서비스 구상 중