

제 15회 BOAZ 컨퍼런스

수어 번역을 통한 위험 상황 속 의사소통 시스템 구축

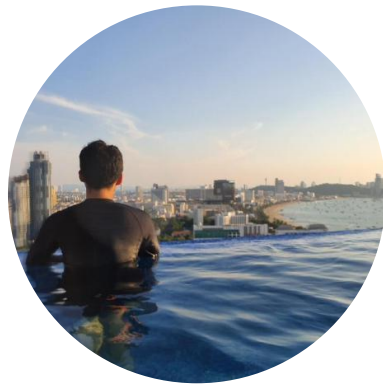
hands-on.

형용사. (말만 하지 않고) 직접 해 보는[실천하는]

Hands-on Contributors.



곽민지



김영민



김영은



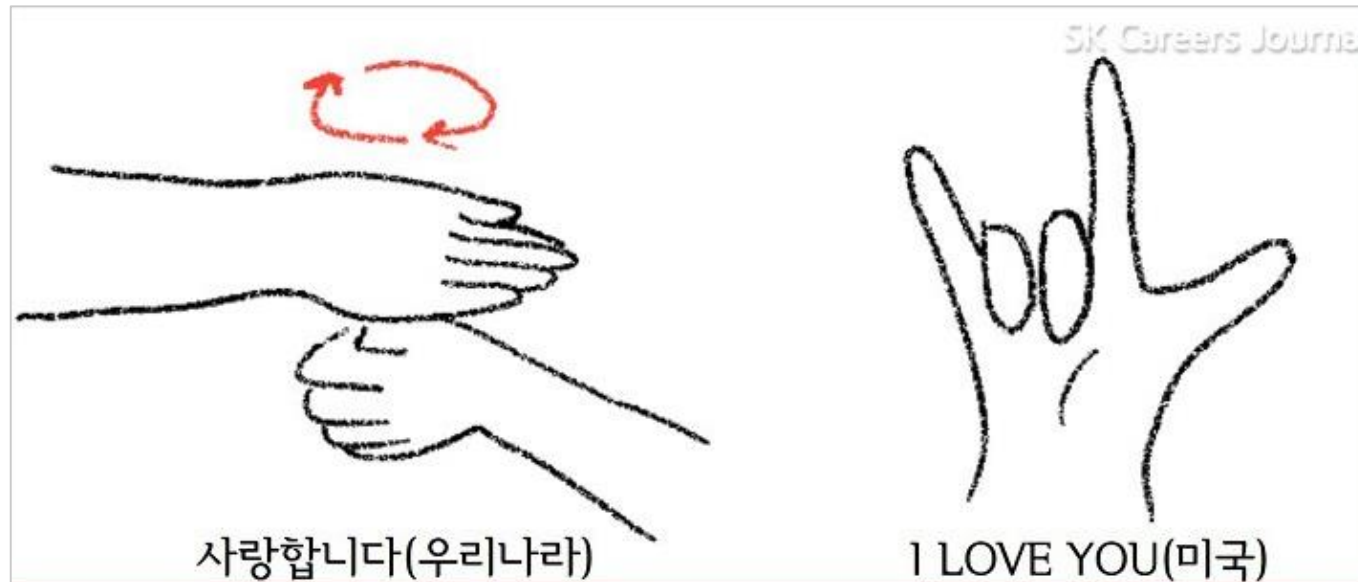
이다인

Contents

- 01 Introduction 주제 선정 배경 | 핵심 아이디어
- 02 Preparation 데이터 수집 | 데이터 전처리
- 03 Modeling Seq2Seq | Experiment | Result | 시연
- 04 Vision 활용방안 및 기대효과 | 프로젝트 개선 방향

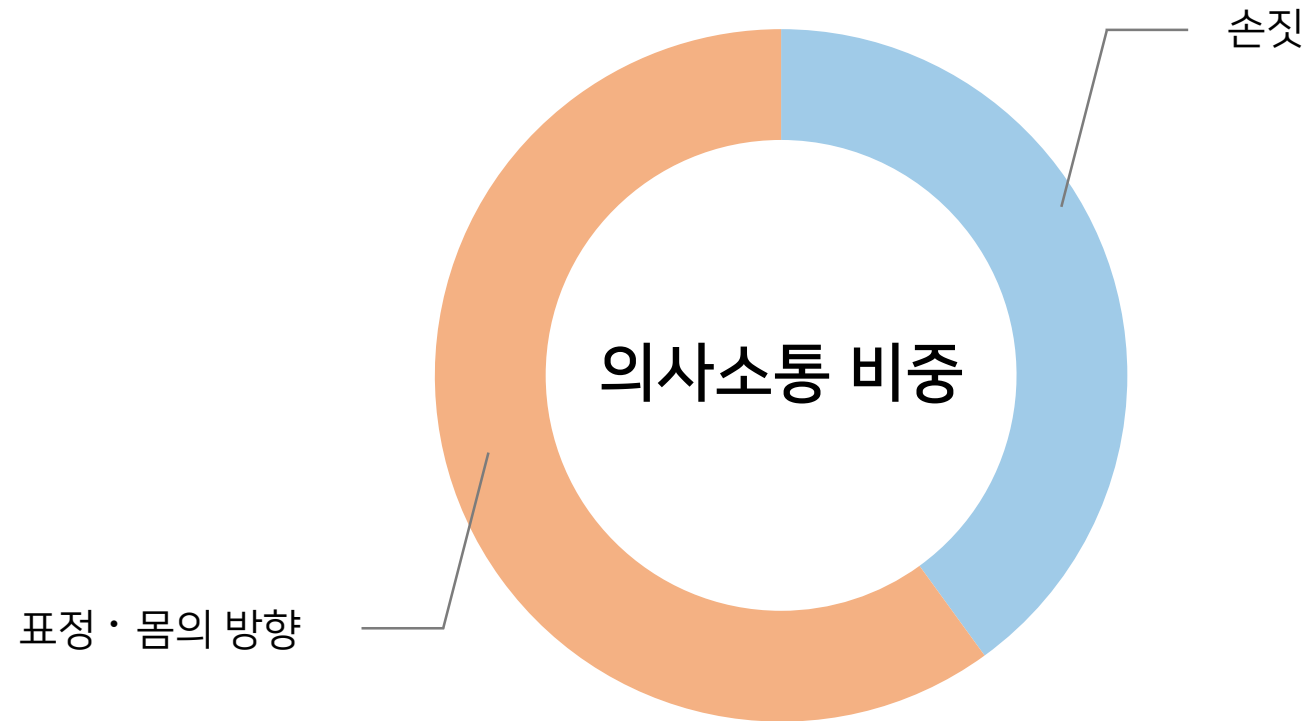
한국 수어(手語)

대한민국 농인의 고유한 ‘보이는 언어’



Introduction

(1) 주제 선정 배경



수어는 약 40%의 손짓과 60%의 다른 요소들로 이루어짐

Introduction

(1) 주제 선정 배경

- 손의 모양은 같으나 표정에 따라 의미가 다름



맛있다



맛있니?

- 손의 모양은 같으나 움직이는 방향에 따라 의미가 다름



오전



오후

위험 상황에서 더 심각한, 농인의 열악한 처지

재난 상황에 장애인이 있을 수 있다는 고려 자체가 없는 상황

위험 상황 인지에 대한 어려움

- 화재 상황에서 대피 알람이나 폭발음 등을 듣지 못하는 경우 多
- 소리 및 진동을 인지하기 어려워 교통수단에 의한 위험성이 높음

응급 상황 전달의 한계

- 응급 상황에서 사고 접수에 대한 어려움 존재
- 수어 통역자의 인력 부족

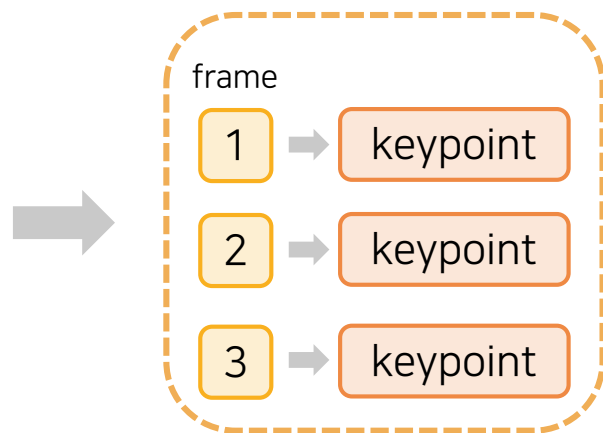
Introduction

(2) 핵심 아이디어

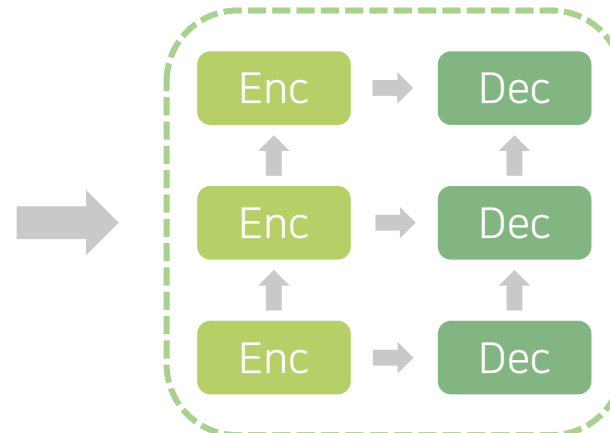
수어를 텍스트로, AI 기반 수어 번역 서비스



수어 데이터 수집



데이터 전처리



모델링 진행



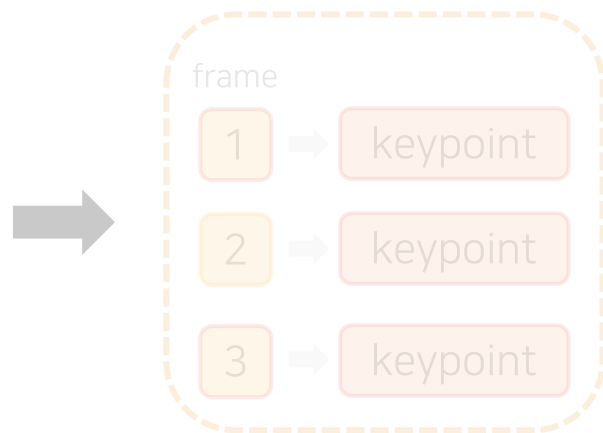
번역 텍스트 추출

Preparation

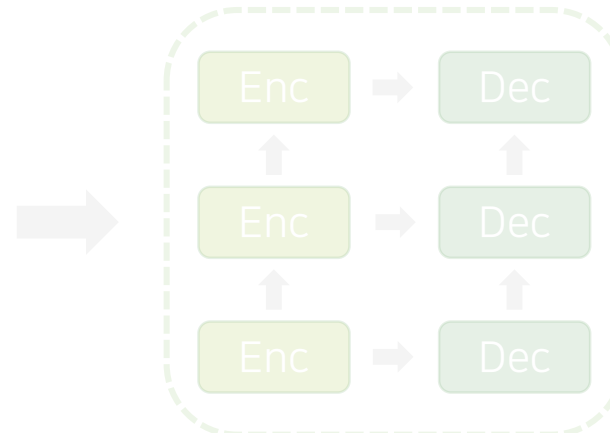
(1) 데이터 수집



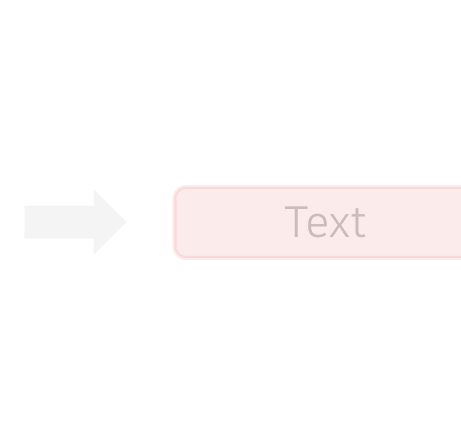
수어 데이터 수집



데이터 전처리



모델링 진행



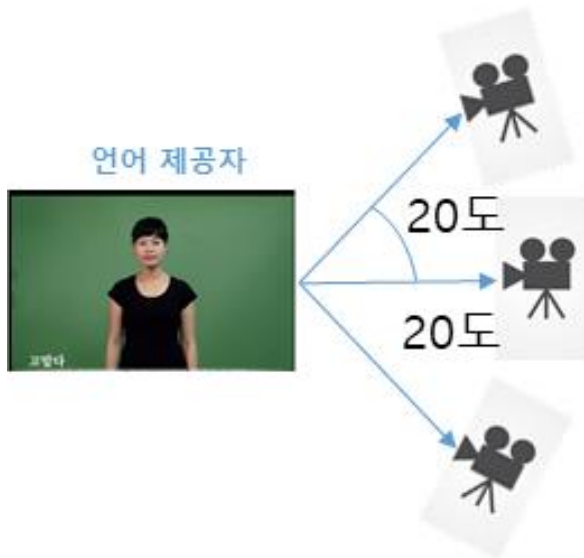
번역 텍스트 추출

Preparation

(1) 데이터 수집

[AI HUB 수어 영상 데이터 셋]

적용 도메인에 맞추어 선별된 단어 419개에 대한 수어 동영상 데이터



데이터 셋 구성

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

적용 도메인: 위험상황

구축 내용: Full HD(1920×1080) 이상의 화질

구축량: KETI_SL_0000043178 ~ KETI_SL_0000043492

Preparation

(1) 데이터 수집

[AI HUB 수어 영상 데이터 셋]

적용 도메인에 맞추어 선별된 단어 419개에 대한 수어 동영상 데이터

대표 도면



데이터 구조

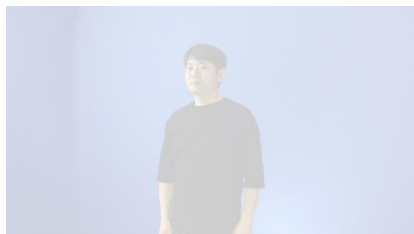
번호	언어 제공자 ID	취득연도	방향	타입(단어/문장)	파일명	한국어
8072	10	2017	측면	단어	KETI_SL_0000008072.MTS	갈히다
8380	10	2017	측면	단어	KETI_SL_0000008380.MTS	화재
3759	5	2017	정면	단어	KETI_SL_0000003759.MOV	화장실
3465	5	2017	정면	단어	KETI_SL_0000003465.MOV	감전
6093	8	2017	정면	단어	KETI_SL_0000006093.MOV	배고프다

CSV 파일 예시.

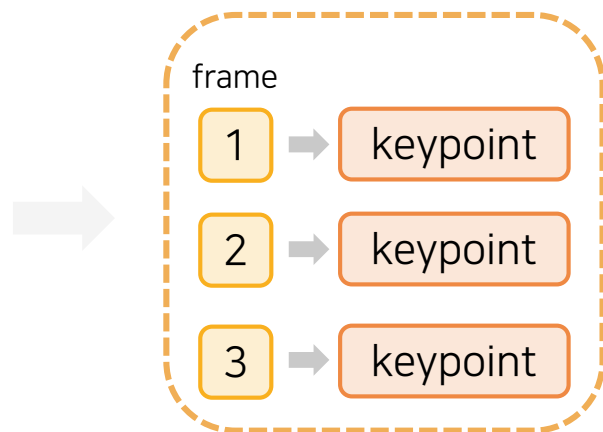
CSV 포맷 파일의 메타 정보로 동영상 연관 정보 전달

Preparation

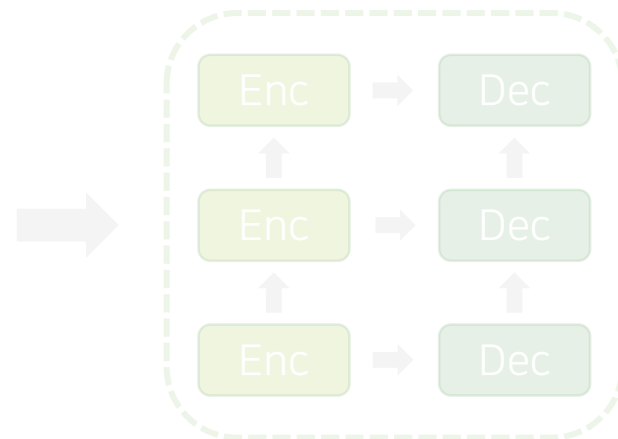
(2) 데이터 전처리



수어 데이터 수집



데이터 전처리



모델링 진행



번역 텍스트 추출

Preparation

(2) 데이터 전처리

각 video를 frame으로 split

약 4초 정도의 수어 영상을 30 fps로 split



KETI_SL_0000042734.mp4



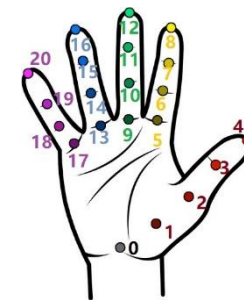
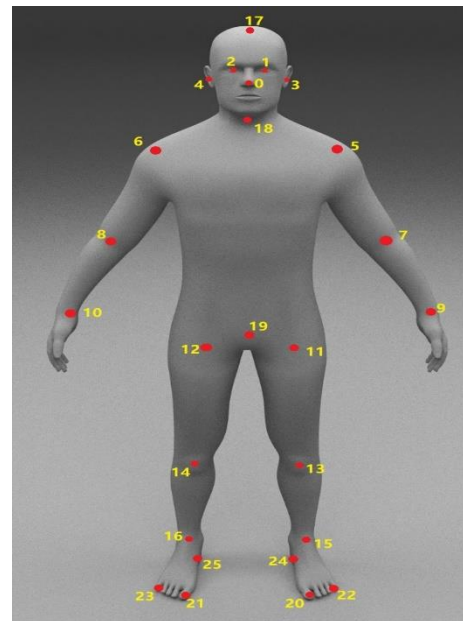
KETI_SL_0000042734_0.jpg ~ KETI_SL_0000042734_81.jpg

Preparation

(2) 데이터 전처리

Pose Estimation

- Alphapose 사용하여 각 frame 별로 keypoint 값 추출.
→ json 파일로 저장
- 전체 keypoint 136개 중 필요 없는 13개 삭제
→ 총 123개 keypoint index 추출
- keypoint matrix



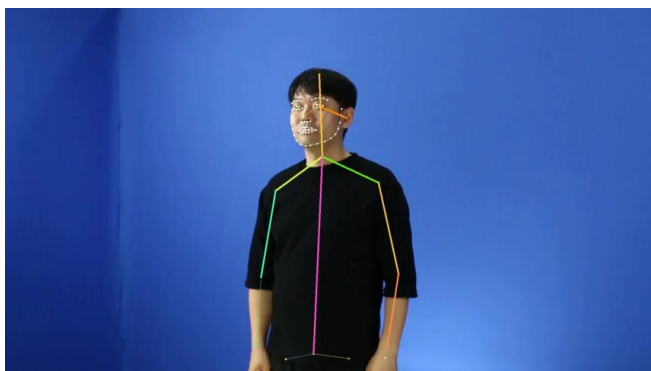
Index	0	1	2	3	4	5	6	7
point	Nose	LEye	REye	LEar	REar	LShoulder	RShoulder	LElbow

8	9	10	11	12	13~80	81~101	102~123
RElbow	LWrist	RWrist	Head	Neck	Face	LHand	RHand

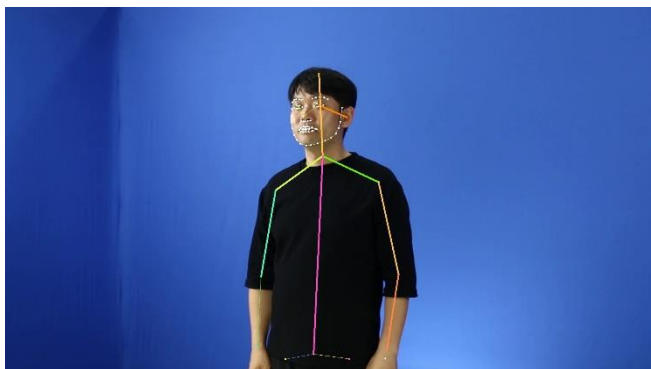
Preparation

(2) 데이터 전처리

Pose Estimation



KETI_SL_0000042734.mp4



KETI_SL_0000042734_0.jpg

- 저장 형식

: 각 body keypoints index 순서대로 [X, Y, Confidence값]



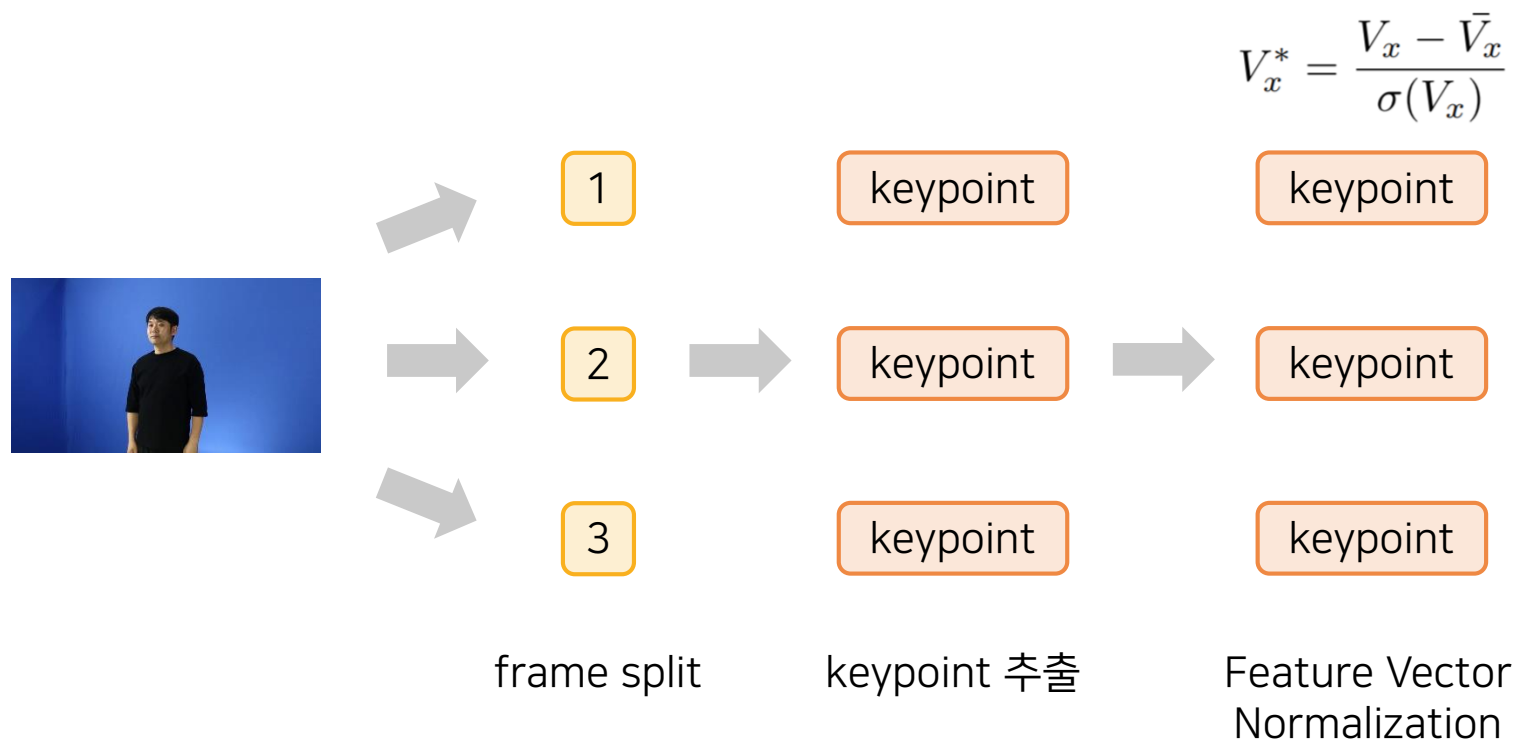
```
{"KETI_SL_0000042734_0.jpg": {"keypoints": [[  
  [864.470947265625, 340.0668029785156,  
  0.3383473753929138], [907.4041137695312,  
  303.8307189941406, 0.36455047130584717], ... , [...]]]}}
```

Preparation

(2) 데이터 전처리

Feature Vector Normalization

Frame마다 keypoint의 절대 위치나 신체 분위의 스케일이 다를 수 있으므로 정규화 진행

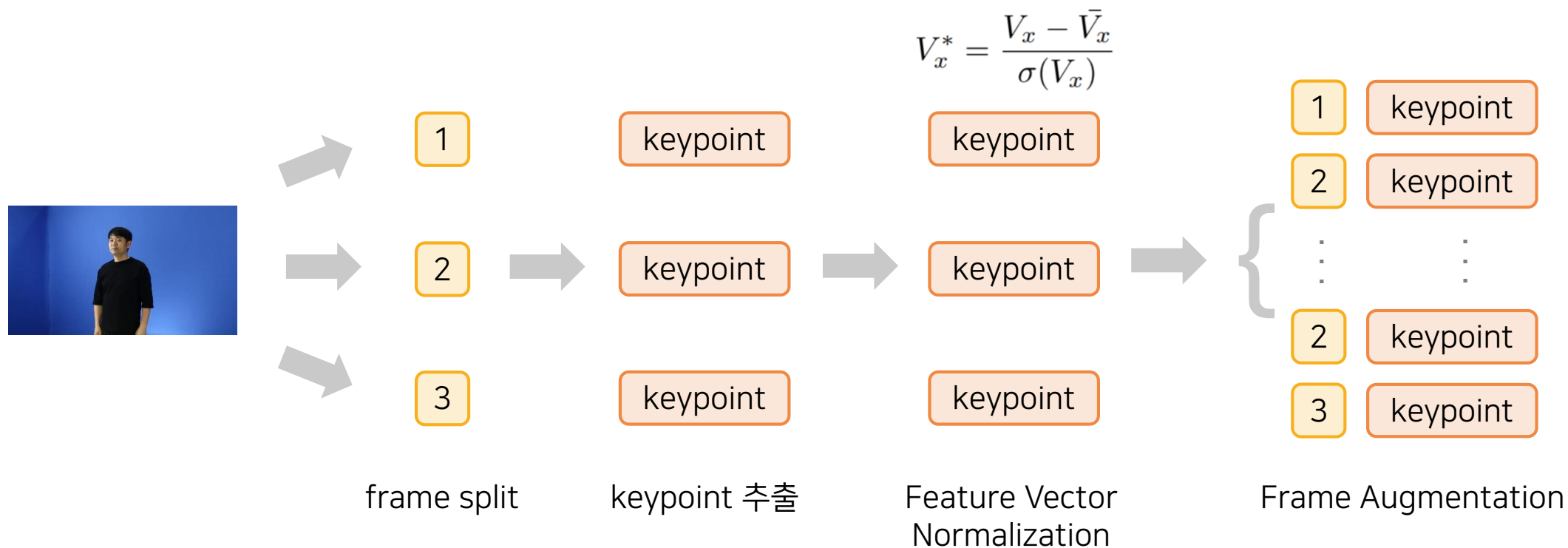


Preparation

(2) 데이터 전처리

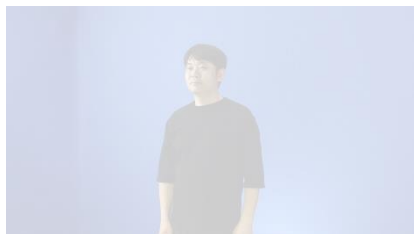
Frame Augmentation

video의 frame을 random하게 선택하여 frame 증강

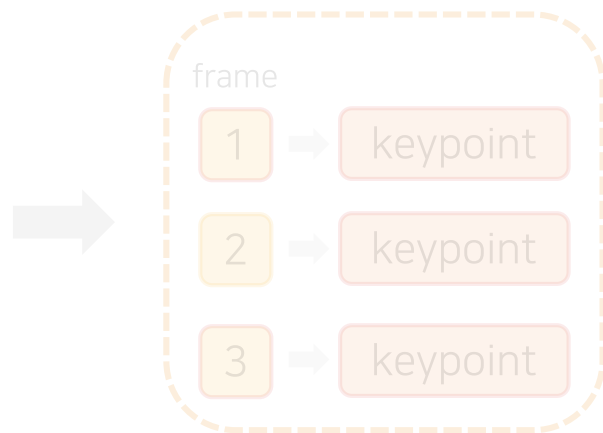


Modeling

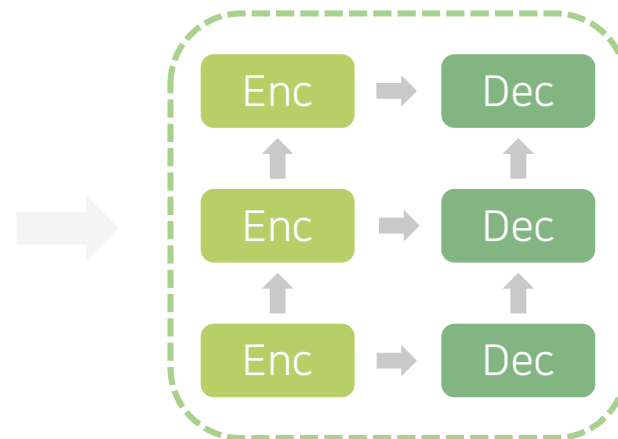
(1) Seq2Seq



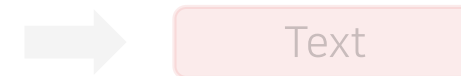
수어 데이터 수집



데이터 전처리



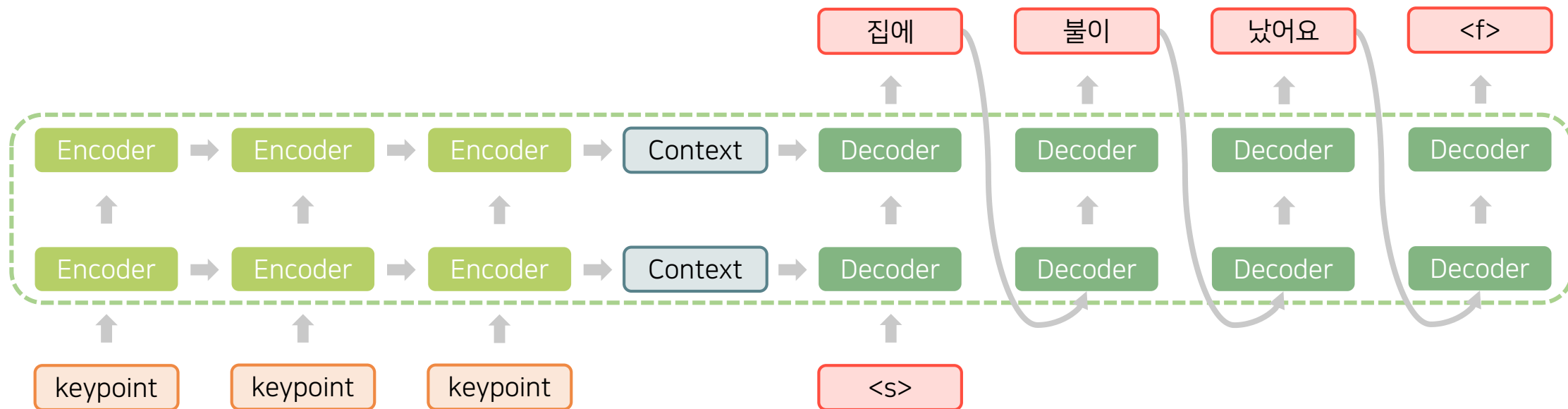
모델링 진행



번역 텍스트 추출

Modeling

(1) Seq2Seq



Modeling

(2) Experiment

	Compare of	Detail
train dataset: 5,703 validation dataset : 1,426 test dataset: 875 Epoch: 100	Augmentation Method	Hand Frame
		All Frame
	Inference Method	Original
		Reverse
	Body keypoint	Hand + Face + Body
		Hand + Body
	Normalization Method	Frame Normalization
		Video Normalization

Modeling

(2) Experiment

Compare of Augmentation Method (Hand Frame(Baseline) vs All Frame)

Model	Hyperparameter	Metrics	Hands Frame	All Frame
GRU-Attention	Adam CrossEntropy	BLEU	90.1	93.1
		Accuracy	90.4	93.3
	AdamW Scheduler	BLEU	90.0	93.7
		Accuracy	89.8	94.2
LSTM	Adam CrossEntropy	BLEU	73.4	61.5
		Accuracy	73.3	61.5
	AdamW Scheduler	BLEU	73.4	52.2
		Accuracy	73.3	52.0

Modeling

(2) Experiment

Compare of Inference Method (Original(Baseline) vs Reverse)

Model	Hyperparameter	Metrics	Original	Reverse
GRU-Attention	Adam CrossEntropy	BLEU	90.1	89.5
		Accuracy	90.4	89.7
	AdamW Scheduler	BLEU	90.0	90.7
		Accuracy	89.8	90.6
LSTM	Adam CrossEntropy	BLEU	73.4	66.3
		Accuracy	73.3	66.4
	AdamW Scheduler	BLEU	73.4	64.9
		Accuracy	73.3	64.8

Modeling

(2) Experiment

Compare of Body Keypoint (Hand + Face + Body(Baseline) vs Hand + Body)

Model	Hyperparameter	Metrics	Hand + Face + Body	Hand + Body
GRU-Attention	Adam CrossEntropy	BLEU	90.1	91.1
		Accuracy	90.4	91.8
	AdamW Scheduler	BLEU	90.0	93.1
		Accuracy	89.8	93.2
LSTM	Adam CrossEntropy	BLEU	73.4	50.7
		Accuracy	73.3	50.8
	AdamW Scheduler	BLEU	73.4	56.9
		Accuracy	73.3	56.9

Modeling

(2) Experiment

Compare of Body Normalization Method (Frame Normalization(Baseline) vs Video Normalization)

Model	Hyperparameter	Metrics	Frame Normalization	Video Normalization
GRU-Attention	Adam CrossEntropy	BLEU	90.1	86.2
		Accuracy	90.4	86.5
	AdamW Scheduler	BLEU	90.0	89.7
		Accuracy	89.8	90.0
LSTM	Adam CrossEntropy	BLEU	73.4	59.6
		Accuracy	73.3	59.4
	AdamW Scheduler	BLEU	73.4	60.8
		Accuracy	73.3	60.8

Modeling

(2) Experiment

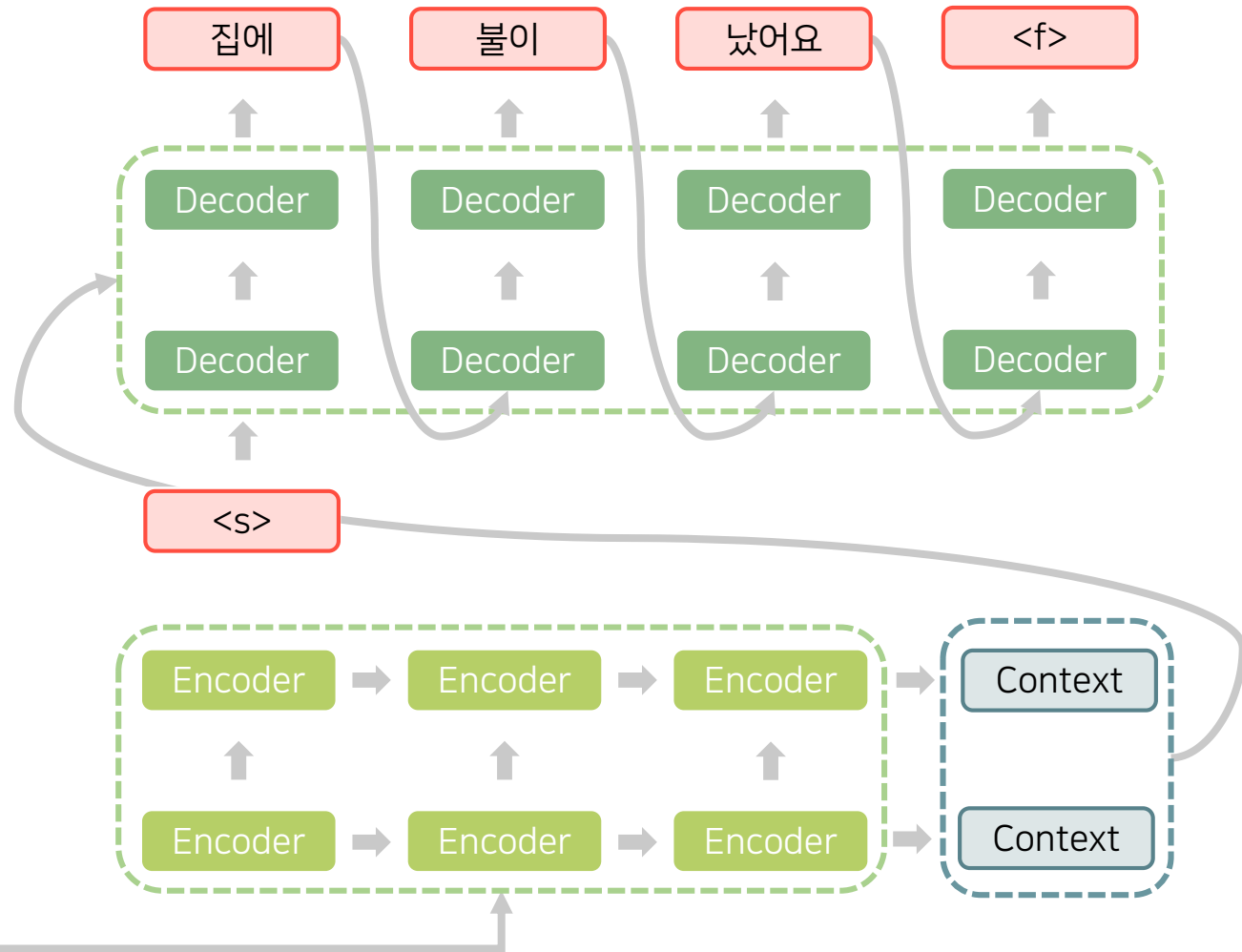
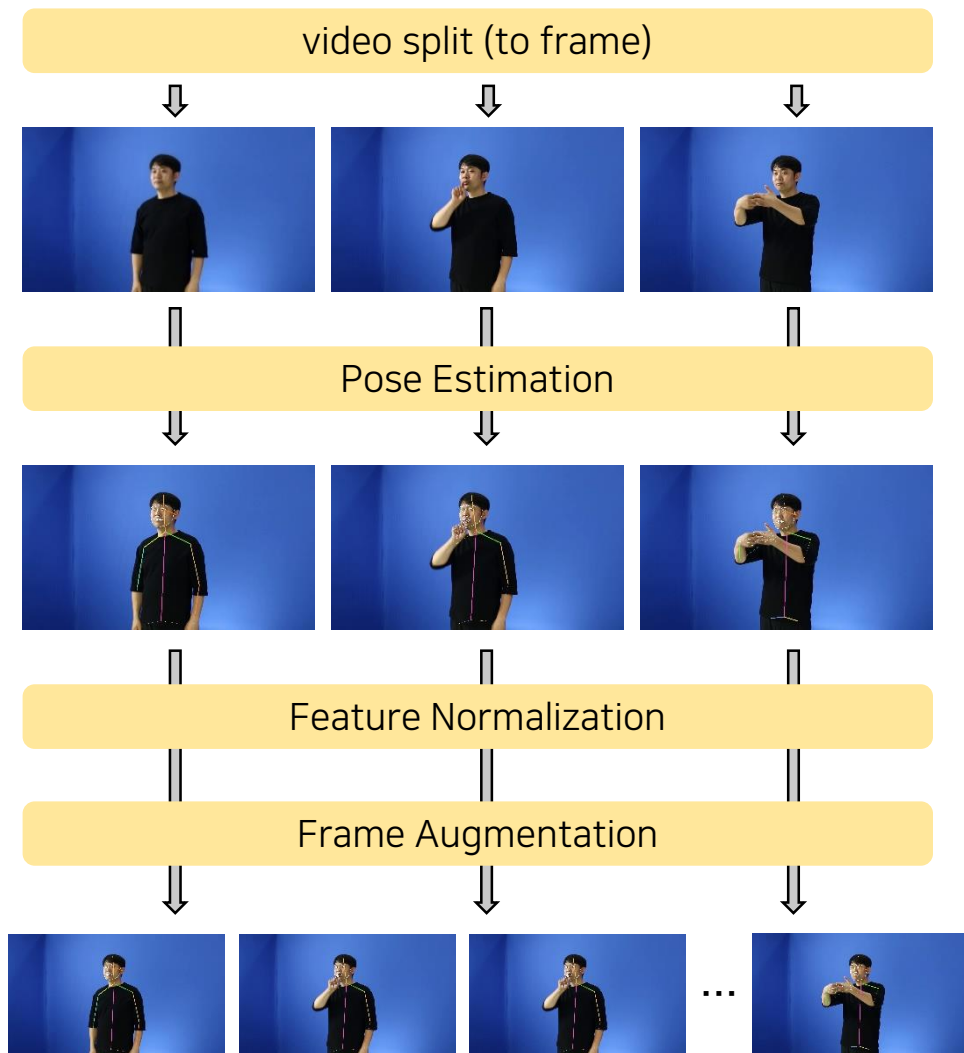
Final Model

(All Frame Augmentation / Reverse / Hand + Body / Frame Normalization)

Model	Hyperparameter	Metrics	Final Model
GRU-Attention	Adam CrossEntropy	BLEU	93.4
		Accuracy	93.5
	AdamW Scheduler	BLEU	95.1
		Accuracy	95.0
LSTM	Adam CrossEntropy	BLEU	49.6
		Accuracy	50.0
	AdamW Scheduler	BLEU	51.5
		Accuracy	51.5

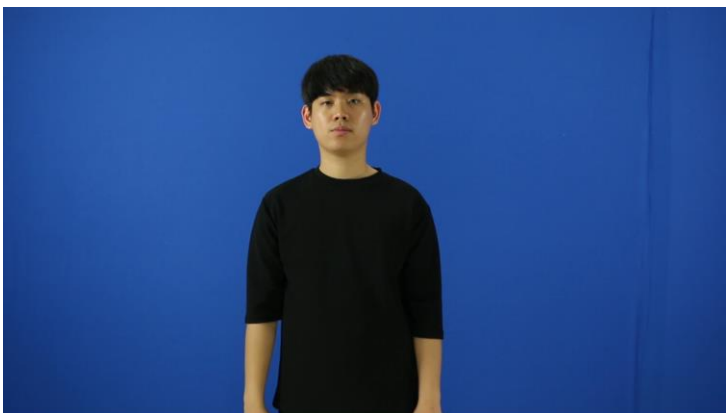
Modeling

(3) Result



Modeling

(3) Result



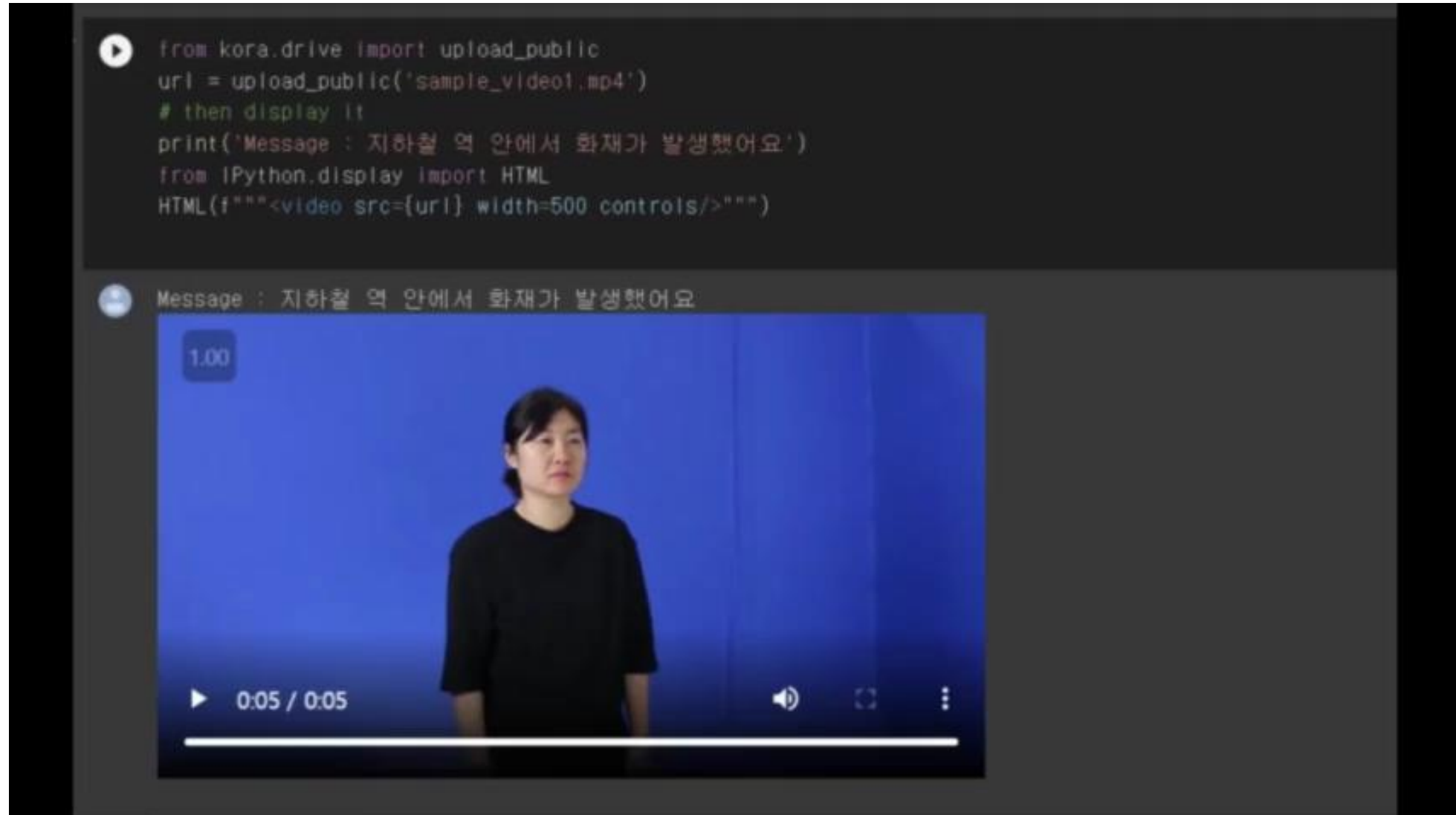
정답: 119 구조대를 불러주세요.

번역 결과: 119구조대를 불러주세요.

dataset	정답	번역 결과
KETI_SL_0000042303	복통이 심해요	복통이 심해요
KETI_SL_0000042304	아파서 못 참을 것 같아요	아파서 못 참을 것 같아요
KETI_SL_0000042305	빨리 도와주세요	빨리 도와주세요
KETI_SL_0000042306	아이가 방에 갇혔어요	아이가 방에 갇혔어요

Modeling

(4) 시연



Modeling

(4) 시연



github

<https://github.com/winston1214/Sign-Langugage-project>

Sign-Langugage-project

Install Module

```
$ pip install pyyaml==5.2
$ pip install scipy==1.1.0
$ pip install pillow==6.2.2
$ pip install torch==1.2.0 torchvision==0.4.0
$ python -m pip install cython
$ sudo apt-get install libyaml-dev
```

Setting(Alphapose)

```
$ git clone https://github.com/winston1214/Sign-Langugage-project.git && cd Sign-Langugage-project
$ python setup.py build develop
```

```
PRO Inference.ipynb
파일 수정 보기 삽입 런타임 도구 도움말
+ 코드 + 텍스트 Drive로 복사

> Setting
[ ] 숨겨진 셀 5개

> Weight file Download
[ ] 숨겨진 셀 1개

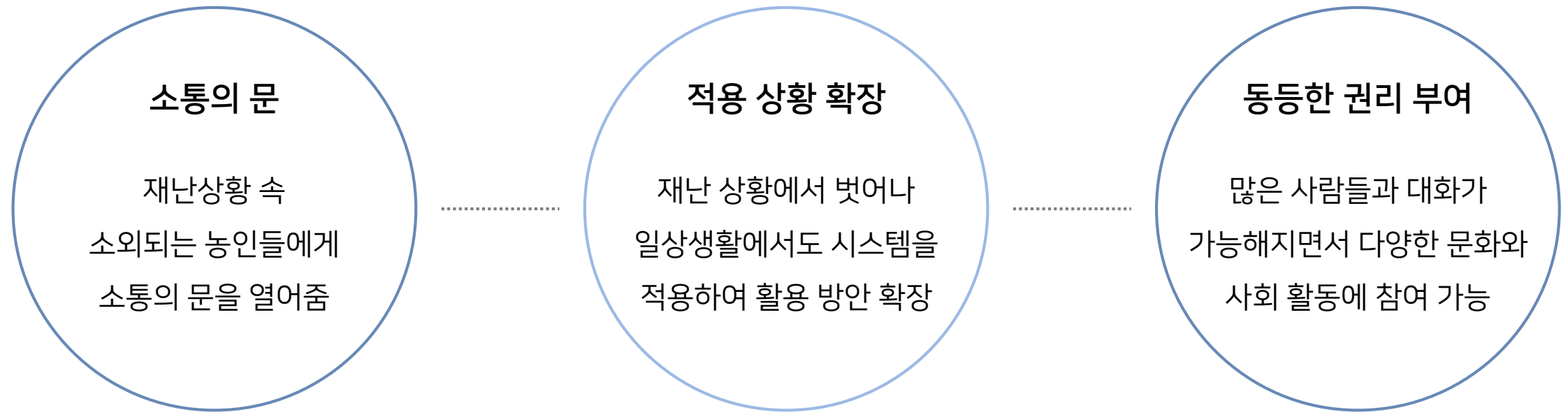
> Sample Video File Download
[ ] 숨겨진 셀 4개

> Inference
[11] !python inference.py --video sample_video1.mp4 --outdir ./ --pt best_GRU_TUNNING.pt --model GRU
video split complete
Loading YOLO model..
Loading pose model from pretrained_models/halpel36_fast_res50_256x192.pth...
23% 37/160 [00:03<00:13, 9.30it/s]===== This video get 159 frames in total.
99% 159/160 [04:27<00:01, 1.68s/it]
===== Finish Model Running.
Results have been written to json.
device : cuda:0
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
Sign Language Translate : 119 구조대를 불러주세요
Time : 281.56188344955444 sec
^C

[12] !python inference.py --video sample_video2.mp4 --outdir ./ --pt best_GRU_TUNNING.pt --model GRU
```

Vision

(1) 활용방안 및 기대효과



Vision

(2) 프로젝트 개선 방향

실시간성

real-time의 시스템을 구축하여 실시간 사용이 가능하도록 함

모델 학습

keypoint 추출 모델을 새롭게 학습시켜 사용

언어 확장

한국 수어 뿐만 아니라 다른 나라의 수어에도 시스템 적용

생성 모델

번역 모델과 더불어 텍스트를 수어로 생성하는 모델을 개발하여 쌍방향 소통이 가능하도록 함



QnA