시청각 콘텐츠 맞춤
AI 기반 수어 통역 영상 제작 솔루션

-인공지능 학습용 데이터 활용 아이디어 공모전-

2021. 01. 20

딥아이 | 김진수 | 김종원

Contents

시청각 콘텐츠 맞춤 AI 기반
수어 통역 영상 제작 솔루션

- 데이터 활용 아이디어 내용 (Problem)
- 실현 서비스 개요 (Solution)
- 학습용 데이터 선정 (Solution)
- 딥러닝 학습 방법 (Method)

제안 동기

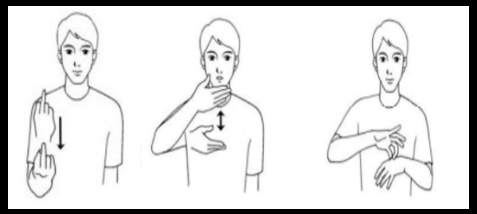
‘손으로 전하는 목소리’ 수어

소리가 아닌 시각으로 소통하는 또 하나의 언어 체계

음성 대신 손의 움직임을 포함한 신체적 신호를 이용하여
의사를 전달하는 시각 언어 Sign Language

동생이 고기를 먹는다

< 한국어 >



동생 - 먹다 - 고기

< 한국 수어 >

01 Sign language

수어는 한국 문법과 다른 문법 체계를 지닌 하나의 언어 체계이며 시각적 정보를 매개로 구현되고 있습니다

02 Natural language

수어를 사용하는 농인들에게 한국어 자막의 문장과 어순은 익숙하지 않은 외국어와 같습니다.

제안 동기

농인들의 '알 권리'와 '언어권 보장'

다양한 분야에서 수화 통역 서비스 제공을 위한 노력, 하지만…

[정부 수어 통역 제공 예시]

[온라인 동영상 스트리밍 서비스 예시]

농인들을 위한 수어 통역

- 정부는 2020년 2월부터 코로나19 현장 브리핑 생중계에 수어 통역 지원 시작
- 대국민 담화나 주요 정책, 재난 상황 발표 현장 등의 방역 대책을 위한 **공공분야에서도 알 권리와 언어권 보장을 위해 수어 통역 제공**

열악한 국내 수어 통역 환경

- 다양한 분야에서 수어 통역 서비스를 제공하려고 노력 중이지만, 전국 기준 **40만명의 농인을 위한 통역사는 600명 수준으로 매우 열악한 상황**
- 인력 및 시간, 비용 등의 문제로 제공되는 분야가 제한적이며 최근 코로나19로 인한 **통역사의 안전 관련 이슈**가 제기되고 있음

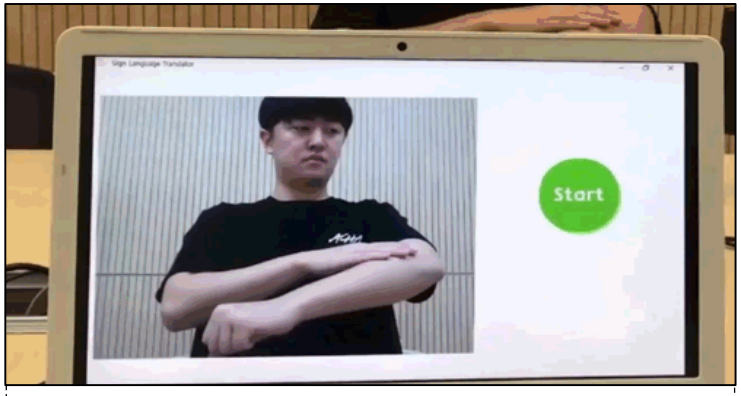
누군가에겐 먼 이야기, OTT 서비스

- 유튜브나 넷플릭스 등에서 제공하는 **온라인 동영상 스트리밍은 보편적인** 서비스가 되며 매년 꾸준히 이용자 수가 증가하며 성장 중
- **청각 장애인들에겐 여전히 불편하고 이용하기 어려운 서비스**이며, 자막, 시청각 보조 자료 등과 관련 정책과 조항이 있지만 여전히 실효성이 낮음

문제 인식

소통과 권리를 위한 현대화된 변화와 혁신이 필요한 시점

소통을 위한 지능형 번역



- 딥러닝을 이용하여 수어를 번역하는 자연어 처리 솔루션 (CNN과 RNN 알고리즘을 응용)
- 시각 언어체계 변환을 위해 영상이 입력되면 이를 텍스트로 변환하는 방식 (vision to text)
- **인물의 포즈, 제스처, 표정, 속도 등의 데이터 기반 해석 기술이 요구되어 구현 난이도가 높음**

택배, 주문 등의 생활 서비스와
관공서의 공공 시설 무인 안내서비스 활용

수어 영상 데이터

번역 ↓

↑ 통역

문자열 데이터

권리 보장을 위한 통역

- 텍스트로 입력되는 언어를 자연어 처리를 통해 수어 체계로 변환 뒤 애니메이션 영상 매칭
- 구문간의 연결성 부족 및 의문, 강조 등의 문법 표현을 수어로 구현하는 기술의 한계를 지님
- **표정이나 자세, 수어의 빠르기 등의 구현 정확도가 떨어져 실제 활용도가 낮음**

재난, 정책 방송 등 국민들의
알 권리를 위한 통역 서비스 활용

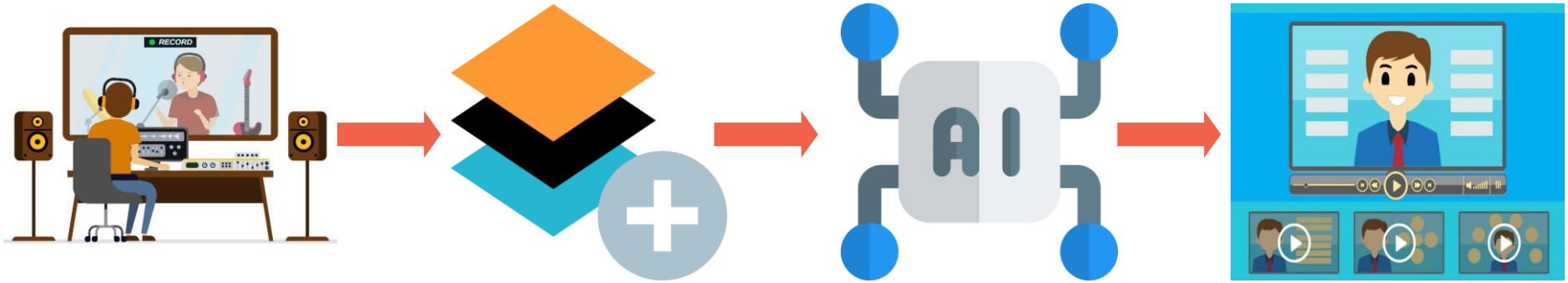
수어는 부가 서비스가 아닌 또 다른 언어,

시청각 콘텐츠 맞춤 AI 기반
수어 통역 영상 제작 솔루션

핵심 아이디어

수어는 부가 서비스가 아닌 또 다른 언어,

시청각 콘텐츠 맞춤 AI 기반
수어 통역 영상 제작 솔루션



영상 콘텐츠
대본 수집

수어 텍스트
데이터 전처리

신경망 학습 모델
기반 영상 생성

콘텐츠 맞춤
수어 영상 제작

핵심 아이디어

AI 수어 영상 제작 솔루션

Flowchart

문자열 데이터

"안녕하세요,
만나서 반갑습니다"

번역 엔진

수어 데이터

기존 수어 통역

[통역사에 의한 수어 통역]

[AI 수어 데이터셋]

수어 스크립트

"안녕하세요, 만나서 반갑습니다"

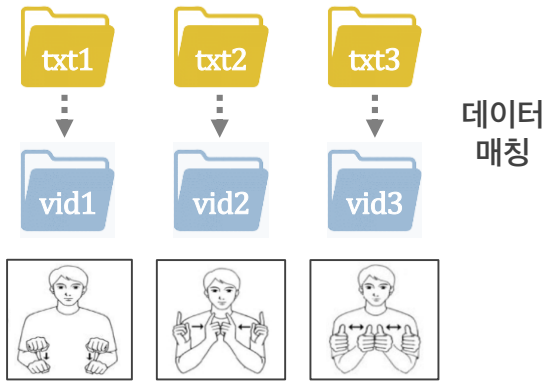
문자열 데이터 전처리

i) 품사 분석

ii) 의존 구문 분석

iii) 영역 개체 추출

"안녕하세요", "만나다", 반갑다"



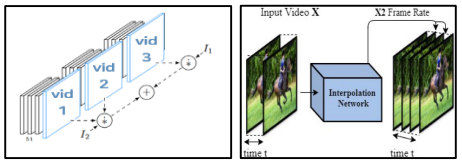
데이터 매칭

[정렬된 영상 클립]

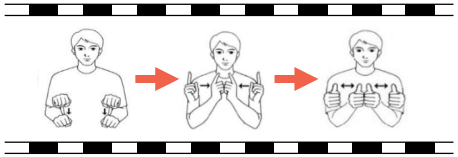
영상 생성

영상 연결

수어 영상 데이터 생성



GAN을 활용한
Frame Connection



GAN을 활용한
Motion Generation

[생성된 수어 통역 영상]

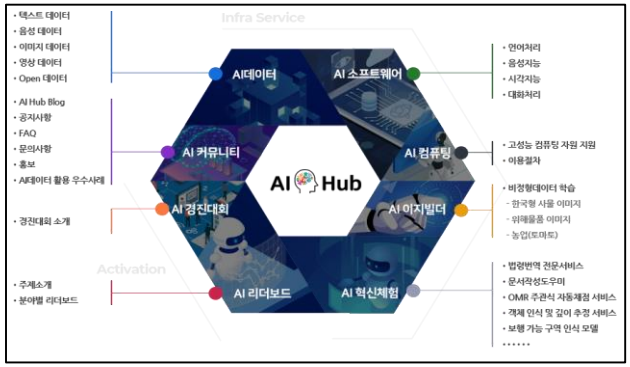
실현 가능성

GAN 신경망 기술의 가파른 성장과 발전

Generative Adversarial Network

GAN을 통한 고해상도 영상 생성 및 인물의 자세 구현 정확도 상승 및
한국형 AI 구축 사업으로 인한 아이디어 실현 가능성 확보

[GAN 기술 활용 예시]



[AI 허브에서 지원하는 다양한 프로젝트 및 데이터]

심층 합성곱 생산적 적대 신경망의 발전

- 2016년 DCGAN (Deep Convolutional Generative Adversarial Network)이 제안된 이후, GAN 관련 연구와 기술은 매우 빠르게 성장 중
- 단순히 새로운 영상을 생성하는 것을 넘어 **흑백 사진 복원, 패션 디자인 보조, 변조 영상 판별 기법 등 산업 비즈니스 모델과 융합되며 산업 분야에서의 활용 가능성이 높게 평가됨**
- 충분한 학습 데이터와 조건만 형성된다면, **영상을 정교하게 구현 가능한 수준까지 도달**

한국판 뉴딜의 핵심, 디지털 뉴딜 사업

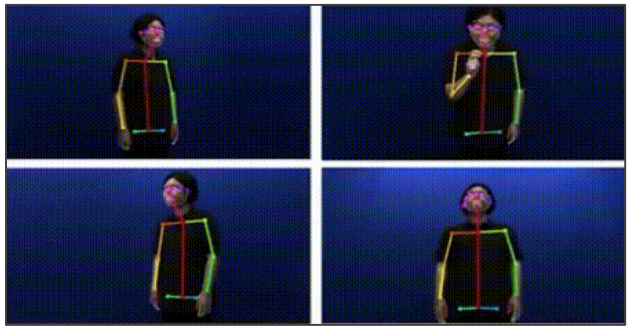
- 제안하는 서비스는 효정과 행동 데이터를 기반으로 수어 영상을 재 생성하는 GAN 신경망 알고리즘이 핵심으로서, **Multimodal 특성을 갖는 수어 영상 데이터가 필수적**
- 수어는 전문성이 요구되는 융복합 자연 언어의 형태이기때문에 데이터 구축이 어려움
- **AI 허브 데이터 구축 사업으로 인해 아이디어의 실현 가능성 확보**

핵심 데이터

수어 영상 AI 데이터

Korean sign language dataset for vision basis AI development

한국 수어 인식 인공지능 기술 및 서비스 개발에 활용 가능한 대규모 한국 수어 데이터이며 영상과 관절 키포인트 및 얼굴의 랜드마크 라벨 정도로 구성



[공개 예정인 수어 영상 AI 데이터셋]

[KETI 지능정보 플래그십 R&D 수어 데이터셋]

다양한 부사 및 의문사 형태의 데이터 포함

- 수어는 얼굴이나 손의 제스처의 차이로 의문문, 강조문 등이 표현됨
- 단일 프레임이 아닌 **연속적인 시계열 데이터** 방식으로 접근해야 하기 때문에 딥러닝 인식 또는 생성 모델 학습을 위한 데이터 구축이 어려움

Multimodal Data

- 클립 영상이외에 수어 인식 및 **통역에 필수적인 관절(키포인트)과 얼굴(랜드마크)** 라벨이 포함된 데이터셋으로서, **딥러닝 기술 분야 융합 서비스 구축 활용가치가 높음**

Korean sign language dataset

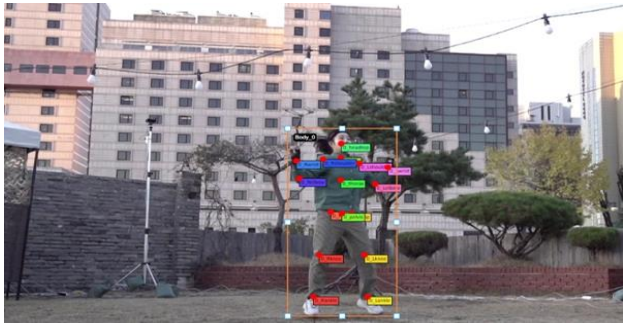
- **총 536,000개의 영상, 2000개의 문장, 3000개의 단어, 1000개의 지숫자/문자로 구성**
- 왼손과 오른손의 포인트 (각 21개), 전체 관절 포인트 (25개), 얼굴 랜드마크 포인트 (68개)
- Json 형태의 메타파일로 제공되며 모든 포인트는 영상 좌표계의 좌표점으로 제공
- **2021년부터 순차적으로 데이터 공개 예정**

융합 데이터

사람 동작영상 데이터 및 VGG 포즈 데이터

Human Motion Video AI Training Dataset & Youtube Pose Dataset

스마트폰 및 일반카메라로 입력된 영상 속 사람 이미지에 대하여 자세추정을
통한 포즈 정보를 획득하기 위한 데이터 및 민간 공개 포즈 데이터

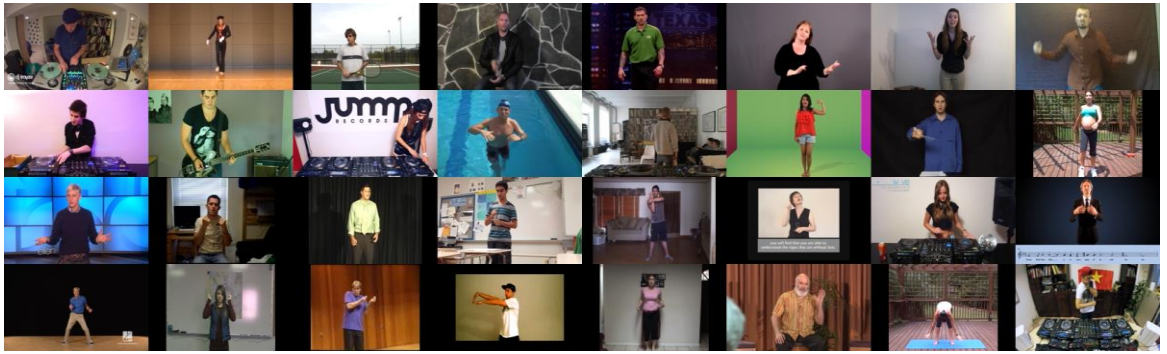


[사람 동작영상 AI 데이터]

기존 모델 평가에 활용되고 있는 데이터

- 기존 모델에 수어 영상 데이터를 최적화하기 위해서는 레퍼런스 모델과 데이터에 대한 이해가 선행되야 하며 **데이터 융합을 통한 Data Argumentation**이 필요함
- 이를 위해, 수어 영상과 유사하고 손의 제스처가 **포함되어 GAN 모델 생성 및 구축에 적합한 사람 동작영상 데이터 및 민간 데이터 VGG 포즈 데이터 활용 가능성** 확보
- 추후 공개되는 수어 데이터 특성에 따라 활용되고있는 기타 민간 데이터 융합 활용 기대

[VGG POSE ESTIMATION 데이터셋]

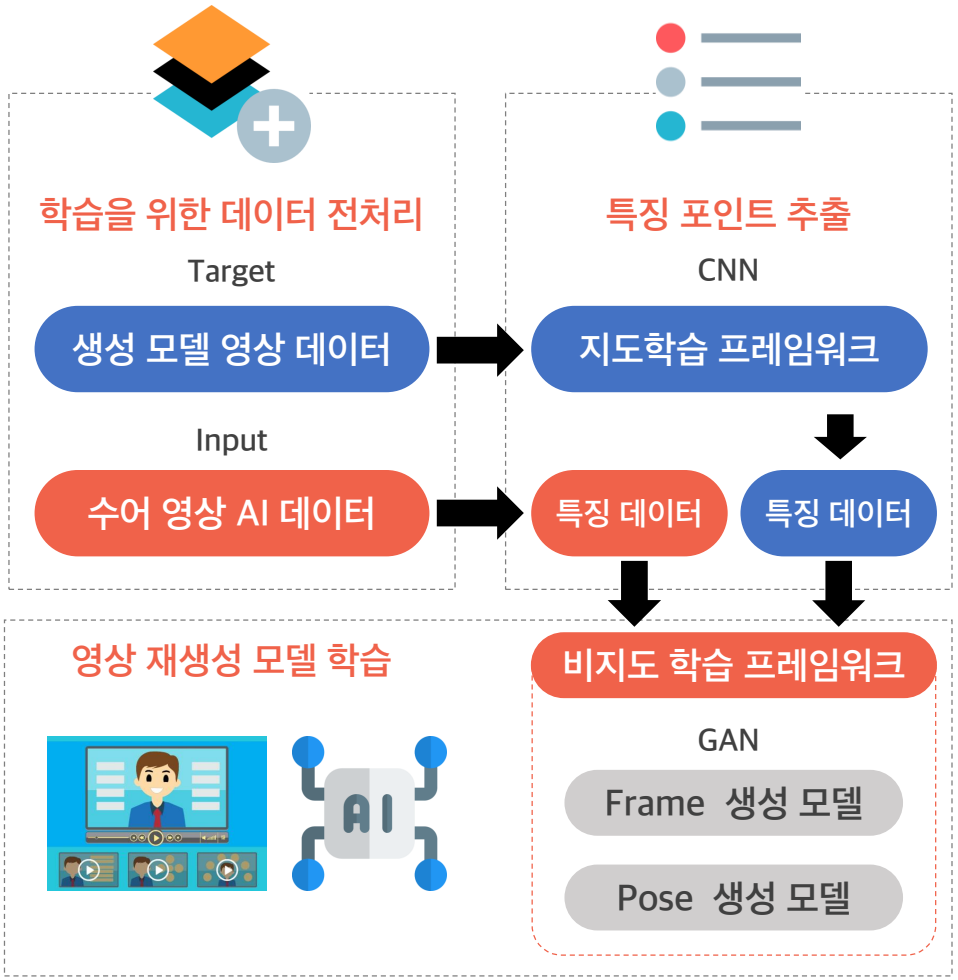


[VGG POSE ESTIMATION 데이터셋]

핵심 메커니즘

데이터와 최종 서비스 특성에 따른 모델 구조 설계

Deep Learning Framework



특징 추출에 특화된 모델

- 객체 탐지 및 예측 효율성이 높은 **심층 신경망 지도학습 프레임워크** 기반 탐지 모델을 통해 GAN 재생성 객체 변환을 위한 데이터 추출
- **단일 인물의 Facial Landmark, Key point 추출**에 뛰어난 DOWN-TOP 방식의 합성곱 신경망 (CNN) 예측 모델 (e.g., **CFA** or **Open Pose**)

영상 융합에 특화된 모델

- 추출된 객체 정보와 목표 클래스 입력으로 예측 프레임의 영상을 생성하고 객체의 키포인트를 최적화하는 적대적 생산적 적대 신경망 (GAN) 생성 모델 (e.g., **Liquid GAN** or **FSGAN**)

딥러닝 학습 핵심 메커니즘

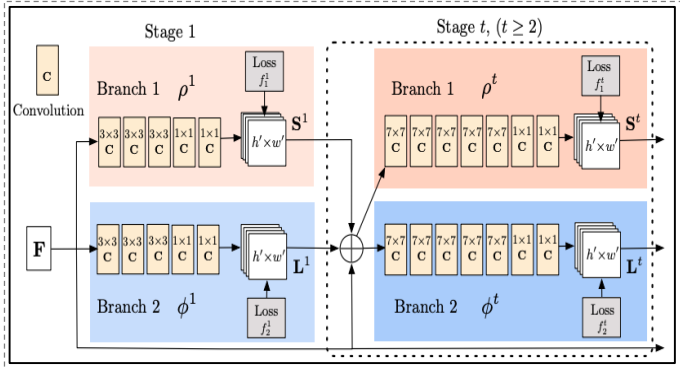
- 목표가 되는 데이터로부터 특징이 추출되는 학습 프레임워크와 추출된 특징 정보를 기반으로 재 생성하는 학습 프레임워크의 2단계로 구성된 **Extraction - Generation 구조**

핵심 메커니즘

데이터와 최종 서비스 특성에 따른 모델 구조 설계

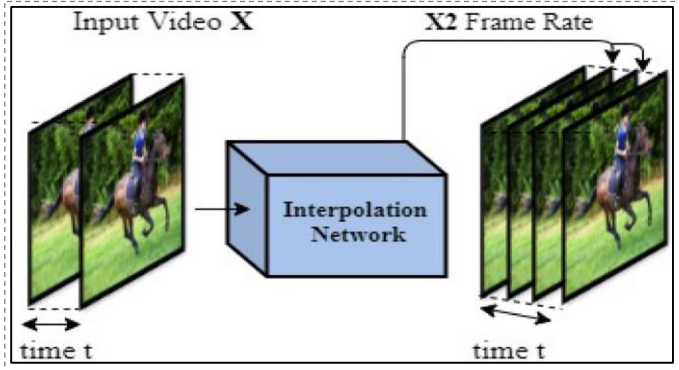
Deep Learning Framework

주어진 영상으로부터 특징점을 추출과 영상 재생성을 잘 할 수 있는 모델을
2단계 신경망 프레임워크로 구축 후 융합하는 방식으로 서비스 구현



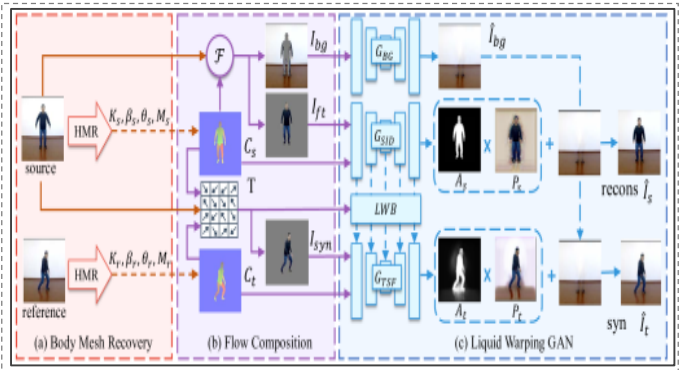
Feature Extraction

- Pose 추출을 위한 CNN 기반 기법 활용
- 수어 영상 데이터에서 분류된 (pose: 25개, hands 21개, face 68개) 추출을 위한 학습 진행
- **기존 모델 가중치를 바탕으로 Find-Turing 학습**
- 제공되는 수어 영상 데이터를 이용한 **K-fold 교차 검증으로 학습 성능 평가**



Video Interpolation

- 영상 클립 교차점의 왜곡을 최소화하기 위한 GAN 기반 가상의 영상 생성 기법 활용
- **예측되는 포인트를 기반으로 기존 모델 가중치 Find-Turing 학습**
- 제공되는 **문장 형태의 수어 데이터를 이용한 검증으로 학습 모델 성능 평가**



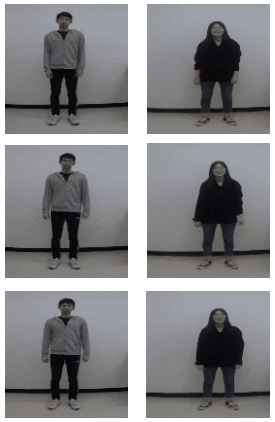
Motion Imitation

- 전처리로 생성된 영상과 수어 모델링 데이터를 입력으로 하는 **GAN Impersonator 모델 활용**
- 최적화된 기존 학습 모델의 가중치를 기반으로 모델(인물)의 수어 영상 재생성
- 구현 정확도를 바탕으로 학습 모델 최종 평가

핵심 메커니즘

실현 가능성 확보를 위한 테스트 MVP 구현

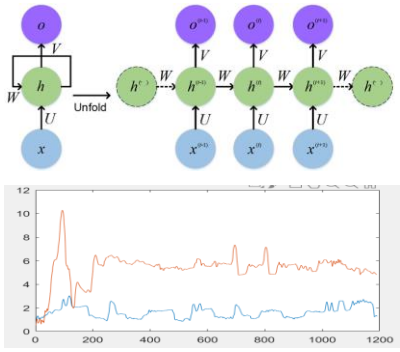
Deep Learning Framework



"안녕하세요"

"만나서"

"반갑습니다"



수어 영상 생성을 위한 데이터 구축 및 전처리 단계

- 기본 문법으로 구성된 수어 영상 데이터 수집을 통한 MVP 모델 설계
- 3개의 문법(클래스)를 구현하는 영상 데이터를 5명의 모델로 수집
- Optical Flow 및 Frame Difference 기법으로 영상 이벤트 발생 및 종료 시점 추정 RNN 메서드 구현 후 불필요 frame 제거

영상 클립 생성을 위한 GAN 이미지 재생성 학습

- Future GAN 기반 VIDEO INTERFORATION 진행

테스트 결과

- **Liquid GAN 기반 모델링 영상 생성 학습 진행**
- 기존 가중치를 기반으로 학습하였으며 별도의 파라미터 조정 및 인물의 키포인트 데이터의 부재로 포즈와 얼굴 랜드마크 구현 정확도가 낮음
- 하지만, 수어 영상 AI 데이터 구축 이후, **Multimodal 데이터셋을 기반으로 생성 모델링 학습을 진행한다면 충분히 가시적인 성능 향상을 가져올 것으로 기대**

VSION

지능형 수어 통역 및 번역 통합 플랫폼 구축

TARGET

제안하는 서비스 실현을 통해 농인들의 알권리 보장과 더불어,
다양한 시청각 콘텐츠 활용하고 소비의 폭을 넓힐 수 있도록 활용되는 것을 최종 목표

- AI 기술을 이용한 콘텐츠 맞춤 수화 통역 서비스 제공
- 공공 기관 내 통역 영상을 제공하는 비즈니스 모델
- 한국어 이외 다양한 문화권 언어 적용 가능성

시청각 콘텐츠 맞춤
AI 기반 수어 통역 영상 제작 솔루션

-끝까지 경청해주셔서 감사합니다 -

2021. 01. 20

딥아이 | 김진수 | 김종원