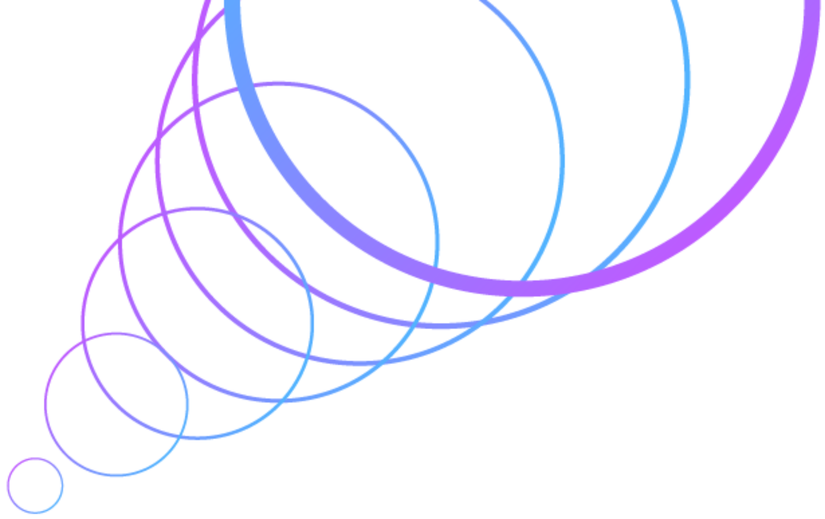
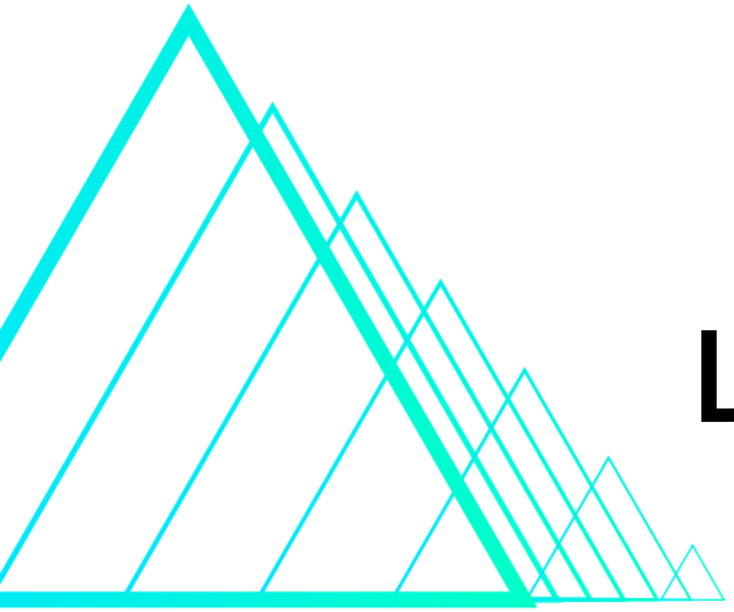




메타버스 시대에 나만의 부캐 만들기

김선태 NAVER CLOVA / Avatar / Leader

CONTENTS

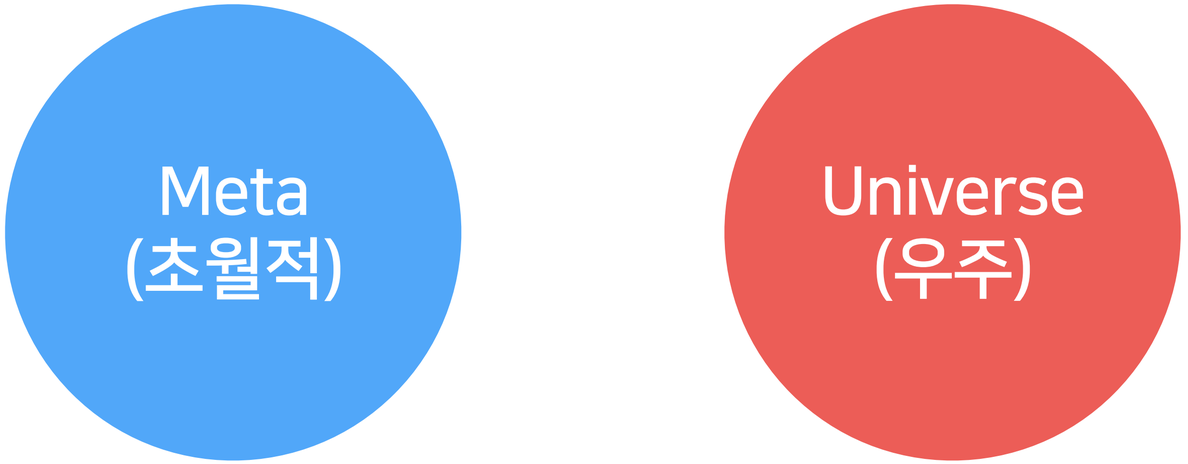
1. 메타버스와 아바타

1.1 메타버스란?

DEVIEW 2021

메타버스란?

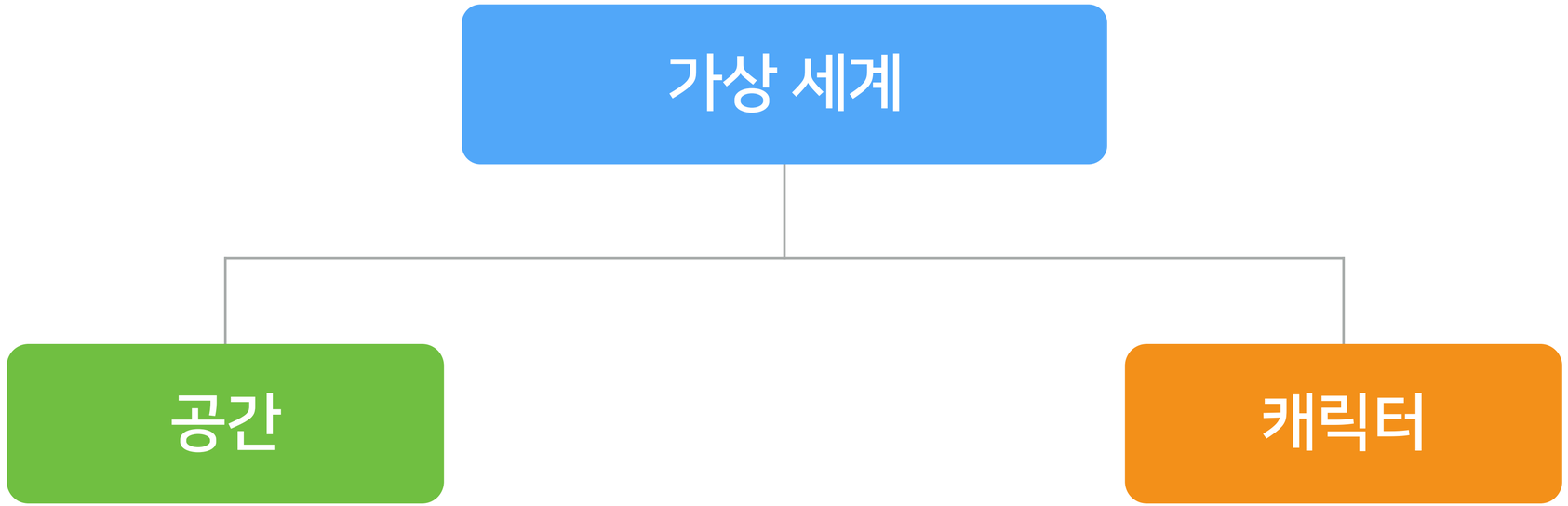
- 가상 공간에서 사회적, 경제적, 문화적 활동 가능
- 제페토, 로블록스 뿐만아니라 Facebook, 배달의민족도 가능



1.2 메타버스 콘텐츠

사용자가 직접 만들거나 다른 사람/회사가 만든 가상 세계

- 공간: 사용자가 접속하는 가상 공간(Virtual World). (집, 편의점, 자동차 등 존재)
- 캐릭터: 공간에서 나를 표현하는 수단, 부캐. (사람, 동물 등 가능)



1.2 메타버스 콘텐츠

DEVIEW 2021

서로 독립적으로 서비스 가능하지만, 같이 있으면 시너지 극대화

- 공간: 모델 하우스, 가상 인테리어, 전시관
- 아바타: 영상통화 캐릭터, Virtual UTuber
- 공간+아바타: 제페토, 로블록스, 게더타운 (행사는 기본이고 게임, 영화도 제작 함)

네이버, 2021년 신규 입사자 입문 프로그램에 제페토 활용

美 '초딩' 장악한 로블록스, 2021년 뉴욕증시 상장

https://www.hankyung.com/finance/article/202103097800i
https://news.mk.co.kr/v2/economy/view.php?year=2021&no=333917

1.3 아바타의 필요성?

DEVIEW 2021

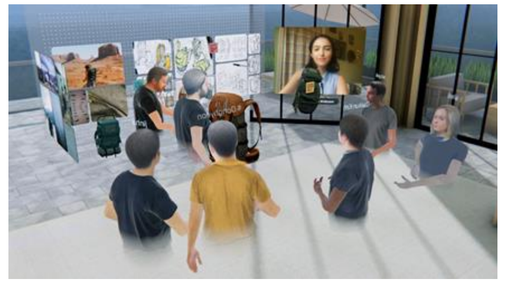
공간: Static

- 정적인 대상: 건물, 기타 사물 등
- 물론 Movable Object도 존재 (자동차, 선풍기 등)

아바타: Dynamic

- 동적인 대상: 인간, 동물, 기타 캐릭터 등
- 가상 세계에서 사람과 똑같음, 대화의 주체.
- 다양한 환경에서 사용자와 인터렉션 가능 (AR, VR, MR)

제페토: 커뮤니케이션



Spatial: 가상 회의

"Interactive Avatar를 만들 수 있는 방법은?"

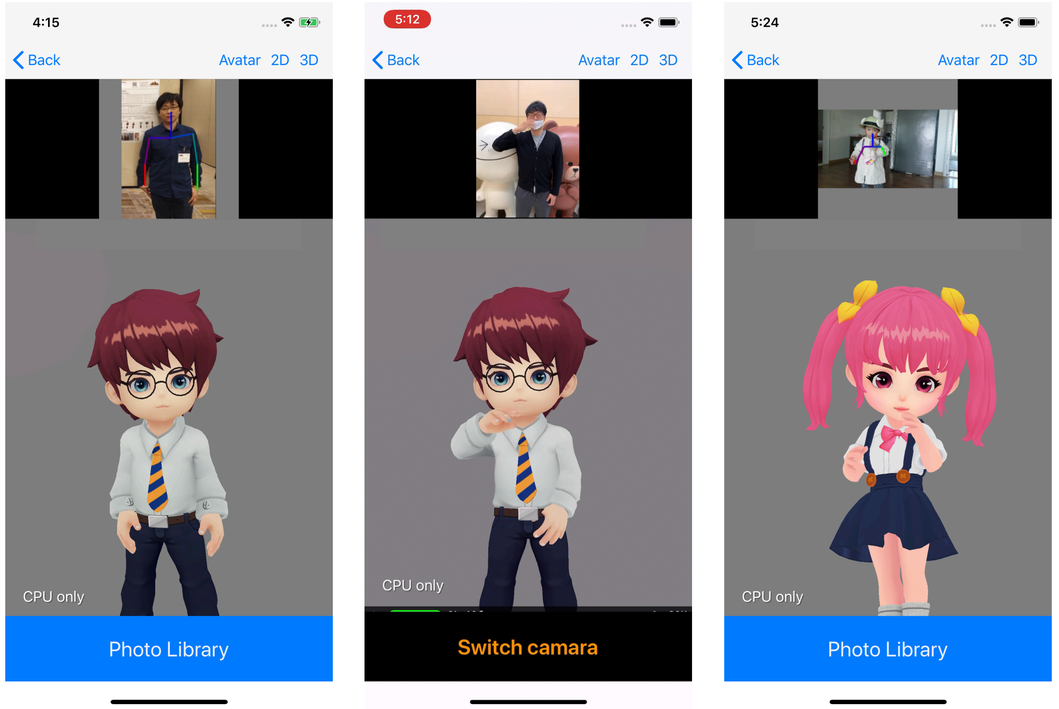
https://www.gamemeca.com/view.php?gid=1636905
https://www.hankookilbo.com/News/Read/202005141534086916

1.4 아바타를 만들 수 있는 다양한 기법들

DEVIEW 2021

DEVIEW 2020: 대근육을 사용하는 아바타

- “나를 따라하는 아바타: 모델 개발부터 모바일에 적용하기까지”
- https://deview.kr/2020/sessions/395

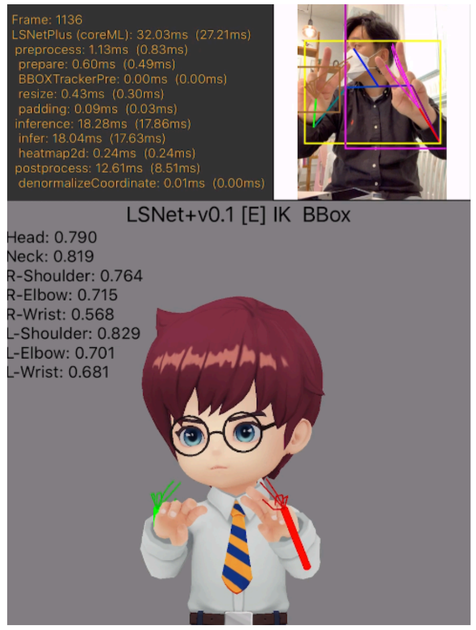


1.4 아바타를 만들 수 있는 다양한 기법들

DEVIEW 2021

DEVIEW 2021: 대근육+소근육까지 사용하는 디테일한 아바타

- “메타버스 시대에 나만의 부캐 만들기”



<Image 기반 캐릭터 생성>

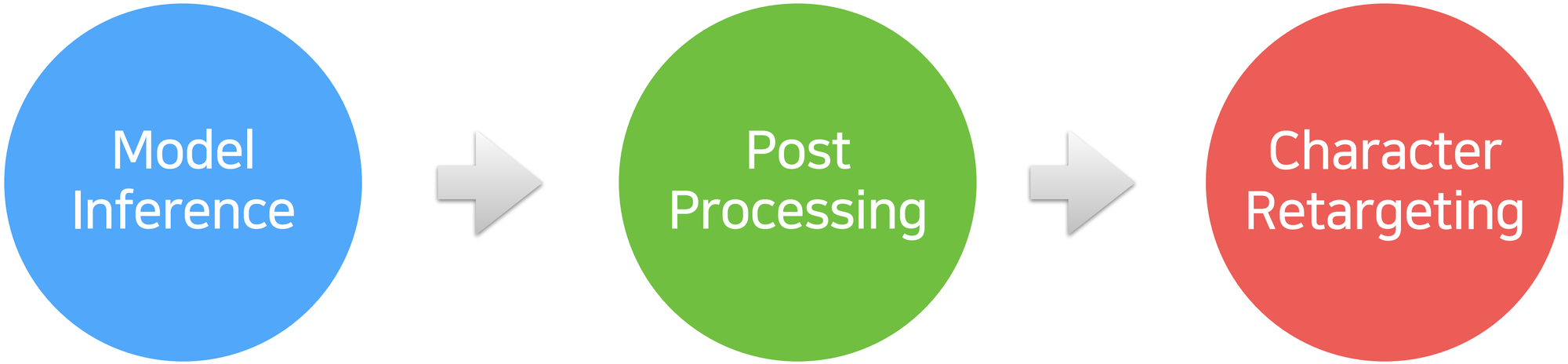
<Text 기반 캐릭터 생성>

2. Image 기반 아바타 생성 기법

2.1 아바타 생성 파이프라인

모바일에서 크게 3단계로 실시간 진행

- Model Inference: 사람의 관절 위치 예측
- Post Processing: 예측된 RAW 데이터 필터링 (IK, 충돌회피, Smoothing)
- Character Retargeting: 3D 캐릭터에 적용

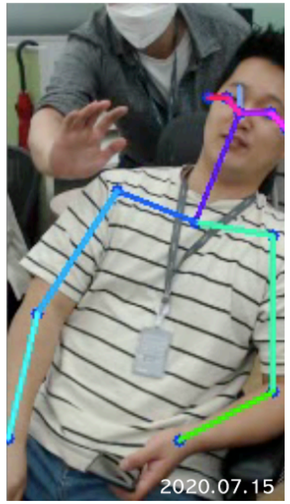


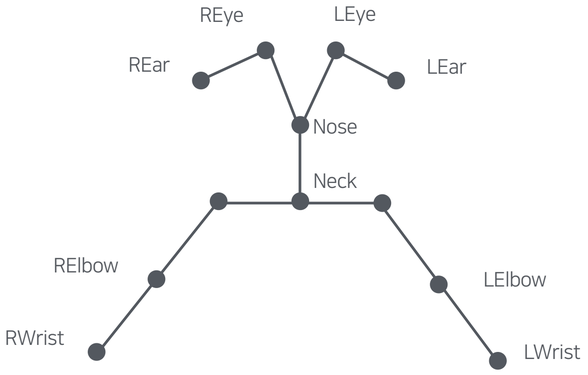
2.2 상반신 + 양손 SDK

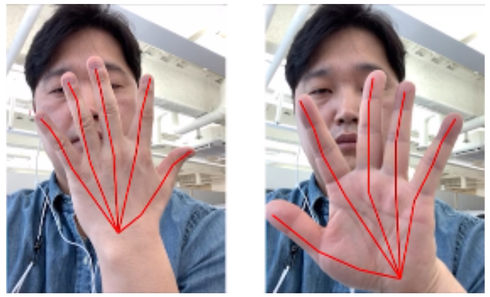
DEVIEW 2021

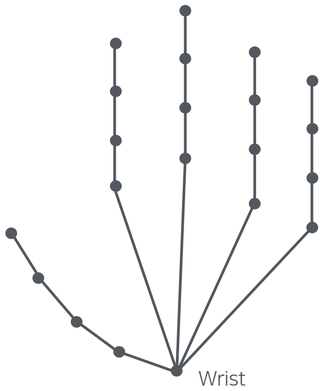
ML 모델 2개 필요

- 상반신 모델: 상반신 Joint + 손 영역 Joint 예측
- 손 모델: 손 영역을 Cropping하여 Joint 예측





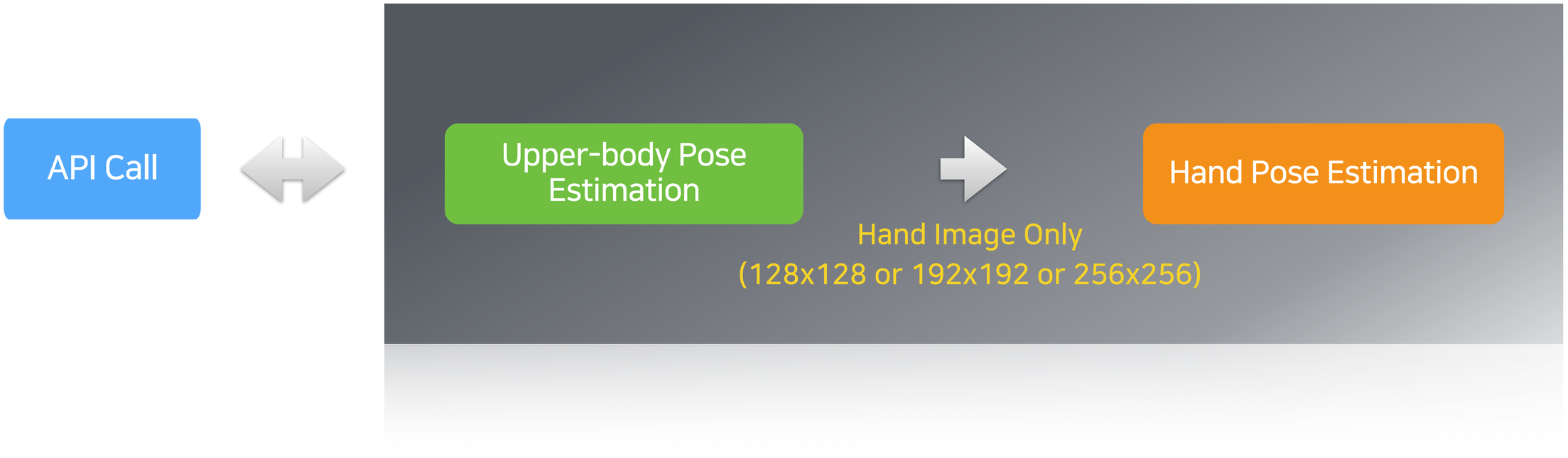




2.2 상반신 + 양손 SDK

End2End “상반신 + 양손” SDK 개발

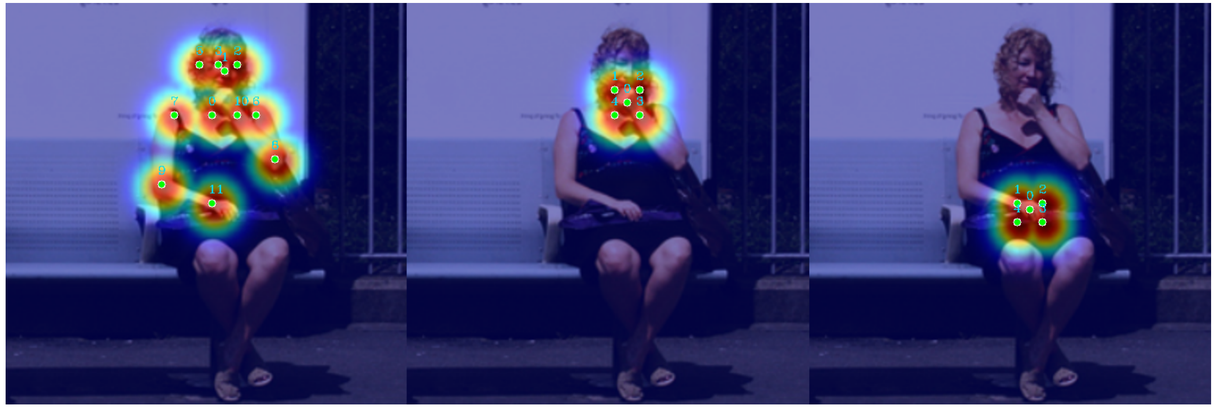
- “사용자는 모델 2개를 각각 사용해야 하는가?” 아니요
- SDK 내부에서 알아서 상반신과 양손 모델 호출

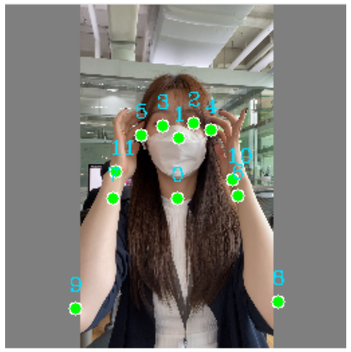


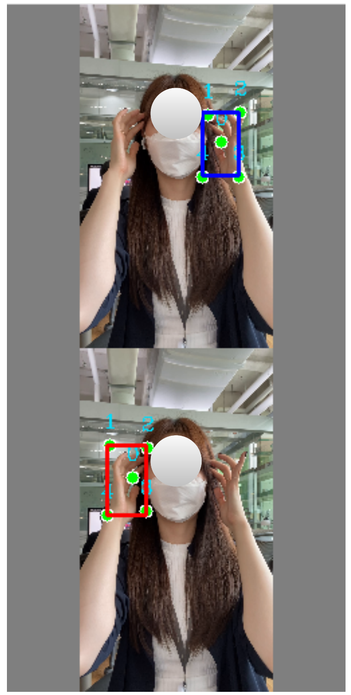
2.3 상반신 모델 with Hand Detection

DEVIEW 2021

- 만약 상반신과 손을 같이 학습하게 되면?
  - 손 조인트 개수 = 42개(21개 X 양손), 상반신 조인트 개수 = 12개
  - 상대적으로 상반신으로 Regression이 잘 안되는 현상 발생
- 상반신에 Hand Detection을 위한 별도의 Heatmap 세팅 및 Dataset 필요



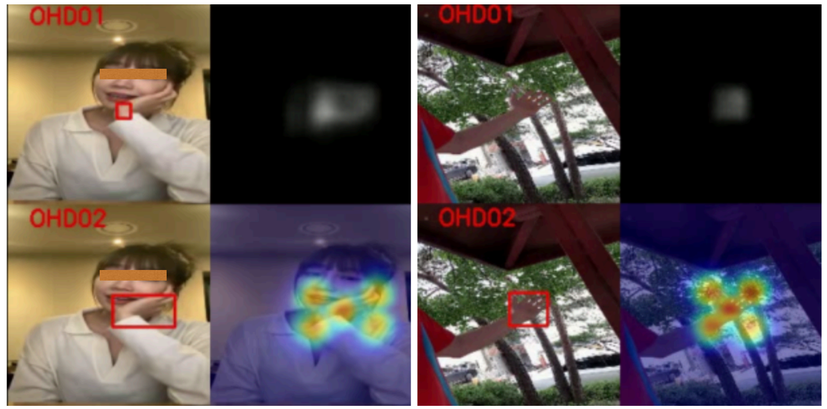




2.3 상반신 모델 with Hand Detection

DEVIEW 2021

- 성능 향상을 위해서 별도 Hand Detection 모델 추가 안함
- 상반신 예측하면서 손 영역 검출

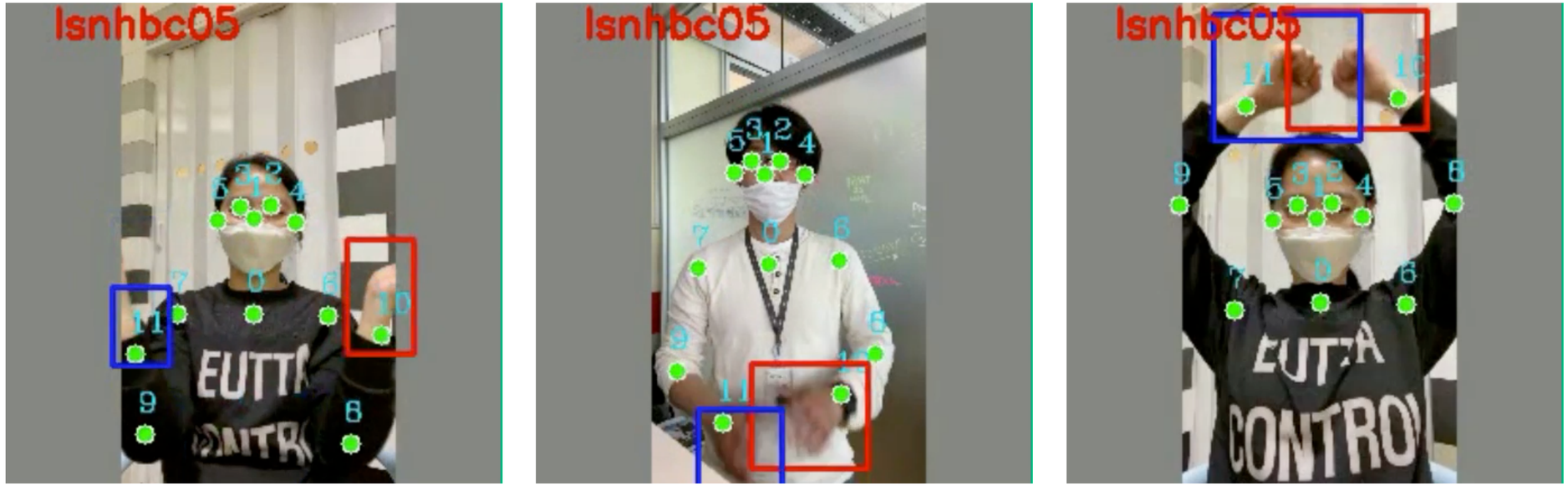


	F1 Score (0.1 <= Scale < 0.3)	F1 Score (0.3 <= Scale < 0.7)	F1 Score (0.7 <= Scale < 0.9)
Box Segmentation —> Heatmap 변경 후	43% 증가	비슷	비슷

2.3 상반신 모델 with Hand Detection

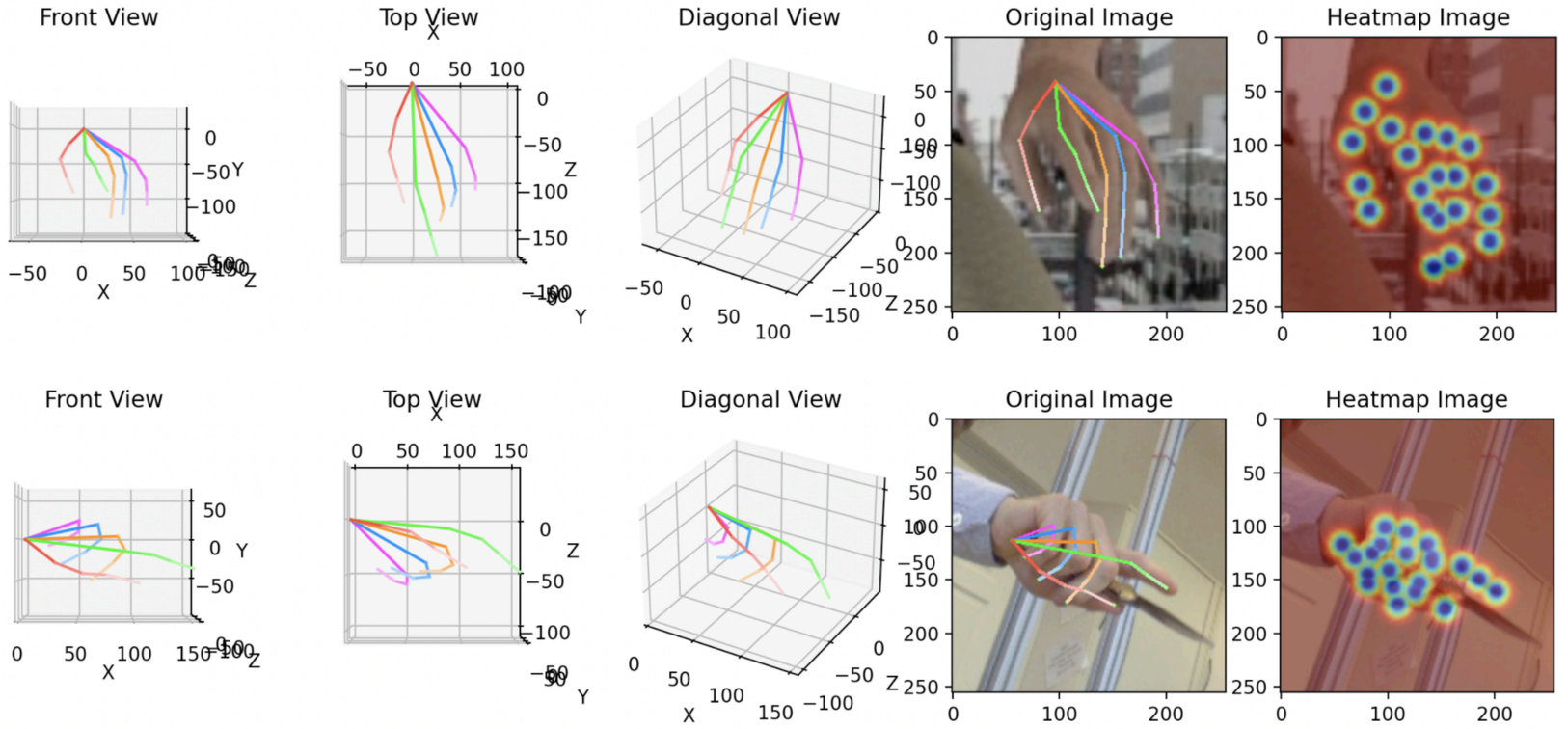
DEVIEW 2021

데모 영상



2.4 손 모델(One Hand)

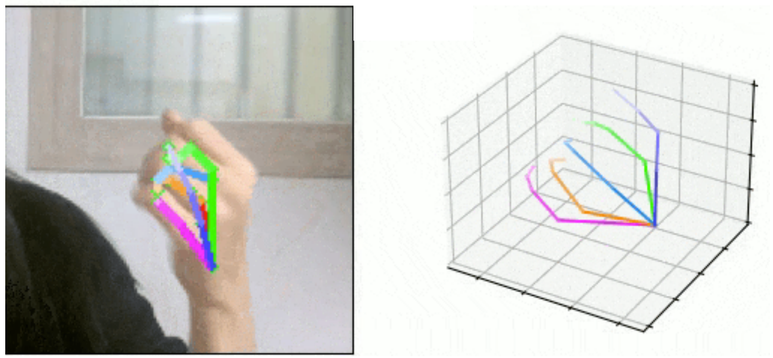
- Lightweight Hand Pose Model
- 2D Heatmap to 3D Pose 학습



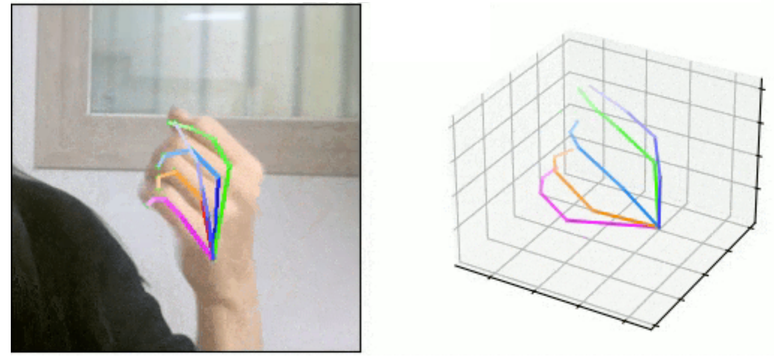
2.4 손 모델(One Hand)

DEVIEW 2021

- 경량화 정도에 따른 Model 개발
- Device 사양에 따라 동적 교체 가능



<Base model>
Parameters: 약 1M 이하

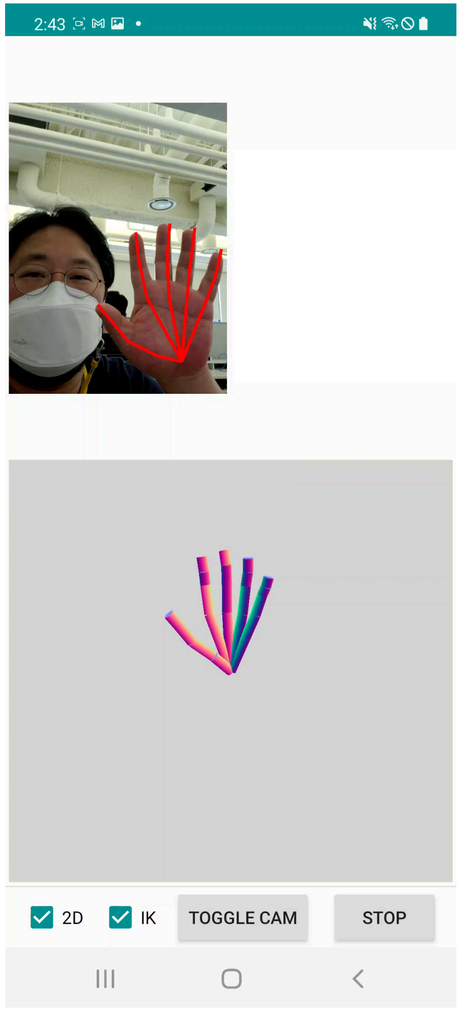


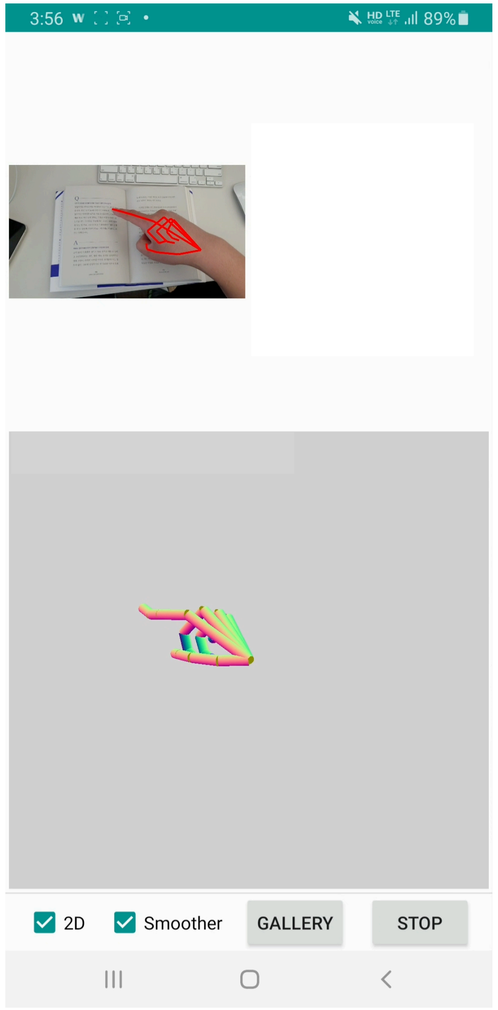
<Big model>
Parameters: 1.5배 증가
FLOPS: 5배 증가
정확도(MPJPE): 15% 향상

2.4 손 모델(One Hand)

DEVIEW 2021

데모 영상

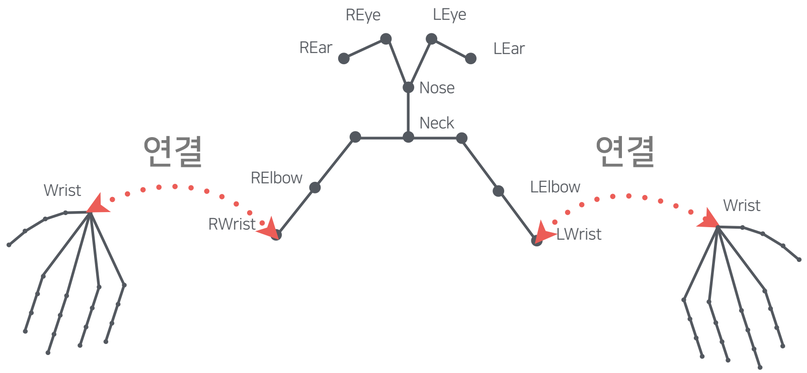


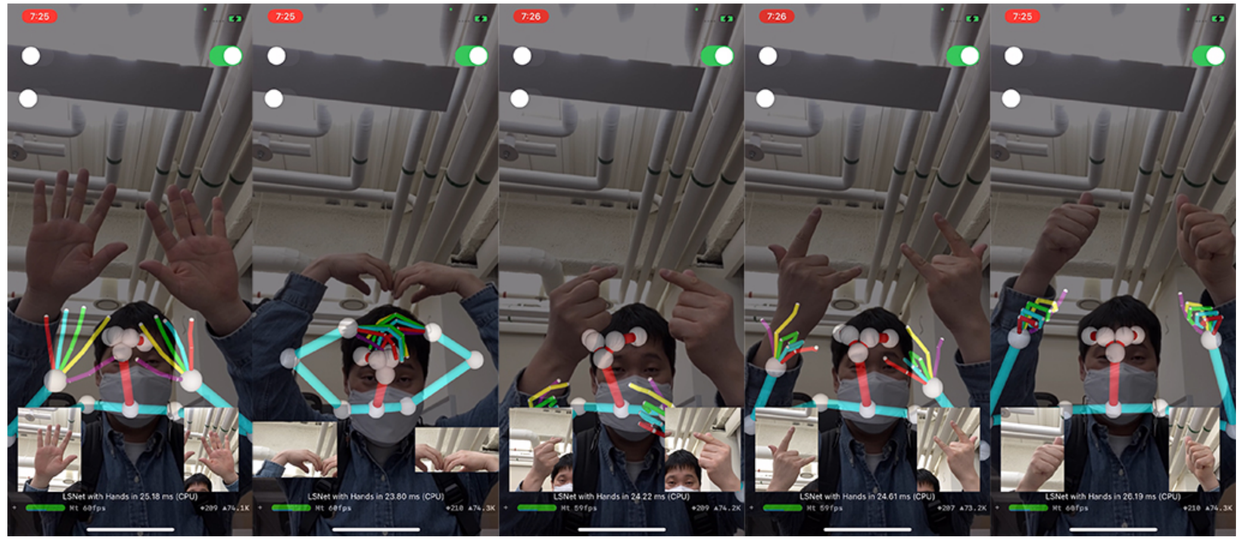


2.5 상반신과 양손 연결하기

DEVIEW 2021

- 상반신의 손목을 원점으로하여 손의 손목 연결
- 왼팔, 오른팔 모두 연결

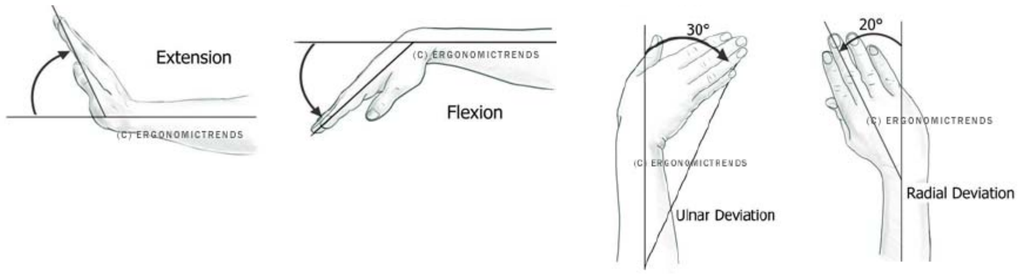


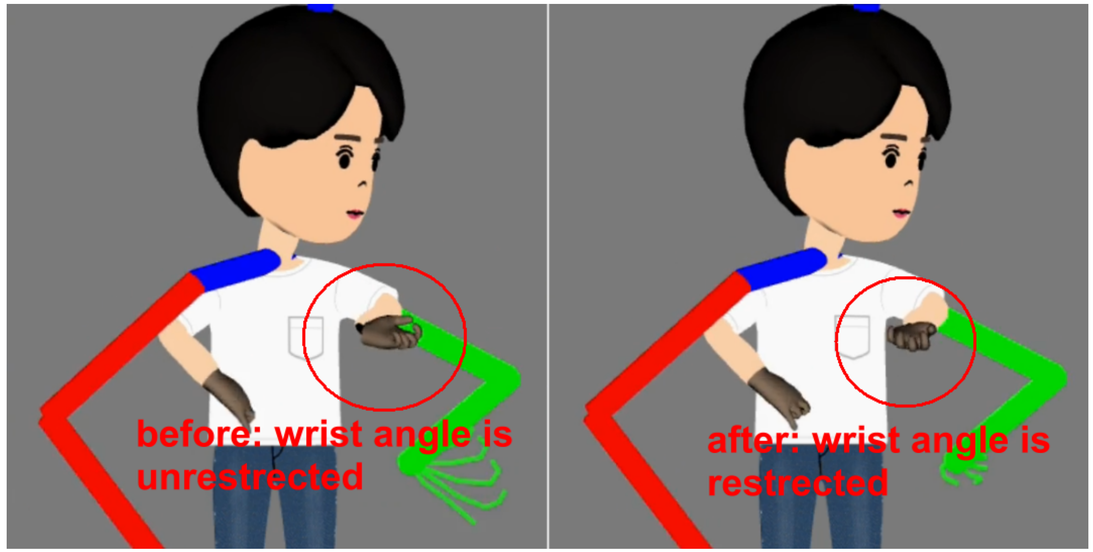


2.6 역기구학(Inverse Kinematics)

DEVIEW 2021

- Constraints for Wrist
- 사람이 취할 수 있는 손목 각도로 제한





https://www.hindawi.com/journals/jr/2013/910961

2.6 역기구학(Inverse Kinematics)

DEVIEW 2021

- Constraints for Hand Joints
- 사람이 취할 수 있는 손가락 각도로 제한
- 최종적으로 각 관절의 각도를 알 수 있음

Initial Pose

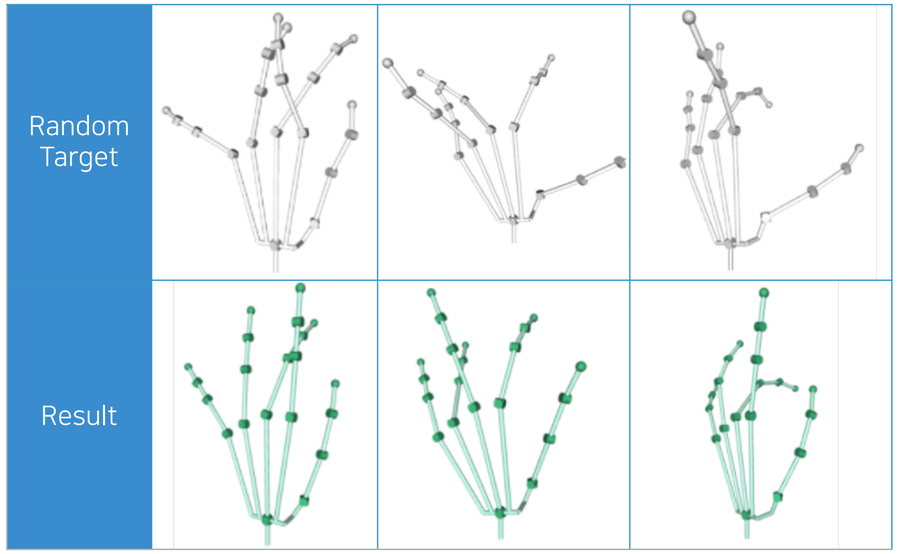


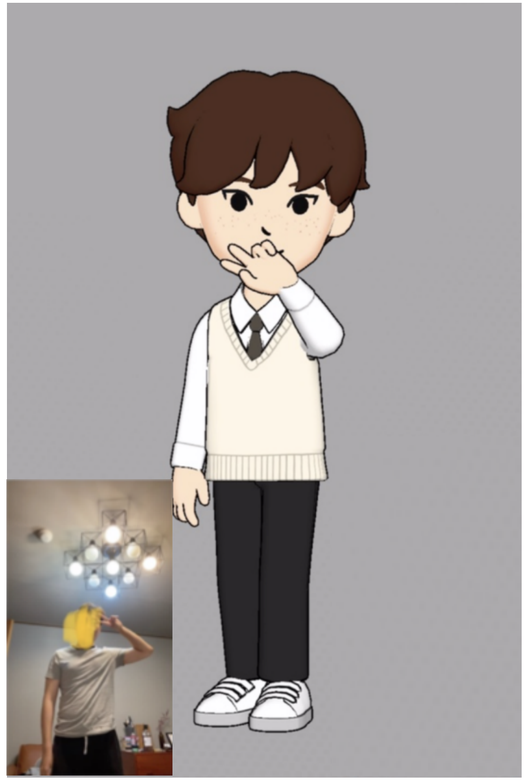
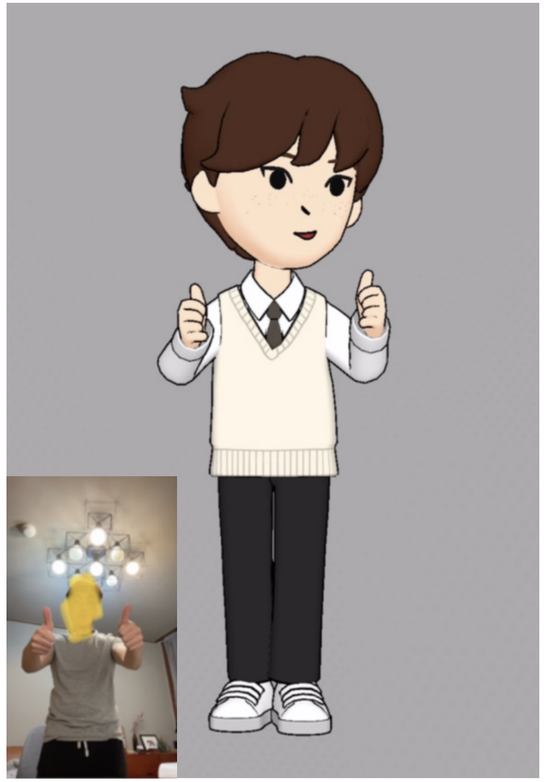
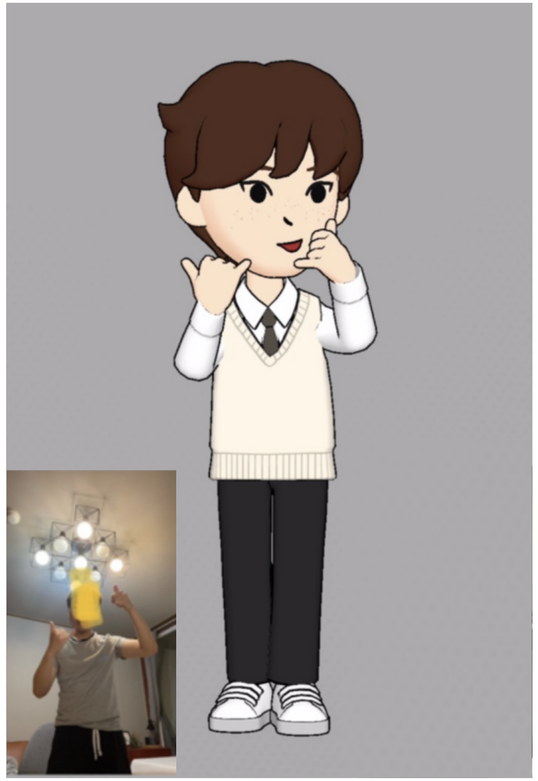
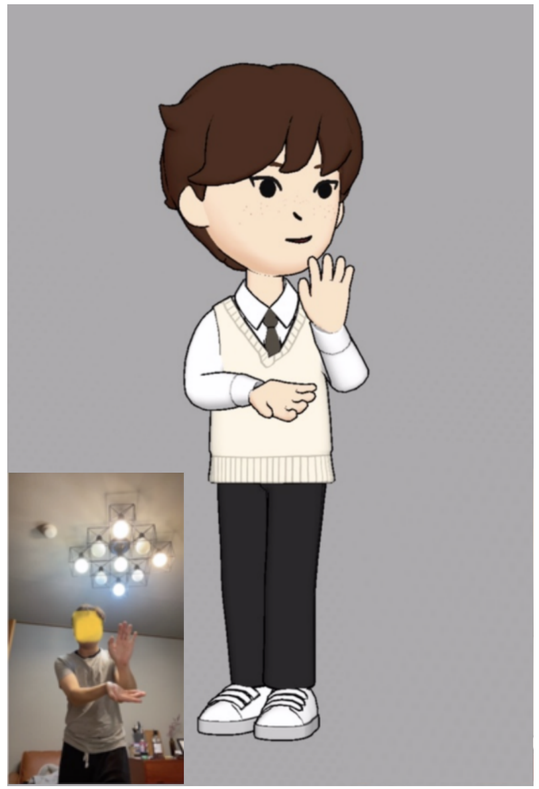
Table 6: Statics constraints [13].

Finger	Flexion	Extension	Abd./add.
Thumb			
TMC	50°–90°	15°	45°–60°
MCP	75°–80°	0°	5°
IP	75°–80°	5°–10°	5°
Index			
CMC	5°	0°	0°
MCP	90°	30°–40°	60°
PIP	110°	0°	0°
DIP	80°–90°	5°	0°
Middle			
CMC	5°	0°	0°
MCP	90°	30°–40°	45°
PIP	110°	0°	0°
DIP	80°–90°	5°	0°
Ring			
CMC	10°	0°	0°
MCP	90°	30°–40°	45°
PIP	120°	0°	0°
DIP	80°–90°	5°	0°
Little			
CMC	15°	0°	0°
MCP	90°	30°–40°	50°
PIP	135°	0°	0°
DIP	90°	5°	0°

https://www.hindawi.com/journals/jr/2013/910961

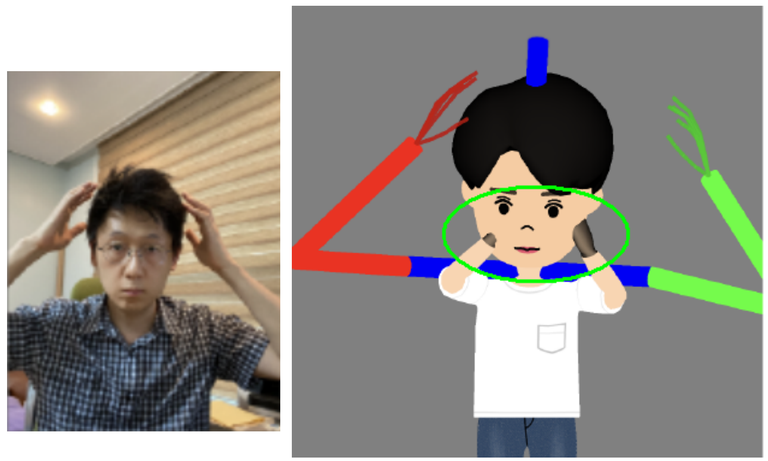
2.7 Hand Retargeting

- LINE 메신저의 아바타 활용 예시
- Wrist 및 각 Joint에 Rotation 적용

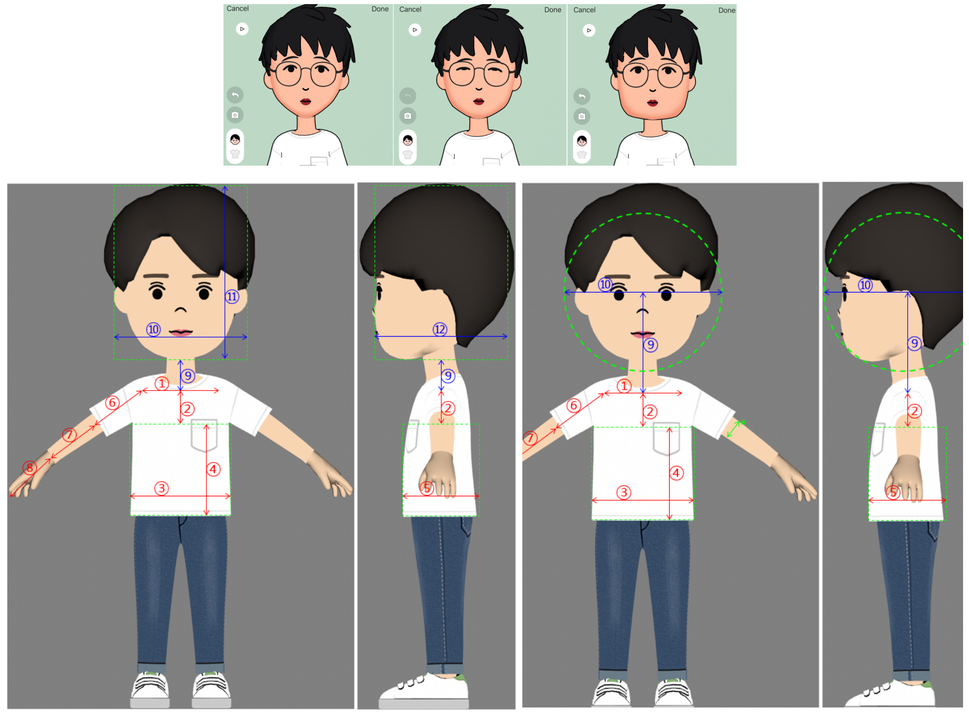





2.8 충돌 회피(Collision Avoidance)

- 3D 캐릭터 특성상 체형이 다를 수 있음 (소두/대두, 저체중/과체중, 짧은팔/긴팔 등)
  - 팔이나 손 등이 몸을 관통하는 경우 발생
- 얼굴, 팔, 손가락에 Collision Shape 세팅 => 실시간 Collision Avoidance 처리



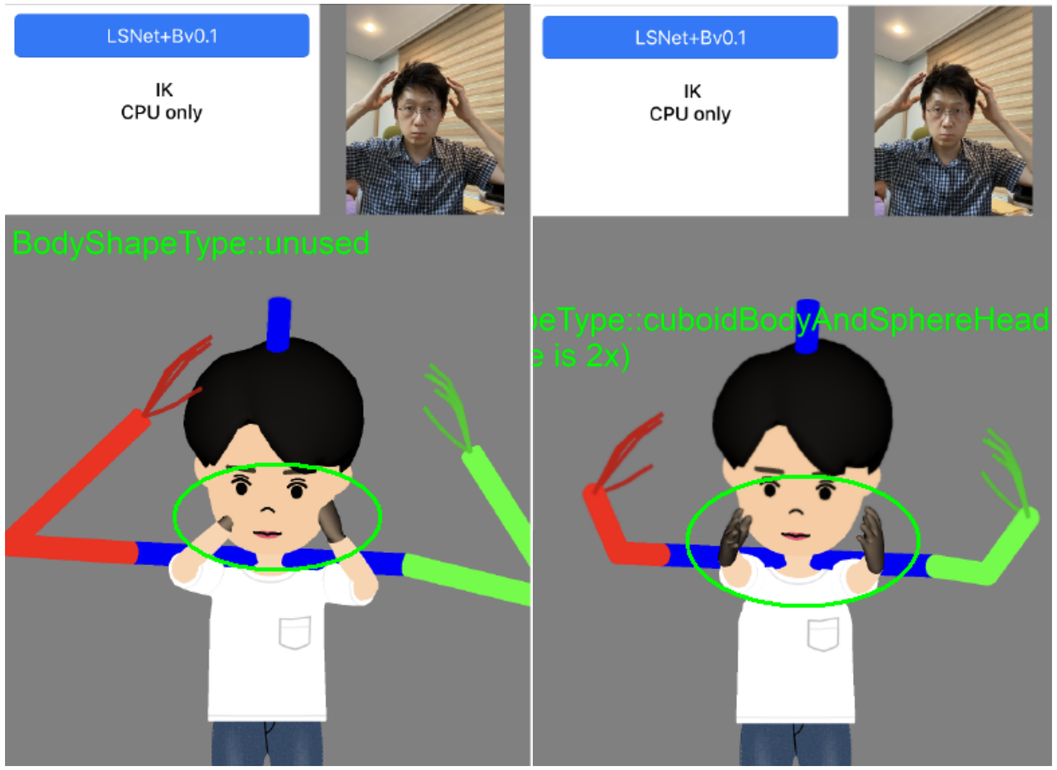
1) 양손이 귀 위에 위치해야 정상
2) 3D 캐릭터의 Bone Length가
Pose 예측으로 계산된 값 보다 작아서 생기는 문제
3) 2D Skeleton과 3D 캐릭터의 관절 각도는 같음
4) 팔이 짧고 대두(머리가 어깨 너비보다 큼)

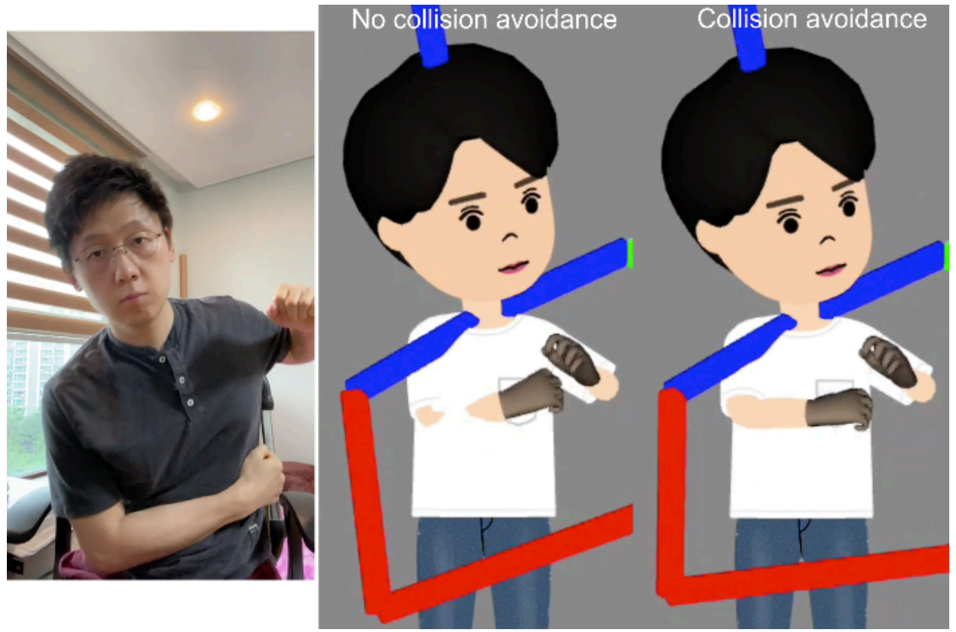


2.8 충돌 회피(Collision Avoidance)

DEVIEW 2021

- 실시간 Collision Avoidance 처리
- Body shape, Bone Length에 따라서 동적으로 충돌 회피 가능

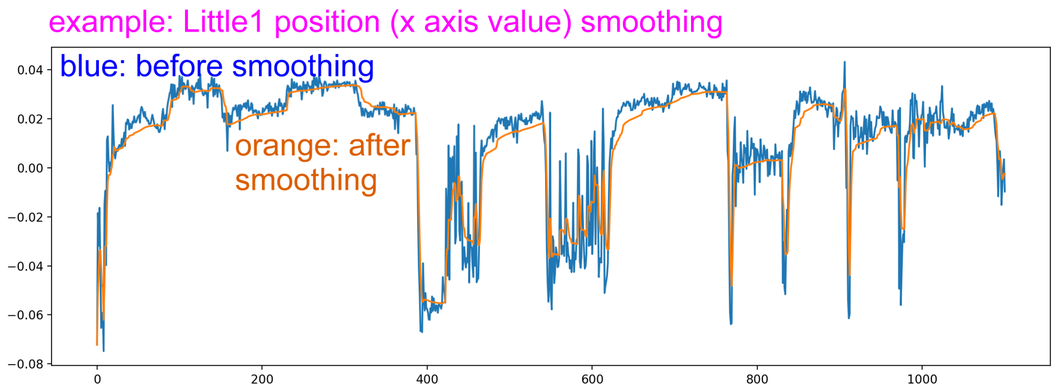


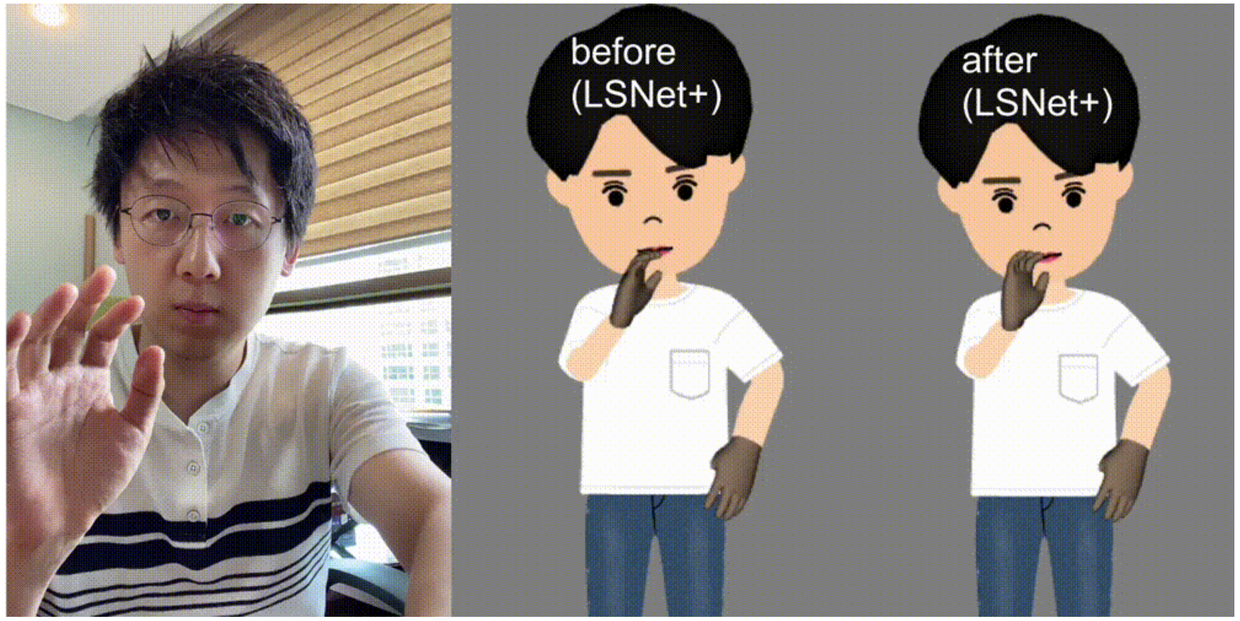


2.9 Smoothing Filter

DEVIEW 2021

- 정지 동작인데 관절이 떠는 경우 발생
- Smoothing 필터로 안정화



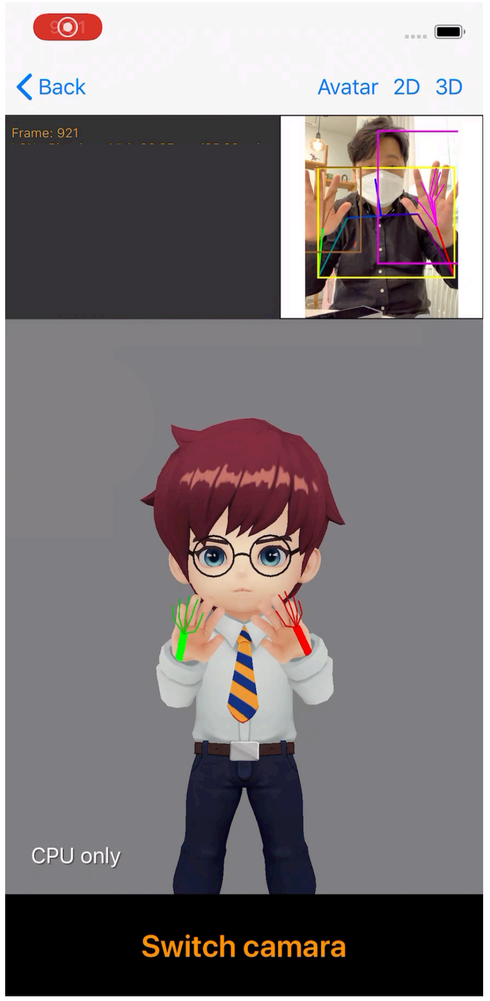


2.10 상반신 + 양손 Demo

DEVIEW 2021

Inference 3번 필요

- 상반신 모델: 1번
- 손 모델: 왼손 1번, 오른손 1번



3. Text 기반 아바타 생성 기법

3.1 기존 연구 실험

- Text To Action
- Text, Audio, Speaker Identity(Multi-model) to Gesture
- Speech To Facial(Lip-Sync) Animation
- Speech To Gesture

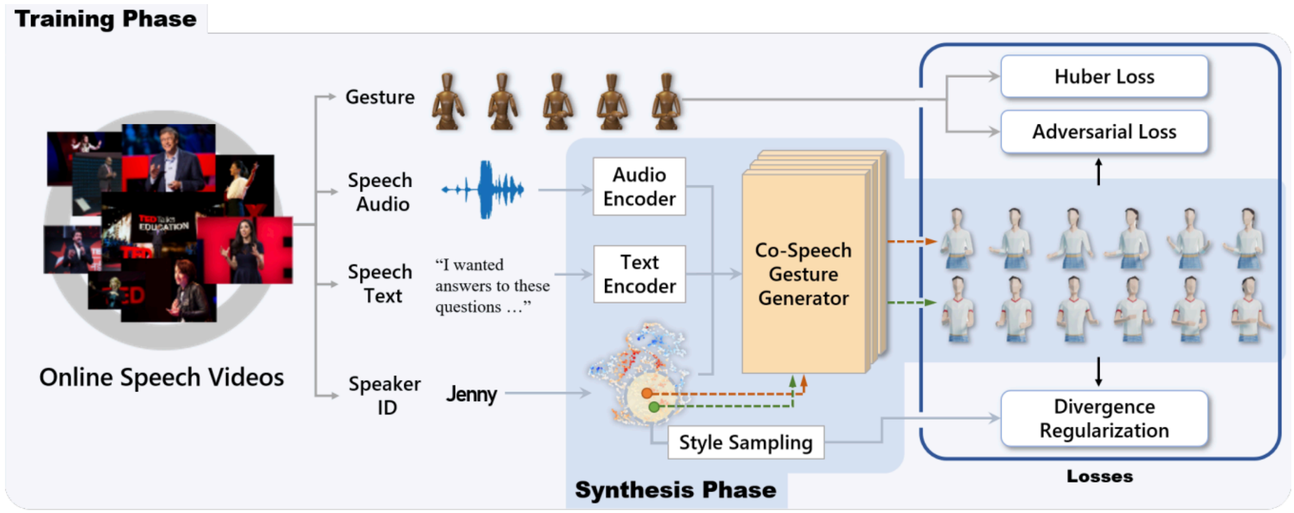
"테스트 시작!"

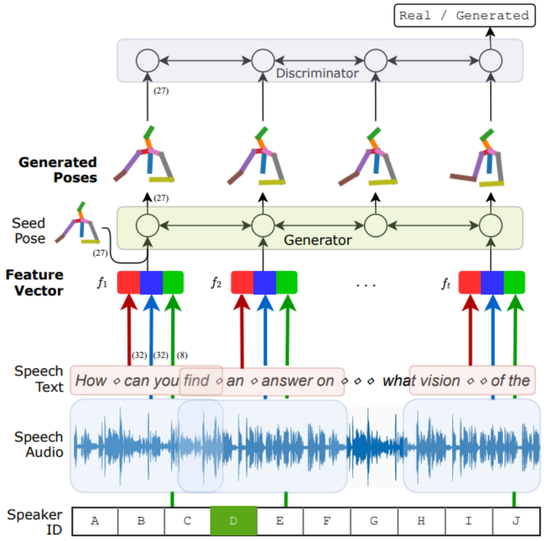
3.1 첫번째 실험

DEVIEW 2021

Text, Audio, Speaker Identity(Trimodal)

"좋은 연구이지만 생성된 자세가 단조로움"



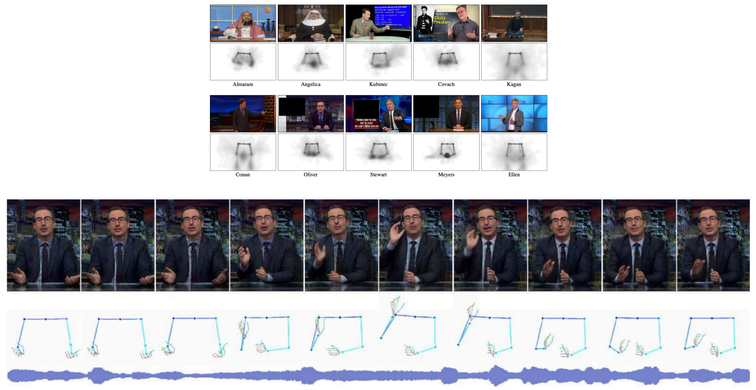


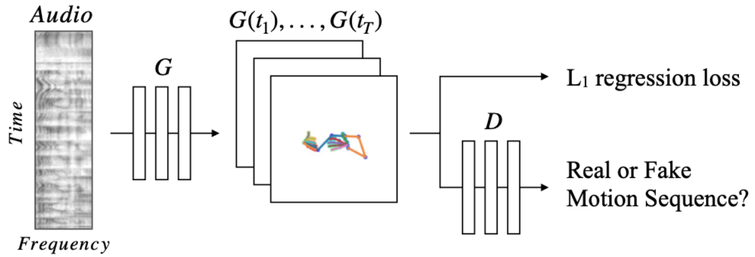
Speech Gesture Generation from the Trimodal Context of Text, Audio, and Speaker Identity (SIGGRAPH Asia 2020)

3.1 두번째 실험

- Speech To Gesture
- 화자별 Audio, Pose 데이터로 학습

"단조롭지 않은 Gesture 생성 가능"

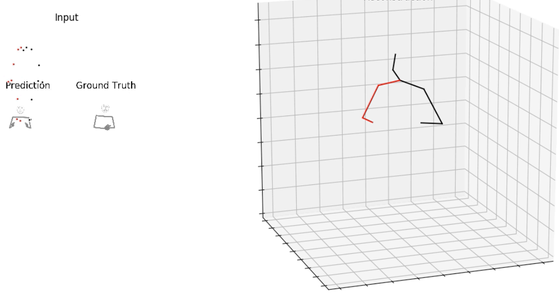




Learning Individual Styles of Conversational Gesture(CVPR 2019)

3.1 두번째 실험

- Speech To Gesture
- 문제: 대근육 모션 합성, 자세가 튀는 구간 발생, Speaker Dependent
- 해결방법: Pose Smoothing 및 소근육 추가 합성해봄 (손 및 고개 끄덕임, Idle 모션 등)



Gesture Generation

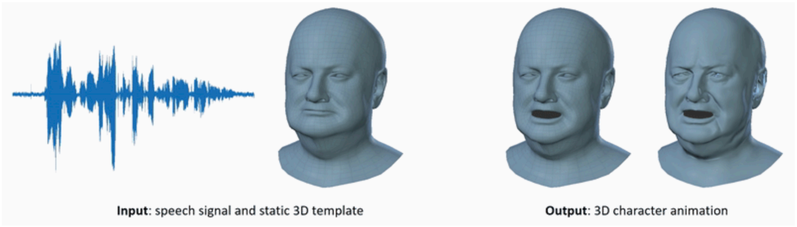
Retargeting

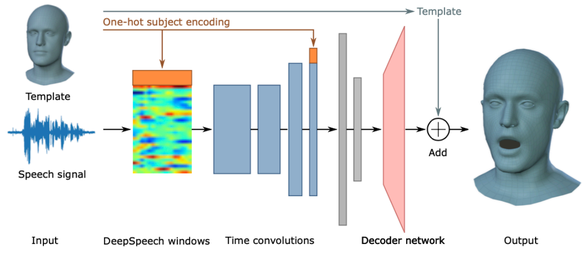
"화자에 종속적이지 않고, 더 자연스러운 방법은 없을까?"

3.1 세번째 실험

DEVIEW 2021

- Speech To Facial Animation
- Audio 기반 End2End 3D 캐릭터 스피치 애니메이션
- 화자에 종속적이지 않고 범용적이고, 다양한 Audio 타입과 Language에 강인





"얼굴 아래쪽(Lower face) 위주의 애니메이션,
감성 표현에 한계 존재"

"범용 모델이라서 다양한 언어에 강인하지만,
발음 및 발화 속도에 Lip-Sync가 부정확한 경우 발생"

Voice Operated Character Animation (CVPR 2019)

3.1 실험 정리

DEVIEW 2021

Gesture

- 대근육을 사용한 애니메이션
- 자연스러운 Gesture 생성 어려움

Face

- 얼굴 표정의 한계 존재
- 정확한 것보다는 범용적인 Lip-Sync.

"가장 큰 문제는 Gesture, Face 모델 2개를 각각 사용했을 때,
얼굴과 바디 애니메이션이 부자연스러울 수 있다."

3.2 추구하는 방향

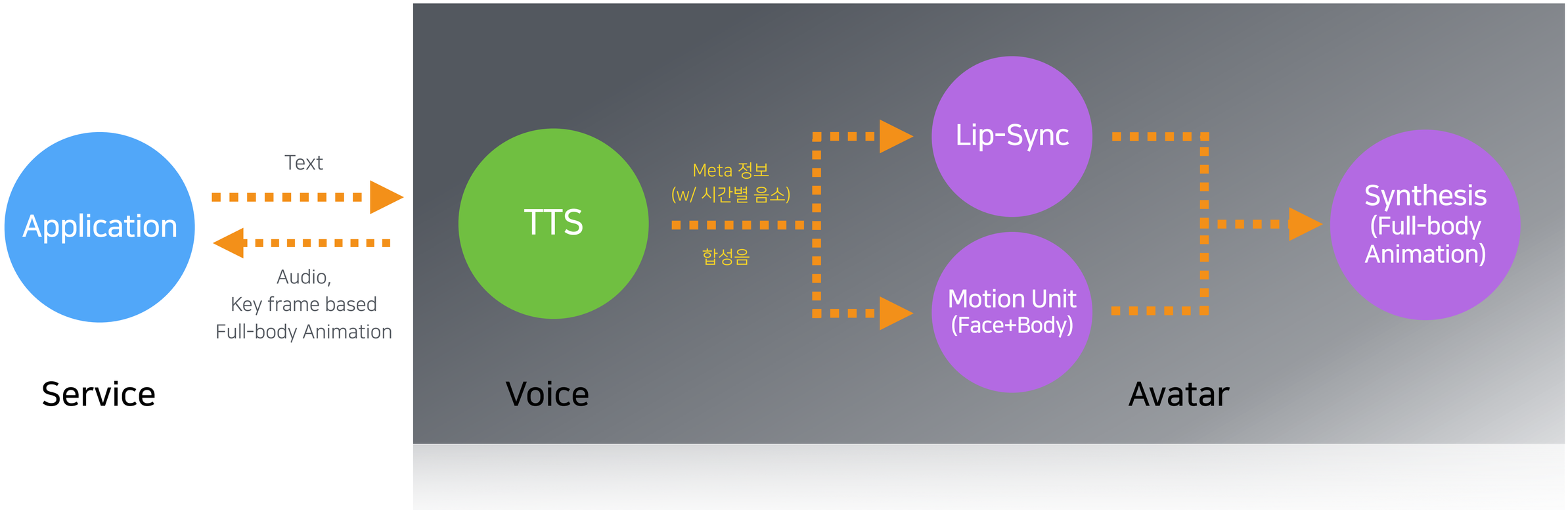
DEVIEW 2021

목표 설정

1. Motion Unit
   - Face와 Body를 하나의 자연스러운 Motion Unit으로 제작
   - Motion Unit을 예측하는 ML 모델 개발
2. Lip-Sync
   - CLOVA TTS 연동으로 또박또박 정확하게 개발
   - Lip-Sync와 Face는 Blending하여 자연스럽게 연출
3. Full-body Animation
   - Lip-Sync, Motion Unit(Face/Body) 합성 —> 다양한 캐릭터에 적용 가능해야 함

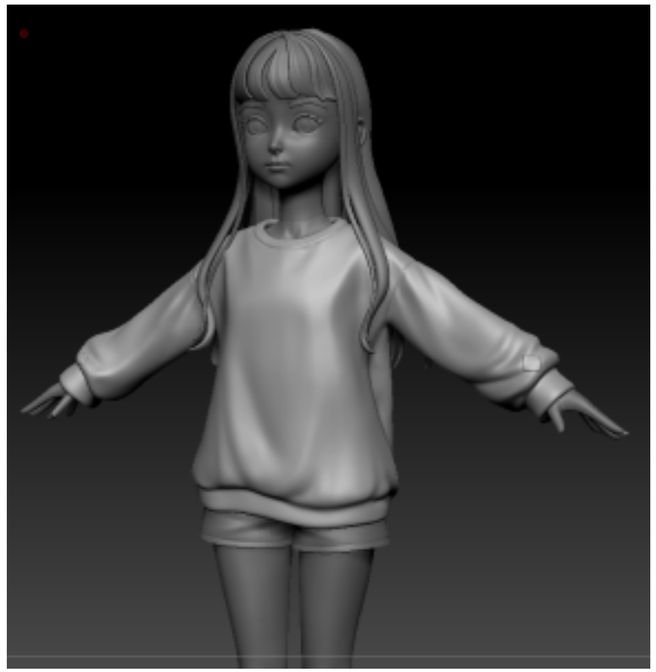
3.2 추구하는 방향

- CLOVA Voice & Avatar팀 기술을 활용한 아바타 생성
- Input: Text
- Output: Full-body Animation + 합성음



3.3 자체 캐릭터 디자인

- 페르소나 기획
- 2D 원화 부터 3D 캐릭터 디자인



3.3 자체 캐릭터 디자인

DEVIEW 2021

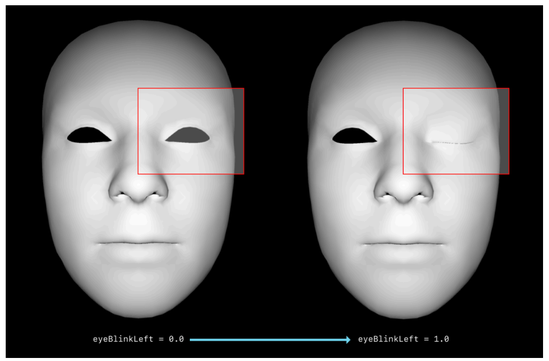
- 다양한 캐릭터에도 적용 가능해야 함
- BlendShape 포멧 정의

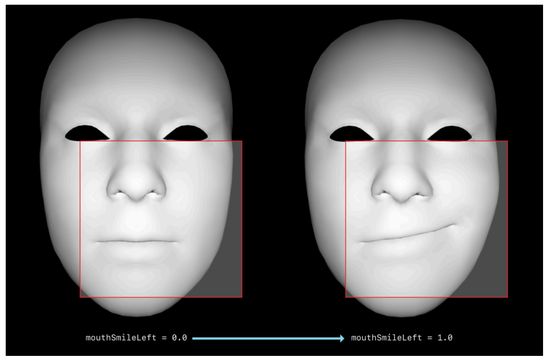
"범용성을 위해서 Apple ARKit BlendShape 포멧 사용"
"립싱크 BlendShape 별도 추가"

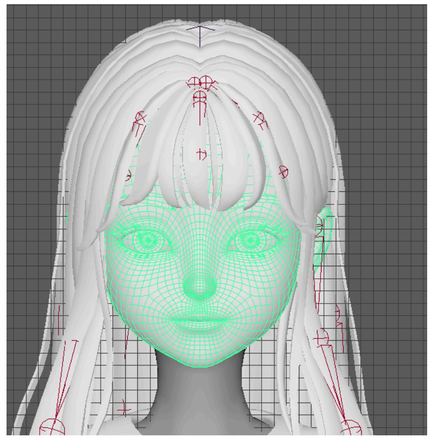
3.4 Morph Target 세팅

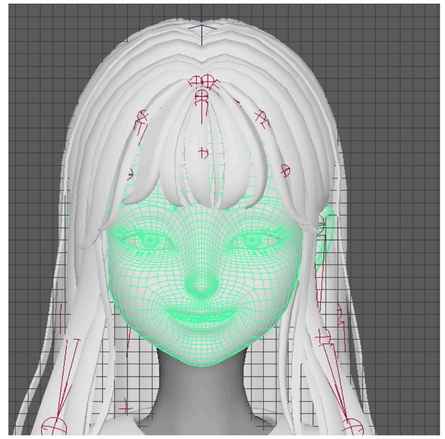
DEVIEW 2021

얼굴 애니메이션을 위한 BlendShape 세팅









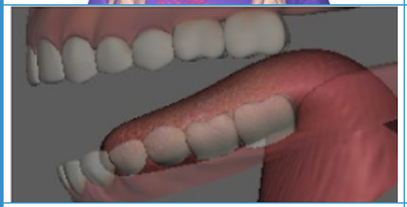
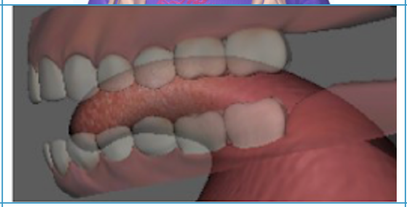
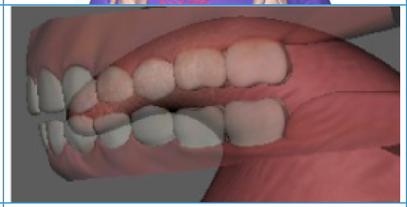
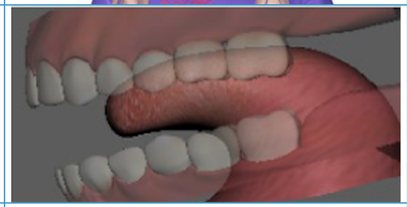
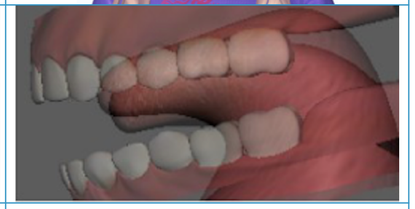
BlendShape Values	전	후
eyeBlinkLeft	0.132	0.31
eyeBlinkRight	0.132	0.31
mouthSmileLeft	0	0.919
mouthSmileRight	0.017	0.913

https://developer.apple.com/documentation/arkit/arfaceanchor/blendshapelocation/2928261-eyeblinkleft
https://developer.apple.com/documentation/arkit/arfaceanchor/blendshapelocation/2928249-mouthsmileleft

3.5 Lip-Sync Animation

DEVIEW 2021

- TTS Phoneme 연동
- BlendShape 세팅

Phoneme	A	E	I	O	U
BlendShape Image					
Shape					
한국어	ㅏ	ㅐ, ㅔ	ㅣ	ㅗ	ㅜ

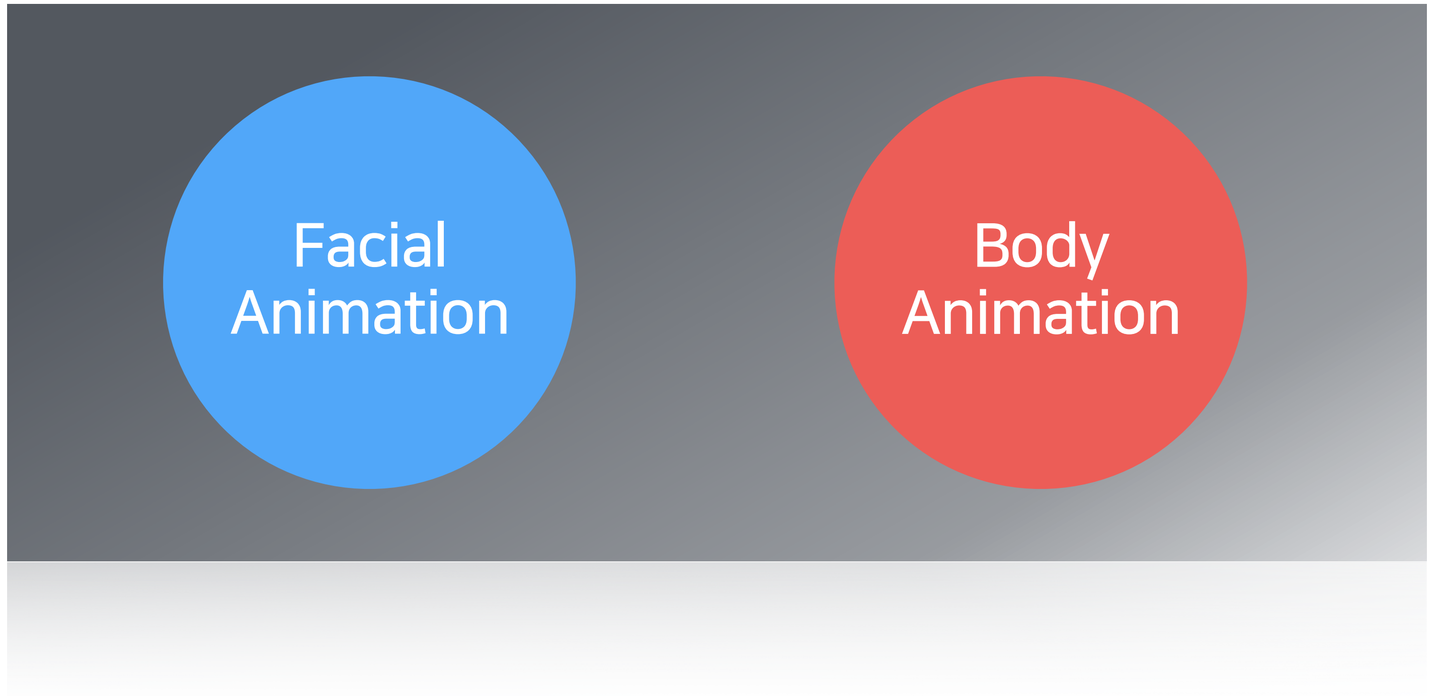
http://journal.cg-korea.org/archive/view_article?pid=jkcgs-26-3-49

3.5 Lip-Sync Animation

- 립싱크 Only 데모(얼굴표정, 바디 애니메이션 제외)
- CLOVA Voice의 TTS 연동

3.6 Motion Unit 예측 모델

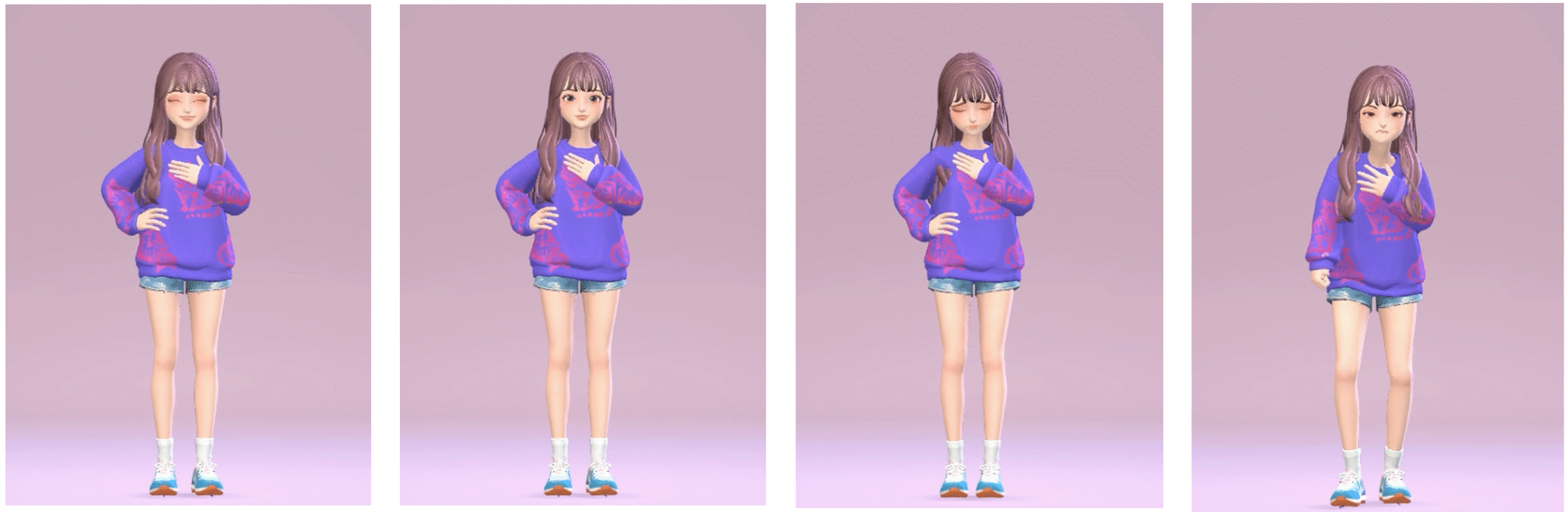
- Face와 Body 애니메이션을 하나의 Motion Unit으로 제작
- Lip Sync. 제외



3.6 Motion Unit 예측 모델

DEVIEW 2021

- Text 기반 Motion Unit 예측
- 같은 동작이더라도 감정별 다양한 Motion 합성 가능



기쁨 중립 슬픔 분노

3.7 Full-Body Animation 합성

DEVIEW 2021

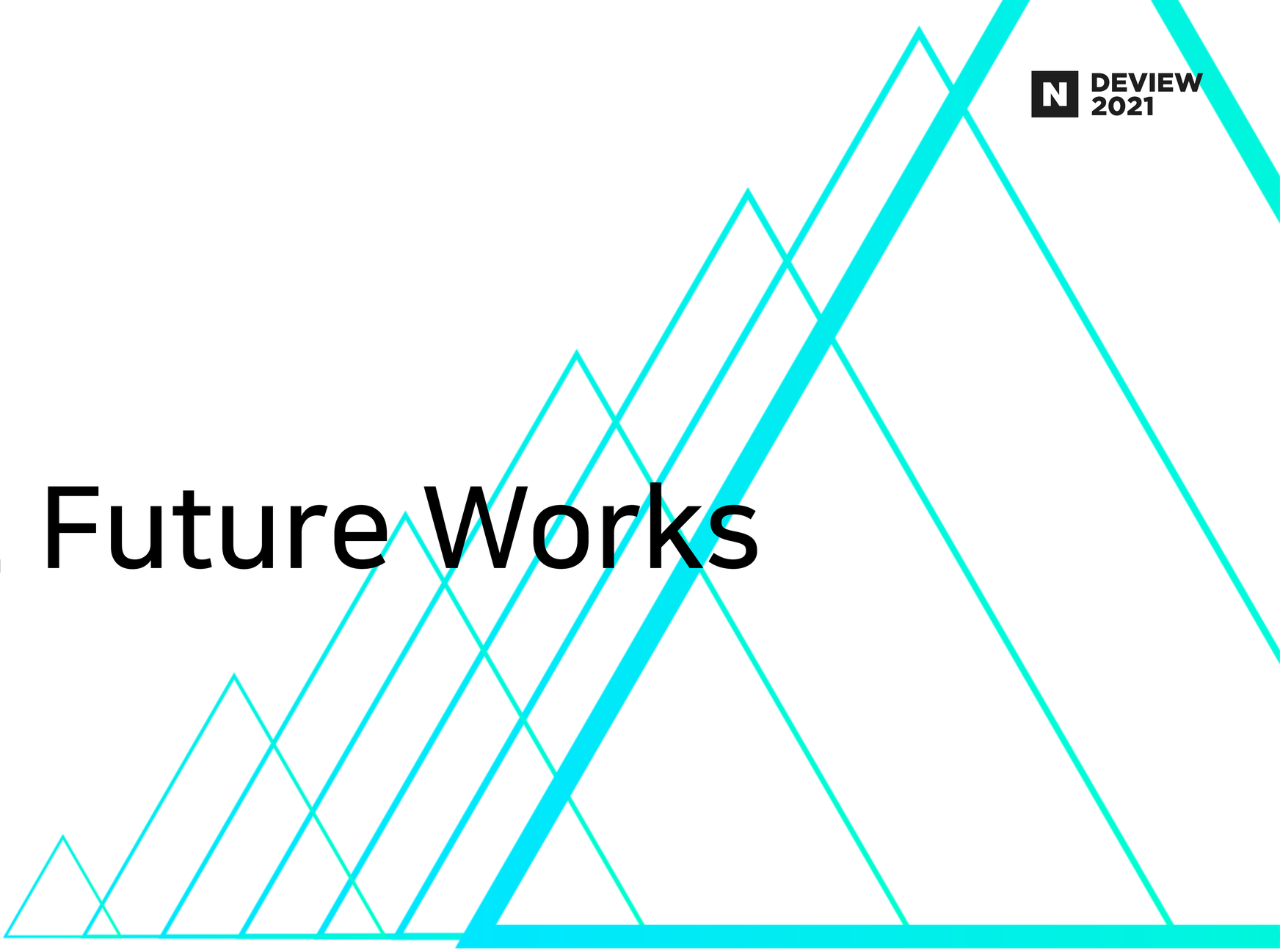
- Text 입력에 따른 아바타 생성 데모

3.7 Full-Body Animation 합성

DEVIEW 2021

- Lip-Sync + Motion Unit(Face, Body) 합성
- 동일 데이터 —> 다양한 캐릭터 적용 가능
- 범용적이고 디테일한 애니메이션 가능

4. Future Works



4. Future Works

DEVIEW 2021

Image 기반 아바타 생성

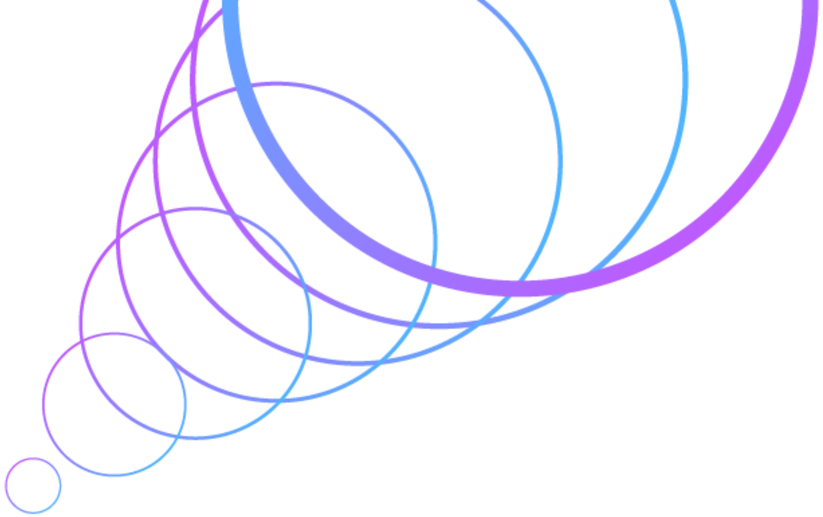
- 외부 배포 기능 개발
- 상용화 (전신 SDK, 상반신+양손 SDK, 한손 SDK)
- Hand Gesture Recognition 개발

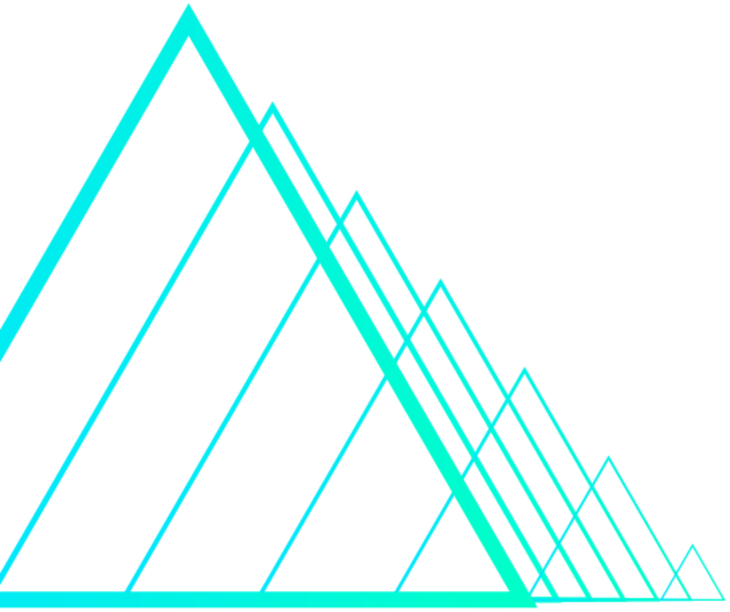
Text 기반 아바타 생성

- 다양한 언어 지원
- 학습 데이터 확대 (Script 및 Motion(Face, Body))
- Style-Controllable 모션 합성

이제 CLOVA Avatar 에서 만나요~!

Recruit(Intern/Regular): suntae.kim@navercorp.com





Q & A

Thank You