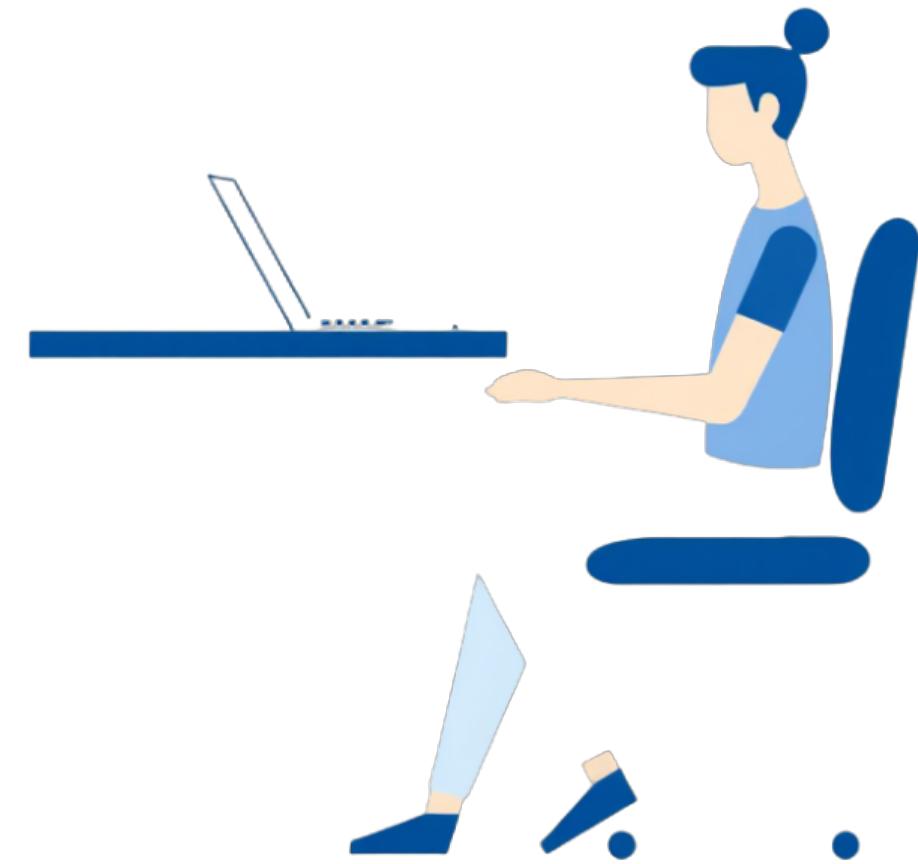


인체 키포인트 기반

자세 인식 시스템



목차

프로젝트 개요

프로젝트 목표, 주요 과정, 응용 분야에 대한 소개

전체 구조

데이터 수집부터 결과 출력까지의
전체 시스템 구조 설명

개발 환경

사용한 IDE, 언어, 하드웨어 및
주요 라이브러리 소개

키포인트 추출 기술

Mediapipe, YOLO를 이용한 키포인트 추출

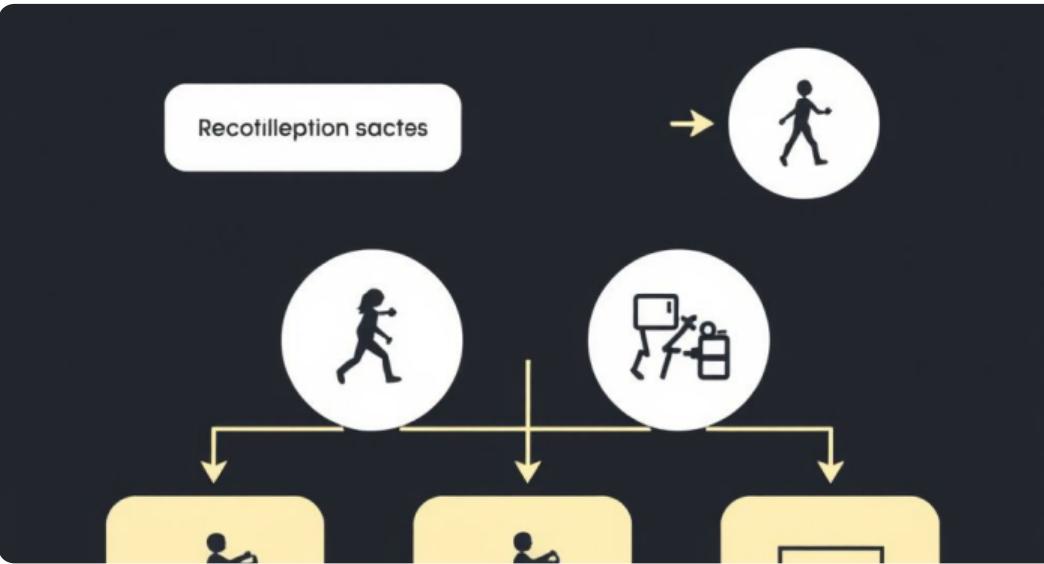
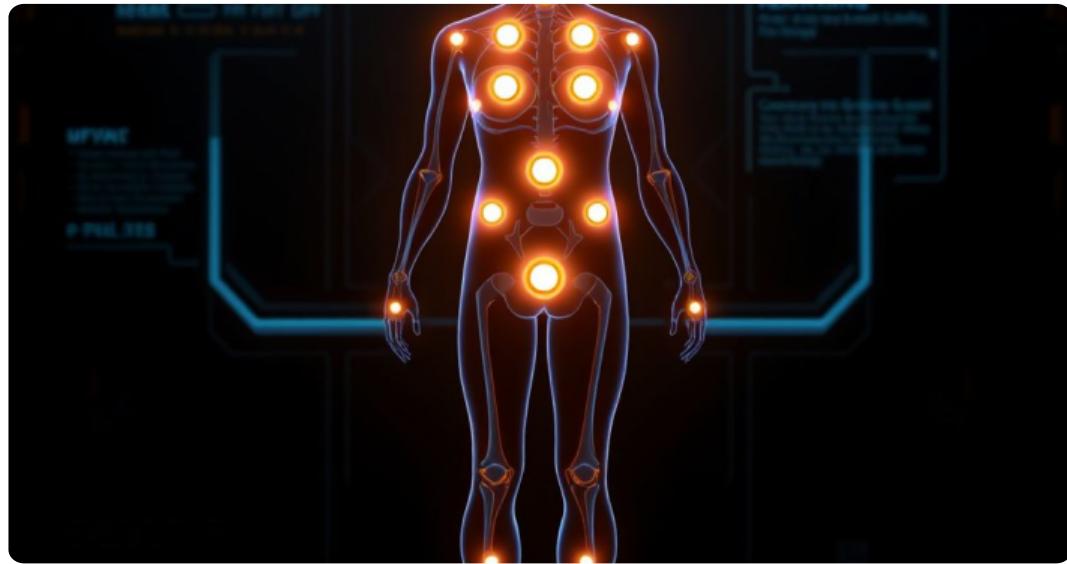
자세 추정 기술

MLP, GRU, Transformer 등
모델 구조를 활용한 자세 추정 기법

데이터셋 및 모델 학습

데이터 처리, 모델 학습 및 평가 방법론 설명

프로젝트 개요



프로젝트 목표

사람의 관절 키포인트를 추출하고, 이를 기반으로 여러 분야에서 자세를 인식하는 AI 시스템 개발

주요 과정

- 사람의 객체 인식 및 키포인트 추출
- 추출된 데이터 기반 자세 분류
- 결과 시각화 및 피드백 제공

응용 분야

- 피트니스: 개인 운동 자세 교정 및 트레이닝 보조
- 스포츠 분석: 선수들의 동작 분석 및 성능 향상
- 재활 의학: 환자의 운동 치료 모니터링
- 게임 및 엔터테인먼트: 모션 기반 게임
- 안전 및 산업: 작업 자세 모니터링



전체 구조

전체 구조



- Mediapipe, YOLO API를 활용하여 사람의 주요 관절 키포인트를 추출 및 인식.
- combination_keypoints 함수로 최종 키포인트 생성
- MLP, GRU, Transformer 모델을 이용해 추출된 키포인트 데이터로 20여 가지 자세 분류.
- 추정된 자세의 정확도 및 관련 정보를 시각적으로 표시. 텍스트 및 그래픽 결과로 사용자에게 피드백 제공

개발 환경

IDE 및 언어

IDE: PyCharm, Google Colab
언어: Python 3.9 (로컬), Python 3.10 (Colab)

로컬 자원

CPU: AMD Ryzen 7900X
GPU: NVIDIA 4060 Ti 16GB
RAM: 32GB

클라우드 자원

GPU: NVIDIA A100 40GB
RAM: 83.5GB

주요 라이브러리

Mediapipe (0.10.18), YOLO (YOLOv8)
PyTorch 2.5.1

키포인트 추출



키포인트 추출 기술



[Mediapipe](#)

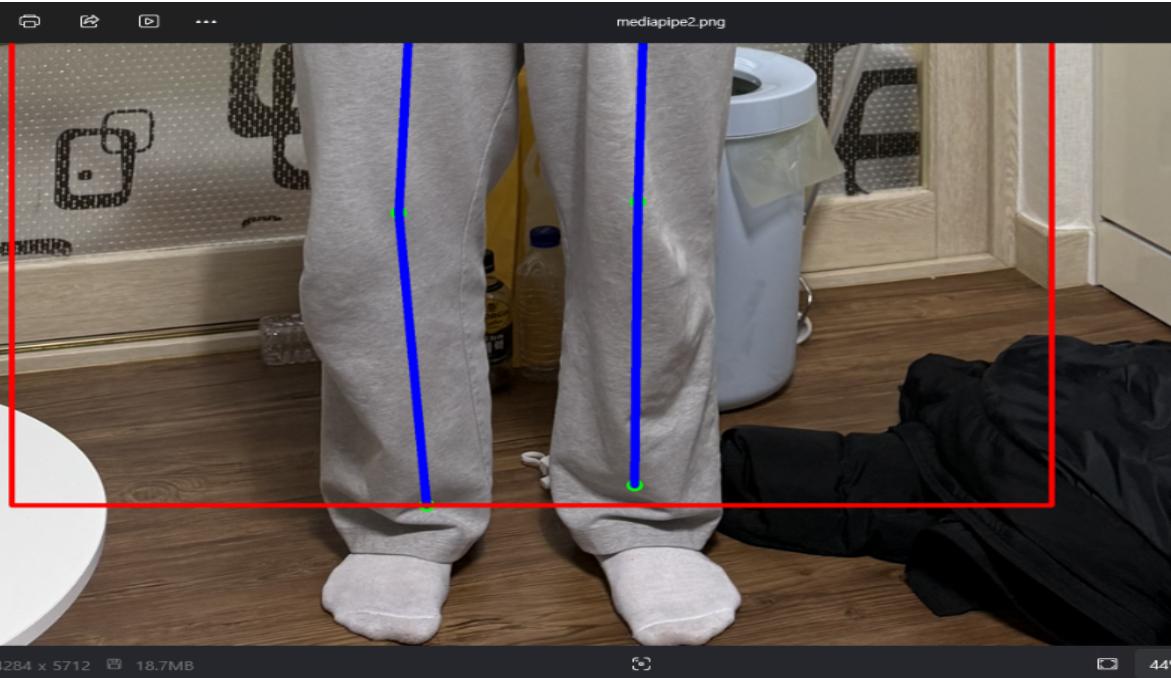
Google에서 개발한 오픈소스 프레임워크로,
실시간으로 사람의 키포인트를 추출할수 있고
다양한 플랫폼에서 사용 가능하며, 빠른 속도와
높은 정확도가 특징



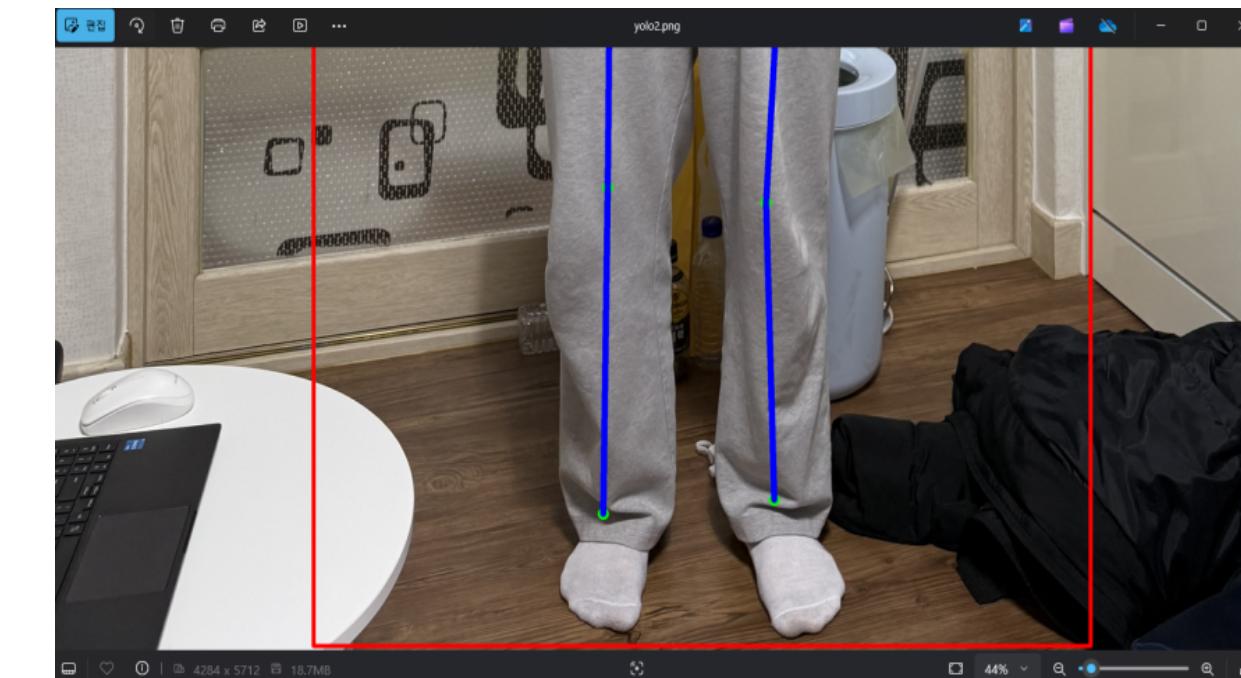
[YOLO](#)

객체 인식에 주로 사용되는 YOLO를 활용하여
사람의 위치를 파악한 후, 키포인트를 추출.
복잡한 환경에서도 높은 정확도를 보임.

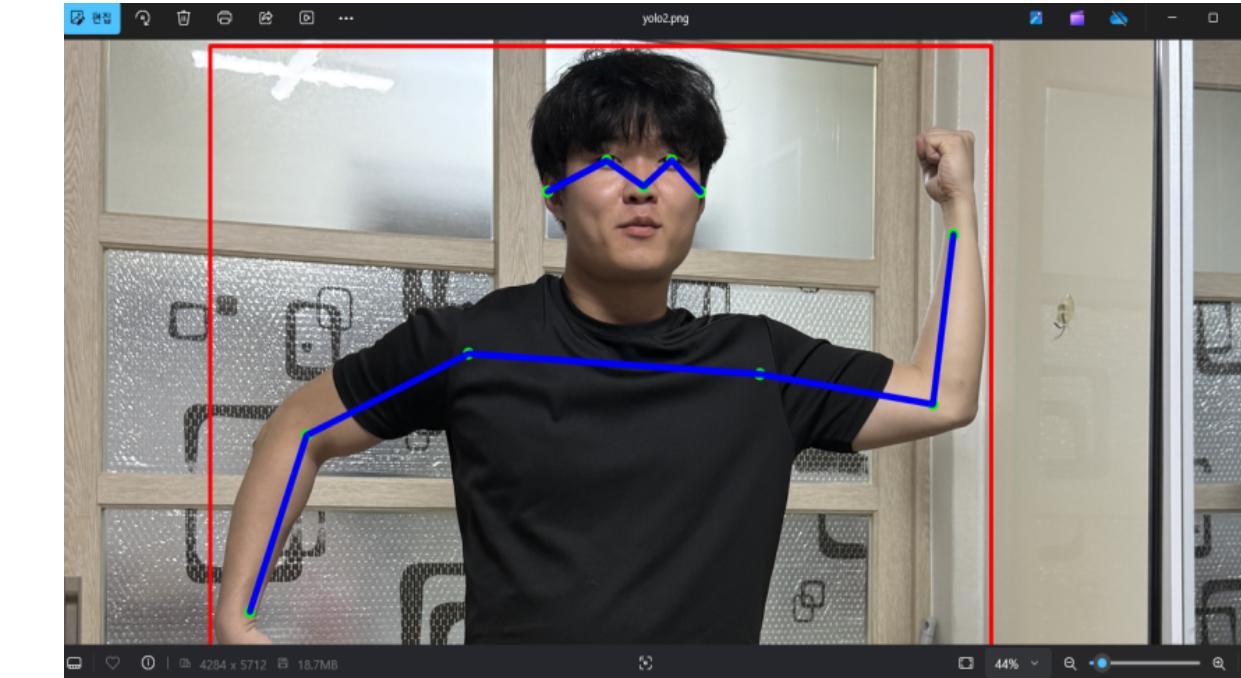
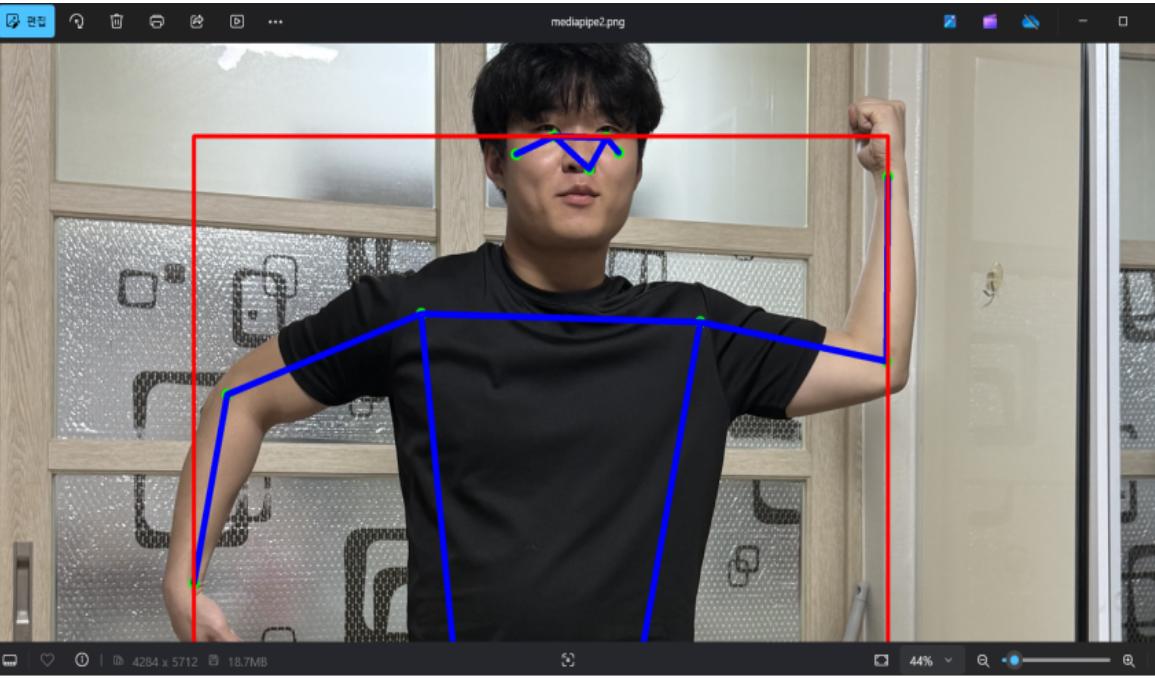
키포인트 조합 이유



Mediapipe



YOLO



사전학습 모델별로 추출을 우수하게 하는 키포인트가 모두 다르기 때문에
키포인트로 별로 모델의 성능을 비교하여 적용하는 것이 중요함

키포인트 추출 기능

01

combination_keypoints

두 모델(Mediapipe, YOLO)에서 추출한 키포인트를 조합하여 최종 키포인트를 생성.
이를 통해 각 모델의 장점을 활용하여 더 정확한 키포인트를 얻을 수 있음

02

normalize_keypoints

추출된 키포인트를 어깨 중심점 기준, 힙 중심점 기준으로 정규화
노멀라이즈는 포즈 데이터를 기준점(골반)으로 옮기고 크기를 어깨 너비 기준으로 통일하는 과정

03

estimate_hidden_keypoints

가려진 키포인트를 추정하여 보정, 해당 기능은 일부 신체 부위가 가려진 상황에서도
전체적인 포즈를 정확하게 추정할 수 있도록 도와줌



자세 추정

데이터 처리

↓ 다운로드



영상이미지

한국인 신체 3D 스캐닝
데이터

2022 | 206.94 GB

① 8,991 ⬆ 46 ⬇ 223

데이터 수집 및 분석

1. AI HUB 한국인 신체 자세 데이터셋 수집
2. 키포인트 JSON 데이터 필드 확인
3. 키포인트 필드 CSV 변환 후 좌표 통계 및 이상치 분석

데이터 변환 전처리 및 증강

1. 변환된 데이터를 수평 뒤집기, 노이즈 추가, 각도 회전 등 다양한 증강 적용
2. 증강 후 약 30만개의 키포인트 > 약 300만개의 키포인트
3. 불필요 컬럼 제거
4. 골반 기준점 기준 정규화 > 어깨너비 기준 스케일링
5. Train, Validation Set 분할

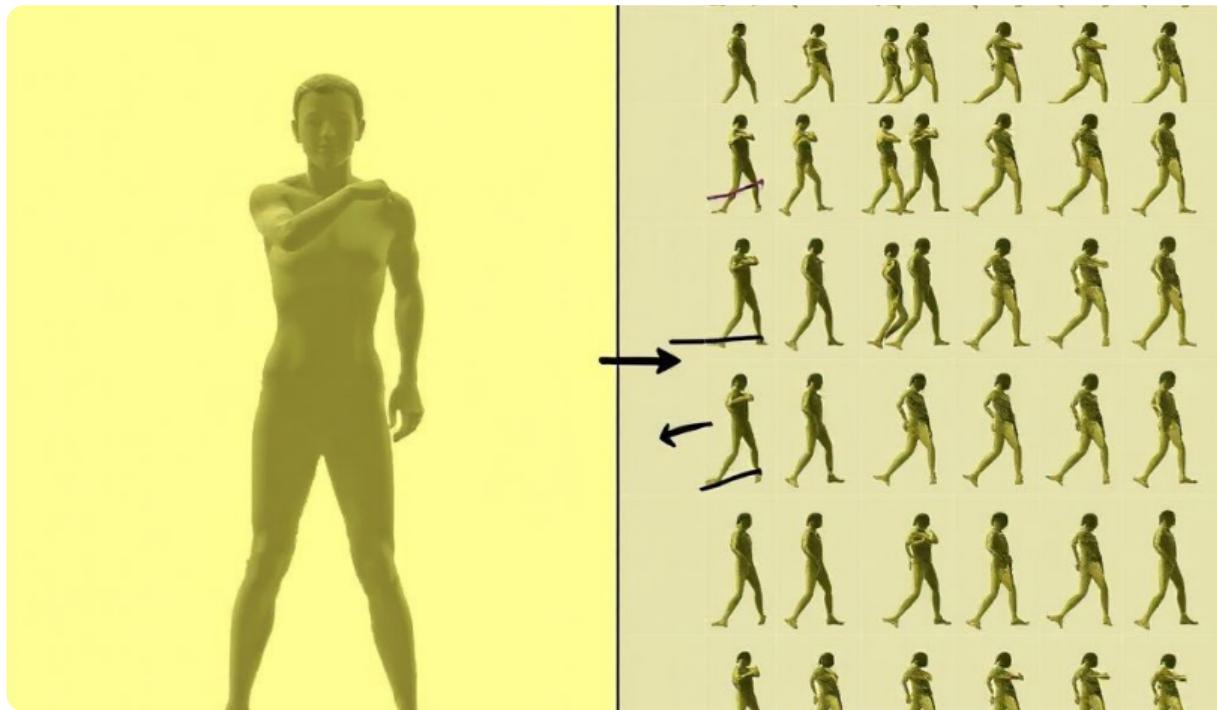
데이터 처리



자세 종류 분류

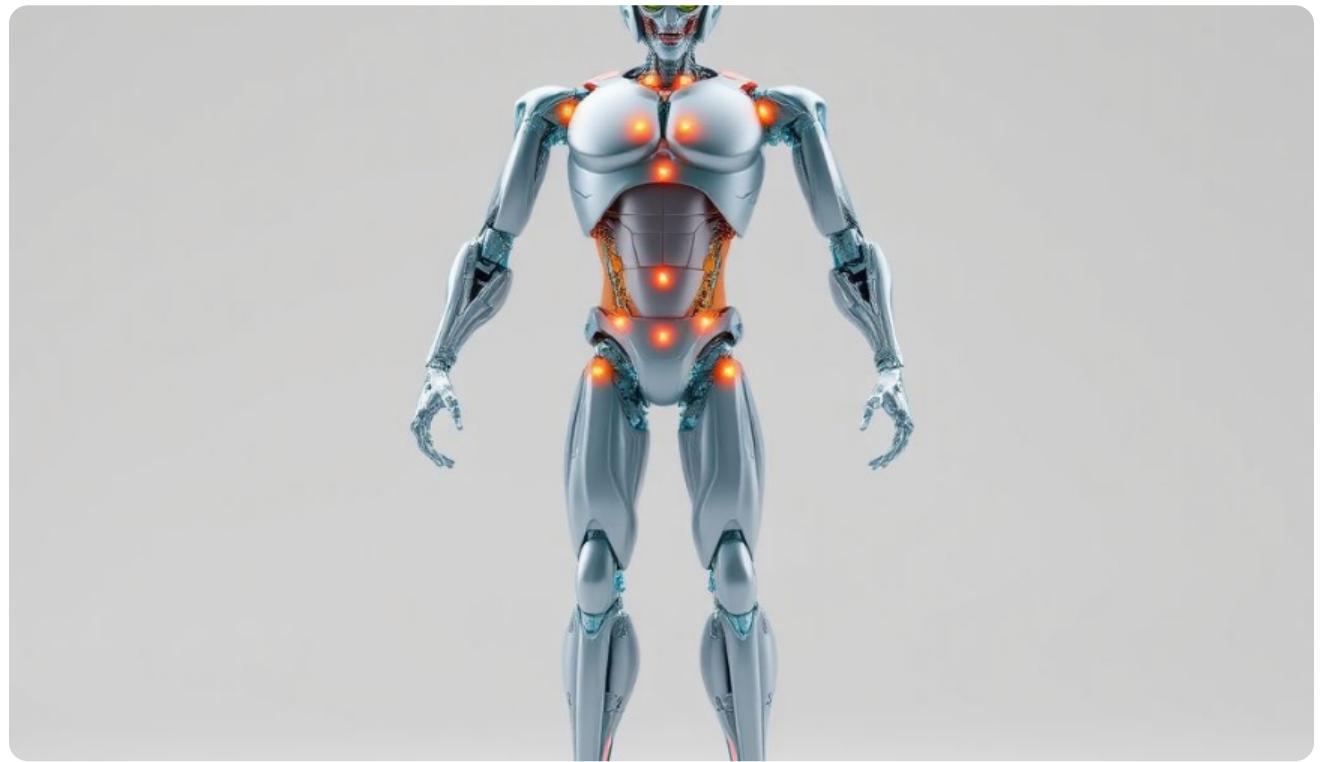
- 자세는 총 20가지 종류
A포즈, I포즈, T포즈, 계단 오르기,
공을 던지려고 힘을 주는 자세, 기지개, 달리기(전력질주),
뒷짐, 막대를 양손으로 잡고 골반 뒤쪽으로 쭉 뻗은 자세,
머리 뒤 깍지를 낀 자세, 몸을 앞으로 숙인 자세, 발레,
벽에 기대어 신발 신기, 의자에 앉은 자세, 조깅, 통화하는 자세,
팔짱, 한 손과 반대편 발을 들며 신난 자세,
한 다리 올리고 편하게 앉은 자세, 허리 회전을 최대로 한 자세

자세 추정 기술



다양한 모델 구조 및 메타 학습기

- MLP, GRU, Transformer, CNN, Vision
Transformer와 같은 모델들이 구조를 활용하여 자세를 정밀하게 분류함.
- 메타 학습기(Meta-learner)를 통해 각 모델의 예측 결과를 통합하여 최종 출력을 생성함. 이를 통해 더욱 정확한 자세 추정이 가능함.



다중 포즈 분류 및 성능 향상

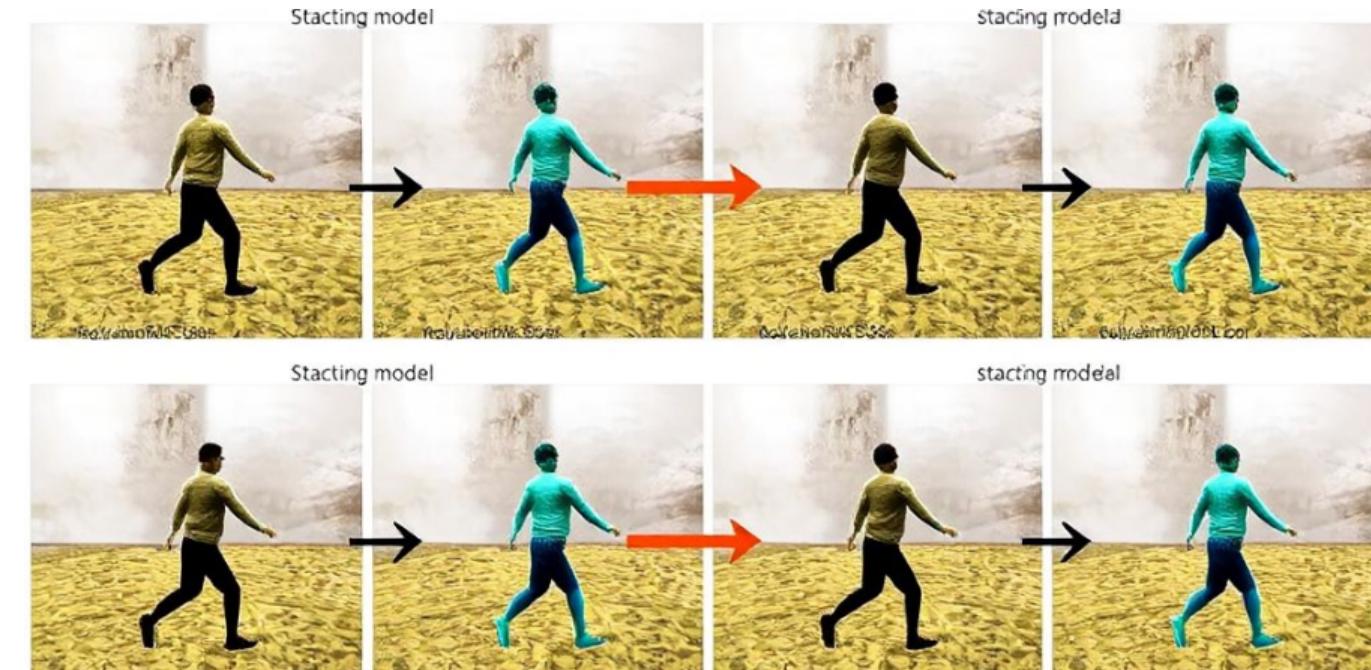
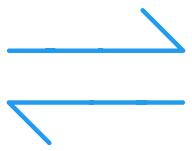
- 다양한 자세를 학습하고 높은 정확도로 다중 클래스를 분류함.
- 교차 엔트로피 손실 함수, AdamW 옵티마이저, ReduceLROnPlateau 스케줄러를 사용하여 모델의 성능을 지속적으로 개선하고 최적화함.

자세 추정 모델 장점



개별 모델의 특징

- MLP: 단순하면서 강력한 기준 모델, 정직 자세에 효과적
- Transformer: 관절 간 상호관계 포착, 복잡한 자세에 강함
- GRU: 연속 프레임의 자세 변화 학습, 효율적
- CNN: 키포인트의 지역적 패턴 추출, 구조적 이해
- ViT: 유연한 특징 추출, 전역/지역 특징 통합



스태킹 모델의 장점

- 다양한 모델의 장점을 결합하여 더 높은 정확도 달성
- 상황에 따라 적응적으로 각 모델의 예측을 결합
- 단순 다수결보다 뛰어난 성능으로 5-6개 클래스에서 도 우수한 결과
- 복잡하고 다양한 자세에 대해 안정적인 인식 능력

모델 학습 및 평가

학습 과정, 평가 지표, 성능 개선 방법

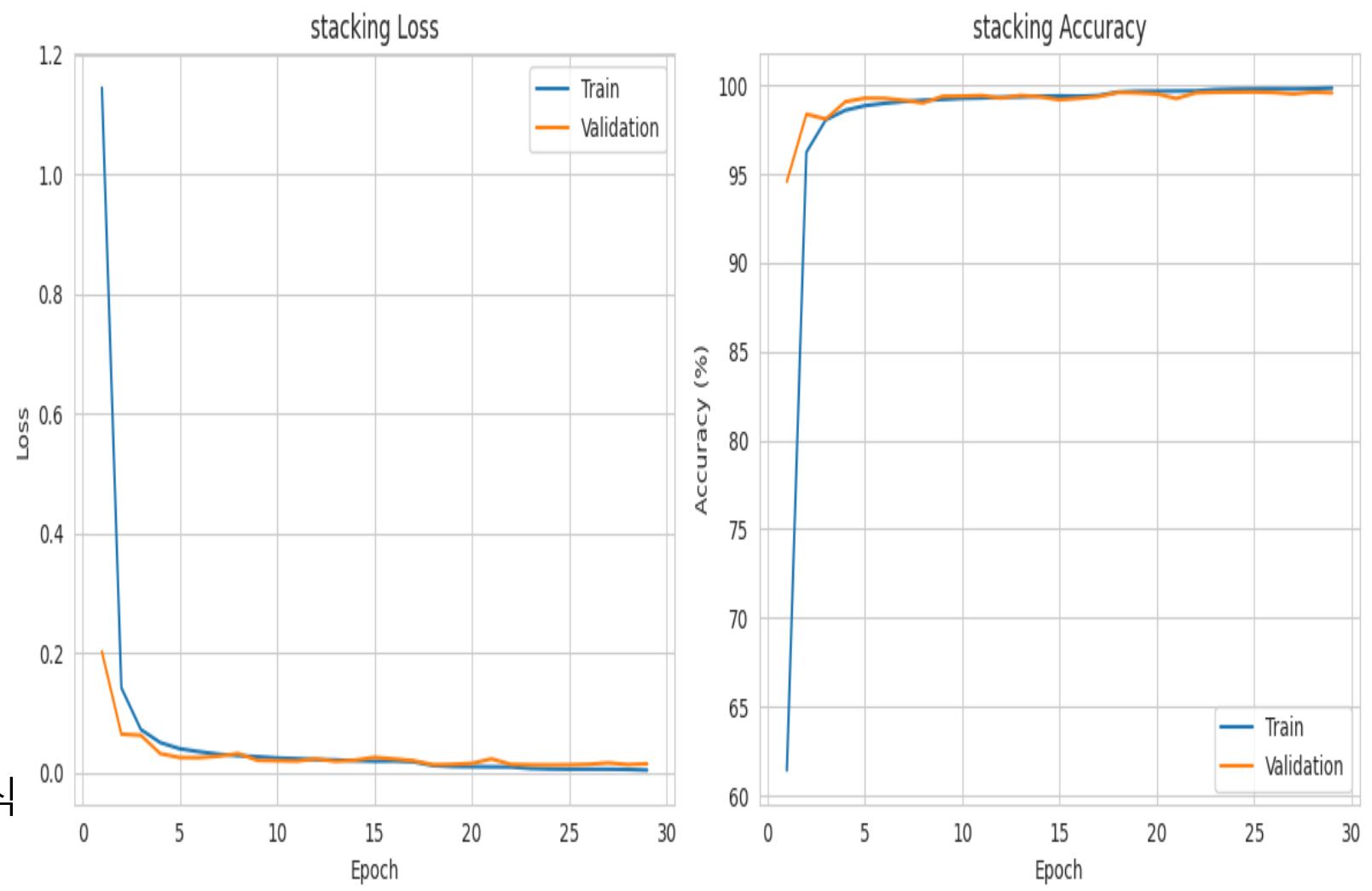
- 다섯 가지 모델(PoseTransformer, PoseMLP, PoseGRU, CNN, Vision Transformer)을 학습하며, 교차 엔트로피 손실 함수, AdamW 옵티마이저, ReduceLROnPlateau 스케줄러를 사용.
- 학습 과정에서 손실과 정확도를 기록하며, 검증 정확도가 가장 높은 모델을 저장.
- 팀원 2인이 촬영한 71개의 이미지 테스트 데이터를 통한 성능 평가.

STACKING Results:

Training - Loss: 0.0039, Accuracy: 99.86%
Validation - Loss: 0.0153, Accuracy: 99.57%
Test - Loss: 0.0375, Accuracy: 99.75%

전체 통계:

- 총 데이터 수: 71개
- 20개의 클래스 중 12개의 자세 100% 인식
- 정확히 예측한 수: 60개
- 오분류된 수: 11개
- 전체 정확도: 84.51% (60/71)



팀 구성 및 역할



이성준

프로젝트의 핵심 역할을 담당. 주요 코딩 작업을 수행하고 최종 프로젝트 발표를 준비. 키포인트 추출 및 자세 추정 알고리즘 개발에 주력.

오승진

프로젝트의 기술적 측면을 보완. 주로 코딩 작업에 참여하며, 키포인트 추출 개발에 주력.

이호성

프로젝트의 문서화 및 자료 정리 담당. 프로젝트 진행 상황을 체계적으로 정리하여 팀 내 정보 공유 촉진.

팀 협업

각 팀원의 전문성을 살려 효율적인 업무 분담 실현. 정기적인 회의를 통해 진행 상황 공유 및 문제 해결. 합숙하면서 서로 코드 보완하고 아이디어 보완하는 방안으로 진행

실행 화면 설명



실행 영상에서는 실시간 비디오 피드에서 사람의 신체 키포인트를 추출하고, 이를 정규화하여 다양한 자세를 정확히 추정하는 과정을 보여줌. Mediapipe, YOLO를 사용해 관절 키포인트를 추출하고, 이를 MLP, GRU, Transformer 모델로 분석하여 자세를 분류. 영상은 추출된 키포인트와 자세 추정 결과를 시각적으로 표시하며, 각 자세별 정확도와 함께 직관적인 그래픽으로 결과를 제공합니다. 실행동영상은 따로 기재

발표시 질문내용 답변 정리

1. 올로 미디어 파이프 결합할때 중복처리하고 좌표결합처리를 어케하는지?

답: 이미지 데이터에 대한 올로와 미디어 파이프의 키포인트를 추출 한 후 스켈레톤을 연결 해 각 모델이 추출의 인간의 시각적으로 정확도와 신뢰도를 구분한 다음 각 모델이 추출을 잘하는 포인트를 선정하여 combination함

2. 실행 동영상에서 같은 동작인데 다른 결과가 나올때가 있는데 그이유

답: 복잡한 자세 혹은 비슷한 자세에서 혼동을 보이는 것은 사실임. 현재 자원의 한계와 더욱 다양한 자세를 학습을 시키는 것에 대한 한계가 있어 현재 단계에서는 조금 어려운 것으로 판단됨. 하지만 증강 기법을 자세마다, 혹은 키포인트마다 조정을 한다면 개선할 여지가 있다고 기대할 수 있음

3. 전체구조 하나로 묶는게 좋을 것 같다 O

4. 노멀라이즈 방법에 대해 기재했으면 좋겠다. O

5. 자세에 대한 종류 몇가지 분류를 하는건지에 대한 내용을 추가하면 좋겠다. O

6. 다중 클래스 다중 포즈로 변경(그리고 분류 종류 먼저 나오게 변경) O

7. 에포크 수 명시 O

8. 학습 결과에서 벨리데이션 ,테스트 , 로스값 찍어봐야함(에포크 수가 증가하는데 벨리데이션 이 트레이닝에 비해 올라감 파라미터가 잘못된것) O

