

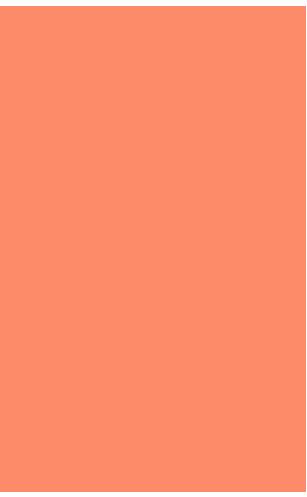
# YOLOv5 및 MediaPipe를 이용한 실시간 3대 운동 AI 자세 교정 서비스에 대한 연구

A Study on Real Time the big three exercises AI posture correction service  
Using YOLOv5 and MediaPipe

---

컴퓨터공학과 3학년 고영민

2023.09.20



# **\_INDEX**

---

**I . 논문 진척사항**

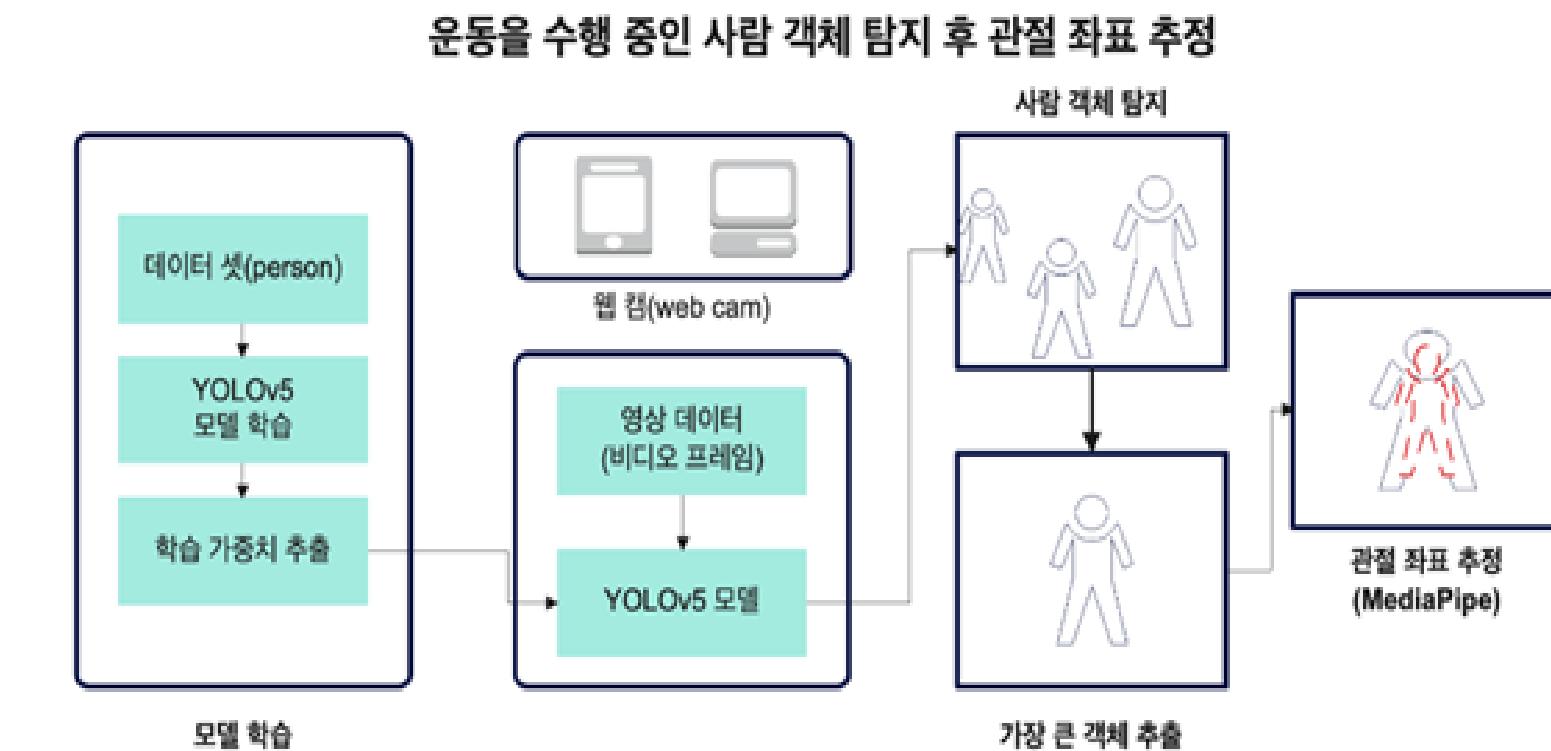
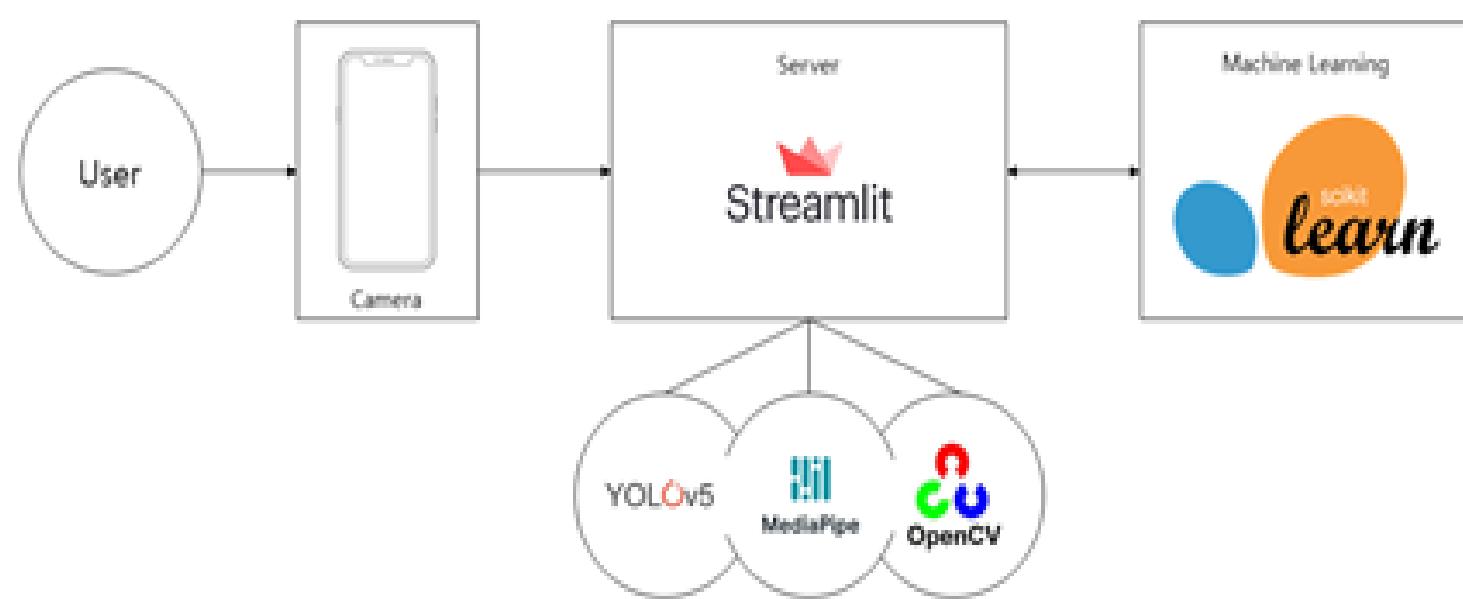
**II . 개발 진척사항**



## | . 논문 진척사항



# I. 논문 진척사항



# I. 논문 진척사항

## YOLOv5 및 MediaPipe를 이용한 실시간 3대 운동 AI 자세 교정 서비스에 대한 연구

고영민

A Study on Real-Time the big three exercises AI posture correction service  
Using YOLOv5 and MediaPipe

Yeong-Min Ko

### 요약

#### Abstract

#### I. 서론

##### 1.1 연구 배경과 필요성

2022년 문화체육관광부의 국민생활체육조사의 결과[1]를 보면 조사대상 중 61.2%는 주 1회 이상 꾸준히 생활체육을 즐겨하고 있으며, 그중에서 보디빌딩은 12.8%로 걷기 다음인 2위로 상당히 높은 비율을 차지하고 있다(그림 1). 보디빌딩의 핵심 운동은 3대 운동인 '벤치 프레스', '데드리프트', '숄더트'가 있다. 하지만 3대 운동은 주로 고중량을 다루는 운동이므로 잘못된 자세로 운동을 수행하는 경우, 큰 부상을 초래할 위험이 크다. 그러나 대다수는 헬스장에 설치되어 있는 거울에만 의존하여 스스로 자세를 파악하며 운동을 수행한다. 하지만 전문적인 트레이너가 자세를 봐주면서 실시간으로 피드백을 하는 경우가 아닌 이상 스스로 자세의 옳고 그름을 판단하기 어렵다.

##### <추가 작성 예정>

본 논문에서는 YOLOv5의 사물 검출 기술과 MediaPipe의 스켈레톤 인식 기술을 융합하여 높은 정확도로 운동 수행 데이터를 수집하고 머신러닝 모델을 구축한 후, 음성합성(TTS) 기술을 활용해 실시간으로 운동 자세에 대해 피드백을 제공함으로써 올바른 운동 자세를 취할 수 있도록 도와주는 서비스를 구현하였다.

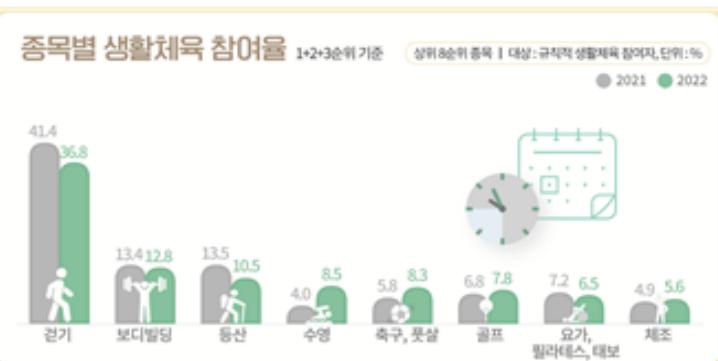


그림 1. 종목별 생활체육 참여율 통계

Fig. 1. Statistics of participation rate in daily sports by sport

#### 1.2 선행 연구 및 한계

옳지 않은 자세로 웨이트 트레이닝을 수행하여 부상을 초래하는 문제를 해결하기 위해 MeidaPipe나 OpenPose와 같은 프레임워크를 활용한 연구가 나타나고 있다. 이러한 문제를 해결하기 위한 방법에는 MediaPipe를 통해 글격 정보를 활용해 관절 점의 각도를 계산하여 기준 범위를 정해 옳고 그름을 판단하는 방법[2], MoveNet을 통해 원본군(사용자)의 글격 정보와 대조군(전문 트레이너 영상)의 글격 정보 얻고 딥러닝을 통해 원본군과 대조군을 비교하여 각 관절별로 옳고 그름을 판단하는 방법[3], 정적인 이미지 파일을 또는 미리 촬영된 비디오 파일을 업로드하여 자세를 판단하는 방법이 있다.[4][5]

기존 연구들은 단순히 전문가와 사용자의 글격 정보를 비교하거나 글격 정보를 활용하여 각 관절의 각도를 구한 값만을 활용하여 자세의 옳고 그름을 판단하여 정확도가 낮다는 한계가 있다. 또한 일부 연구에서는 정적 이미지나 촬영된 비디오 파일에 대해서만 피드백하는 방식으로 실제로 부상 위험이 가장 높은 운동을 수행하는 동안에는 실시간으로 자세에 대한 피드백 할 수 없다는 한계가 있다.

본 논문에서는 이러한 한계를 극복하기 위해 MediaPipe를 통해 추출한 관절 좌표 데이터 중에서 각 운동별로 사용되는 주요 관절만을 사용할 뿐만 아니라 주요 관절 간의 각도를 계산한 값을 후처리하여 지도학습을 진행함으로써 모델의 정확도를 높일 뿐만 아니라 Streamlit과 TTS(Text-To-Speech) 기술을 활용하여 실시간으로 운동 자세에 대해 교정해주는 인공지능 서비스를 구현하는 것을 목표로 한다.

#### II. 배경 지식

##### 2.1 YOLOv5

YOLOv5(You Only Look Once version 5)는 컴퓨터 비전 및 객체 감지 분야에서 중요한 역할을 하는 딥러닝 기반의 객체 감지 알고리즘이다. YOLOv5는 YOLO 시리즈 중 비교적 최근에 제안된 5번째 버전으로, 빠른 속도와 높은 정확도를 제공하여 다양한 응용 분야에서 사용되고 있다. 이 알고리즘의 주요 특징 중 하나는 실시간 객체 감지를 가능하게 하는 높은 처리 속도로 인해 실시간 비디오 스트림에서 실시간 객체 검출을 수행하는 데 매우 유용하다. 그림 2는 YOLOv5의 모델별 성능을 나타낸 것이다. 본 논문에서는 실시간 처리에 가장 적합한 YOLOv5s 모델을 사용한다.

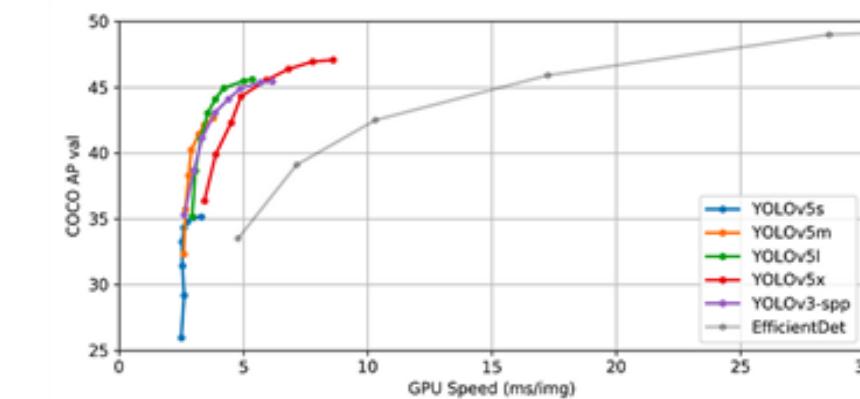


그림 2. YOLOv5 모델별 성능

Fig. 2. Performance of YOLOv5 models

# I. 논문 진척사항

미디어파이프(MediaPipe)는 구글(Google)에서 개발한 오픈소스 프레임워크로, 비전 및 머신러닝을 기반으로 한 컴퓨터 비전 응용 프로그램을 개발하는 데 사용되며 특히 실시간 데이터 스트리밍에서 작동하도록 설계되어 있다. 미디어파이프는 객체 감지, 자세 추정, 얼굴 감지, 손 추적 등과 같은 다양한 비전 작업을 지원한다. 본 논문에서는 그중에서 자세 추정(Pose Estimation)을 활용하여 그림 3과 같이 33개의 관절 좌표 데이터를 추출하여 활용한다.

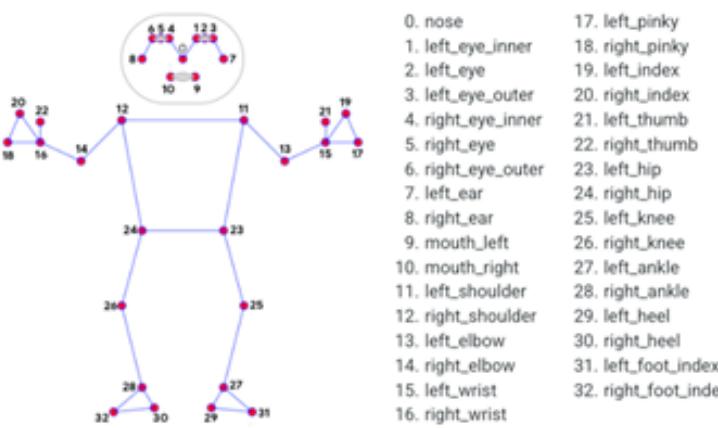


그림 3. MediaPipe Pose Landmark  
Fig. 3. MediaPipe Pose Landmark

## 2.3 OpenCV

OpenCV(Open Source Computer Vision Library)는 컴퓨터 비전 및 이미지 처리 작업을 위한 오픈소스 라이브러리로 다양한 프로그래밍 언어에서 사용할 수 있다. OpenCV는 이미지 및 비디오 처리, 객체 감지, 얼굴 인식, 자동차 번호판 인식, 캘리브레이션 등 다양한 컴퓨터 비전 작업을 지원하며, 컴퓨터 비전 연구 및 응용 프로그램 개발에 널리 사용된다. 본 논문에서는 파이썬 프로그래밍 언어를 사용하여 OpenCV 라이브러리를 활용한다.

## 2.4 Streamlit

Streamlit은 데이터 과학 및 웹 애플리케이션 개발을 위한 파이썬 기반의 오픈소스 라이브러리이다. Streamlit을 사용하면 데이터 시각화, 웹 애플리케이션, 대시보드 등을 빠르고 간편하게 개발할 수 있으며, 데이터 과학자 및 엔지니어들이 데이터를 시각화하고 공유하는데 매우 유용한 도구이다. 본 논문에서는 Streamlit을 통해 실시간으로 이미지 처리를 수행함으로써, 운동 자세에 대한 피드백을 사용자에게 실시간으로 제공한다.

## III. 제안하는 실시간 3대 운동 AI 자세 교정 서비스

### 3.1 시스템 아키텍처

본 논문에서 제안하는 시스템 구성도는 그림 4와 같다. 제안하는 실시간 3대 운동 AI 자세 교정 서비스는 Streamlit을 이용하여 서버를 구축한다. 사용자의 모바일 카메라로 촬영된 모든 프레임을 OpenCV를 통해 실시간 영상으로 처리한다. 서버에서는 전달받은 영상의 모든 프레임에 대해서 YOLOv5를 이용해 학습시킨 사람 객체를 탐지하는 모델을 통해 바운딩 박스가 가장 큰 객체를 실제 운동 중인 객체로 판단하고, 가장 큰 바운딩 박스 내부에 있는 객체에 대해서 MediaPipe 프레임워크를 활용해 운동을 수행하는 사람의 관절 좌표 정보를

추정한다. 마지막으로 각 운동별로 올바른 자세와 옳지 않은 자세를 학습시킨 머신러닝 모델을 통해 자세에 대한 피드백을 제공한다.

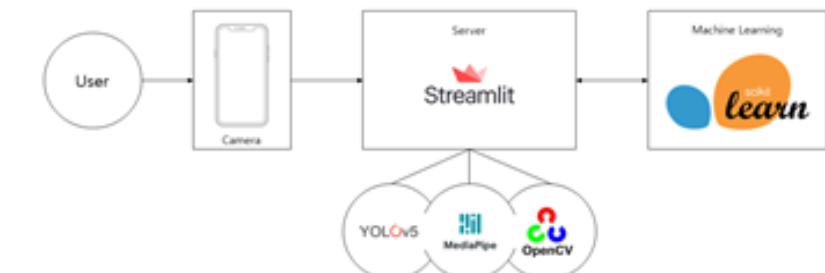


그림 4. 시스템 아키텍처  
Fig. 4. System Architecture

### 3.2 운동을 수행 중인 사람 객체의 관절을 추정하는 알고리즘

MediaPipe는 모바일에서도 실시간 추론을 할 수 있도록 매우 빠른 모델이지만 여러 사람이 아닌 오직 한 사람만 추적하여 관절 좌표를 추정할 수 있다. 하지만 헬스장과 같이 여러 사람이 함께 촬영될 때는 자세 추정의 정확도가 낮아질 수 있으며 정확하지 않은 피드백을 제공할 위험이 있다. 따라서 본 논문에서는 이러한 한계를 극복하기 위해 그림5와 같은 운동을 수행 중인 사람 객체의 관절을 추정하는 알고리즘을 제시한다.

본 논문에서 제안하는 운동을 수행 중인 사람 객체의 관절을 추정하는 알고리즘은 사람과 관련된 데이터셋을 YOLOv5 모델로 학습시킨 후 그 가중치를 사용하여 사용자의 웹캠으로부터 입력된 영상 데이터 중 사람 객체를 탐지한다. 그리고 가장 큰 바운딩 박스를 가진 객체를 실제 운동 중인 사람 객체라고 판단하고 해당 바운딩 박스 내부에 있는 객체에 대해서 MediaPipe 프레임워크를 활용해 관절 좌표를 추정한다.

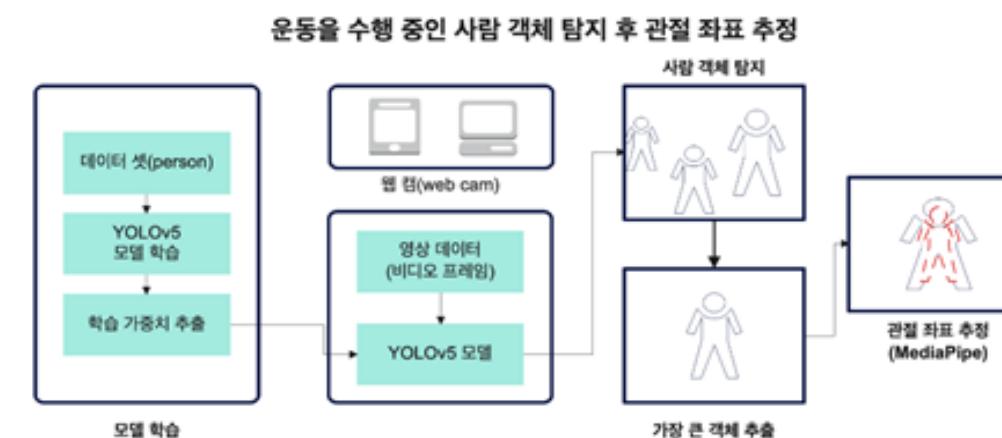


그림 5. 운동을 수행 중인 사람 객체 탐지 후 관절 좌표 추정  
Fig. 5. Estimating joint coordinates after detecting a person object exercising

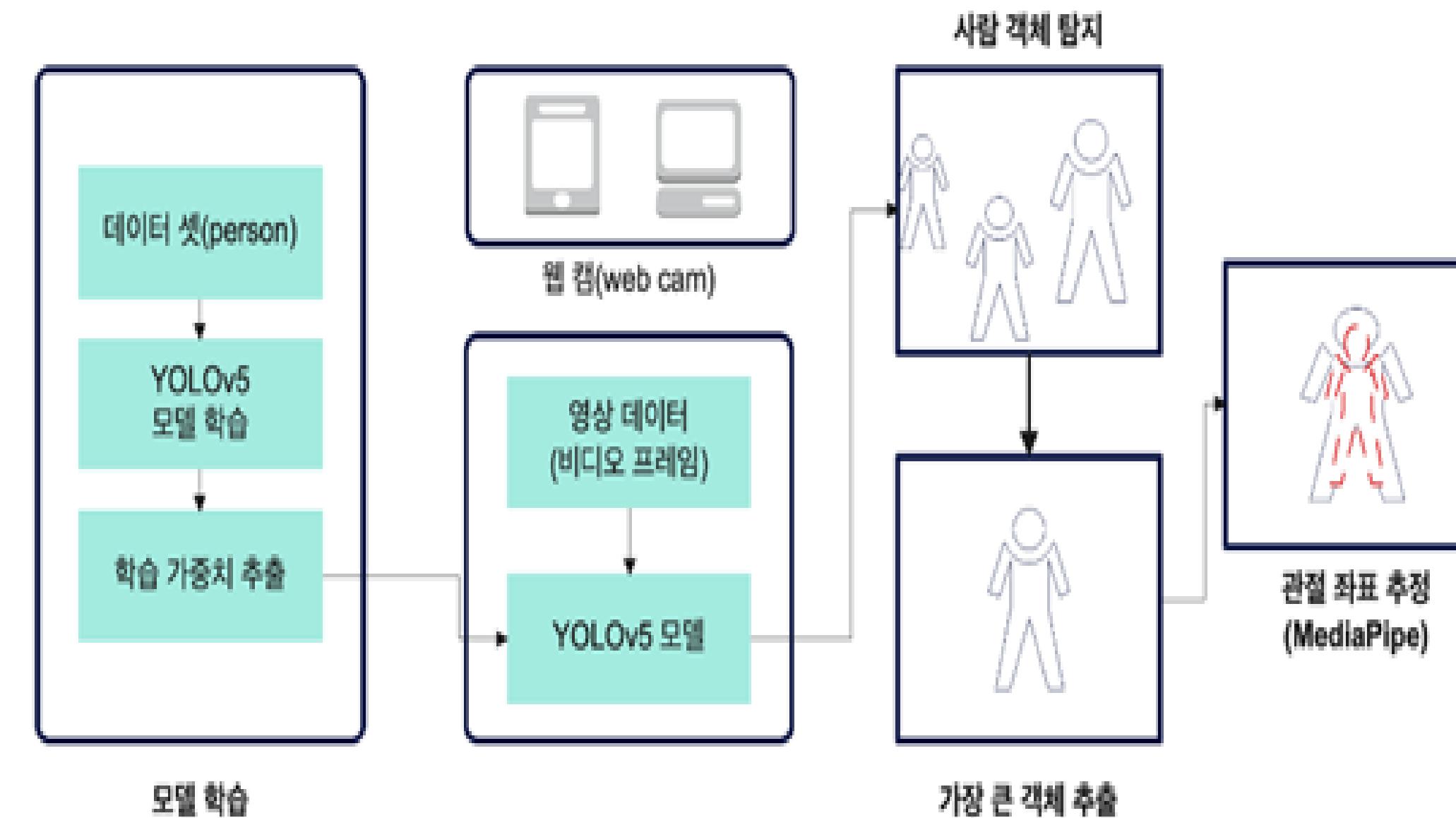


## II. 논문 및 개발 진척사항



## II. 개발 진척사항

### 운동을 수행 중인 사람 객체 탐지 후 관절 좌표 추정



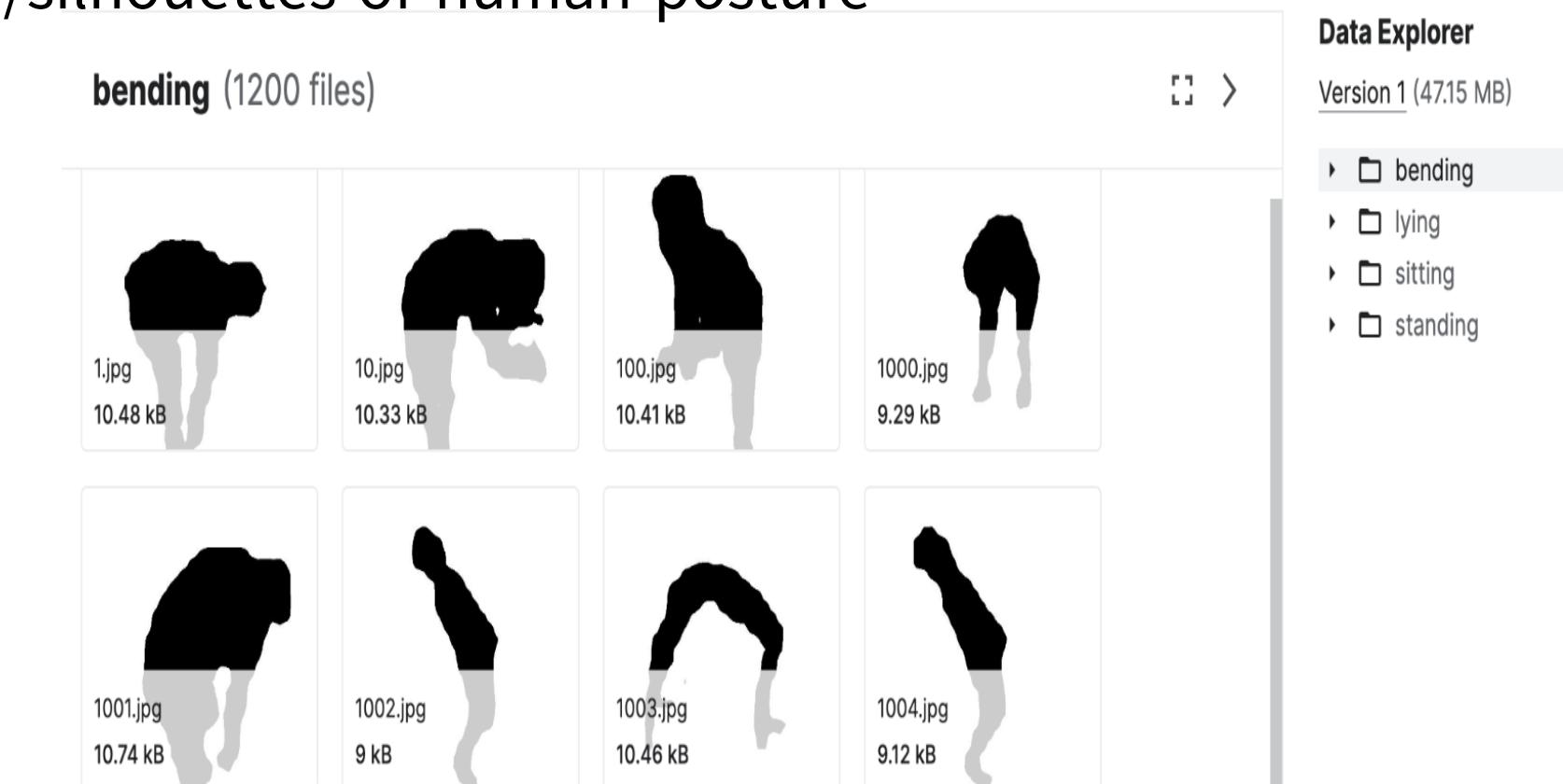
## II. 개발 진척사항

- YOLOv5\_Detect\_Only\_Person
- 총 데이터셋 805장(Google Images Scraping & \*Roboflow)
  - Keywords: 벤치 프레스, 스쿼트, 데드리프트, Standing, Lying Down Pose

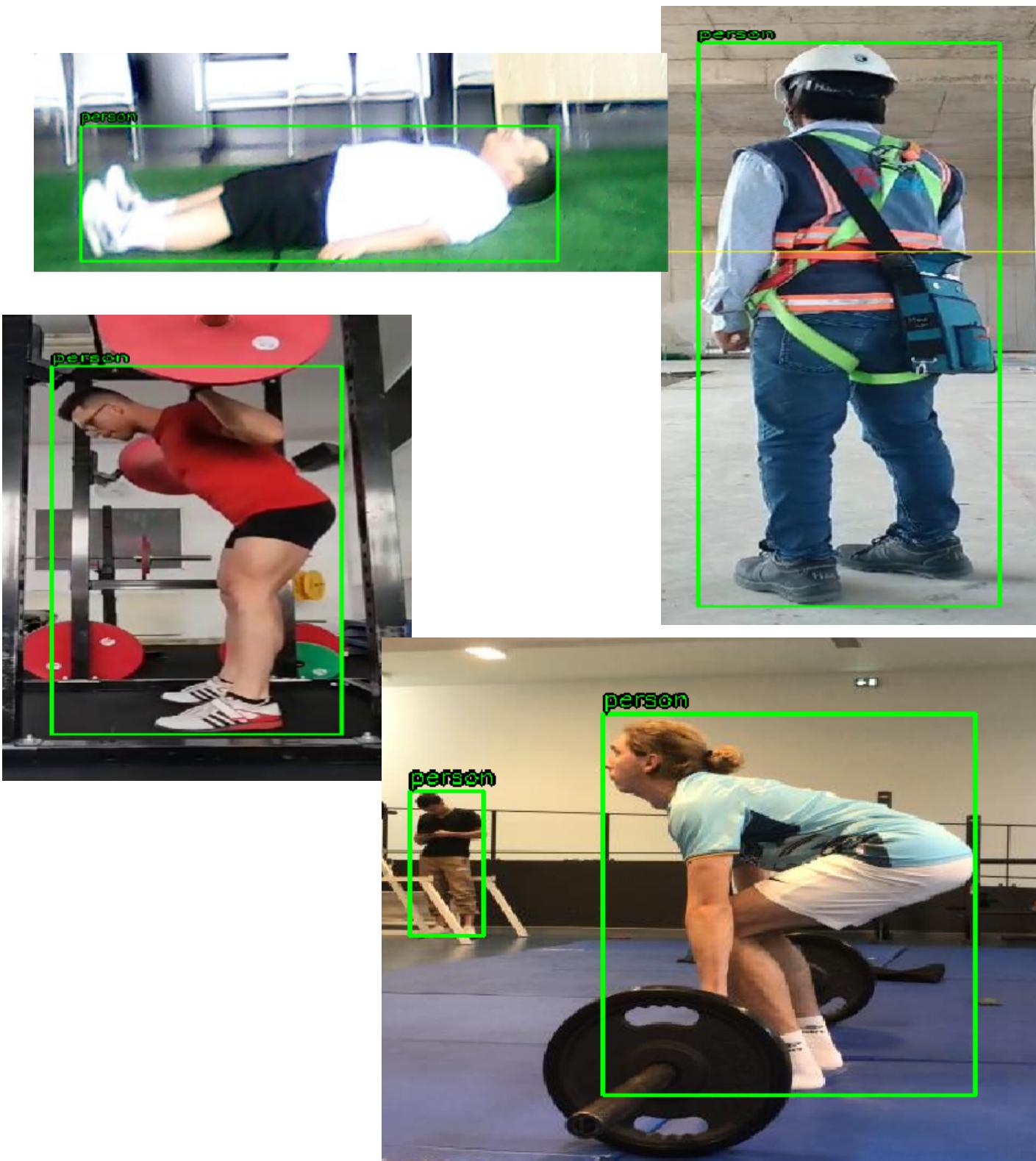
+

- 총 데이터셋 4,800장(Kaggle) ~ used for only training model
  - Silhouettes of human posture
    - <https://www.kaggle.com/datasets/deepshah16/silhouettes-of-human-posture>
  - Class: bending, lying, sitting, standing

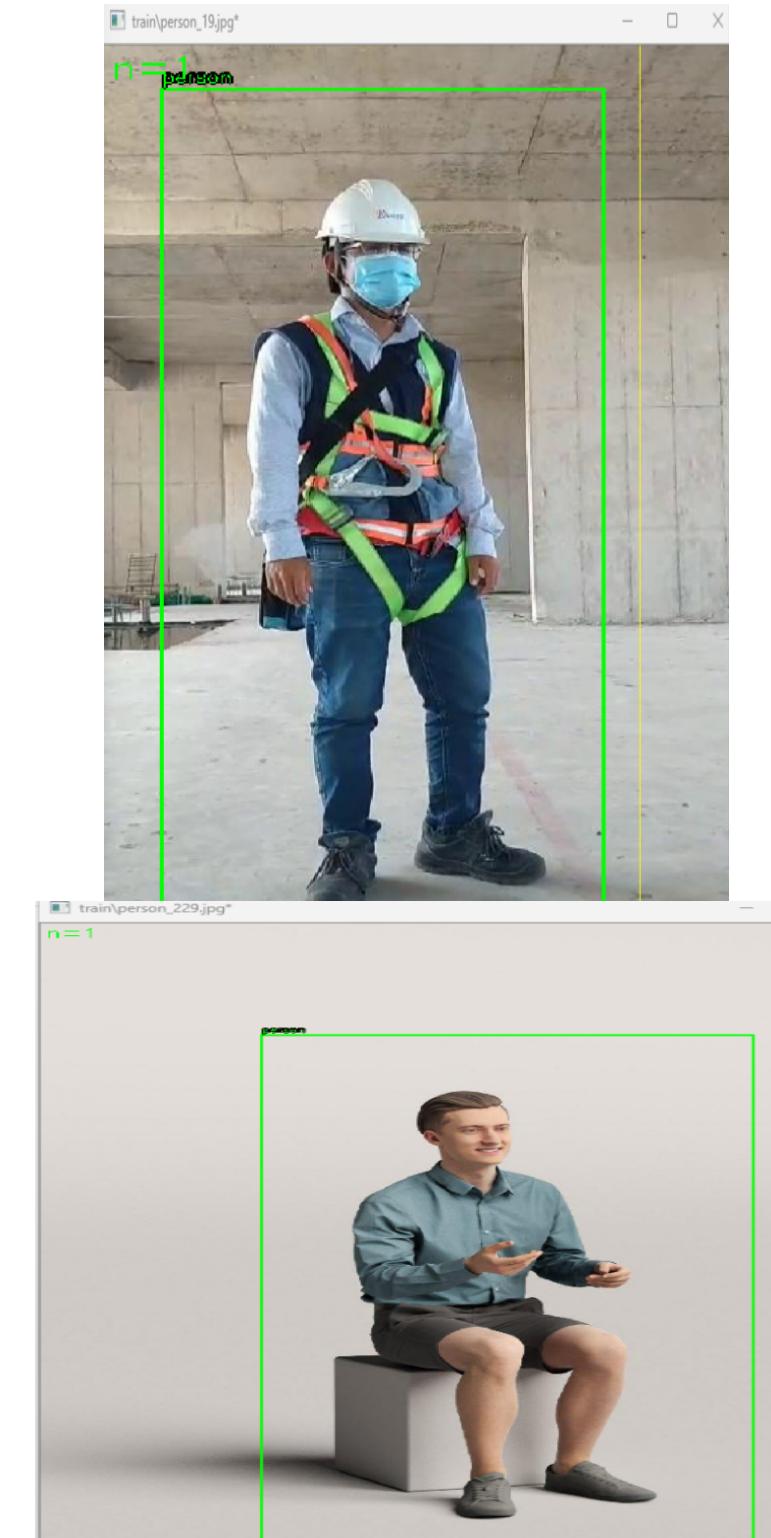
\*Roboflow: 컴퓨터 비전 및 이미지 처리 작업을 위한 플랫폼



## II. 개발 진척사항 - Relabeling



• 2주차



• 3주차



## II. 논문 및 개발 진척사항

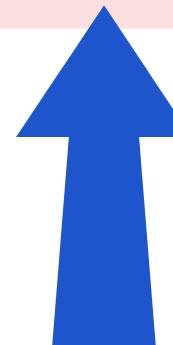
### ii. 개발

- YOLOv5\_Detect\_Only\_Person

- batch: 16 / epochs: 200(but, early stop: 167)

- 2주차

```
# epochs 100
python ./yolov5/train.py --batch 16 --epochs 100 --data person.yaml --cfg ./yolov5/models/yolov5s.yaml --weights yolov5s.pt --img-size 640 --model yolov5s --name results_person --cache
Fusing layers...
YOLOv5s summary: 157 layers, 7012822 parameters, 0 gradients, 15.8 GFLOPs
          Class   Images Instances      P      R    mAP50    mAP50-95: 0% | 0/6 [00:00
<?, ?it/s]          Class   Images Instances      P      R    mAP50    mAP50-95: 17%|#6 | 1/6 [00:00
<00:01, 3.05it/s]          Class   Images Instances      P      R    mAP50    mAP50-95: 33%|##3 | 2/6 [00:00
<00:01, 2.49it/s]          Class   Images Instances      P      R    mAP50    mAP50-95: 50%|#### | 3/6 [00:01
<00:01, 2.52it/s]          Class   Images Instances      P      R    mAP50    mAP50-95: 67%|#####6 | 4/6 [00:01
<00:00, 2.54it/s]          Class   Images Instances      P      R    mAP50    mAP50-95: 83%|#####3 | 5/6 [00:01
<00:00, 2.83it/s]          Class   Images Instances      P      R    mAP50    mAP50-95: 100%|##### | 6/6 [00:01
<00:00, 3.17it/s]
          all     162     204    0.995    0.984    0.994    0.843
Results saved to yolov5\runs\train\results_person
```



- 3주차

```
Validating yolov5\runs\train\results_person2\weights\best.pt...
Fusing layers...
```

```
YOLOv5s summary: 157 layers, 7012822 parameters, 0 gradients, 15.8 GFLOPs
```

	Class	Images	Instances	P	R	mAP50	mAP50-95: 0%	0/6 [00:00<?, ?it/s]
s]	Class	Images	Instances	P	R	mAP50	mAP50-95: 17% #6	1/6 [00:00<00:01, 3.19it/
s]	Class	Images	Instances	P	R	mAP50	mAP50-95: 33% ##3	2/6 [00:00<00:01, 2.39it/
s]	Class	Images	Instances	P	R	mAP50	mAP50-95: 50% ####	3/6 [00:01<00:01, 2.41it/
s]	Class	Images	Instances	P	R	mAP50	mAP50-95: 67% #####6	4/6 [00:01<00:00, 2.43it/
s]	Class	Images	Instances	P	R	mAP50	mAP50-95: 83% #####3	5/6 [00:01<00:00, 2.68it/
s]	Class	Images	Instances	P	R	mAP50	mAP50-95: 100% #####	6/6 [00:01<00:00, 3.06it/
	all	162	186	0.989	1	0.994	0.705	

```
Results saved to yolov5\runs\train\results_person2
```

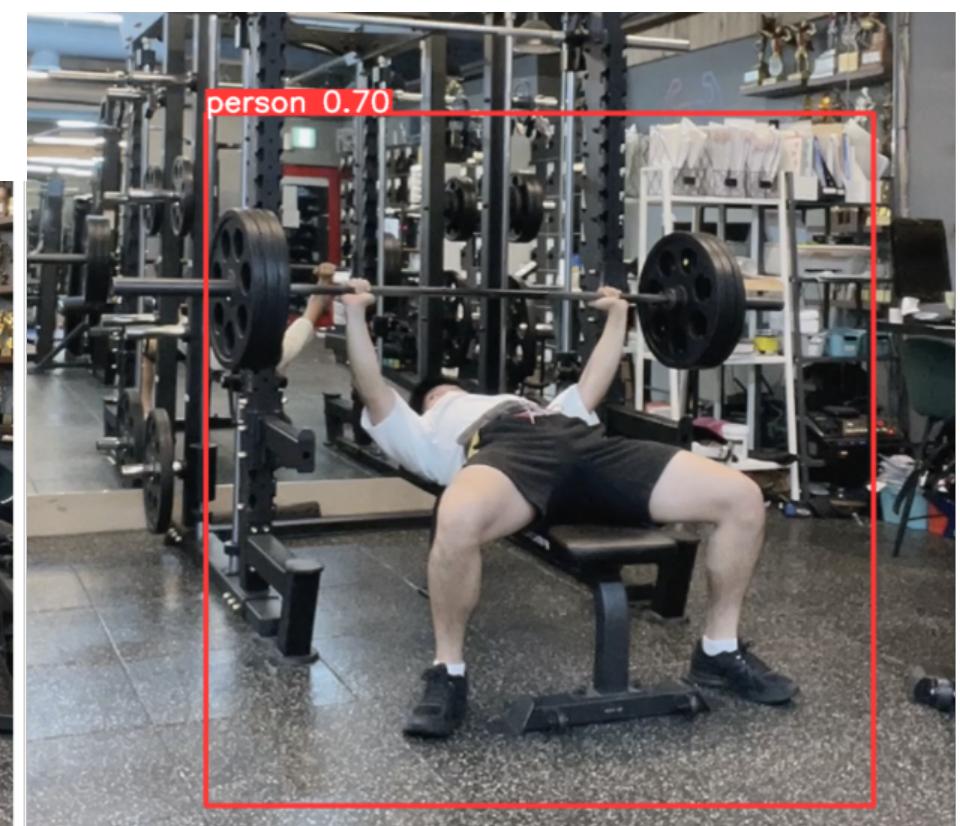
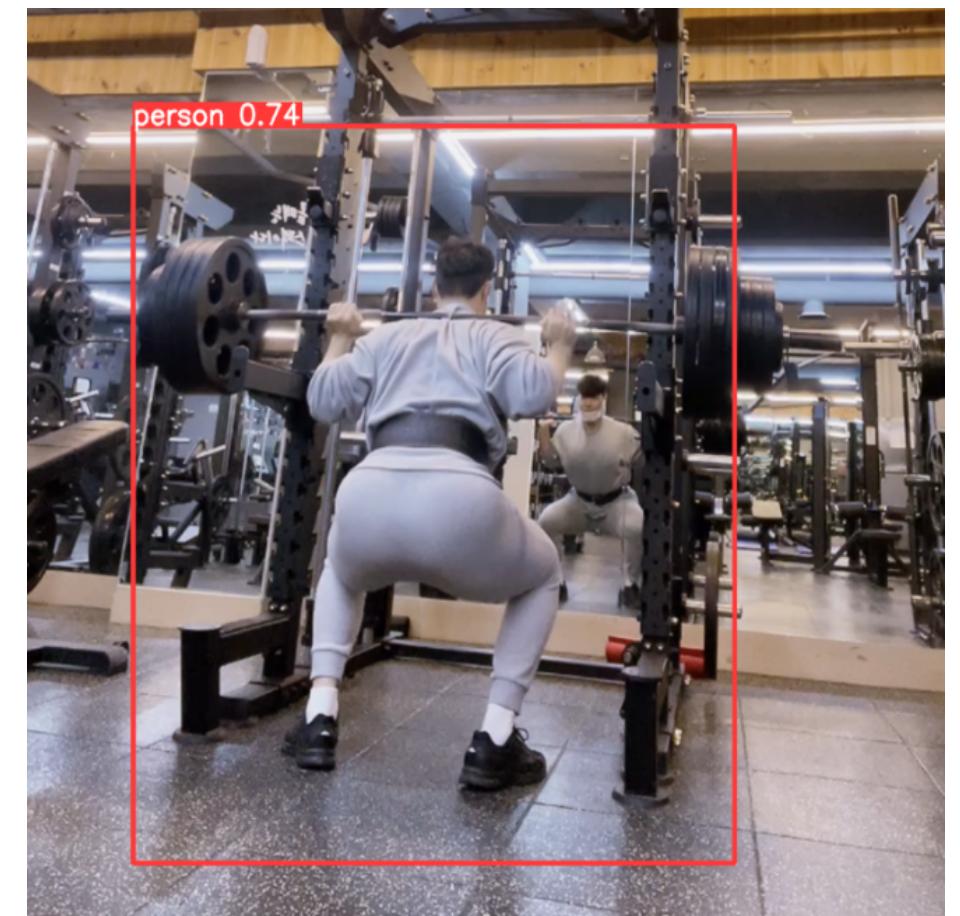
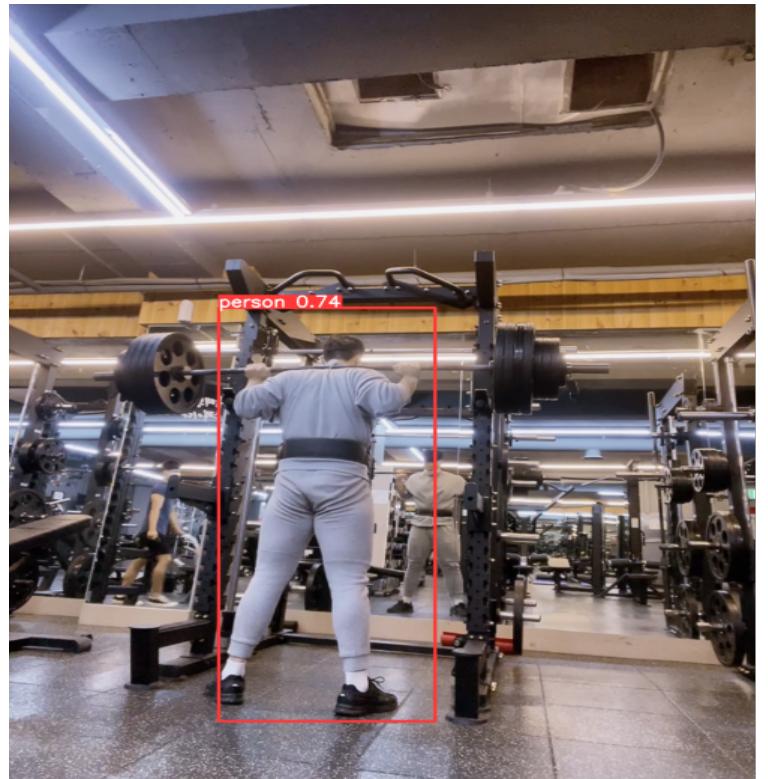


• 3주차

## II. 논문 및 개발 진척사항

### ii. 개발

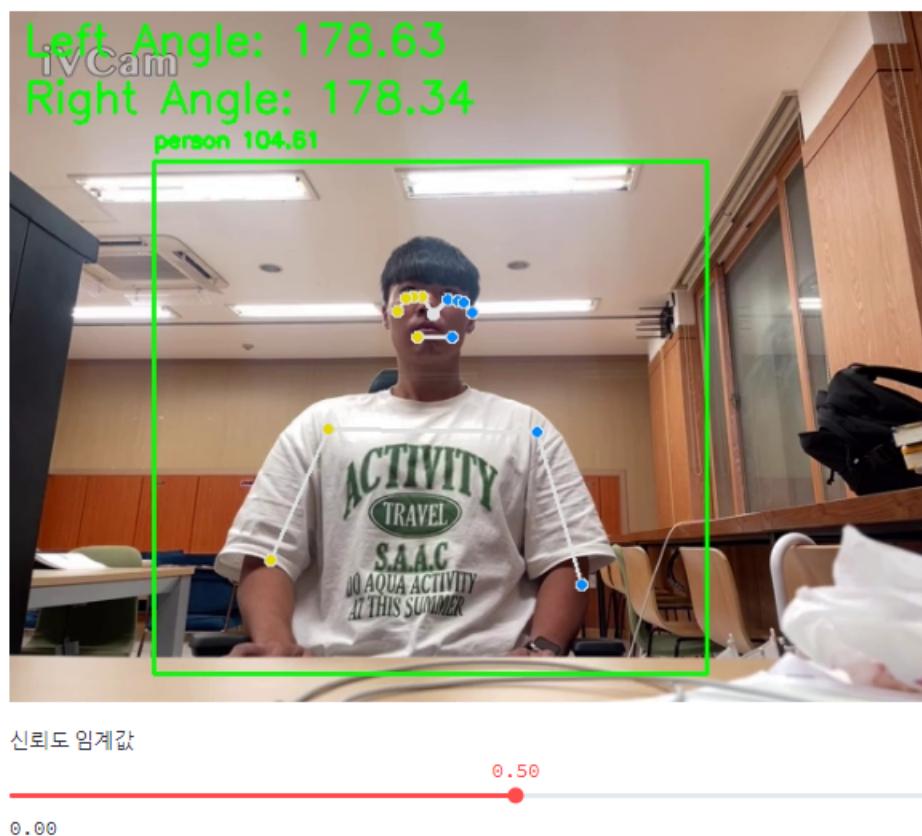
• 2주차



## II. 논문 및 개발 진척사항

### ii . 개발

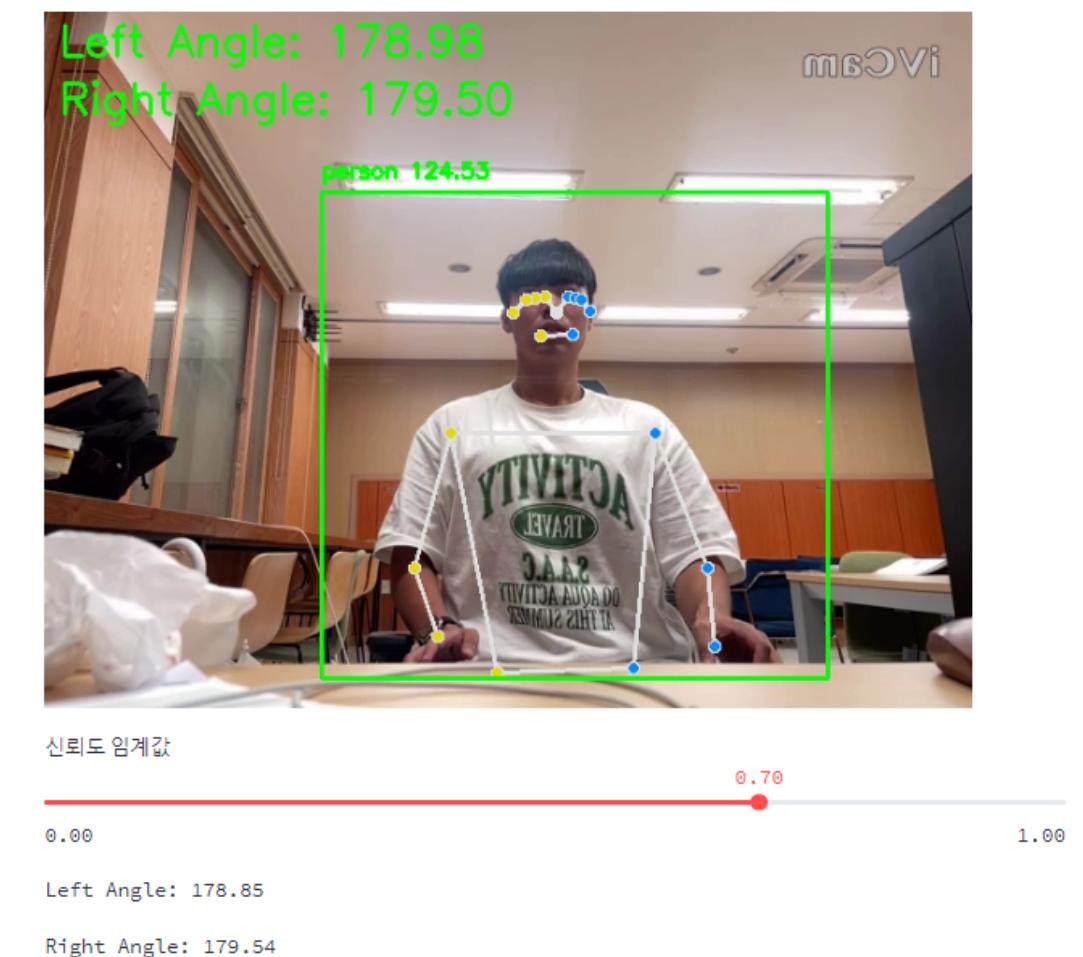
실시간 3대 운동 AI 자세 교정 서비스



실시간 3대 운동 AI 자세 교정 서비스

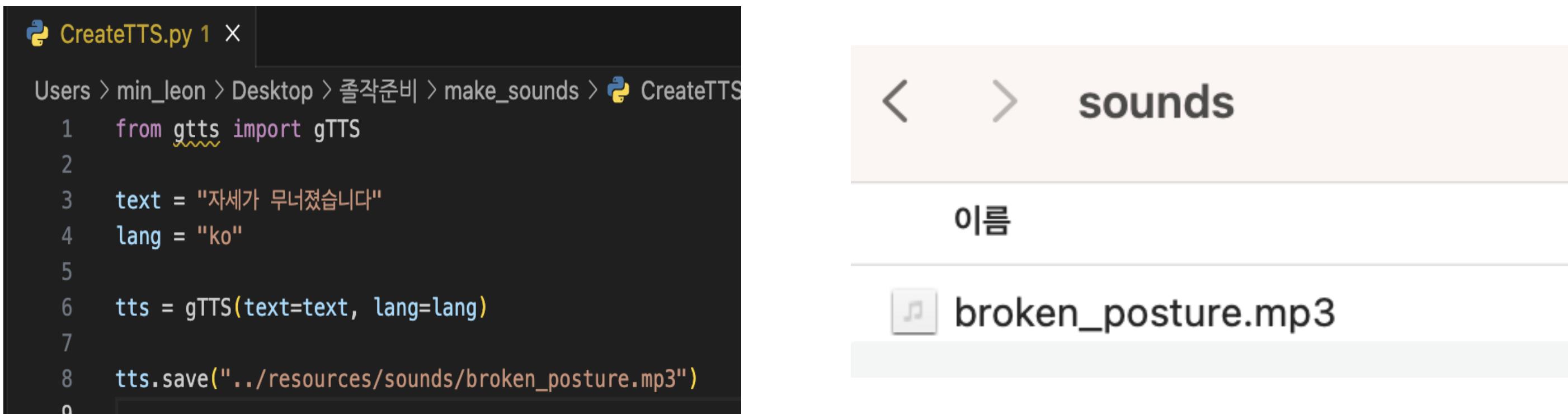


실시간 3대 운동 AI 자세 교정 서비스



## II. 논문 및 개발 진척사항

### ii . 개발



The terminal window shows the following Python script:

```
>CreateTTS.py 1 ×  
Users > min_leon > Desktop > 졸작준비 > make_sounds > CreateTTS.py  
1 from gtts import gTTS  
2  
3 text = "자세가 무너졌습니다"  
4 lang = "ko"  
5  
6 tts = gTTS(text=text, lang=lang)  
7  
8 tts.save("../resources/sounds/broken_posture.mp3")
```

The file browser shows a folder named "sounds" containing a file named "broken\_posture.mp3".

# II. 논문 및 개발 진척사항

## ii . 개발

☰ README.md

### AI\_Exercise\_Pose\_Feedback

#### Description

A Study on the big three exercises AI posture correction service Using YOLOv5 and MediaPipe  
Yeong-Min Ko

#### development environment

- OS: MAC m1 & Windows 11
- Frameworks & Libraries: YOLOv5, MediaPipe, OpenCV, Streamlit
- Device: iPhone 12 Pro(WebCam using iVCam)

#### Data

- YOLOv5: [Detect only one Person](#)
  - Scraping Images from Google & Roboflow
    - Bench Press
      - Google Search Keyword: Bench Press
      - [Faller Computer Vision Project](#) with lying down
    - Squat
      - [Squat-Depth Image Dataset](#)
      - [HumonBody1 Computer Vision Project](#) with Standing
    - Dead Lift
      - [SDT Image Dataset](#)
    - More(bending, lying, sitting, standing)
      - [Silhouettes of human posture](#)

- Exercise Posture Correction
  - Bench Press
  - Squat
  - Dead Lift

#### How to Use

- Open your terminal in mac, linux or your command prompt in Windows. Then, type "Streamlit run Streamlit.py".

#### Major project records

- 2023/09/10: The project detecting only one person using yolov5 was completed.
- 2023/09/11: As a result of using it in combination with mediapipe, the accuracy was lower than expected. Therefore, we plan to do labeling by adding extra space around people.
- 2023/09/16: The bounding box to train the model was significantly relabeled, and as a result, pose estimation was finally successful with high accuracy when using yolov5 and mediapipe together. And [the streamlit file](#) was impleted to estimate holistic pose after detecting only person closest to the camera using yolov5.

#### Project Progress

- Week1: Requirement Analysis
  - [Read More](#)
- Week2: Prototype Development & Mini Test
  - [Read More](#)
- Week3: Retrain the model detecting only person and Estimate holistic pose after detecting only person closest to the camera using yolov5
  - to be added.

YOLOv5 및 MediaPipe를 이용한 실시간 3대 운동 AI 트레이닝 서비스에 대한 연구

**THANK YOU**



컴퓨터공학과 3학년 고영민

2023.09.20