



INTELLIGENCE FITNESS MONITOR

Kavya P^a, Kosal Ram G^a, Sangavi S^a, Dr. J.B.Jona^b, Mr.S.A.Gunasekaran^c

^aMsc. Decision and Computing Sciences, Department of Computing, Coimbatore Institute of Technology, Coimbatore, India

^bAssociate Professor, Department Of Computer Applications, Coimbatore Institute of Technology, Coimbatore, India.

^cAssistant Professor, Department Of Computer Applications, Coimbatore Institute of Technology, Coimbatore, India.

ABSTRACT:

Exercise is very beneficial for personal health and fitness. Exercise mistakes are made when the user does not use the correct form, or puts the current fitness work focused on promoting dietary supplements, exercise access, exercise clothing does an excellent job in improving fitness. Meanwhile, the current situation makes it difficult to monitor the training of a trainee. Encouraged by the development of health care such as robot surgery, automatic integration of resilience is the motto and is achieved through computer vision and Neural Networks. This paper introduces AI fitness monitoring consisting of various modules that help control stamina and focuses on a self-paced exercise monitoring system that incorporates a recurring exercise calculation procedure - Bicep curl, Squats and Pull-ups, in practice. Using a media pipe to track a person's movements, measuring a person's yoga posture using a simple dense pose classifier, Fitness chatbot using LSTM and a BMI calculator from a person's face using MCTNN and transfer learning. Absolutely this module helps people monitor their fitness without the help of personal trainers

Keywords: Fitness monitor, BMI, Workout counter, media pipe, Chatbot, Pose estimation classifier.

1. Introduction:

In these unprecedented times we should be at home without much contact around the world. Staying busy while staying home is a daunting task. Practicing exercise or yoga can improve overall health. The results will be exponential with the right guidance or mentor. However, the current situation makes it impossible to direct physical activity with a video call as the perfect position is not visible in the 2D frame. The latest technology is fit, which can enhance the effects of exercise in less time. The technology can be applied to fitness such as gym applications, wearables, video conferencing and mentors. They analyze from data collected in operating systems, fitness equipment and improved methods can produce better results. All available applications play as a support for exercise, improving post-workout exercise. This is the maximum limit for applications available based on eligibility. Exercising at home is good but without a trainer, standing, breathing during an outing can be bad. There are many side effects such as body aches, muscle aches due to increased muscle stress, poor posture and more. There are many ways to eliminate this problem in the mobile app as well as do live teaching, to explain the best shape in detail. Live training is possible, time flexibility can also be achieved. In live training individual care should be taken while exercising or yoga but it can be expensive and a lot of time is required in individual care. We suggest the feature of providing exercise guidance in real time. Video conference guidance can be done with a lot of exercise. This is a very effective method. But it is not suitable for the whole body. Some of the tests available on social media can be done without training but proper posture should be maintained while performing the task. It is not always possible to advise when training. Real-time suggestions are needed to improve the moment and get a great result. We recommend a Fitness instructor, an app to guide people when they do exercise and monitor overall fitness.

2. Related Works

2.1. Fitness Application

Exercise can have a positive effect on the body and is permanent. Exercise not only improves a physical aspect but also is part of the mental health of a person. Exercising offers a great deal of relief from stress. It increases productivity and sounds fresh in everyday activities. It increases intelligence, mental health, life expectancy and also reduces health problems. The study by Steinberg et al tells us that exercise improves positive mood and reduces negative moods. Existing fitness apps can help with weight loss, nutrition and diet, relaxation and general exercise knowledge. Studies suggest that participating in high-quality sports often results in a longer lifespan. Regular exercise can reduce almost all physical problems and give you the strength to deal with the world mentally and physically. Deep learning for Health Researches in the healthcare field using AI are booming. Naïve takes a variety of forms such as diagnostics, classification, analysis of pre-existing medical data, integrating health care with robotics using deep learning that opens the way to saving lives. Deep learning is also used in fitness programs such as fitbit, smart watches and mobile work monitoring apps. This data can help

predict specific suggestions for achieving their goals. We take the concept of using deep learning in fitness to make a real impact on lives.

3.Dataset

Yoga Pose Classifier

The yoga poses dataset has been collected from Kaggle dataset consists of 3 yoga asanas – mountain pose, warrior pose and tree. Since this is a baseline approach to build a pose classifier, only 3 poses have been taken. There are 250 samples per class.



Fig 1 Yoga Poses

Chat Bot

Dataset consists of questions and responses. And these data are manually created by us in order to give customized trainer suggestions. For the Fitness trainer chatbot the model has been focused on 12 categories like Beginners, biceps, BMI, diet, goodbye, greeting, payments, pushups, thanks, trainer, weight, workout

```
{
  {
    "tag": "diet",
    "patterns": [
      "Can you share a diet plan ?",
      "Shall I cheat my diet once in a week?",
      "Can You suggest a diet plan which includes chocolates?"
    ],
    "responses": [
      "A high-protein, low-carbohydrate diet is the most effective diet plan",
      "There is no specific guideline for when or how frequently your cheat meal or day should occur"
    ]
  },
  {
    "tag": "payments",
    "patterns": [
      "Do you take credit cards?",
      "Do you accept Mastercard?",
      "Can I pay with Paypal?",
      "Are you cash only?"
    ],
    "responses": [
      "We accept VISA, Mastercard and Paypal",
      "We accept most major credit cards, and Paypal"
    ]
  },
  {
    "tag": "workout",
    "patterns": [
      "Do i need to workout every day",
      "How long should i work out for?"
    ],
    "responses": [
      "No, you do not need to work out every day.",
      "your workouts do not need to be long to be effective"
    ]
  },
  {
    "tag": "biceps",
    "patterns": [
      "What is the best biceps exercise for beginners",
      "How to build biceps at home",
      "Suggest any easy Biceps workout"
    ],
    "responses": [
      "you have to do exercises like Wall Supported Dumbbell Curl, Hammer Curl, Concentration curl, Walter Curl",
      "Dumbbell curls alternate 3 sets 10-15 reps"
    ]
  }
}
```

Fig 2 Chatbot Query dataset

BMI Calculator

The data used for training was taken from Kaggle. This dataset comprises 1530 records consisting of Facial images, sex, race. Gender and age.

	age	race	sex	bmi	index	
 mma_watson	0	54	Black	Male	39.047808	248038.jpg
 hilary_clinton	1	26	Black	Male	29.285856	250464.jpg
 kent_zheng	2	24	White	Male	42.820836	263711.jpg
 kim_jong-un	3	21	Black	Male	23.025951	270020.jpg
 leonardo	4	18	White	Male	19.004109	279426.jpg

Fig 3 BMI Calculator dataset

4.Pre – processing and Feature extraction

Workout Counter

Training a model will limit the possibility of dynamically counting the repetitions of a workout.

In case of using computer vision the key points are detected using the posenet model which also tracks them simultaneously. Using these key points, the angle between the key points is calculated. This approach is comparatively faster and efficient as it doesn't depend on the trained images.

MediaPipe Pose is a ML solution for high-constancy body pose tracking, inferring 33 3-D landmarks and heritage segmentation masks at the entire body from RGB video frames making use of our BlazePose research that still powers the ML Kit Pose Detection API. Modern techniques are specifically depending on computer packages for discovery, at the same time as our technique achieves real-time overall performance on many cutting-edge phones, desktops / laptops, through python or even the web.

PoseNet is a Deep Learning TensorFlow model that allows you to measure and track a person's posture (known as "pose estimation") by finding body parts such as elbows, hips, wrists, knees, and ankles.

It uses the joints of these body parts to determine the shape of the body. Today, many industries are using this technology to improve their performance, as well as technologies such as augmented reality experiences, animation and games, and robotics. The emergence of human-like robots, visual game simulation, movement tracking, and body language interpretation can be done using these types of deep h learning models on PoseNet.

PoseNet is trained in MobileNet Architecture. MobileNet is a Convolutional neural network evolved through google that's skilled at the ImageNet dataset, majorly used for Image type in classes and goal estimation. It is a light-weight version which makes use of depthwise separable convolution to deepen the network and reduce parameters, computation cost, and improved accuracy. There are lots of articles associated with MobileNet that might be discovered on google. The pre-trained models run in our browsers, that's what differentiates posenet from different API-established libraries. Hence, each person with a confined configuration in a laptop/computing device can make use of it without problems employing such models and building top projects. The video is directly embedded by posenet in order to the angles and key points.

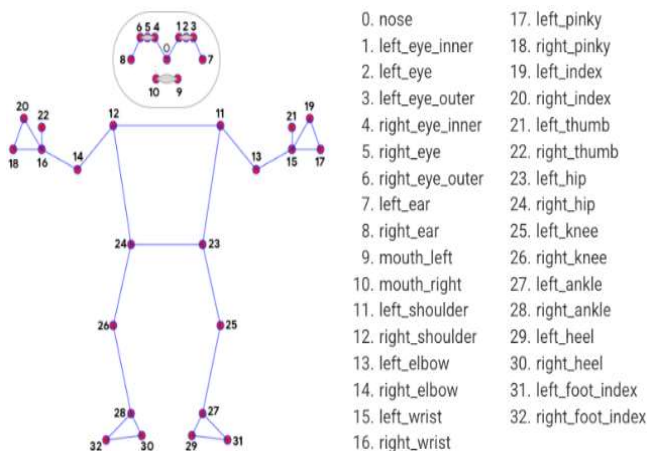


Fig 4 Mediapipe posenet

Yoga Pose Classifier

The yoga data set describing the image files will be uploaded to the python as a Data Frame and the images will be attached to their labels and unnecessary or poorly matched data will be deleted. Once the images are uploaded, they are pre-processed using the OpenCV library and prepared for model training and testing.

Images will be resized and reshaped, as the images become darker. Images will be transmitted as vectors of arrays as features to the model and the target will be transmitted as a list of label encoded columns. Once these steps are completed, the images are transmitted over a posenet that gives us a total of 33 key points that we can use, from eye to ear to knee and ankle and their links are saved as CSV.

Chat Bot

Preliminary processing of the data was performed prior to modeling. It is compulsory to clean the corpus with the help of Regular Expressions. Then different input sequence lists and targeted sequences will be created and it is necessary to create lists of unique tokens (input tokens and targeted tokens) in the dataset. For the target sequence, we will add '<START>' at the beginning of the sequence and '<END>' at the end of the sequence so that the model knows where to start and end the text generation.

There must be unique input tokens and targeted tokens in the dataset. Then the input includes a dictionary that will store input tokens, as it creates pairs of the key value, the word is key and the value is index. Similarly, in targeted tokens, a dictionary of targeted features is generated. The features of the dictionary will help to encode sentences into one-hot vectors. Contrasting sentences include a dictionary that keeps the index as a key and the word as the value.

The sequence to sequence predicts a word given in the user input and then each of the next words is predicted using the probability of likelihood of that word to occur. In building the Generative chatbot this approach for *text generation* given in the user input is used.

The encoder removes the last state (memory) vector which becomes the first decoder state. A method called teacher forcing is a decoder training that enables it to predict the following words in the target sequence given in the past words. States are transferred to the server in each decoder layer. 'Hi', 'how', 'are', and 'you' are called input tokens while 'I', 'am', and 'fine' are called targeted tokens. The probability of an 'am' token depends on past names and encoder states. By adding the '<END>' token to tell our decoder when it stopped.

BMI Calculator

The input preprocessing pipeline are as follows:

- load the image, resize to 224 x 224 and convert to array, which forms the features (X)
- map the labels (y: {BMI, Age, Gender}) from meta-data
- random sample from train and valid dataset to build the generator for model fitting

In order to Detect multiple faces MTCNN is used. Multi-Task Cascaded Convolutional Neural Network is a modern tool for face detection, leveraging a 3-stage neural network detector. It is a framework developed as a solution for both face detection and face alignment. The process consists of three stages of convolutional networks that are able to recognize faces and landmark locations such as eyes, nose, and mouth.

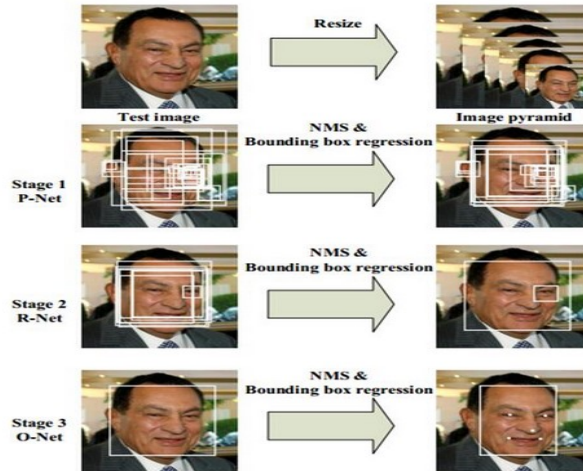


Fig 5 MTCNN

After detecting faces, the feature map has been extracted using transfer learning models for extracting all minute features from the image

5.Objective of the study

- To use a suitable Deep Learning Algorithm and bring automation on fitness monitoring module and provide beneficial insights.
- To track the human body and count the no of repetitions for the workouts
- To classify the yoga poses
- To Create a Chatbot that assist in gym maintenance
- To calculate BMI value using human Face

6.Model Implementation

Workout Counter

As said above there's no precise model concerned on this module alternatively by calculating the angle between the key factors counting system happens

- To count the repetitions, the algorithm monitors the chance of a target pose class.
- To count the repetitions of an exercise, the movement of the problem was analyzed frame by frame. And what was taken into consideration as a repetition was a change of state

Let's take sit-ups with its "up" and "down" terminal states:When the possibility of the "down" pose class passes a certain threshold for the first time, the algorithm marks that the "down" pose class is entered.Once the possibility drops under the threshold, the algorithm marks that the "down" pose class has been exited and increases the counter.To avoid instances while the probability fluctuates around the threshold (e.g., while the consumer pauses between "up" and "down" states) inflicting phantom counts, the threshold used to detect while the state is exited is actually barely lower than the one used to detect when the state is entered. It creates an interval in which the pose class and the counter can't be changed.

A biceps curl counter is primarily based on human pose estimation. The aim of this module is to make the person aware about inaccurate range of movement in real-time in order that he/she will make corrections in it at once without the guidance of a coach.Often as a novice, we're practicing with wrong workout postures which subsequently ends in no benefit or withinside the worst conditions, undesirable harm and tear to the muscles. The counter doesn't tick if the range of motion during a bicep curl is within 40 degrees and 160 degrees. Hence to get the counter going, the user has to curl and uncurl his/her biceps in a full range of motion. So no cheating now!

For push ups when the person moves from an "Up move" to a "Down move". This did not work immediately so a tolerance had to be built due to the accuracy of the model not being 100%. The current state of the person was then changed to a specific movement if X consecutive frames were

classified with that movement. This value of X can be changed exercise wise, and have found that push ups and pull ups work better with different X values.

Yoga Pose Classifier

After acquiring the key points for distinct yoga poses in csv, for preliminary validation and Baseline, KNN is used for training the gesture classifier. K-Nearest Neighbor is one of the most effective Machine Learning algorithms primarily based totally on Supervised Learning technique. Simple Dense Model is used as an improvised model for Yoga pos classifier

Sequential is the easiest way to build a model in Keras. It allows you to build a model layer by layer. Each layer has weights that correspond to the layer that follows it. Dense layer is the regular deeply connected neural network layer. It is the most common and frequently used layer. Dense layer does the below operation on the input and returns the output.

Output = activation (dot (input, kernel) + bias)

where,

- input represent the input data
- kernel represent the weight data
- dot represent numpy dot product of all input and its corresponding weights
- bias represent a biased value used in machine learning to optimize the model
- activation represents the activation function.

Advantage of this is that the features are automatically deduced and optimally tuned for desired outcome.

Chat Bot

LSTMs are explicitly designed to get rid of the long-term dependency problem. LSTMs additionally offer solutions to the Vanishing/Exploding Gradient problem. Remembering information for the long term is their default behavior. All recurrent neural networks have a series of repeating modules of the neural network. In RNNs, the repeating module may have a simple structure, inclusive of a single tanh layer

The concept is to outline some LSTM networks with versions in hyperparameters as a part of the ensemble version. The member models work collectively in parallel, and each of their individual outputs is aggregated to generate the output overall model. As a fine-tuning measure, the idea of Pruning has also been employed. Segmentation, Vector Space Model (VSM), Classification algorithm & Response the basic components of the chatbot of the chatbot. After the output class is predicted, the Chatbot returns it to the user's question.

The LSTM-based encoder-decoder structure has proven to be extra robust, cleaner, and quicker than an ANN version. This model additionally suggests promise as a generative-based model which previous generative approaches had been now no longer capable of showing. The sentence structure is dynamic albeit nevertheless basically meaningless as it is free from context whilst constructing sentences. But the largest benefit proven with the aid of using this model is the relative ease of constructing. All this model desires to work properly with a lot of information in contrast to older procedures which require consistent human intervention to monitor properly. This smooth technique is suited and beneficial for huge usages of chatterbots. The proposed LSTM architecture tries at improving the consumer level with the aid of supplying an experience of the continuance of context in a series of conversations. The algorithm does so with the aid of generalizing the features vital to creating the verbal exchange human

BMI Calculator

Face detection is accomplished with the aid of using MTCNN, that's capable of stumbling on more than one face inside a picture and drawing the bounding box for every face. Multi-task Cascaded Convolutional Networks (MTCNN) is a framework evolved as a solution for both the above mentioned tasks. After detecting faces from MCTNN, its modeling pipeline is as follows

To face prediction VGGFace is used and additionally an comparative study has been performed with RESNET50

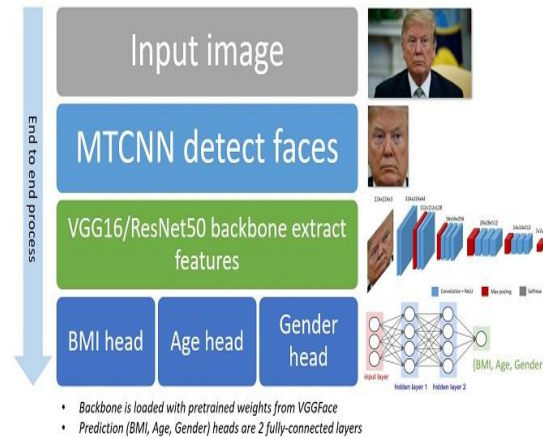


Fig 6 BMI Architecture

Face detection and Face alignment. The procedure includes 3 levels of convolutional networks which are capable of recognizing faces and landmark regions inclusive of eyes, nose, and mouth. It serves the following two functions for this project:

- Pre-process and align the facial features of the picture.
- Prior model training, every picture is pre-processed with the aid of using MTCNN to extract faces and crop pictures to recognize the facial part. The cropped pictures are stored and used to train the model in later parts.

Transfer learning from VGGFace, with VGG16 backbone has been done. Multi-task learning to three modules collectively permit prediction for more than one individual within the same picture. In the inference phase, faces may be detected from the input picture. For every face, it's going to undergo the same pre-processing and make the predictions.

In CNN architecture, convolutional blocks are followed by the simple dense layers to output the prediction. In a naïve implementation, we will construct three models to predict BMI, age and gender individually. However, there may be a sturdy disadvantage that three models are required to be taught and serialized separately, which appreciably will increase the upkeep efforts.

Since we're going to predict BMI, Age, Sex from the same picture, we will proportion the equal backbone for the 3 exclusive prediction heads and subsequently simplest one model have to be maintained.

Input image □ **VGG16** □ **Separate dense layers x 3** □ **weighted [BMI], [AGE], [GENDER]**

This is the maximum simplified multi-task learning structure, which assumes impartial tasks, and subsequently separate dense layers have been used for every head. Other studies, inclusive of Deep Relationship Networks, used matrix priors to model the connection among tasks.

7. Results and Discussion

Workout Counter

Using angle rather than a model was very robust and accurate that it gives clear counts of sit ups, pullups and bicep curls and the demo are as follows

Demo

SITUPS



Fig 7 Situps counter

PULL UPS



Fig 8 Pull - ups

Bicep Curl

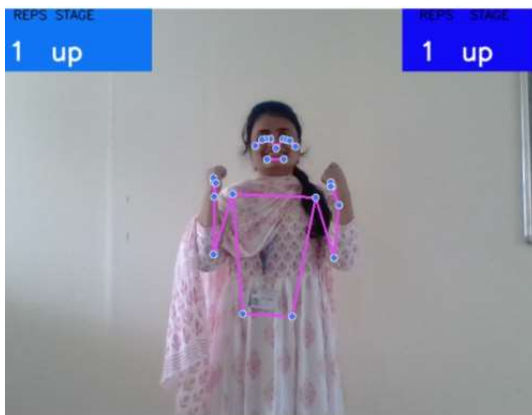


Fig 9 Bicep up



Fig 10 Bicep down

Yoga Pose Classifier

The Pose-net Keypoints with Simple Dense model gives an accuracy of about 97% and its confusion matrix is as follows:

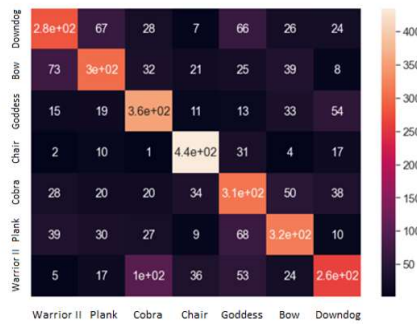


Fig 11 Confusion Matrix



Fig 12 Warrior pose

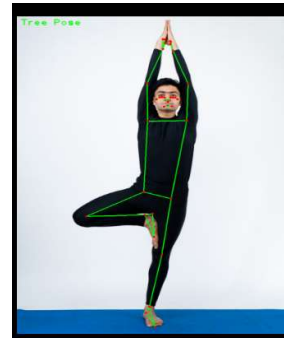


Fig 13 Tree pose

Chat Bot

The Gym chatbot is as follows, as for now we have initially added 12 tags, with this data it gives about an accuracy of 80% and the Demo chat with gymbot is as follows.

```

let's chat (from: user - to: bot)
bot: Hi
user: Hi there, what can i do for you?
bot: how to lose weight
bot: do workouts regularly and follow high protein diet
bot: suggest a diet plan
bot: just take more fruits, vegetables, brown rice, and chicken.
bot: need more details
bot: you have to do exercises like wall supported handball curl,hammer curl,concentration curl,seiter curl
bot: suggest an exercise plan for beginner
bot: As a beginner, initially for warming up the body, do the following exercise for a week, later specific workout for parts can be started: Jumping jacks: 2*8 reps,walking lunges
bot: i like to do that
bot: i do not understand...
    
```

Fig 14 Chatbot dialogues

Further it can be built robust by adding more data with custom options.

BMI Calculator

The model VGG 16 model outperforms resnet50 and gives a lesser RMSE value for all the 3 modules. And accuracy of about 89%

Model	BMI (RMSE)	Age (RMSE)	Sex (AUC)
vgg16	4.56	5.66	0.99
vgg16_fc6	4.99	6.04	0.99
resnet50	5.21	7.02	0.99

Fig 15 BMI Comparitive Model



Fig 16 BMI, Age, Sex Detector

8. Conclusion And Future Work

Nowadays lifestyles are turning into busier and we hardly ever discover time in our schedules to be healthful and healthy and exercising daily. This has brought about many sicknesses and fitness issues. Implementation of Deep learning within the area of health can remedy many problems. The fitness associated programs and gadgets are making our lives less complicated and eases our health journey. Individuals can use this of their very own workouts, for this reason making them greater efficient.

There is lots of scope of improvement on this assignment. The assignment may be upgraded to assist greater physical activities. A User interface may be brought for smooth navigation thru the physical activities. The instructor will recommend an exercise session plan and its depth in line with the frame kind and weight. This may be evolved right into an entire android/ios software for ease of use.

From the short perception supplied above, it suggests that exercise assistant and health manual makes use of a few ideas of pose estimation, calls for a digital digicam to seize the frame pose as enter to the gadget generated and with the assist of pose estimator, will offer the stats of exercising remember as output in human-readable form.

Future paintings can also additionally encompass the motion of the digital digicam vertically and horizontally to seize some other huge style of physical activities or it can encompass using more than one camera to seize the frame pose from diverse angles so one can feed the template of different physical activities.

9. Scope

- It can estimate the poses and count repetitions for a limited number of exercises as pose estimation using computer vision for some exercises and postures can be difficult.
- The can be developed as a cross-web application and can be used as a mobile android/ios application.

10. Limitation

This project has been developed as a baseline and it contains only a few workouts. In order to make the module generalized for all workouts, all other exercises have to be integrated to this module

REFERENCES:

1. Robust 3d hand pose estimation in single depth images: from single-view CNN to multi-view CNNs by L.Ge, H.Liang, J.Yuan, and D.Thalmann. IEEE conference on computer vision and pattern recognition, 2016
2. Robust articulated-icp for real-time hand tracking by A.Tagliasacchi, M.Schroder, A.Tkach, S.Bouaziz, M.Botsch, and M.Pauly. In Computer Graphics Forum, volume 34 Wiley Online Library, 2015
3. Martinez-Martin, E.; Costa, A.; Cazorla, M. PHAROS 2.0-A PHysical Assistant RObot System Improved. Sensors
4. Sulikowski, P.; Zdziebko, T. Deep learning-enhanced framework for performance evaluation of a recommending interface with varied recommendation position and intensity based on eye-tracking equipment data processing. Electronics
5. <https://google.github.io/mediapipe/solutions/pose.html>