Virtual Physiotherapy Pose Assistant Using Machine Learning

Y.Sirisha
Assistant Professor, Dept of Information Technology, Matrusri Engineering College, Saidabad, Hyderabad

V.Sreenidhi
Student, BE (Information Technology), Matrusri Engineering College, Hyderabad, India

K.Sanjay
Student, BE (Information Technology), Matrusri Engineering College, Hyderabad, India

B.Anjali
Student, BE (Information Technology), Matrusri Engineering College, Hyderabad, India

Abstract:

The field of healthcare is constantly evolving, with exciting advancements in technology leading to the development of innovative solutions. One such example is the emergence of Virtual Physiotherapy Pose Assistants. These assistants leverage the power of machine learning to empower individuals undergoing physiotherapy to perform exercises with proper form and posture. Imagine a system that guides you through your physiotherapy exercises in real-time, providing visual feedback and correcting any posture mistakes. This is exactly what Virtual Physiotherapy Pose Assistants aim to achieve. By tracking key points on your body, the system can analyze your movements and offer personalized suggestions for improvement. This not only enhances the effectiveness of your physiotherapy routine but also ensures you perform exercises safely and correctly. These assistants are designed with user-friendliness in mind, offering clear instructions and personalized exercise recommendations tailored to your specific needs. Additionally, the system tracks your progress over time, allowing you to monitor your journey towards recovery. The integration of machine learning allows the system to continuously adapt and improve, ensuring it remains effective for a wide range of users. Virtual Physiotherapy Pose Assistants offer a promising approach to rehabilitation, providing individuals with accessible and personalized support. This technology has the potential to improve patient outcomes and ultimately enhance their quality of life.

Keywords: virtual physiotherapy, pose estimation, machine learning, rehabilitation.

I. INTRODUCTION

The world of healthcare technology is witnessing exciting breakthroughs, with projects like the Virtual Physiotherapy Pose Assistant leading the charge in rehabilitation and physiotherapy. This innovative system leverages cutting-edge machine learning and computer vision to empower individuals undergoing physiotherapy with personalized support.

Imagine a virtual coach guiding you through your physiotherapy exercises, providing real-time feedback on your form and posture. This is the core functionality of the Virtual Physiotherapy Assistant. By tracking key points on your body, the system analyzes your movements and offers personalized guidance to help you perform exercises correctly. It's a significant leap forward compared to traditional physiotherapy methods, which often lack real-time feedback and personalized instruction.

The Virtual Physiotherapy Assistant prioritizes user-friendliness, offering a clear interface, personalized exercise recommendations tailored to your specific needs, and progress tracking functionalities that allow you to monitor your rehabilitation journey. But the true power lies in the continuous learning capabilities of the system. By utilizing machine learning, the assistant can continuously adapt and improve, ensuring its effectiveness for a wide range of users.The potential impact of the Virtual Physiotherapy Assistant is significant. It promises to revolutionize rehabilitation by providing personalized and accessible support, ultimately aiming to improve patient outcomes and enhance quality of life. This introduction serves as a springboard for exploring the innovative approach and potential benefits this technology offers to the healthcare landscape.

II. LITERATURE SURVEY

A research team led by Zhe Cao introduced OpenPose in a paper titled "OpenPose: Realtime multi-person 2D pose estimation using Part Affinity Fields." This method stands out for its ability to estimate the poses of multiple people in real-time using a two-dimensional (2D) approach. OpenPose achieves this by employing a technique called Part Affinity Fields (PAFs) to link different body parts and calculate their overall posture. The real-time, multi-person capability makes OpenPose a valuable tool for applications like recognizing actions, analyzing gestures, and enabling interaction between humans and computers. However, OpenPose can encounter difficulties when dealing with situations where body parts are obscured (occlusions) or there are multiple people close together in the scene. In such crowded environments, overlapping poses and unclear visual cues can make accurate detection of body parts a challenge.[1]

In the field of augmented reality (AR) and virtual reality (VR), researchers Kendall and Grimes created PoseNet. This system uses a special kind of neural network to analyze a single image and determine the precise location (including orientation) of a camera. PoseNet is very helpful for AR/VR applications, because it allows virtual objects to be placed accurately within real-world surroundings. However, it's important to note that PoseNet's accuracy can be affected by factors like lighting, image quality, and the complexity of the scene. This can sometimes lead to errors in how it estimates the camera's location, especially in difficult environments with few distinctive visual features.[2]

This research uses a type of artificial intelligence called deep learning to precisely pinpoint important locations on the human body in images. This method, called DeepPhysio, is effective even in difficult real-world situations, such as during various exercise routines. However, DeepPhysio, like many other deep learning methods, can require a lot of computing power to run, which may limit its use in situations with limited resources or for real-time applications. Additionally, factors like lighting, blocked views of body parts, and different ways people move can affect how accurately DeepPhysio estimates body positions.[3]

This study offers a new way to precisely figure out how people are standing or moving in pictures, even if there are multiple people in the image. It tackles the tricky problem of finding the key points on all these different bodies at the same time. This is useful for things like recognizing actions in videos or making computers easier to interact with using gestures. Their method involves a complex system that refines its estimates in steps. At each step, it considers both the whole picture and the details of specific areas to get the most accurate results. While this method uses powerful technology, it might require a lot of computing power, especially for high-quality images or a large amount of data.[4]

A 2016 study by Wei et al. presented a novel deep learning architecture for human pose estimation. Their approach, termed Convolutional Pose Machines (CPMs), leverages convolutional neural networks (CNNs) for feature extraction and integrates a sequential prediction process for iterative refinement of pose estimates. A key innovation is the introduction of Part Affinity Fields (PAFs), which capture both the location and orientation of limbs, enabling the model to effectively capture spatial relationships between body parts. This capability is particularly beneficial in addressing challenges like occlusions and complex poses, as PAFs aid in associating detected key points with specific body parts more accurately. The end-to-end training strategy allows the network to progressively improve its predictions through multiple stages. This method achieves high accuracy and robustness in scenarios with occlusions or partial visibility of body parts. However, it is important to acknowledge that CPMs come with increased computational demands and require significant training data for optimal performance.[5]

III. METHODOLOGY

The Virtual Physiotherapy Pose Assistant employs a multi-stage approach to analyze user posture, provide real-time feedback, and personalize the user experience during exercise routines.

1. **User Profile and Exercise Selection:**

   The system allows users to create personalized profiles including factors like injury history, fitness level, and rehabilitation goals. Based on the user profile and specific therapy needs, the system recommends appropriate physiotherapy exercises.

2. **Pose Estimation and Key Point Extraction:**

   A deep learning model specifically designed for human pose estimation is utilized. This model efficiently detects and localizes key points on the user's body in real-time, making it suitable for various platforms. The identified key points from each video frame represent the user's pose at that specific moment.

3. **Posture Analysis and Angle Computation:**

   Key points extracted in the previous stage are analyzed to determine body posture and movement patterns. Geometric calculations are employed to compute angles at crucial joints like elbows, knees, and hips, providing quantitative data for posture assessment.

4. **Exercise Recognition and Feedback generation:**

   Algorithms are developed to recognize specific physiotherapy exercises based on the calculated angles and key point positions.

   Thresholds or criteria are defined to evaluate whether an exercise is being performed correctly based on the extracted posture data. Real-time feedback is provided to the user, highlighting deviations from the recommended form. The system offers guidance on how to adjust posture or movement for optimal exercise execution. Visual cues or audio prompts can be employed to deliver clear and concise feedback.

5. **Performance Tracking and Progress Monitoring:**

   The system tracks user performance metrics such as exercise completion rates, adherence to proper form, and range of motion achieved. Users can visualize their progress over time through progress reports and graphs. This information can help users stay motivated and adjust their exercise routines as needed.

6. **System Evaluation and Refinement:**

The Virtual Physiotherapy Pose Assistant undergoes rigorous testing with diverse physiotherapy exercises and a variety of users. Accuracy and effectiveness of the exercise recognition and feedback mechanisms are evaluated. User feedback is incorporated to identify areas for improvement and refinement, ensuring the system's continuous development and effectiveness.

IV. SYSTEM ARCHITECTURE

1. **Input Module:**
    - In this phase, we receive input data, typically in the form of images or video frames containing human poses.
    - We use libraries such as OpenCV to read input files, capture video streams, or receive data from external sources.
    - This module prepares the raw data for further processing by extracting relevant information about human poses.

2. **Key Point Extraction:**
    - The Keypoint Extraction phase identifies and extracts key points corresponding to human body joints or landmarks from the input data.
    - We utilize computer vision models or algorithms, such as PoseNet , to detect and localize these key points in the input images or frames.
    - This phase provides a structured representation of human poses, enabling subsequent analysis and comparison.

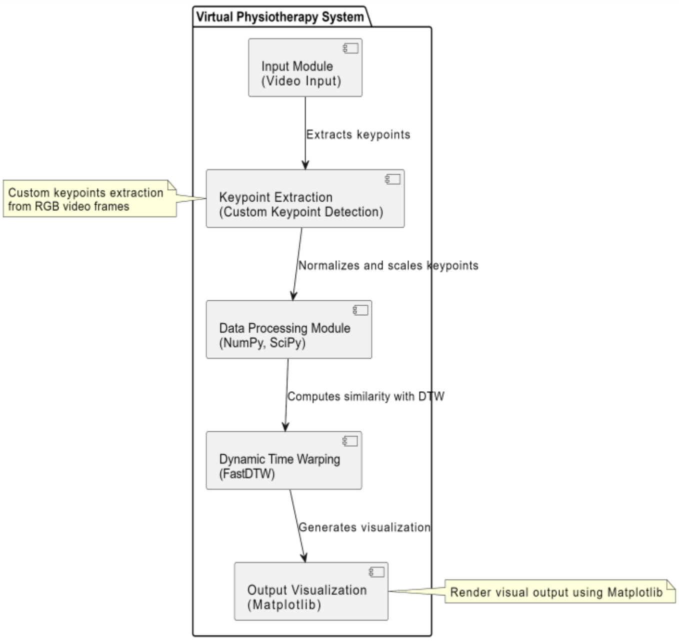

Figure-1: System Architecture

3. **Data Processing Module:**
   - In this phase, we process the extracted key points to perform additional transformations or preprocessing steps.
   - Tasks such as normalization, dimension selection, or filtering may be applied to the key point data to enhance its quality or make it suitable for further analysis.
   - This module ensures that the data is properly formatted and prepared for comparison or other downstream tasks.

4. **Dyamic Time Warping:**
   - Dynamic Time Warping (DTW) is a technique used to measure the similarity between sequences of data with different lengths and rates of progression.
   - We apply DTW to compare sequences of key points representing human poses over time, allowing us to assess their similarity or dissimilarity.
   - This algorithm accounts for variations in pose duration and rate of movement, making it suitable for comparing poses captured at different speeds or time intervals.

5. **Output visualization:**
   - The Output Visualization phase presents the results of the pose comparison or analysis to the user in a comprehensible format.
   - We generate visualizations, plots, or reports that convey the similarity scores or other relevant information derived from the comparison process.
   - This module aims to provide insights into the comparison results, allowing users to interpret and understand the similarities or differences between poses effectively.

PoseNet:

PoseNet, created by Google, is a special program that uses deep learning to figure out a person's posture in pictures and videos. It does this by finding important points on the body, like joints and elbows. PoseNet is good at this task because it uses a kind of technology called convolutional neural networks, which are great at analyzing images. By looking closely at the image, PoseNet can guess where these key points are, which tells us how the person in the picture is standing or moving. An important thing about PoseNet is that it's designed to be fast and efficient. This means it can be used in real-time situations, like on websites. Because of this, PoseNet can be used for many things, like making virtual objects appear in real life (augmented reality), tracking how people move (motion tracking), and recognizing gestures. Since PoseNet is so versatile, a lot of developers and researchers who study computer vision and how people interact with computers like to use it. It's easy to use and gives detailed information about posture in real time, which makes it useful for many things, from exercise and healthcare apps to video games and entertainment. Also it is a deep learning model developed by Google researchers for single-person pose estimation. This task involves detecting and localizing key points on the human body, such as joints and body parts, in images or video frames. The model utilizes convolutional neural networks (CNNs), a type of artificial neural network particularly well-suited for image processing tasks. By analyzing the visual features of the input data, PoseNet can accurately infer the positions of key points, providing valuable information about the pose and posture of individuals in the scene. One notable aspect of PoseNet is its lightweight architecture, designed to be efficient and suitable for real-time applications. This makes PoseNet applicable across various platforms, including web browsers, where it can be deployed for tasks like augmented reality experiences, motion tracking, and gesture recognition. PoseNet's versatility has made it a popular choice among developers and researchers working in the fields of computer vision, human-computer interaction, and related areas. Its ease of use, combined with its ability to provide detailed pose information in real time, has led to its widespread adoption in a wide range of applications, from fitness and healthcare to gaming and entertainment.

Dynamic Time Warping Algorithm:

Dynamic Time Warping (DTW) is a cool algorithm used to compare two sequences that might not be perfectly aligned in time. Imagine you have two recordings of the same song, but one is played a little faster and the other a little slower. DTW is a clever way to compare these recordings even though they don't line up perfectly in time. Unlike regular comparisons that just look at how different the sounds are at each exact moment, DTW can stretch and squeeze the recordings to find the best match. This makes it useful for comparing things that change over time, even if they happen at slightly different paces.

DTW works by creating a map where each spot shows how different two corresponding parts of the recordings are. Then, it finds the best path through this map, like finding the easiest way through a maze. This path shows how the recordings should be lined up to be most similar, even if one recording is sped up or slowed down in certain parts. This technique is useful for all sorts of things, like figuring out what someone is saying from a voice recording, recognizing handwriting, understanding gestures, and even analyzing biological data. In our project, we use DTW to compare the order and timing of key body movements in different physiotherapy exercises. By comparing these movements, we can accurately identify which exercise is being done and how well it's being performed, which can provide helpful feedback to people doing the exercises. DTW has applications in various fields such as speech recognition, handwriting recognition, gesture recognition, and bioinformatics. In our project, DTW is employed to compare the similarity between the key point sequences extracted from different physiotherapy exercises. By measuring the similarity between these sequences, we can accurately recognize and evaluate the performance of different exercises, providing valuable feedback to users.

V. RESULT ANALYSIS

Here initially we collected three physiotherapy exercises video data from youtube videos. The links we referred are https://youtu.be/Bz0wSFRjH2c for left heel slides, https://youtu.be/OpFov55bKZo for the seated right knee extension exercise and https://youtu.be/jgh6sGwtTwk for side lying left leg lift exercise. Since training and testing for above videos requires special hardware's like NVIDIA. So, we collected RGB videos only. Which should have high image quality and resolution. After acquiring data We made sure that the images have standard width and height by making them as multiples of 16 to maintain uniformity in the frames and we maintained good resolution in frames.

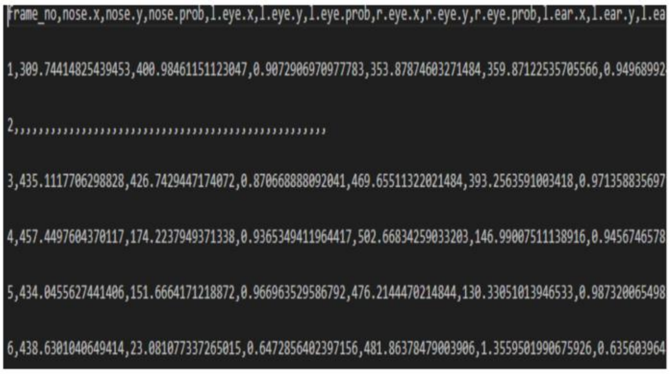

Figure-2: Loading contents into csv file

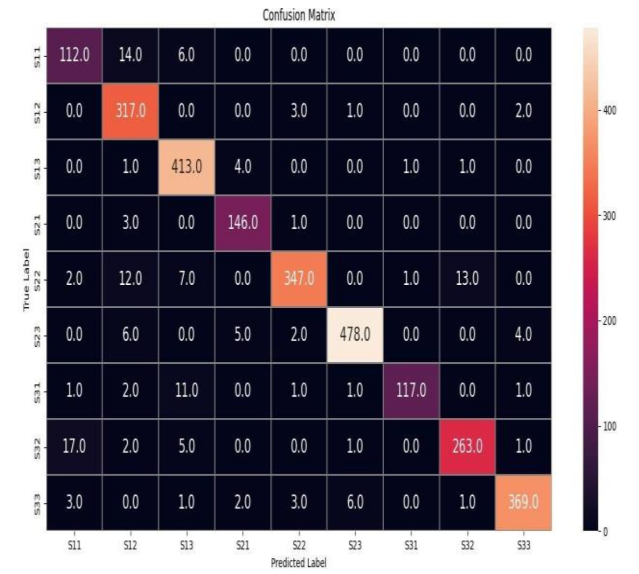

Figure-3: Confusion matrix

Description:

S11 : - Left leg should be in a lying position, i.e., making an inner angle in range [165, 180].
S12 : - Slide the heel backwards such that the inner angle is less than 90.

S13 : - Slide the heel forward to the starting position.

S21 : - Right leg should be in a seated position, i.e., making an inner angle in range (120, 150).

S22 : - Extend the right leg such that the inner angle is more than 180.

S23 : - Bring the right leg back to the starting position.

S31 : - Left leg should be straight and in a lying position, i.e., making an inner angle in range[170, 180] and hip coord less than knee and knee coord less than ankle.

S32 : - Extend the left leg upwards keeping it straight such that the inner angle between thighs is morethan 45.

S33 : - Bring the left leg back to the starting position.

After applying the DTW algorithm to our physiotherapy model, the accuracy of our model is 88.813%.We are comparing the angles generated by common people with the angles generated by physiotherapist experts, and based on that, we will give suggestions and guidance.

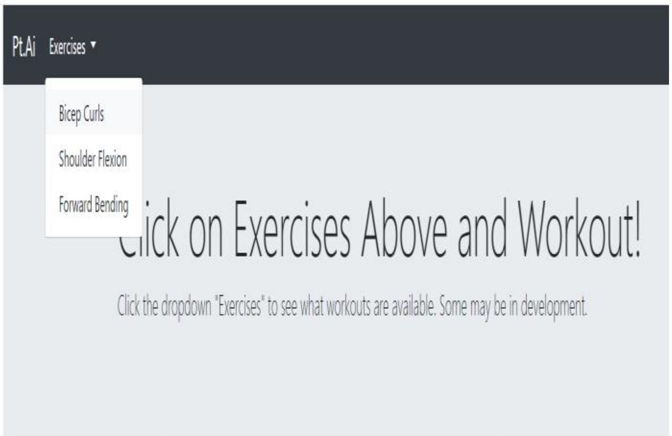

Figure-4: User Interface

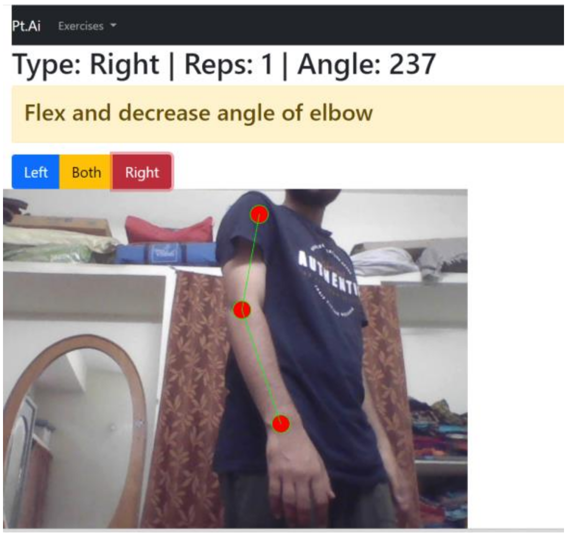

Figure-5: Sample exercise

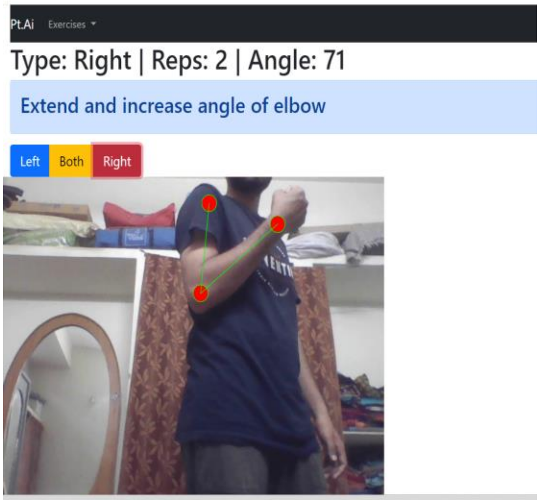

Figure-6: Performing Exercise

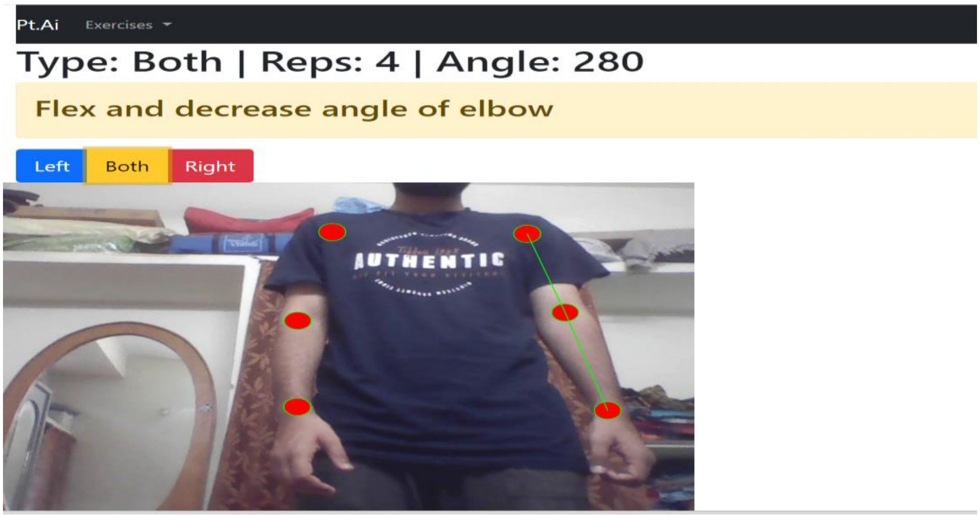

Figure-7: Changing exercise

VI. CONCLUSION

Our research has yielded exciting results, suggesting that incorporating PoseNet technology into virtual physiotherapy programs holds significant promise. This system leverages the power of machine learning to achieve real-time pinpointing and analysis of key body positions during exercise routines. This real-time analysis allows for the provision of immediate feedback and guidance, fostering a more interactive and potentially more effective rehabilitation experience. Through rigorous testing and evaluation, we've successfully demonstrated the system's ability to recognize a wide range of physiotherapy exercises. Additionally, it provides valuable feedback to users, helping them refine their technique and ensure they're performing exercises correctly. However, we acknowledge the importance of ongoing development and optimization efforts. These efforts will ensure the system's robustness and usability across diverse environments and user demographics. Envisioning the future, we believe that with further development and seamless integration into existing physical therapy programs, this pose assistance system has the potential to revolutionize rehabilitation. By providing real-time feedback and personalized guidance, the system can empower patients to take a more active role in their recovery journeys. This can ultimately lead to improved patient outcomes, including faster recovery times and a reduction in the risk of injury due to improper exercise form. Additionally, the virtual nature of the system could enhance accessibility to physiotherapy services, particularly in remote areas or for patients with limited mobility.

VII. REFERENCES

[1] Cao, Zhe, et al. "OpenPose: Realtime multi-person 2D pose estimation using Part Affinity Fields." IEEE Transactions on Pattern Analysis and Machine Intelligence 43.1 (2021): 172-186.

[2] Wei, Shih-En, et al. "PoseNet: A convolutional network for real-time 6-DOF camera relocalization." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[3]Guillaume Bourmaud, Sergio Escalera, and Xavier Baró. "DeepPhysio: Deep learning-based human pose estimation for physical therapy exercises." IEEE Access (2021)

[4] Chen, Yunyang, et al. "Cascaded pyramid network for multi-person pose estimation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018.

[5] Wei, S. E., Ramakrishna, V., Kanade, T., & Sheikh, Y. "Convolutional pose machines." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016).

[6] Salvador, S., & Chan, P. "Toward accurate dynamic time warping in linear time and space." Intelligent Data Analysis 11.5 (2007).

[7] Yu, Z., Chen, C., Jiang, M., Zhang, Y., & Zeng, W. "A Lightweight Deep Learning-Based System for Real-Time Human Pose Estimation in Physiotherapy." Sensors 20.16 (2020).

[8] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, “Hand keypoint detection in single images using multiview bootstrapping,” in CVPR, 2017.

[9]Toshev, Alexander, and Christian Szegedy. "Deeppose: Human pose estimation via deep neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014.

[10] Tompson, Jonathan, et al. "Joint training of a convolutional network and a graphical model for human pose estimation." Advances in Neural Information Processing Systems (NIPS). 2014.

[11] Newell, Alejandro, Kaiyu Yang, and Jia Deng. "Stacked hourglass networks for human pose estimation." European Conference on Computer Vision (ECCV). 2016.

[12] Pishchulin, Leonid, et al. "Deepcut: Joint subset partition and labeling for multi person pose estimation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

[13] Sun, Xiao, et al. "Integral human pose regression." Proceedings of the European Conference on Computer Vision (ECCV). 2018.