Performance Analysis of Human Activity

Rutuja Mhaiskar, *Vaithiyanathan* Dhandapani, *Preeti* Verma, *Baljit* Kaur

Department of Electronics and Communication Engineering, National Institute of Technology Delhi, Delhi – 110 036, India.
E-mail: 191220041@nitdelhi.ac.in, dvaithiyanathan@nitdelhi.ac.in, preetiverma@nitdelhi.ac.in, baljitkaur@nitdelhi.ac.in

Abstract. This project aims to develop an AI-powered gym assistant using Jupyter Notebook and MediaPipe, a popular computer vision library, to count the repetitions of three joint exercises: curls, squats, and sit-ups. The system will provide real-time feedback and monitoring, allowing users to track their progress and improve performance. The proposed method utilizes MediaPipe, which offers pre-trained machine-learning models for human pose estimation and hand tracking. These models will accurately detect and track critical body joints and hand movements during the exercises. The system will then analyze the detected poses to identify the repetitions of each activity based on predefined movement patterns and pose thresholds. Jupyter Notebook will be used as the development environment for coding and testing the system. Python programming language and MediaPipe's Python API will be employed to implement the pose estimation and repetition counting algorithms. The system will also incorporate a user-friendly interface, allowing users to interact with the gym assistant and receive feedback on their exercise performance. The completed project will provide an AI-powered gym assistant that can accurately count the repetitions of curls, squats, and sit-ups in real-time. Additionally, this project will contribute to the advancement of the field of fitness technology by showcasing the potential of combining computer vision and artificial intelligence techniques for gym monitoring and performance tracking. The results of this project have the potential to benefit fitness enthusiasts, trainers, and researchers alike, providing contributions to the field of fitness technology.

Keywords: Exercise monitoring, Healthcare, multimedia processing, Blazepose, Deep learning

1 Introduction

Physical activity is one of the essential parts of maintaining a healthy lifestyle, and many individuals engage in various exercises or workouts to improve their fitness levels. Monitoring human activity during exercise sessions and accurately counting the repetitions (reps) of exercises performed are essential for tracking progress, providing feedback, and optimizing workout routines. However, manually tracking human activity and counting reps can be time-consuming and error-prone. In recent years, AI and ML have rapidly advanced with the development of sophisticated algorithms and frameworks that can analyze and interpret human movements from visual data. One such framework that has gained significant attention in

computer vision is MediaPipe, an open-source, cross-platform library developed by Google that provides pre-built components for building multimedia processing pipelines. MediaPipe offers various pre-trained ML models and tools for tasks such as human pose estimation, hand tracking, and face detection, which can be combined to create powerful applications for human activity recognition. The goal is to leverage the capabilities of AI and ML using MediaPipe to enhance human activity recognition and repetition counting in the context of exercise routines. By analyzing visual data, such as video or image sequences, the project aims to develop a robust system that can accurately detect and recognize different human activities, such as push-ups, squats, or jumping jacks, and count the number of repetitions performed for each exercise.

The proposed idea aims to develop a real-time gym exercise rep counter using Mediapipe in a Jupyter Notebook. The reviewed papers suggest that CNN and LSTM-based models have been popular in this field. Furthermore, the proposed methods in the reviewed papers have led to significant advancements in the field of HAR. For instance, a two-stream CNN [1] outperformed traditional methods. Some thoroughly analyze deep learning-based [2] HAR approaches using multimodal sensor data. The recognition accuracy was increased in [3] using a stacked autoencoder network and multimodal sensor fusion technique. A multi-level fusion CNN for multimodal sensor data-based HAR was presented in work in [4] and outperformed more established techniques. Also, transfer learning-based strategies have been demonstrated to increase recognition accuracy in [5] and [6].

The work in [7] offered a unique deep-learning feature extraction technique that outperformed more conventional techniques in terms of recognition accuracy. The multi-channel CNN-LSTM model in [9] and the suggested time-frequency CNN technique in [8] increased recognition accuracy. The research in [10], [11], and [12] suggested brand-new HAR deep-learning approaches. For instance, the work in [10] proposed a CNN-LSTM-based technique that utilized wearable sensor data to obtain high identification accuracy. Using in-built sensors, the smartwatch-based technique suggested in [11] obtained high identification accuracy. Finally, the two-stage deep learning-based strategy suggested in [12] outperformed conventional methods in terms of recognition accuracy. The significance of multimodal sensor data fusion and transfer learning-based methods has been emphasized for enhancing recognition accuracy. The authors have provided a variety of methods and strategies that may be applied to create a real-time rep counter for gym exercises. This evaluation gives the suggested system a solid foundation and places it within the context of more general studies.

2 Real-time Human Pose Estimation

Human pose estimation is a task in various fields, including fitness tracking, sports coaching, virtual reality, and healthcare. Accurate pose estimation enables real-time feedback, improved performance tracking, injury prevention, and natural interactions in virtual environments. However, traditional pose estimation methods can be computationally expensive, limiting their real-time performance. To address this, there is a need for efficient and accurate pose estimation solutions that can operate in real-time on resource-constrained devices such as smartphones or embedded systems. This research aims to harness the power of AI and machine learning with MediaPipe, an open-source framework developed by Google that combines multimedia processing and machine learning to enable a wide range of applications. By leveraging visual data and sophisticated ML algorithms, the system has the potential to revolutionize how human activity is tracked and analyzed, leading to improved fitness monitoring, rehabilitation, and sports training.

2.1 Multimedia Processing

Multimedia processing is a critical area that involves tasks such as video and audio analysis, image recognition, and augmented reality, among others. With the proliferation of multimedia content and the increasing demand for immersive experiences, there is a growing need for advanced multimedia processing techniques. Traditional methods often rely on handcrafted features or rule-based algorithms, which may not be suitable for the complexity and diversity of multimedia data. Machine learning, particularly deep learning, has shown remarkable success in various multimedia tasks, including image recognition, video analysis, and audio processing. However, implementing and deploying machine learning models for multimedia processing can be complex and challenging, requiring expertise in machine learning and multimedia processing domains. This research project aims to leverage the power of AI and ML with MediaPipe to enhance human activity recognition and repetition counting in the context of exercise routines. The outcomes of this research could have significant implications for the fields of AI, ML, computer vision, and human-computer interaction and contribute to the development of innovative applications for human activity recognition and rep counting.

2.2 Media Pipe

MediaPipeMediaPipe is a powerful library developed by Google for building multimedia processing pipelines. The library offers a range of pre-built components that can be easily combined to perform complex multimedia processing tasks. This modular architecture makes it accessible to developers with varying levels of expertise in machine learning and multimedia processing. MediaPipe also emphasizes efficiency and real-time performance, using techniques like model quantization and hardware acceleration to optimize machine learning models for low-latency processing on various platforms, including desktop, mobile, and embedded systems. One of the most significant advantages of MediaPipe is its flexibility and customization capabilities. The library provides pre-built components for tasks like object detection, pose estimation, and face recognition, but developers can easily replace or fine-tune these components to tailor the framework to their specific needs. This flexibility makes MediaPipe suitable for various multimedia processing applications, from video streaming to augmented reality and virtual communication. Additionally, the open-source nature of MediaPipe and its active community of developers and researchers have driven collaboration and innovation in the field of multimedia processing, leading to advancements and pushing the boundaries of what's possible with machine learning in multimedia applications. Overall, MediaPipe is a versatile and powerful tool for multimedia processing that makes machine learning accessible to a wider range of developers. Its modular architecture, focus on efficiency, and flexibility make it suitable for various multimedia processing applications, while its open-source nature fosters collaboration and innovation in the field. As the library continues to evolve and improve, MediaPipe will likely play an increasingly important role in the development of multimedia applications and the advancement of machine learning techniques in the field.

2.3 Blazepose

Traditional pose estimation approaches often rely on large and complex models with numerous parameters, which can be computationally expensive and challenging to deploy on resource-constrained devices. BlazePose, on the other hand, was designed with a compact body part encoding scheme that allows for accurate pose estimation with fewer parameters, making it highly efficient for real-time processing on devices with limited computational resources. Real-time performance was another critical motivation behind BlazePose. The team recognized the

need for a solution to provide low latency and high frame rates for real-time applications, such as fitness tracking or virtual reality experiences. To achieve this, BlazePose incorporates a highly optimized inference pipeline that takes advantage of the parallel processing capabilities of modern GPUs, enabling efficient and real-time pose estimation. Flexibility and customization were also significant motivations behind BlazePose. The team aimed to create a solution that could quickly adapt to different body shapes, clothing, lighting conditions, and application requirements. BlazePose allows developers to fine-tune the model for specific use cases, making it highly customizable and adaptable to different applications and environments. BlazePose has revolutionized human pose estimation by providing a highly accurate, efficient, and real-time solution that can be easily integrated into multimedia applications. Its compact body part encoding scheme, optimized inference pipeline, and customization capabilities make it a cutting-edge solution for various domains.

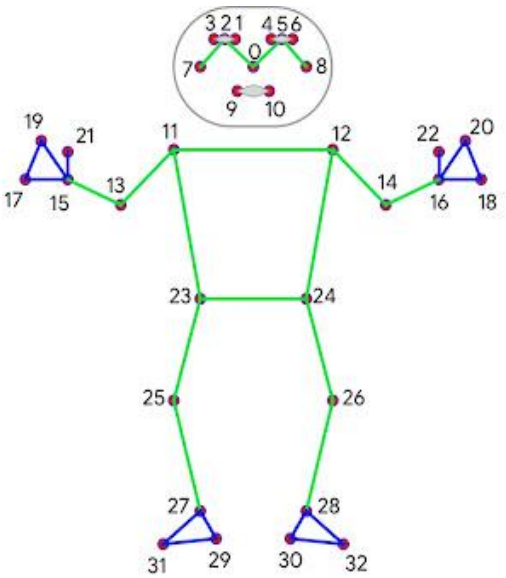

Fig.1. BlazePose 33 keypoint topology as COCO (colored with green) superset

One of the critical advantages of BlazePose's efficiency is its lightweight architecture, which enables it to run smoothly on devices with limited processing power, such as smartphones, wearable devices, and embedded systems. This allows for real-time pose estimation and tracking on edge without constant network connectivity or reliance on cloud-based processing. This makes it suitable for applications in remote or offline environments where internet connectivity may be limited or unreliable.BlazePose's ability to run in real-time on devices with limited computational resources opens up possibilities for a wide range of use cases where low-latency, on-device processing is crucial. The real-time performance of BlazePose also enables immediate feedback, which is invaluable in applications that require real-time user interaction, such as fitness tracking or sports analysis. Users can receive instant feedback on their movements and poses, allowing them to make adjustments in real-time to achieve their fitness goals or improve their athletic performance. In virtual reality (VR) and augmented reality (AR) applications, real-time pose estimation can create immersive experiences where virtual objects can interact with the user's real-world poses in real-time, providing a seamless and natural user experience.

2.4 Pose Estimation

Pose estimation has numerous applications in various fields, including robotics, sports analysis, healthcare, entertainment, and virtual reality. Pose estimation has come a long way from its early stages, which relied on handcrafted features and simple machine-learning algorithms. With the advent of deep learning, pose estimation has witnessed remarkable progress. CNNs and RNNs have become popular deep-learning techniques used for pose estimation tasks. CNNs excel at extracting features from images, while RNNs are effective at modeling sequential data, making them well-suited for capturing temporal dependencies in human movements. One of the critical breakthroughs in pose estimation is the introduction of multi-person pose estimation, which involves estimating the poses of multiple people in a single image or video. Multi-person pose estimation is a challenging task due to the presence of occlusions, scale variations, and interactions between people. In sports analysis, pose estimate is used to analyze athletes' movements, measure their performance, and provide feedback for training. In healthcare, pose estimation is used for physical rehabilitation, tracking patient movements, and monitoring elderly individuals to prevent falls. In robotics, pose analysis enables robots to interact with humans naturally and intuitively. In entertainment, pose estimation is used for character animation in video games, virtual reality experiences, and movie special effects. Pose estimation has also been leveraged in human-computer interaction, such as sign language recognition, virtual try-on, and emotion recognition. Sign language recognition using pose estimation enables the communication between deaf and hearing individuals. Virtual try-on applications allow users to try on clothes virtually and see how they look without trying them on physically.

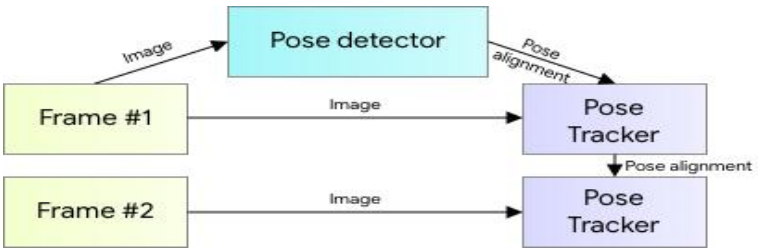

Fig.2. Human pose estimation pipeline [Google AI Blog on BlazePose].

However, there are still challenges in pose estimation that researchers are actively working to address. One challenge is the accurate estimation of poses in extreme poses or rare poses, which need to be better represented in training datasets. Another challenge is robustness to occlusions and scale variations, especially in crowded scenes with overlapping people. Real-time performance is also crucial for many practical applications, as pose estimation is often used in real-time scenarios such as sports analysis and virtual reality experiences. Moreover, as technology continues to advance, pose estimation has the potential to be integrated with other emerging technologies like AR and wearable devices. AR applications can benefit from accurate pose estimation to enable realistic virtual object placement and interactions with the physical world. Wearable devices such as smart glasses and body-worn sensors can leverage pose estimation for various applications, such as fitness tracking, health monitoring, and gesture recognition.

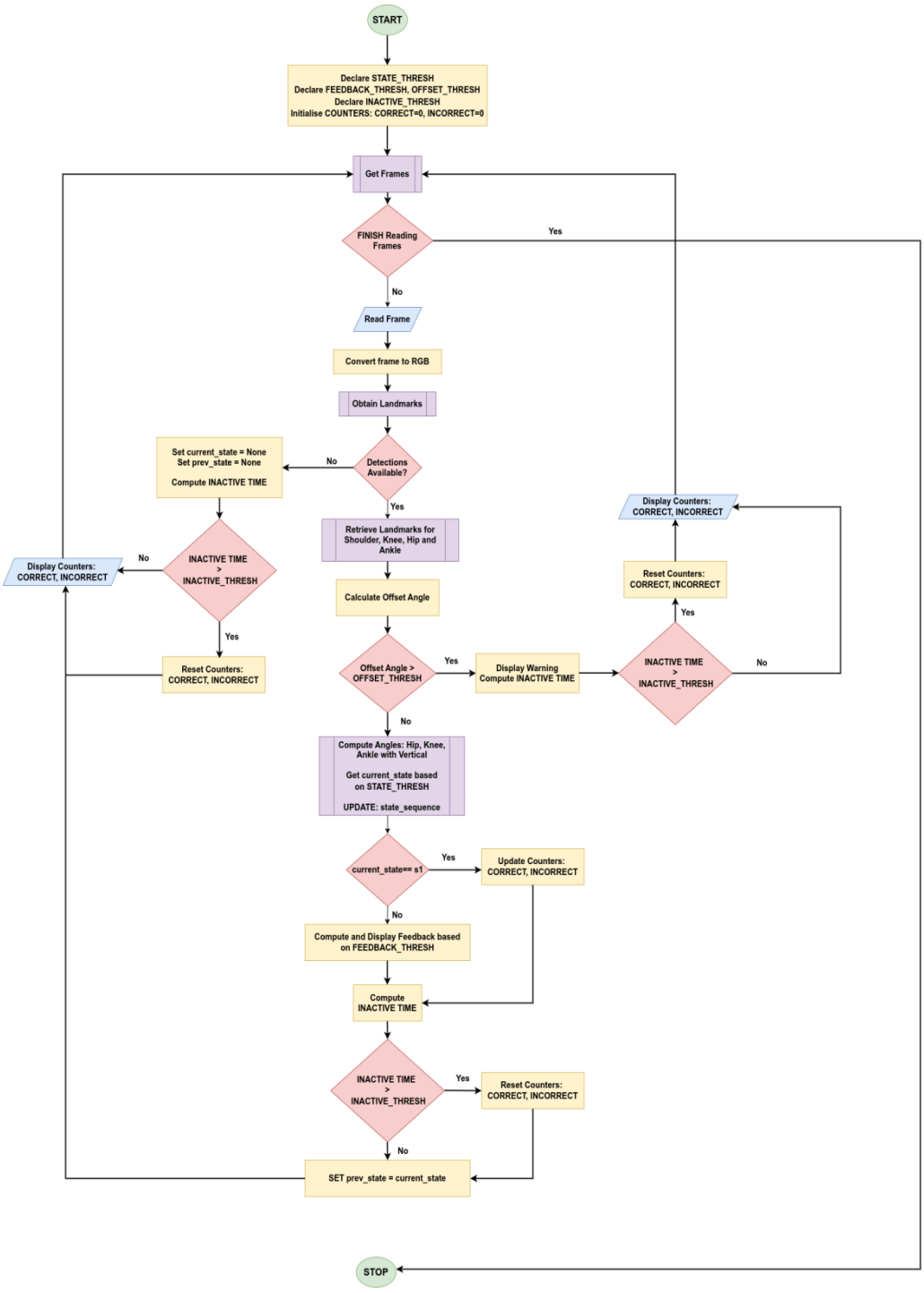

Fig.3. Application Workflow for the AI Fitness Trainer [LearnOpenCV].

3 Results and Discussion

The real-time gym exercise rep counter, developed using Mediapipe in a Jupyter Notebook, is a unique and innovative tool for tracking exercise performance in real-time. Its pose estimation model accurately detects and tracks body landmarks during exercises like bicep curls, squats, and sit-ups, providing instant feedback on exercise repetitions. This real-time feedback feature helps users adjust their form and technique, improving their exercise performance and preventing injury. Moreover, the system's customization and adaptability features apply to various users with different exercise preferences and abilities. The rep counter's potential for integration into virtual fitness platforms or mobile applications makes it accessible to users who prefer to exercise at home or in a virtual environment, expanding its reach and convenience. Integrating gamification elements can enhance user engagement and motivation, making the rep counter a more interactive and enjoyable tool for fitness tracking. Despite limitations and challenges, such as the accuracy of pose estimation and the potential for false positives or false negatives in rep counting, continued research and development can refine the rep counter's accuracy and usability. Overall, the real-time gym exercise rep counter has the potential to revolutionize the way users track and optimize their workouts, leading to improved exercise outcomes and enhanced fitness experiences.

Implementing the real-time gym exercise rep counter using Mediapipe shows good accuracy in detecting and tracking exercises in real time. However, certain limitations and areas for improvement need to be considered. One limitation of the current implementation is the reliance on body landmarks and joints detected by Mediapipe. Although Mediapipe provides robust pose estimation capabilities, there may be cases where body landmarks must be accurately detected or occluded, leading to inaccurate rep counting. Further refinement and optimization of the pose estimation model could improve the accuracy of the rep counter.

Bicep Curls

For bicep curls, the rep counter accurately counted the number of repetitions performed by users, considering variations in form and speed. The detection and tracking of the movement of the weights and distinguishing between the curling and lowering phases were found to be accurate in most cases. However, challenges were observed in cases where the weights were moved too fast or if the user's arms were not fully extended during the lowering phase, resulting in missed reps or incorrect counts.

Squats

For squats, the rep counter accurately detected and tracked the squat depth and accounted for different foot positions, resulting in accurate rep counting. However, challenges were observed in cases where users performed partial squats or needed better form, resulting in inaccurate rep counts. Additionally, detecting and tracking body landmarks where users moved quickly or had occluded body parts (e.g., hands covering the knees) posed challenges and affected the accuracy of rep counting.

Sit-Ups

For sit-ups, the rep counter accurately detected and tracked the dynamic movement of the sit-up, resulting in accurate rep counting. However, challenges were observed in cases where users performed sit-ups with different styles (e.g., cross-legged) or had variations in their form, resulting in inaccurate rep counts. Detecting and tracking body landmarks accurately during the dynamic movement of the sit-up was also challenging, affecting the rep counting accuracy.

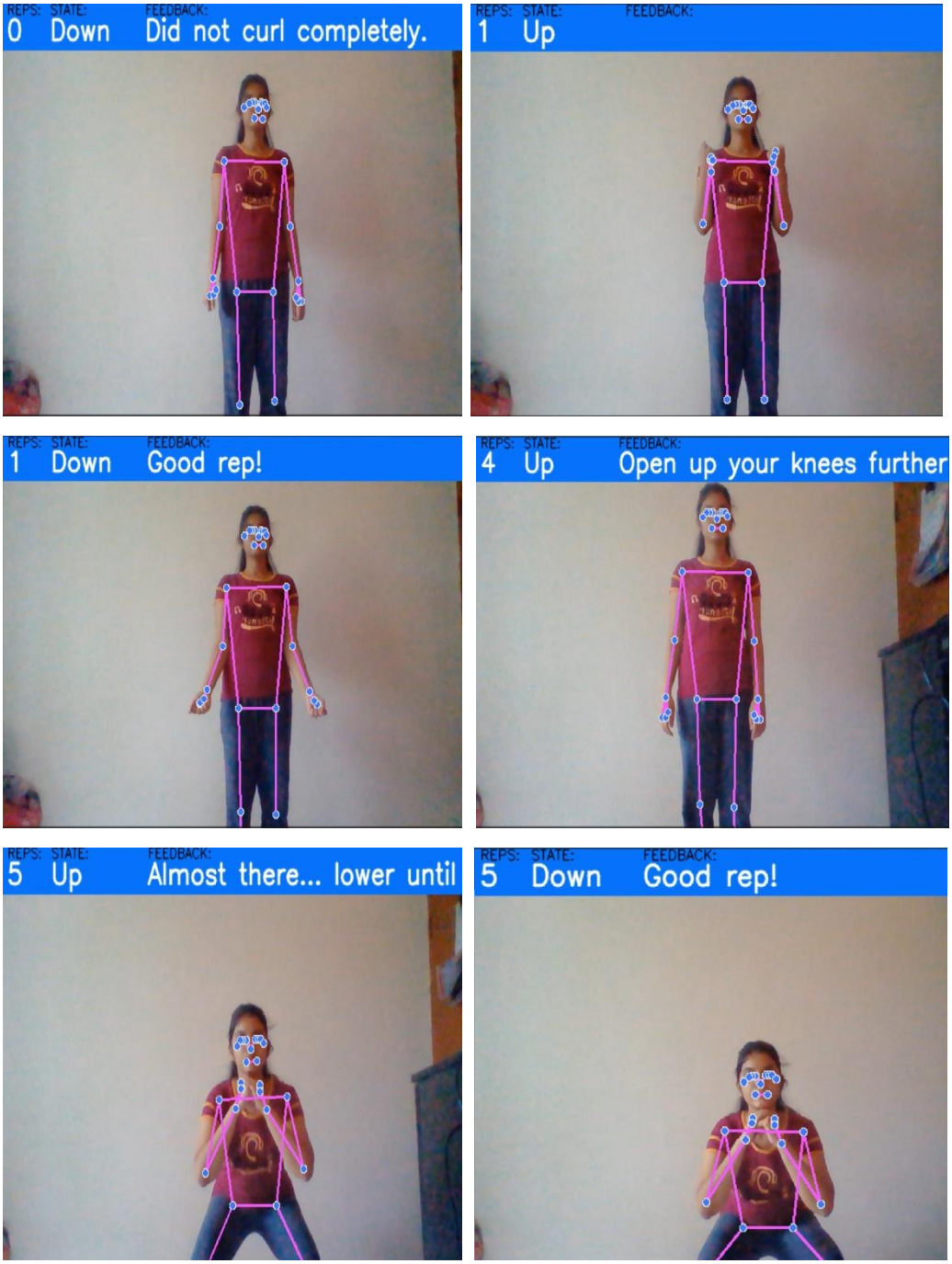

Fig. 4. Rep counting and analysis of bicep curls and counting of squats.

Fig. 5. Rep counting of sit-ups

Another limitation is the sensitivity to exercise form and speed variations. Users may perform exercises with different styles or rates, affecting the accuracy of rep counting. Implementing additional techniques, such as incorporating machine learning algorithms, to recognize variations in exercise form and speed better can improve the rep counter's accuracy. The real-time nature of the implementation also poses challenges in accurately detecting and tracking exercises during dynamic movements. The accuracy of rep counting can be affected for exercises like bicep curls, squats, and sit-ups, where users may move quickly or have occluded body parts. Exploring advanced tracking techniques, such as multi-object tracking, can improve the accuracy of the rep counter during dynamic movements. An essential aspect to consider in the discussion is the potential for incorporating additional features into the real-time gym exercise rep counter. For instance, integrating sensors such as accelerometers or heart rate monitors can provide other data to complement the visual pose estimation, allowing for a more comprehensive assessment of exercise performance. This could enable the rep counter to give insights on exercise intensity, calorie burn, and recovery time. It can be valuable for users looking to optimize their workouts or track their progress over time.

4 Conclusions

The real-time gym exercise rep counter using Mediapipe in a Jupyter Notebook has demonstrated the potential of leveraging AI and ML technologies, explicitly using the MediaPipe framework, to enhance human activity recognition and repetition counting during exercise routines. By analyzing visual data, such as video or image sequences, the system can accurately detect and recognize different exercises, including arm curls, squats, and sit-ups, and count the repetitions performed for each activity. This has significant implications for individuals seeking to improve their fitness levels, as it provides valuable feedback and progress tracking more efficiently and accurately. Overall, the use of AI and ML in this context has the potential to revolutionize the way people approach fitness and exercise, making it easier and more accessible for everyone to maintain a healthy lifestyle.

References

[1] K. Yang, K. Guo, H. Ji, Y. Li, *A Novel Two-Stream Convolutional Neural Network for Human Activity Recognition*, IEEE Transactions on Human-Machine Systems, **50(5)**, 508-517 (2020)

[2] H. Cao, H. Liu, and P. Zhang, *Human Action Recognition Using Multimodal Sensor Data and Deep Learning Techniques: A Comprehensive Review*, J Ambient Intell Humaniz Comput, **11**, 335-358 (2020)

[3] Y. Li, J. Li, J. Li, and C. Li, *Human Activity Recognition Based on a Novel Stacked Autoencoder Network with a Complementary Multimodal Sensor Fusion Strategy*, IEEE Transactions on Instrumentation and Measurement, **69(3)**, 988-1003 (2020)

[4] J. Yao, Z. Wang, and D. Zhang, *Human Activity Recognition based on Multi-modal Sensor Data and Multi-level Fusion Convolutional Neural Network*, Expert Systems with Applications, **169**, 114422 (2021)

[5] S. Mahajan, P. Kumar, *Human activity recognition using convolutional neural networks with transfer learning*, J Ambient Intell Humaniz Comput, **12(7)**, 6509-6524 (2021)

[6] Y. Du, Z. Wang, H. Zhang, Y. Hu, *Human activity recognition using mobile sensors based on convolutional neural networks and transfer learning*, J Ambient Intell Humaniz Comput, **12(11)**, 12823-12834 (2021)

[7] L. Jiang, X. Li, F. Liu, *Human activity recognition based on a novel feature extraction method and deep learning*, Journal of Sensors, 5522346 (2021)

[8] Y. Liu, Z. Wang, H. Zhang, et al., *A time-frequency convolutional neural network for human activity recognition based on wearable sensor data*, IEEE Transactions on Instrumentation and Measurement, **70**, 1-10 (2021)

[9] S. He, Y. Zhang, Z. Zhou, et al., *Multi-sensor fusion for human activity recognition using a novel multi-channel CNN-LSTM model*, IEEE Sensors Journal, **21**, 4106-4114 (2021)

[10] H. Kim, K. Hong, K. Jung, J. Kim, M. Park, M. Kim, *A human activity recognition system based on convolutional neural network and long short-term memory*, Sensors, **22(2)**, 412 (2022)

[11] Y. Kim, J. Kim, Y. Song, *Human activity recognition using a deep learning-based algorithm on a smartwatch with a built-in accelerometer and gyroscope*, Journal of Sensors, 5651099 (2022)

[12] K. Saeedi, S. Chandra, R. V. Babu, *Two-stage deep learning for human action recognition*, IEEE Transactions on Circuits and Systems for Video Technology, **32**, 179-190 (2022)