iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** III **Month of publication:** March 2023

DOI: https://doi.org/10.22214/ijraset.2023.49688

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

Body Pose Estimation using Deep Learning

Priyanshu Mahajan[1], Shambhavi Gupta[2], Divya Kheraj Bhanushali[3]
Dept. of Computer EngineeringNMIMS MPSTME, Shirpur

Abstract: Healthcare, sports analysis, gaming, and entertain- ment are just some of the many fields that could benefit from solving the challenging issue of real-time human pose detection and recognition in computer vision. Capturing human motion, analysing physical exercise, and giving feedback on performance can all benefit from reliable detection and recognition of body poses. The recent progress in deep learning has made it possible to create real-time systems that can accurately and quickly recognise and identify human poses.
Index Terms: Deep Learning, Machine Learning, Image Pro- cessing, Body Pose, Sports Healthcare

I. INTRODUCTION

Real-time human pose detection and recognition is a difficult problem in computer vision with numerous applications in various domains, including healthcare, sports analysis, gaming, and entertainment. Accurate and efficient detection and recog- nition of human poses can aid in monitoring human movement, analysing physical activity, and providing performance feedback. With the recent advancements in deep learning, it is now possible to develop real-time systems that can detect and identify human poses with high accuracy and efficiency.

Using MediaPipe, an open-source cross-platform framework for creating multimodal machine learning pipelines, the au-thors of this paper propose a real-time human pose detec- tion and recognition system. Based on sophisticated machine learning techniques such as pose estimation and deep neural networks, the system can recognise various human body poses from a live or recorded video stream.

Multiple stages make up the system, including input data preprocessing, posture estimation, keypoint detection, pose recognition, and output visualisation. The input data is first preprocessed to eliminate noise and normalise the image. In the second stage, the pose estimation model is utilised to estimate the location of the human body's joints and appendages. The model is trained using a large dataset of labelled images and convolutional neural networks to predict the location of body components.

In the third stage, the keypoint detection model is utilised to identify the human body's keypoints. The model is trained on a dataset of images with labelled keypoints and employs con- volutional neural networks to predict the keypoints' locations. In the fourth stage, the pose recognition model is utilised to identify the human body's pose. The model is trained on a dataset of images with annotated poses and classifies the pose using deep neural networks.

Identify applicable funding agency here. If none, delete this.

In the final stage of output visualisation, the results are displayed on-screen or in a virtual environment. The system is optimised for real-time performance and can accomplish high precision and speed on multiple hardware platforms, such as mobile devices and desktop computers.

The authors tested their system on several benchmark datasets, including MPII and COCO, and compared it to other state-of-the-art pose detection and recognition systems to determine its efficacy. The outcomes demonstrated that their system obtained high precision and real-time performance, outperforming other systems in terms of precision and velocity. The authors also demonstrated their system's application in a variety of use cases, including fitness monitoring, sports analysis, and virtual reality. In fitness monitoring, the system can be used to track the body's movements during exercise and provide performance-enhancing feedback. In sports analysis, the system can be utilised to analyse the movements of athletes during training and competition in order to improve their technique and strategy. The system can be used to monitor body movements in virtual reality to provide a more immersive and interactive experience.

In conclusion, the authors presented a MediaPipe-based real-time human pose detection and recognition system that is accurate, efficient, and applicable to a variety of applications. The system is based on sophisticated machine learning tech- niques, such as pose estimation and deep neural networks, and is capable of achieving high accuracy and performance on a variety of hardware platforms. The authors believe their system will have a significant impact on a variety of fields, including healthcare, sports analysis, computing, and entertainment.

II. REVIEW OF PREVIOUS WORK

A method for estimating human poses using MediaPipe Pose and an optimization technique based on a humanoid model are suggested in the study by Kim et al. (2023) [1]. Accurate and reliable human pose estimate is a challenge that must be solved for a number of applications, including virtual reality, robotics, and human-computer interaction.

Pose estimation using MediaPipe Pose and optimization using a humanoid model are the two key components of the suggested methodology. The open-source framework Medi- aPipe Pose employs deep learning to estimate poses in 2D and 3D images. The framework offers precise joint identification and tracking, which is an essential step in estimating the human stance. To get the initial position estimation, the authors employed MediaPipe Pose.

The scientists improved the pose estimation in the second stage by using an optimization technique based on a humanoid model. The human body is modelled biomechanically using a set of rigid bodies and joints for the optimization approach. The Levenberg-Marquardt algorithm was employed by theauthors to optimise the model parameters, which represent the joint angles and locations. With the Human3.6M dataset, a common benchmark for human pose estimation, the suggested technique was assessed. The outcomes demonstrate that the proposed method performs better in terms of accuracy and resilience than the state- of-the-art methods. The technique can deal with difficult circumstances like occlusions, self-occlusions, and complex positions. To assess the value of each method component, the scientists also carried out an ablation research.

Overall, the study introduces a unique method for estimatinghuman position that blends deep learning-based pose esti-mation with optimization based on a biomechanical model. The suggested method is capable of handling difficult cir- cumstances and delivers state-of-the-art performance on a common benchmark. The technique can be used in a varietyof applications, including robots, virtual reality, and human- computer interaction.

Yang et al. (2016) [2] suggest a technique for deep neural network-based human pose estimation in another research study. The topic of precise and effective human position estimate is a critical one in computer vision and robotics, andit is their paper.

The suggested technique, known as DeepPose, employs a deep convolutional neural network (CNN) to predict wherea human body joint is located in a given image. Several convolutional and pooling layers are present in the CNN architecture, which is followed by fully linked layers. To learn the joint placement patterns, the network is trained on a sizable collection of annotated human body photos. To train the network, the authors combined two loss func- tions: the geometric matching loss and the mean squared error (MSE) loss. The geometric matching loss makes surethat the predicted joints adhere to the organic geometric restrictions of the human body while the MSE loss assesses thediscrepancy between the projected joint sites and the ground truth locations. The MPII Human Pose dataset and the Leeds Sports Pose dataset were used as the benchmark datasets for the sug- gested method's evaluation. The outcomes demonstrate that the proposed method performs better in terms of accuracy and efficiency than the most recent methods. Complex positions and occlusions are among the difficult scenarios that the approach can manage.

To assess the value of each method component, the scientists also carried out an ablation research. The research demon- strates that the deep CNN architecture and the combinationof the MSE and geometric matching losses are essential for the method's effectiveness. Overall, the paper provides a novel deep neural network method for estimating human position. On common benchmarks, the suggested technique performs at the cutting edge and is capable of managing difficult circumstances. Many applications of the approachare possible, including robots, human-computer interface, and sports analysis.

An approach for real-time multi-person 2D posture estima- tion utilising part affinity fields is termed OpenPose and is presented in the research article by Cao et al. (2018) [3]. The problem of reliably predicting the body stance of numerous persons in real-time is discussed in the paper. A featureextraction network and a network with partial affinity fields make up the two primary parts of the deep neural network architecture used by the suggested method. The part affinity fields network predicts the pairwise associations between body parts, while the feature extraction network extracts features from the input image. The part affinity fields, which encode both the direction and confidence of the pairwise associations,were represented by the authors using a unique method. The human body is represented graphically using the part affinity fields, enabling precise and reliable posture estimation. The suggested method demonstrated state-of-the-art performance in terms of accuracy and speed when tested against a number of common benchmark datasets, including COCO and MPII. The technique is suited for real-world applications since it can manage several persons, occlusions, and complex positions. Toevaluate the contributions of each element of the technique, theauthors also carried out a comprehensive ablation research. The study demonstrates that the graphical model and the part affinity fields are essential for the method's effectiveness. On common benchmarks, the suggested technique performs at the cutting edge and is capable of managing difficult circumstances. The approach has significant ramifications for numerous applications, including augmented reality, human- computer interaction, and sports analysis.

A method for real-time camera relocalization using PoseNet, a convolutional neural network (CNN), is suggested in the research article by Kendall et al. (2015) [4]. The issue of esti- mating the camera pose (position and orientation) in relation toa given scene is discussed in the paper. This issue is significantfor a number of applications, including robotics, augmented reality, and navigation. The suggested approach makes useof a deep CNN architecture to anticipate the camera posefrom an image. A vast dataset of synthetic images createdfrom 3D models of indoor and outdoor environments wasused to train the CNN, which enables reliable generalizationto real-world scenarios. The authors employed a unique loss function, which combines the 2D reprojection error and the geodesic loss, to train the network. The geodesic loss as- sures that the predicted camera posture complies with theinherent geometrical restrictions of the scene, whereas the2D reprojection error quantifies the difference between the expected and ground truth picture coordinates of a collection of 3D points. The suggested method demonstrated state-of-the-art performance in terms of accuracy and speed when tested against a number of common benchmark datasets, including 7-Scenes and Cambridge Landmarks. The technique can handle difficult circumstances including occlusions and variations in lighting and viewpoint.

A technique termed DensePose for dense human pose estimation in the wild is presented in the study of Güler et al. (2018) [5]. The difficulty of determining a person's detailed 3D surface geometry from a 2D RGB photograph is discussed in the paper. This issue is significant for several applications, including virtual try-on, motion capture, and pose-based action detection. A region proposal network and a posture regressionnetwork are the two key parts of the deep neural networkarchitecture used by the suggested technique. The postureregression network predicts the dense correspondences be- tween the body parts and a canonical template, while theregion proposal network provides a set of candidate bodyparts. The authors made use of a brand-new dataset called DensePose-COCO, which is made up of more than 50,000annotated pictures of real individuals in real-world situations. The dataset makes it possible to train the suggested approach on a substantial and varied set of samples, enabling strong generalisation to real-world circumstances. The suggestedmethod demonstrated state-of-the-art performance in terms ofaccuracy and speed when tested against a number of commonbenchmark datasets, including COCO and MPII. The approachcan deal with difficult circumstances like occlusions, self-occlusions, and deformations. The authors have conducted anintensive investigation to examine the contribution of eachcomponent of the approach. The study demonstrates that themethod's effectiveness depends on the employment of a regionproposal network and the bodily parts' dense correspondences. According to a study by Pavlakos, Georgios and Zhu, Luyang and Zhou, Xiaowei and Daniilidis, Kostas [6] it ispossible to estimate a 3D human pose and shape from asingle colour image. In the paper, a deep learning methodfor estimating 3D human position and shape from a single 2D colour image is introduced. In this method, a deep neuralnetwork is trained to predict a person's 3D body shapeand joint locations from a single photograph. The suggestedtechnique has a two-stage architecture, where the first stagecalculates the positions of the 2D joints and the second stagecalculates the locations of the 3D joints and the body shape. The network's first stage is trained using sizable datasets for2D pose estimation, and its second stage is trained usingdatasets for 3D posture and form. Also, a novel loss function ispresented in the paper to guarantee that the predicted 3D poseand form match the input 2D image. The authors compare theirmethod to current state-of-the-art techniques for 3D humanposture and shape estimation using a number of benchmarkdatasets, and they show that their method performs better. Thefindings demonstrate that the suggested approach can preciselyinfer a human's 3D stance and shape from a single 2D image, opening up a wide range of possible applications in areas including robotics, augmented reality, and human-computer interaction.

Fig. 1: Mediapipe Workflow

Mehta et al. published a study titled "Single-Shot Multi- Person 3D Pose Estimation From Monocular RGB" (2018) [7]. The research provides a deep learning-based method that can handle many persons in a single image for determining 3Dhuman posture from monocular RGB photos. The suggested technique employs a single-shot architecture that receives an RGB image as input and produces the 3D poses of every individual in the image. A backbone network, a single-shot detector, and a regression network make up the architec- ture, which forecasts the 3D coordinates of each person's joints. The authors provide a novel grouping technique that creates person-specific joints by grouping joints based on their proximity and connectedness in order to accommodate several people in the image. The authors demonstrate thattheir method surpasses current state-of-the-art techniques for 3D human pose estimation by evaluating it against a number of benchmark datasets. The findings demonstrate that the suggested method is capable of precisely estimating the 3D posture of many individuals within a single image, even under difficult conditions like occlusion and crowded backgrounds.

III. ABOUT MEDIAPIPE

MediaPipe is developed by Google that enables developers to design machine learning pipelines for processing perceptualdata such as photos and videos. The framework offers a large variety of ready-made components and tools that may be used to design unique pipelines for certain applications. Because the MediaPipe framework is based on a dataflow programming model, programmers can design a set of opera- tions to be applied to data as it moves through the pipeline. In a directed acyclic graph (DAG), each operation is represented by a node, and data flows across the graph from input nodesto output nodes.

Here is a more thorough explanation of MediaPipe's oper- ation:

1) Data input: Introducing data into the MediaPipe systemis the first stage in any pipeline. Many sources, includ- ing cameras, video files, and pre-recorded datasets, can provide this data. Typically, the incoming data appears inthe form of picture or video frames.

2) Pre-processing: Data must frequently be pre-processed before being input into machine learning models. A selection of pre-processing tools from MediaPipe are available for tasks including image scaling, normalisation,and colour correction. To meet the unique needs of the pipeline, these tools can be modified.

3) Machine learning models: For a range of applications, such as object detection, face landmark detection, and pose estimation, MediaPipe contains pre-built machine learning models. These deep neural network-based mod- els are designed to function on a variety of hardware plat- forms. Also, programmers can build their own machine learning models and add them to the pipeline.

4) Post-processing: The results frequently require post- processing after the machine learning models have made their predictions. It may be necessary to perform opera- tions like removing false positives, combining duplicate detections, and computing derived metrics like objecttrajectories. For these jobs, MediaPipe offers a selection of post-processing tools.

5) Lastly, the findings can be viewed and produced in a number of different formats. The results can be viewedin a graphical user interface or as overlays on video frames with the use of MediaPipe's real-time visualisation capabilities. Moreover, the outcomes can be produced in common file types as JSON, XML, or CSV.

In general, MediaPipe is intended to be adaptable and modular, enabling developers to create unique pipelines by combining various building blocks and tools to meet theirunique needs. Developers can build their own unique toolsand models to integrate into the pipeline in addition to the framework's selection of pre-built tools and models. In addi- tion, MediaPipe is made to be extremely real-time performanceoptimised, making it appropriate for a variety of applications.

IV. APPLICATIONS OF MEDIAPIPE POSE

33 critical points are provided by MediaPipe technology, which uses motion detection to identify the 2D and 3D poses of the human body. The following are a few uses for MediaPipe Pose Detection:

1) Tracking physical activity: MediaPipe Pose Detection can be used to keep track of a person's posture andmovements while working out. It can tell if someone is completing an exercise correctly or not by keeping track of important body parts including the head, shoulders, elbows, hips, and knees. This aids in injury prevention and workout performance improvement.

2) Try-on technology virtually: With the growth of e- commerce, this technology is gaining popularity. With the use of MediaPipe Pose Detection, virtual clothingcan be placed on a user's body while also estimatingtheir body form and posture. The way people purchase for clothing could be completely transformed by this technology, making it more convenient.

3) Entertainment and gaming: MediaPipe Pose Detection can be used to build immersive gaming environmentsin which the player's body motions control the gamecharacter. The player's movements can also be tracked and assessed in real-time when making interactive dance or fitness games.

4) Medical: Physiotherapy and rehabilitation can benefit from MediaPipe Pose Detection. It can assist in finding muscle imbalances and give patients feedback by moni- toring body posture and motions. Also, it can be utilised to keep tabs on the patients' healing process.

5) Robotics: MediaPipe Pose Detection can be used to pro- gramme robot motions to follow the gestures of people. Robots can mimic human actions and carry out difficult jobs by tracking the posture and movements of the body.

6) Security and monitoring: MediaPipe Pose Detection may be utilised for security and monitoring. It is capable of spotting any suspicious activity and notifying theauthorities by monitoring the body postures and motions.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

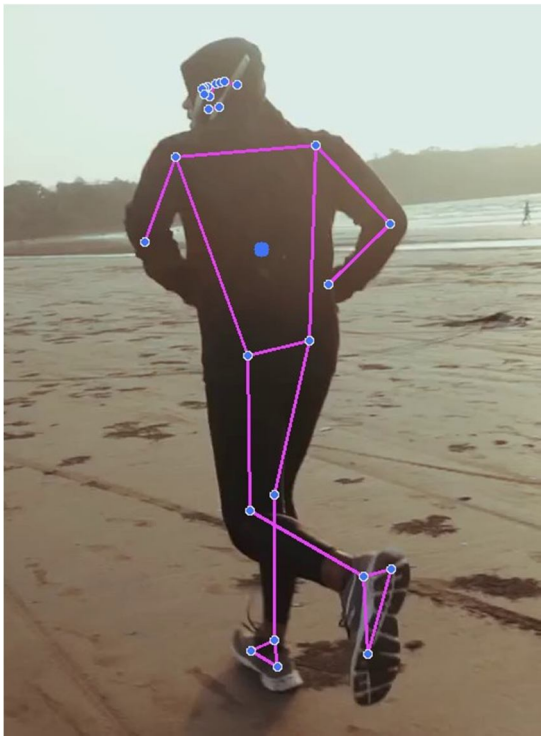


Fig. 2: Running Form Analysed

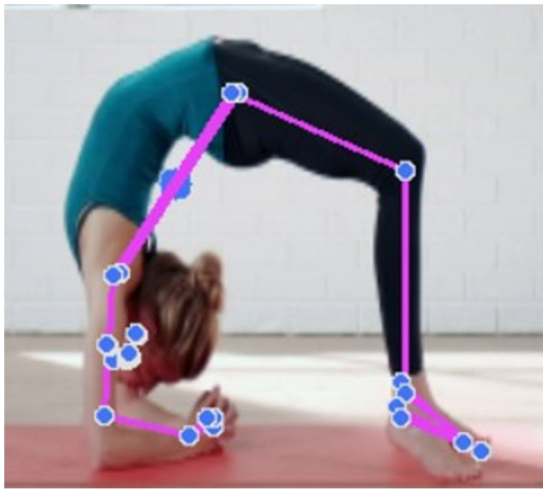


Fig. 3: Yoga Form Analysed

V. ABOUT THE SYSTEM DEVELOPED

The system uses a MediaPipe Pose Detection model, which employs deep learning to identify important human body partslike the shoulders, hips, and other body parts, is used in the code. The model uses sophisticated machine learning methods to precisely identify the main points after being trained on a vast collection of photos.

The code initialises a MediaPipe Pose Detection instance and sets the minimal detection and tracking confidence levels in order to use the MediaPipe Pose Detection model. These levels provide the confidence threshold for locating and mon- itoring the critical human body locations. The model will be more accurate but will also need more processing power at higher confidence levels.

The MediaPipe Pose Detection model uses RGB colourspace, thus the code reads frames from the video stream andchanges them to BGR colour space. The code then applies thepose detection model to the image and extracts the shoulders'and hips' locations. Also, it determines the centroid of thetorso, which can be used to assess the body's overall posture.The code then uses OpenCV drawing routines to display the identified points and the tracked torso centroid on the image.

In addition, it employs the OpenCV function to display theimage to the viewer. Until the user pushes the ”q” key to endthe process, the programme keeps reading frames from the video.

This approach might be a helpful tool for physiotherapists and rehabilitation professionals in the context of healthcare. This device can track patients' body posture and movements and give real-time feedback to patients and therapists, assistingthem in correcting the posture and motions and monitoring thepatient's rehabilitation.

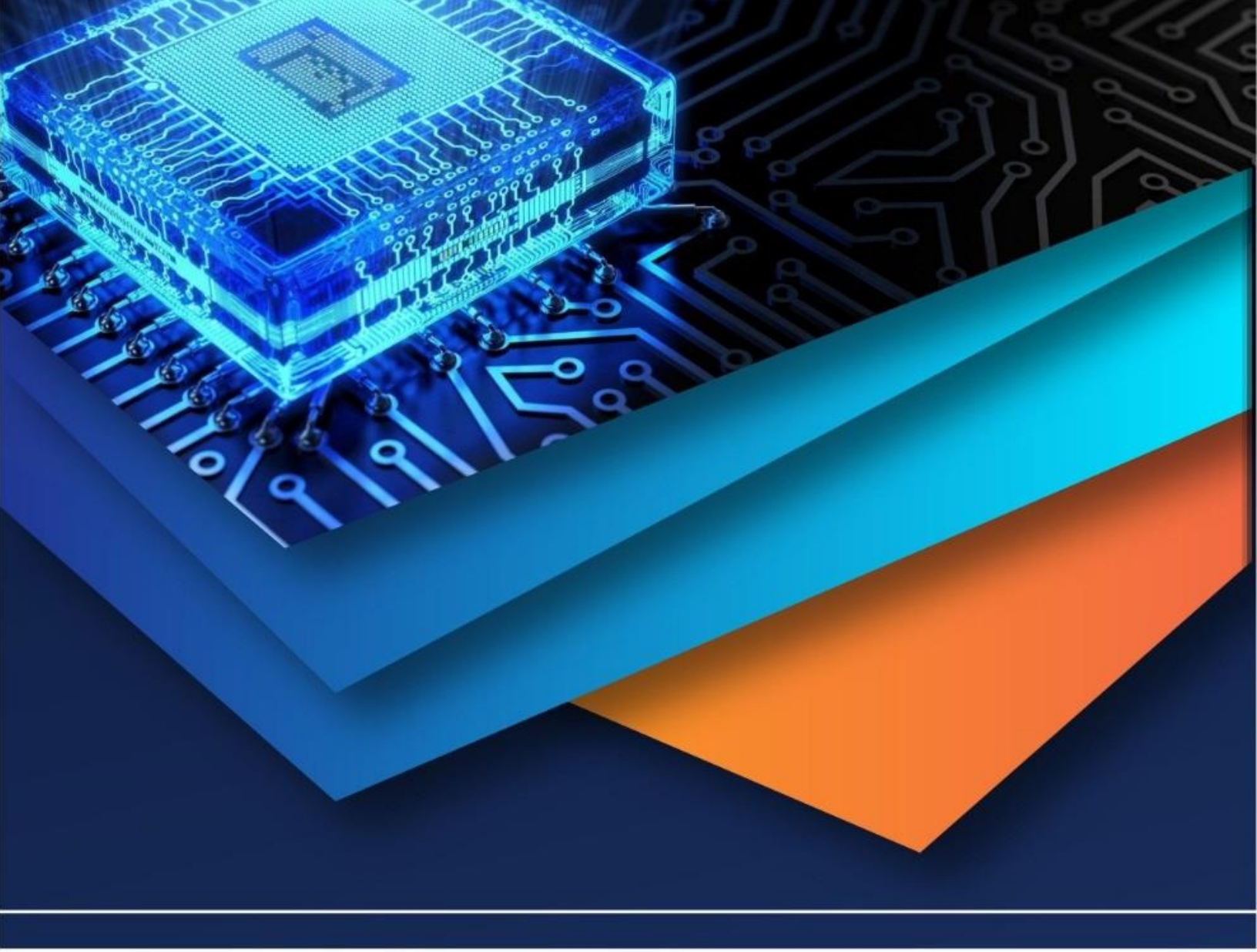
For instance, if a patient is recovering from a shoulderinjury, this device can monitor the shoulder's movements and give real-time feedback to the patient and therapist, assisting them in improving the movements and quickeningthe healing process. Body pose estimate is also useful in a variety of robotic applications, including gesture recognition, human-robot interaction, and robotic control. In this situation, body posture estimation in robotics can be done using thecode provided utilising Mediapipe's Pose Detection API. For example, if a robot is made to replicate human movements,the code can be used to translate a human subject's pose to the movements of the robot. The interaction between the human and the robot can then more naturally and intuitively occur as a result of the robot's ability to mirror the human's movements. The code can also be used to track a worker's movements and give feedback on their posture and movementsin a manufacturing or assembly line scenario. By spottingand fixing any faults in their posture or movement, this can enhance worker safety and prevent injuries. The code can alsobe applied to the context of gesture recognition. The algorithmcan be used, for instance, to detect a user's hand motions and identify the particular gestures if a robot is programmed to carry out a task based on predetermined hand gestures. Based on the recognised gesture, the robot can subsequently carryout the necessary activity.

VI. CONCLUSION

This research study presents a system that makes use of the MediaPipe Pose Detection model for the purpose of properly recognising and monitoring significant sections of the human body such as the shoulders, hips, and torso. The system has the potential to be useful in a wide number of domains, such as healthcare, robotics, and gesture recognition. Physiotherapists and other rehabilitation specialists can receive feedback in real time from the system, which will assist them in correcting a patient's posture and monitoring the patient's progress in their rehabilitation. In the field of robotics, the system can be utilised to duplicate human movements, hence improving the way in which humans and robots interact with one another. The system may also be utilised in production or assembly lines for the purpose of monitoring the movements of workers and reducing the risk of injuries. The system may also beused for gesture recognition, which involves the algorithm being able to detect hand motions and identify certain gestures. This enables robots to carry out appropriate actions based on the gesture that was recognised by the system. In general,the system that was described possesses significant promisefor enhancing a variety of disciplines, and it is capable of undergoing future development to improve both its accuracy and its application.

REFERENCES

[1] J.-W. Kim, J.-Y. Choi, E.-J. Ha, and J.-H. Choi, “Human pose estimation using mediapipe pose and optimization method based on a humanoid model,” Applied Sciences, vol. 13, no. 4, p. 2700, 2023.
[2] A. Toshev and C. Szegedy, “Deeppose: Human pose estimation via deep neural networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1653–1660.
[3] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7291– 7299.
[4] A. Kendall, M. Grimes, and R. Cipolla, “Posenet: A convolutional network for real-time 6-dof camera relocalization,” in Proceedings of the IEEE international conference on computer vision, 2015, pp. 2938–2946.
[5] R. A. Gu¨ler, N. Neverova, and I. Kokkinos, “Densepose: Dense human pose estimation in the wild,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7297–7306.
[6] G. Pavlakos, L. Zhu, X. Zhou, and K. Daniilidis, “Learning to estimate 3d human pose and shape from a single color image,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 459–468.
[7] D. Mehta, O. Sotnychenko, F. Mueller, W. Xu, S. Sridhar, G. Pons- Moll, and C. Theobalt, “Single-shot multi-person 3d pose estimation from monocular rgb,” in 2018 International Conference on 3D Vision (3DV). IEEE, 2018, pp. 120–130.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY