International Journal of
Optical Innovations & Research

Volume 3, Issue 1, 2025
January–June
DOI (Journal): 10.37591/IJOIR

https://journals.stmjournals.com/ijoir

Research IJOIR

3D Hand Interaction in Virtual Space

Vedant Gawner*, Smita Badarkhe, Sanket Fulpagare, Adity Ghode

Abstract

In the realm of augmented reality (AR) and virtual reality (VR), it is necessary to facilitate users' interplay with the digital world. In virtual environments, the system will be able to track the user's hand movements in three dimensions, providing the impression that they are essentially there. The system will adopt a camera component for input, a Python library called Open-CV for real-time video streaming, and a media-pipe library to record the user's hand movements. The Unity Gaming Engine will be adopted because it is required to create a virtual environment in instructions to interact with the 3D world. The system tracks hand positions more accurately and reduces noise by using specific filters in addition to sophisticated deep learning techniques to identify hand motions. Unity's robust toolset and its support for popular hand tracking devices, such as the Leap Motion Controller and Oculus Quest, facilitate the development of these systems. Additionally, Unity's flexible architecture and scripting capabilities allow for the customization of hand interactions, making it possible to tailor the user experience to specific applications, from gaming to virtual simulations and educational tools. Unity's extensive asset store and integration with popular VR SDKs (Software Development Kits) facilitate rapid prototyping and deployment of hand tracking features, making it accessible to both indie developers and large studios. To manage complicated activities and maintain a quick and responsive system, we also investigate the use of cloud computing. The system's functionality was evaluated in a variety of virtual environments. The outcomes demonstrated how much more sensitive and accurate our hand tracking is, which is crucial for interactive VR activities like gaming, remote collaboration, and virtual training.

Keywords: Python, C++, unity engine, machine learning

INTRODUCTION

A fascinating and rapidly developing field, 3D hand tracking in virtual environments has drawn a lot of attention because of its possible applications in an extensive range of industries. For the 3D hand and body position estimation job in depth picture, a novel anchor-based technique called Anchor-to Joint Regression Network (A2J) with end-to-end learning capability is proposed. A potential solution to this problem is the use of 3D hand tracking technology. It enables users to control and manipulate objects with their hands in virtual surroundings.

Accurate three-dimensional monitoring of hand and finger movements allows users to interact with computer interfaces like they would in the real world by pointing, gesturing, and reaching out and holding items. In mandate to improve immersion and user involvement in virtual reality (VR), 3D hand tracking in VR environments is very related because it enables for natural and intuitive interactions. It is an invaluable technology that will influence how immersive experiences and human-computer interplay will evolve in the future. It has applications in gaming, education, healthcare, design, accessibility, and more.

***Author for Correspondence**
Vedant Gawner
E-mail: vsgawner@gmail.com

Student, Department of Electronics & Telecommunication, Smt. Kashibai Navale College Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India

Received Date: July 17, 2024
Accepted Date: March 27, 2025
Published Date: April 08, 2025

Citation: Vedant Gawner, Smita Badarkhe, Sanket Fulpagare, Adity Ghode. 3D Hand Interaction in Virtual Space. International Journal of Optical Innovations & Research. 2025; 3(1): 13–19p.

LITERATURE REVIEW

Guo *et al*. suggest a unique computer vision method for efficiently capturing and assessing the 3D hand movements of Parkinson's Disease (PD) patients [1]. This method makes use of depth cameras and spatial-temporal 3D hand pose estimation. Their approach incorporates a temporal encoding module that is intended to improve the A2J deep learning architecture. This module specifically addresses the issue of pose jittering, which has the potential to undermine movement analysis accuracy. Furthermore, the method incorporates a pose refining procedure that lessens the reliance on big datasets, increasing the effectiveness and accessibility of hand movement evaluation. By addressing serious shortcomings in conventional evaluation techniques, this novel framework seeks to increase the objectivity and reliability of hand movement assessments in patients with Parkinson's disease.

The Anchor-to-Joint regression network (A2J), a novel anchor-based method for 3D hand and body position prediction from depth pictures, is introduced by Xiong *et al.* [2]. Anchor points are densely distributed on depth pictures and are used as local regressors for joint positions in an end-to-end learning architecture. These anchor points collectively forecast joint positions in an ensemble fashion by incorporating global-local spatial context information, which improves the generalization capacity of the model.

The suggested A2J paradigm is different from the current state-of-the-art approaches, which include point-set based techniques, 3D convolutional neural networks (CNNs), and fully convolutional networks (FCNs) based on encoder-decoder pairs. This novel method tackles the challenges of 3D articulated posture estimation and outperforms conventional techniques in capturing the subtleties of hand and body movements.

An advanced technique known or ideas of "Anchor-to-Joint Regression Network", is suggested by the writers. The A2J method possibly uses deep learning and regression approaches to predict, utilizing anchor points or reference features in the depth image, the 3D joint positions of articulated objects or people. The findings are possibly going to benefit PC vision, object recognition, and human pose estimation.

Takala and Heiskanen draw attention to the fact that scholars are becoming more and more interested in virtual reality (VR) avatars and the phenomena of virtual body ownership [3]. They have improved the RUIS for Unity toolkit by adding new capabilities that facilitate the creation of VR apps with customized and adaptable avatars, building on their earlier work with avatars. This development seeks to expand our understanding of body ownership in virtual reality contexts by enabling a more personalized and immersive user experience in virtual surroundings. The revised toolset, which considers the changing field of virtual reality research, gives developers the resources they need to produce more individualized and captivating VR experiences.

The application of auto-scaled full-body avatars in virtual reality (VR) environments is examined in the paper "Auto-scaled Full Body Avatars for Virtual Reality" by Takala and Heiskanen from Aalto University, Finland [3]. This study looks on creating avatars that can change dynamically to fit the user's physical characteristics, improving how immersive virtual reality can be. The study explores several possible applications, such as interactive virtual body alteration, and describes the technical components of avatar scaling. By emphasizing avatar adaptation, this work highlights the importance of personalized representations in virtual reality and advances our understanding of human embodiment and interaction in virtual settings.

Teleb and Chang of Kean University's Department of Computer Science conducted a study titled "Data Glove Integration with 3D Virtual Environments" that investigates the use of data gloves as input devices for navigating and engaging with 3D virtual environments [4]. The study looks at the technical frameworks and approaches required to integrate data gloves in an efficient manner, highlighting how they can improve user immersion and engagement in 3D settings like games, simulations, and virtual reality apps. By concentrating on this integration, the study draws attention to the important part data gloves can play in enhancing the user experience in virtual environments that are immersive, opening the door to more natural and interesting interactions in a range of applications.

COMPUTER VISION

In computer vision, information from a motionless video camera is transformed into judgments or fresh representations with the goal of accomplishing particular goals. In essence, it uses computational methods to alter images into representations that can be decision data, analytical parameters, information-specific graphics, or other formats. Picture processing techniques, which change picture data through techniques including enhancement, restoration, and analysis, are intimately related to this field. Furthermore, computer vision can record and analyze patients' hand movements without physical contact by using depth cameras and spatial-temporal 3D hand posture estimation.

A new vision-based 3D PD hand dataset was constructed, and the model achieved 81.2% classification accuracy, surpassing individual clinicians [1, 2]. There exist many solutions for providing hands-on interactions with virtual environments. The most common approach is to control 3D hand models that directly reproduce, in a virtual form, the position and posture of the user’s hands [5–7].

VIRTUAL ENVIRONMENT

3D environment design is a crucial process. It involves creating immersive virtual environments that users can explore and interact with the objects (3D Blocks). A global reference system that allows for a coherent specification of virtual content and camera pose is made possible by an efficient abstraction of several tracking systems [8]. Regardless of the tracking techniques used, this framework enables designers to place virtual objects in the actual world with accuracy. It enables the simulation of applications before they are deployed, permits real-time remote debugging of tracking performance, makes it easier to define complex interactions between users and virtual content, and offers insights into user behavior inside an augmented reality (AR) experience [9].

With the help of this system's virtual reality (VR) interface, users may watch and engage with every aspect of the augmented reality (AR) experience from a distance as shown in Figure 1. This consists of ambient maps, live camera feeds, 3D representations of actual buildings, and extensive sensor data from every user within the augmented reality setting.

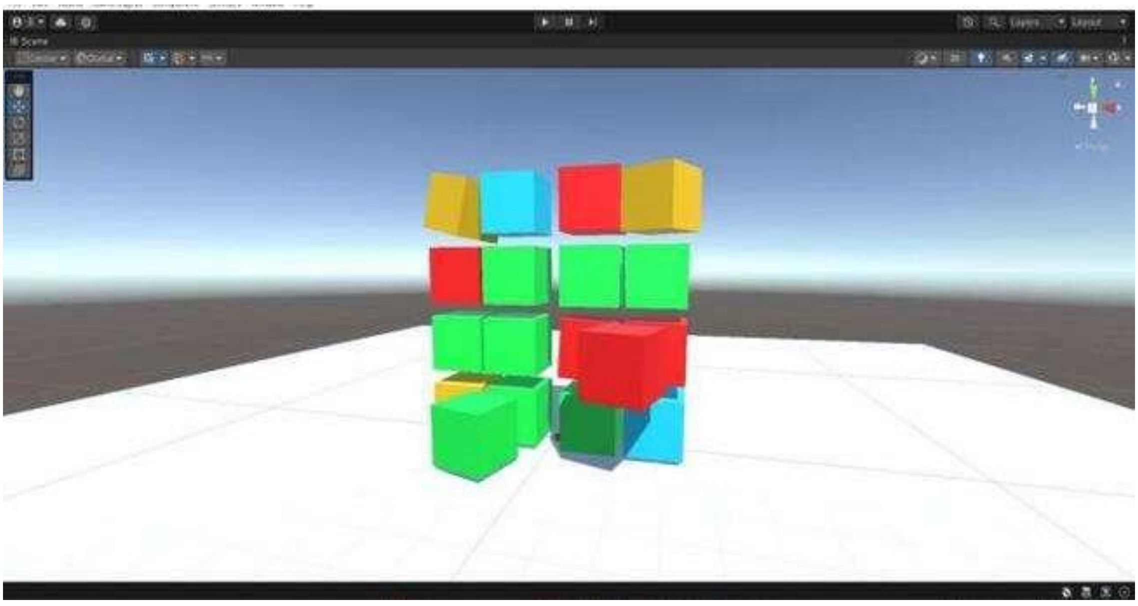

Figure 1. Virtual environment in unity gaming engine.

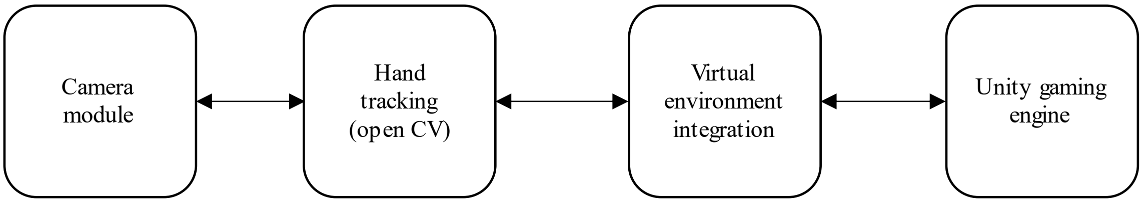

Figure 2. Block diagram of 3D hand tracking in VE system.

For this system, Unity Gaming Engine simplifies hand tracking implementation, making it easier for developers to create engaging experiences within virtual environments. Whether you are building games, simulations, or interactive applications, hand tracking adds a new dimension of interaction.

SYSTEM DESCRIPTION

The captured video feed will be used as the input for the hand tracking process. The captured video feed from the camera module is processed using the Open-CV library for hand tracking. Open-CV provides computer vision algorithms and functions that allow the detection and tracking of the user's hand [10]. Virtual Environment Integration is responsible for integrating the tracked hand's position and movements into a virtual environment.

It entails translating hand gestures in the physical world into equivalent actions in the virtual world. The Unity Game Engine receives the tracked hand's data from the previous block and translates it into actions within the virtual environment. The flow starts with the camera module capturing the real-world scene.

The captured video feed will be used as the input for the hand tracking process. The captured video feed from the camera module is processed using the Open-CV library for hand tracking. Open-CV provides computer vision algorithms and functions that allow the detection and tracking of the user's hand. Virtual environment integration is responsible for integrating the tracked hand's position and movements into a virtual environment. It entails translating hand gestures in the physical world into equivalent actions in the virtual world.

The Unity Game Engine receives the tracked hand's data from the previous block and translates it into actions within the virtual environment. The flow starts with the camera module capturing the real-world scene. Block diagram of 3D hand tracking in VE system as shown in Figure 2.

RESULTS AND DISCUSSION

3D Hand Tracking in virtual environments; we have now arrived at a juncture where the machine can recognize the structure of hand or human body. The system displays different human body parts and analyzes how they work. It interacts with virtual reality to ensure smooth operation, making adjustments based on how the machine is functioning. As we embark on this journey through the results and engage in thoughtful discussion, we aim to unravel the significance of our findings, exploring the transformation potential of 3D hand tracking in redefining human-computer interaction, immersive experiences.

In 3D hand tracking in virtual environment, we have seen the working of software and the virtual environment which we used in the system for the completion of the system required. The plan has carefully chosen technology stack, including Unity for 3D graphics and Python for scripting, to build the virtual environment.

- Data collection and preprocessing steps: ensure a diverse dataset for accurate hand tracking.
- Machine learning models: The integration of hand tracking logic within the 3D graphics environment, facilitated by Unity, is a central focus.

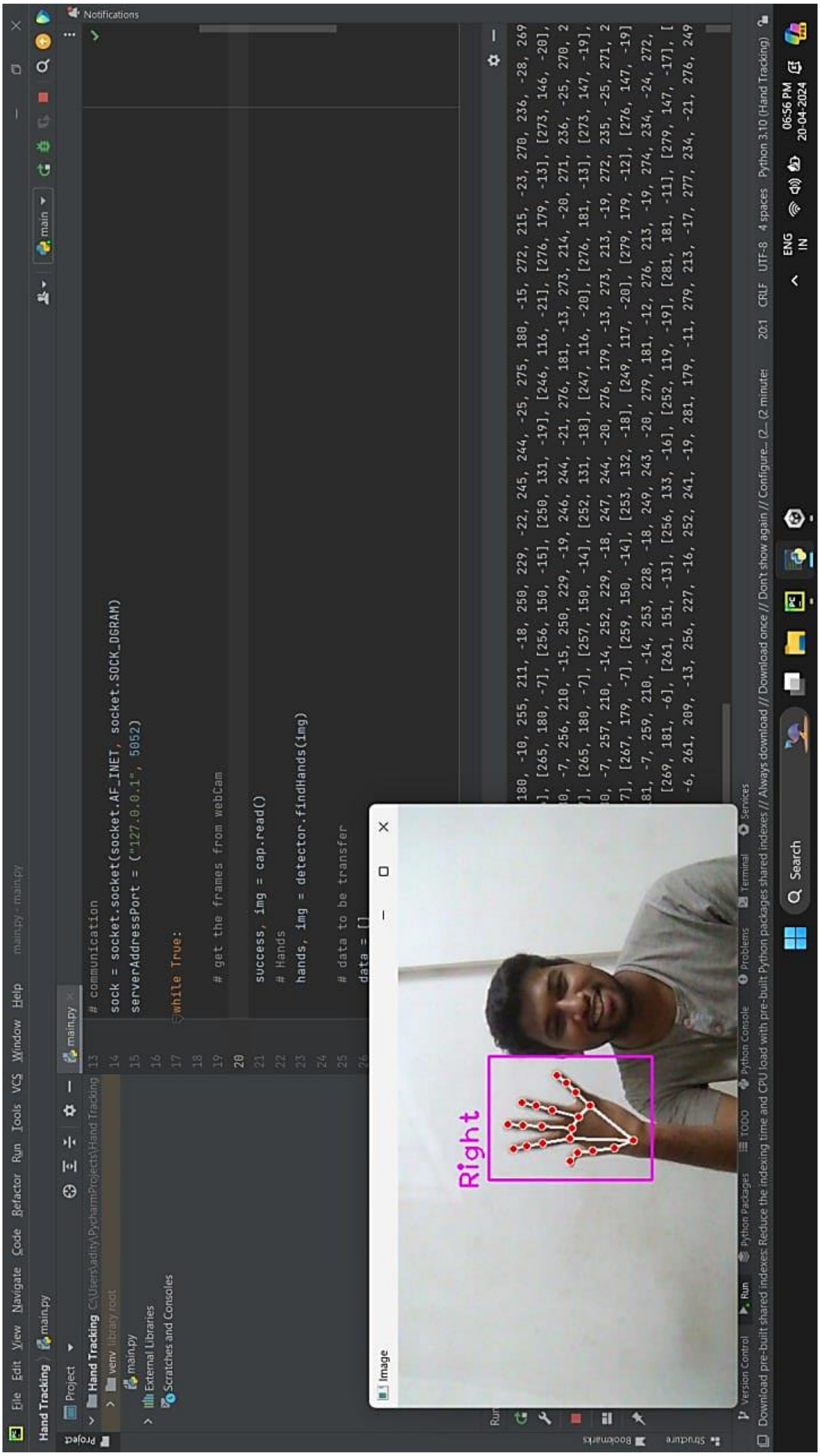

Figure 3. Coordinates interaction in pycharm.

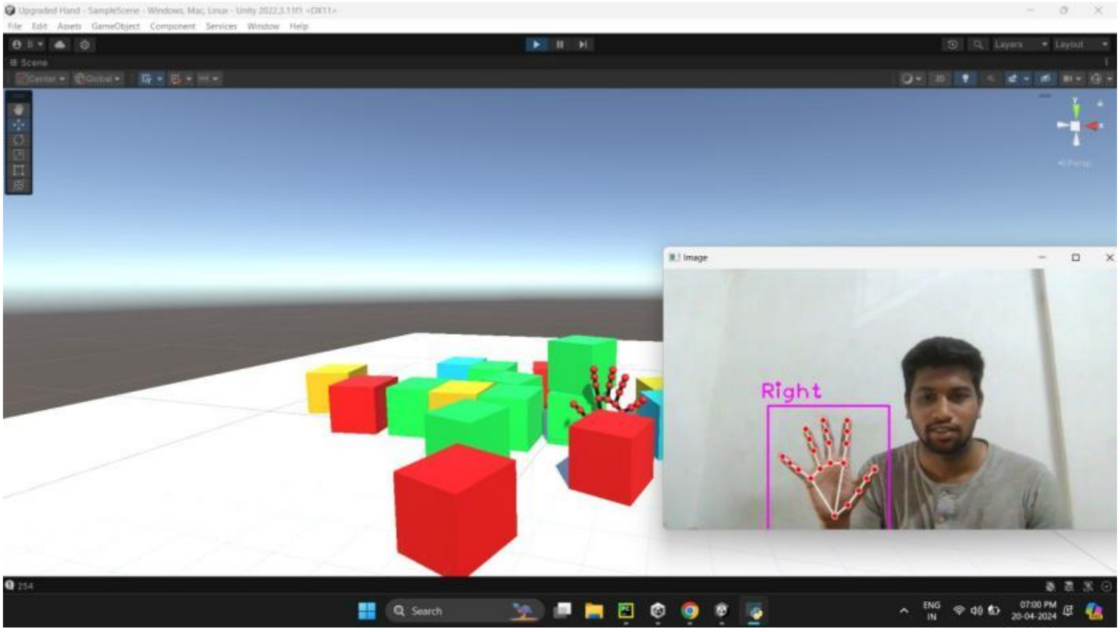

Figure 4. Object interaction inside unity.

Thorough testing, debugging, and user feedback sessions drive iterative improvements, refining the system's performance and responsiveness.

Continuous transformation of information with the machine and its application and handling the resources on its working is positive approach of all group the first objective has already done the interface with the hand tracking in virtual environment; it gave us the coordination of the hand which are tracked in our OpenCV. Now, the next part of our system is to interact it with the unity gaming engine for enter it into the virtual reality or environment for testing and implementing the system.

Implementing 3D hand tracking in Unity for virtual environments involves integrating devices with Unity's game engine to accurately capture and interpret hand movements within the virtual space. Which has created by developers to interact with it with user friendly configuration. The interaction of Unity Engine with user is also integration of py-charms Open-CV libraries which are inbuilt inside of it (Figures 3 and 4).

CONCLUSION

The integration of IoT in diabetes management has the potential to revolutionize healthcare by providing innovative, efficient, and sustainable solutions. IoT-enabled devices, such as Continuous Glucose Monitoring (CGM) systems and wearable technologies, empower patients to manage their condition effectively while enabling healthcare providers to deliver timely and personalized care. These systems not only improve patient outcomes but also reduce the strain on healthcare infrastructure, particularly in underserved regions. By minimizing resource consumption and optimizing care delivery, IoT contributes significantly to the goals of sustainable healthcare. Its alignment with the vision of "Viksit Bharat" underscores its relevance in addressing challenges such as accessibility, affordability, and resource efficiency in India's healthcare system. Despite its promise, the widespread adoption of IoT faces hurdles, including data security concerns, interoperability challenges, and high costs. Addressing these issues through technological innovation, regulatory frameworks, and public awareness campaigns is essential to realizing the full potential of IoT in healthcare. As India strides toward a more inclusive and sustainable future, leveraging IoT in healthcare represents a pivotal step in achieving equitable and efficient healthcare delivery. Continued investment in IoT research and development, along with collaborative efforts to overcome existing barriers, will be key to unlocking its transformative potential.

REFERENCES

1. Guo Z, Zeng W, Yu T, Xu Y, Xiao Y, Cao X, Cao Z. Vision-based finger tapping test in patients with Parkinson's disease via spatial-temporal 3D hand pose estimation. IEEE J Biomed Health Inform. 2022 Mar 29; 26(8): 3848–59.
2. Xiong F, Zhang B, Xiao Y, Cao Z, Yu T, Zhou JT, Yuan J. A2j: Anchor-to-joint regression network for 3d articulated pose estimation from a single depth image. In Proceedings of the IEEE/CVF international conference on computer vision. 2019; 793–802.
3. Takala TM, Heiskanen H. Auto-scaled full body avatars for virtual reality: Facilitating interactive virtual body modification. In 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). 2018 Mar 18; 1–1.
4. Teleb H, Chang G. Data glove integration with 3d virtual environments. In 2012 IEEE International Conference on Systems and Informatics (ICSAI2012). 2012 May 19; 107–112.
5. Emery KJ, Zannoli M, Xiao L, Warren J, Talathi SS. Estimating gaze from head and hand pose and scene images for open-ended exploration in VR Environments. In 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). 2021 Mar 27; 554–555.
6. Narasimhaswamy S, Nguyen T, Huang M, Hoai M. Whose hands are these? hand detection and hand-body association in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022; 4889–4899.
7. Gallotti P, Raposo A, Soares L. v-Glove: A 3D virtual touch interface. In 2011 IEEE XIII Symposium on Virtual Reality. 2011 May 23; 242–251.
8. Zhang L, Zhou W, Zhang X, Lou X. An end-to-end computer vision system architecture. In 2022 IEEE International Symposium on Circuits and Systems (ISCAS). 2022 May 27; 2338–2342.
9. Zhang R, Jing X, Wu S, Jiang C, Mu J, Yu FR. Device-free wireless sensing for human detection: The deep learning perspective. IEEE Internet Things J. 2020 Sep 16; 8(4): 2517–39.
10. Tian Y, Pan G, Alouini MS. Applying deep-learning-based computer vision to wireless communications: Methodologies, opportunities, and challenges. IEEE Open J Commun Soc. 2020 Dec 23; 2: 132–43.

