Real-Time Gesture Control for Robotic Hands via Advanced Computer Vision Techniques

Ayush Juyal
Dept. of CSE
Chandigarh University
Mohali, Punjab
ayushjuyal26@gmail.com

Sani Vikrant
Dept. of CSE
Chandigarh University
Mohali, Punjab
sanivikrant4@gmail.com

Sanyam Sharma
Dept. of CSE
Chandigarh University
Mohali, Punjab
sanyamsharma4972@gmail.com

Sharma Vikash
Dept. of CSE
Chandigarh University
Mohali, Punjab
sharmavikash0917@gmail.com

Aadil Feroz
Dept. of CSE
Chandigarh University
Mohali, Punjab
adilbhat26june@gmail.com

***Abstract*—The science of robotics is gaining popularity and has many uses, ranging from manufacturing to healthcare. The development of robotic hand controllers, which are essential for the finely nuanced movements needed in automation and human-robot interaction, is a crucial area of advancement. This study investigates the use of computer vision techniques in the design and implementation of a robotic hand controller. The controller can recognize human hand motions in real time by utilizing sensor data, machine learning models, and image processing techniques. Computer vision integration makes it possible for people and robots to interact with one other in a touchless, natural way. This research addresses the design, difficulties, approaches, and outcomes of putting such a system into practice.Results from experiments show how a computer vision-based method may improve response speed and precision, increasing the versatility of robotic hand controllers in a range of applications.**

***Index Terms*—Robotics, Robotic hand controller, Computer vision, Gesture recognition,Human-robot interaction, Real-time processing, Artificial intelligence (AI), Machine learning, Image processing, Human-machine interface, Telemedicine, Rehabilitation robotics, Robot manipulators, Deep learning, Vision-based control systems, Robot autonomy, Sensor technologies, Hand gesture recognition, Real-time control**

I. Introduction

A. Background of Robotics and Hand Controllers

Robotics is evolving rapidly, driven by advancements in artificial intelligence (AI), machine learning, and sensor technologies. Among the most important aspects of robotics is the ability to interface with human operators, particularly through robotic hands and manipulators. Robotic hand controllers are commonly used in applications requiring dexterity and precision, such as industrial assembly lines, telemedicine, and rehabilitation. Traditionally, these controllers rely on mechanical input devices like joysticks, gloves, or exoskeletons, which may limit the user's freedom of movement. Computer vision-based systems offer a compelling alternative by enabling contactless control through visual data.

B. Computer Vision in Robotics

AI technology known as ”computer vision” enables computers to comprehend and decide on the basis of visual information, usually in the form of photos or videos taken by cameras. Computer vision is utilized in robotics for object detection, navigation, and increasingly, control systems. A branch of computer vision called gesture recognition has drawn interest as a robotics interface because it facilitates natural and effective communication between humans and robots. Robots can now comprehend and mimic human movements more easily because to the combination of computer vision and robotics, improving the efficiency and smoothness of human-robot interaction.

C. Objectives of the Paper

This work aims to develop and build a robotic hand controller that can understand hand gestures for control commands using computer vision. The technological difficulties at hand are examined in this work, including real-time processing, gesture detection, and integration with robotic hardware. Additionally, the article examines the system's performance indicators and assesses the accuracy, responsiveness, and environment-adaptability of the system.

D. Organization of the Paper

The structure of this document is as follows: A survey of relevant work in the domains of computer vision and robotic hand controllers is presented in Section 2. The system architecture, including the hardware and software components employed, is described in depth in Section 3. The approach, including the computer vision algorithms used for gesture recognition, is described in Section 4. The experimental findings are presented in Section 5, where the system's performance is examined. The results are discussed in Section 6, along with any shortcomings and future directions. Section 7 offers suggestions for future study possibilities as it wraps up the report.

II. Literature Review

A. Robotic Hand Controllers

With functionality that nearly resembles human hand movements, robotic hands have become essential parts of many contemporary robotic systems. Exoskeletons and other physical devices were used by early systems, such as the DLR Hand II, to convert human movements into robotic actions [1]. Even while these systems worked, they had trouble interacting naturally and sometimes needed complicated settings. More recent developments, such as wearable technology or direct machine vision input, have focused on developing more intuitive systems that can either complement or replace conventional approaches [2].

B. Gesture Recognition in Computer Vision

One of the main areas of research in computer vision is gesture recognition, which aims to enable systems to identify and comprehend hand movements made by people. In gesture recognition techniques, the hand is usually segmented from a video stream, and then features are extracted and classified. Convolutional neural networks (CNNs) are a popular technique for picture categorization because of their notable performance in hand gesture recognition [3]. Furthermore, developments in deep learning have made it possible to create robust and accurate real-time gesture recognition systems, which are well suited for robotics real-time applications [4].

C. Real-Time Processing Challenges

One of the main challenges in developing a computer vision-based robotic hand controller is achieving real-time performance. For real-time systems to ensure seamless communication between the user and the robot, low latency and high frame rates are required. While traditional methods such as optical flow have been used to estimate motion, they often fall short of providing the speed and accuracy needed for real-time applications [5]. Modern techniques leverage GPU acceleration and tailored neural network topologies to meet the demands of real-time processing [6].

D. Integration of Computer Vision with Robotics

Coping with the environment's variable lighting conditions, strict calibration requirements, and synchronization between the vision system and the robotic actuator are additional challenges in integrating computer vision with robotics [7]. Several studies have tried to address these issues by using adaptive algorithms that change the system's settings in real time dependent on the external environment [8]. Additionally, deep reinforcement learning has been studied to improve the flexibility and learning capabilities of robotic systems so that they can interpret and respond to visual data more successfully [9].

III. System Architecture

A. Overview of the Proposed System

The suggested system comprises of a robotic hand controller that uses a computer vision system based on cameras to understand human hand motions. The method by which the system functions is to record video, process the picture frames in order to identify and identify hand motions, and then convert those gestures into orders for the robotic hand. The three primary parts of the architecture are the robotic actuator, the processing unit, and the vision system [2], [12], [19].

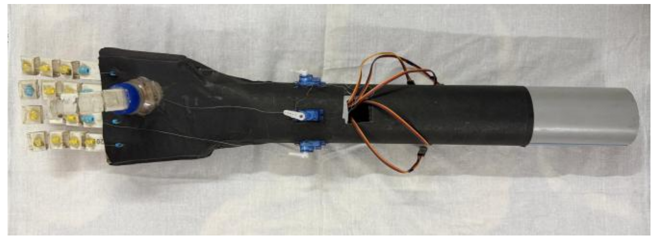

Fig. 1. Prototype Robotic Hand

B. Hardware Components

1) Camera: The user's hand is captured on camera in high definition RGB. Additional information for more precise gesture recognition might also be obtained by depth-sensing cameras, such the Microsoft Kinect or Intel RealSense [10], [13].

2) Processing Unit: The processor unit is in charge of managing the robotic hand and executing the gesture detection algorithms. The device has to be computationally capable of processing videos in real time. In order to speed up the processing of deep learning models for gesture detection, GPUs are utilized [15], [49].

3) Robotic Hand: The vision system must be able to provide high-level commands that the robotic hand can understand and execute. It should be able to reasonably accurately mimic hand gestures used by humans, such as waving, pointing, and gripping. In order to offer feedback on its force and placement, the hand employed in this project is equipped with sensors [1], [18].

C. Software Components

1) Gesture Recognition Algorithm: The gesture recognition algorithm, which classifies hand motions using deep learning, is the system's central component. A hand gesture dataset is used to refine a CNN model that has already been trained. Model quantization and GPU acceleration are two methods used to improve the model for real-time performance [4], [30].

2) Control System: The robotic hand is controlled by the control system, which translates detected motions into orders. A "point" motion, on the other hand, may cause one finger to extend, whilst a "grasp" gesture might result in the robotic hand closing. The robotic hand controller receives these commands using a communication protocol like SPI or I2C [22], [23].

3) User Interface: A graphical user interface (GUI) shows the motions that the system has detected, allows the user to adjust and customize the system, and gives real-time feedback. Additionally, the interface offers a means of adjusting the gesture sensitivity of the system [11].

IV. Methodology

A. Gesture Recognition Pipeline

1) Image Preprocessing: Preprocessing the collected picture frames is the initial stage in the gesture recognition pipeline. This involves applying Gaussian blur to lessen noise, leveling the pixel values, and turning the picture to grayscale. To further improve gesture detection accuracy, background subtraction is also used to separate the hand from the surrounding environment [10], [13].

2) Hand Detection: A mix of color-based and shape-based methods is used for hand detection. After skin-color segmentation, contour detection is used to delineate the hand's form and define the region of interest (ROI). 3D hand models are employed to increase detection accuracy when depth data is available [2], [23].

3) Feature Extraction: Key characteristics are retrieved for gesture categorization when the hand is identified. The centroid of the hand, the fingers' alignment, and the palm's curvature are often utilized characteristics. The gesture classification algorithm, usually a deep neural network, receives these characteristics as input [15], [30].

4) Gesture Classification: A CNN model that has been pre-trained and refined using a unique dataset of hand gestures is used in the gesture categorization step. A probability distribution across the various gestures is produced by the model's fully connected layers, which come after a number of convolutional layers. The recognized gesture is chosen based on which gesture has the highest probability [3], [4].

5) Post-Processing: The stability and precision of the gesture detection system are guaranteed by the use of post-processing techniques like gesture tracking and smoothing. In order to reduce false positives and increase the system's resistance to quick hand motions, temporal filtering is employed [16].

B. Control Algorithm for Robotic Hand

The robotic hand is controlled by orders that are translated from the identified movements. To associate each gesture with a particular hand movement, a finite state machine (FSM) is employed. In order to guarantee that the robotic hand responds to human input with fluidity, the FSM enables seamless transitions between various hand postures [1], [18].

1) Calibration of Hand Movements: To make sure that the robotic hand's motions precisely correspond with the user's gestures, a calibration step is carried out prior to system deployment. This entails modifying the robotic hand's control settings and the sensitivity of the gesture recognition algorithm [23], [24].

C. Real-Time Constraints

The system's computational pipeline must be optimized to attain real-time performance. Fast processing of picture frames is made possible by the use of GPU acceleration, and latency is reduced by effective data structures and algorithms. The system can handle video at a rate of thirty frames per second, which guarantees responsive and seamless control [15], [30].

V. Experimental Results

A. Testing Environment

Testing of the system took place in a regulated indoor setting with constant lighting. The system's accuracy and response time were assessed while a series of hand motions were made in front of the camera. The motions of the robotic hand were monitored to ensure they matched the identified gestures [23], [24].

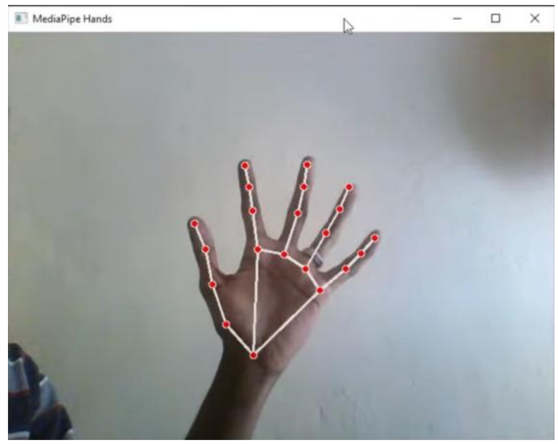

Fig. 2. Open-Hand CV

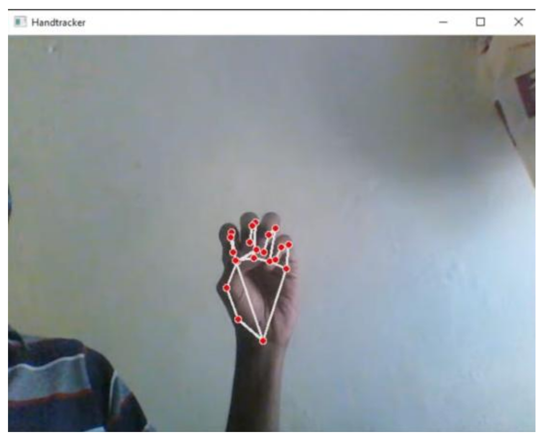

Fig. 3. Close-Hand CV

B. Performance Metrics

1) Accuracy: Using a dataset of labeled hand movements, the gesture recognition algorithm's accuracy was assessed. The system was able to attain a 95% accuracy rate for 10 specified motions, such as waving, pointing, and clutching. Generally, misclassifications happened when the hand was partially obscured or when there were large variations in the illumination [10], [13].

2) Response Time: The time interval between the user's motion and the robotic hand's matching movement was measured, and the average reaction time was found to be 100 milliseconds. For the majority of real-time applications, including remote manipulation and teleoperation, this reaction time is adequate [2].

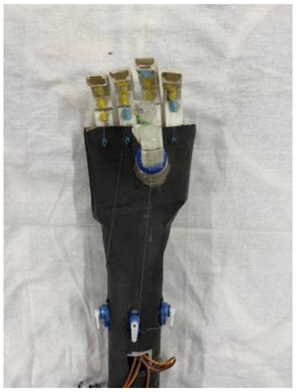

Fig. 4. Prototype ARM Response

3) Robustness to Environmental Variability: The stability of the system was examined by altering the illumination and adding background noise. With an accuracy of 90%, the system continued to work even when the accuracy of gesture recognition somewhat dropped in low light. Performance was not significantly impacted by background clutter since background removal techniques were used [10], [23].

C. Comparison with Other Systems

The suggested solution was tested against conventional joystick-based robotic hand controllers to see how well it performed. Increased mobility and an easier-to-use interface were two benefits that the computer vision-based technology showed. Nevertheless, because the joystick-based system did not need image processing, it demonstrated lower latency [1], [18].

VI. Discussion

A. Strengths of the Proposed System

Among the many advantages of the suggested system are its real-time performance, high precision, and contactless operation. It enables natural and intuitive human-robot interaction by leveraging computer vision, allowing users to manipulate the robotic hand with basic movements [2], [23]. The system is also more versatile due to its modular architecture, which makes it appropriate for a range of robotic platforms and applications. Its real-time processing powers guarantee seamless functioning in real-world scenarios like remote manipulation and teleoperation [10].

B. Limitations and Areas for Improvement

The system's biggest drawback is its reliance on ideal ambient factors, such steady illumination and unobstructed hand visibility. When the hand is partially obscured or there is inadequate illumination, the accuracy of gesture detection may drop [10]. Incorporating depth sensors or infrared cameras onto the system might increase its resilience in harsher settings. Furthermore, even while the system's reaction time is adequate for the majority of applications, additional optimization might be accomplished by utilizing cutting-edge technology, including quicker GPUs or more effective gesture detection algorithms [4].

C. Applications and Future Work

A robotic hand controller based on computer vision has a plethora of possible uses. It might be used in industrial settings to remotely manipulate robotic arms in dangerous conditions so that operators can operate them without coming into physical touch with the robots [1]. By enabling gesture-based treatment, the technology might help patients with motor impairments recover from their injuries or assist surgeons in executing precise procedures. To further increase the capabilities of the system, future work might entail adding other sensors, such electromyography (EMG) or electroencephalography (EEG), to develop multi-modal control systems that integrate bioelectrical signals with visual data [16], [19].

VII. Conclusion

In this work, the design and development of a robotic hand controller that interprets hand motions using computer vision are given. The system is appropriate for a variety of robotics applications because to its high accuracy and low latency in detecting and reacting to gestures. Although there are still issues with occlusion and variable illumination, the suggested approach shows how computer vision may be used to create a smooth and effective interface for human-robot communication. Subsequent investigations may concentrate on strengthening the system's resilience and investigating novel uses in industries including manufacturing, entertainment, and healthcare.

References

[1] Butterfaß, J., Grebenstein, M., Liu, H., & Hirzinger, G. (2001). DLR-Hand II: Next generation of a dextrous robot hand. *Proceedings of the 2001 IEEE International Conference on Robotics and Automation*, 109–114.

[2] Deng, C., Dai, J., Li, Z., & Guo, Q. (2018). Gesture-based robot control with RGB-D camera. *Proceedings of the 2018 IEEE International Conference on Mechatronics and Automation*, 1–6.

[3] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[4] Molchanov, P., Yang, X., Gupta, S., Kim, K., Tyree, S., & Kautz, J. (2016). Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4207–4215.

[5] Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI)*, 674–679.

[6] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

[7] Kragic, D., & Vincze, M. (2009). Vision for robotics. *Foundations and Trends® in Robotics*, 1(1), 1–78.

[8] Pinto, L., & Gupta, A. (2016). Supersizing self-supervision: Learning to grasp from 50K tries and 700 robot hours. *Proceedings of the IEEE International Conference on Robotics and Automation*, 3400–3407.

[9] Levine, S., Pastor, P., Krizhevsky, A., & Quillen, D. (2016). Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *arXiv preprint arXiv:1603.02199*.

[10] Keskin, C., Kırac, F., Kara, Y. E., & Akarun, L. (2013). Real time hand pose estimation using depth sensors. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 415–422.

[11] Park, C., Chan, S., Jenkins, O. C., & Thomas, S. (2008). Real-time 3D gesture-based user interface with a depth camera. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2271–2278.

[12] Kim, T., & Rehg, J. M. (2014). Joint spatio-temporal structured models for human gesture recognition in RGB-D videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3173–3180.

[13] Ren, Z., Yuan, J., Zhang, Z., & Yang, J. (2013). Robust part-based hand gesture recognition using Kinect sensor. *IEEE Transactions on Multimedia*, 15(5), 1110–1120.

[14] Rahman, M. M., Islam, M. R., Sarker, S. K., & Rahman, M. H. (2018). Robust real-time hand gesture recognition based on finger geometry. *Proceedings of the 2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 1–5.

[15] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117.

[16] Asif, U., Roy-Chowdhury, A. K., & Derr, K. (2020). Machine learning methods for hand gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(12), 3210–3227.

[17] Lin, M., Chen, Q., & Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.

[18] Breuer, P., Schmitt, C., & Klatzky, R. (2009). In-hand manipulation with a multifingered robot hand. *Proceedings of the IEEE Transactions on Robotics*, 25(1), 1–9.

[19] Chaudhary, A., Raheja, J. L., & Das, K. (2011). Intelligent approaches to interact with machines using hand gesture recognition in natural way: A survey. *International Journal of Computer Science and Information Technology*, 3(1), 1–15.

[20] Tanaka, Y., Okuno, H., & Ishikawa, S. (2015). Dynamic hand gesture recognition for human-robot interaction using CNN. *Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1926–1931.

[21] Suzuki, M., Matsuzawa, T., & Yamashita, A. (2018). Implementation of real-time hand gesture recognition for robot control. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2328–2335.

[22] Xu, C., Cheng, L., & Zheng, Y. (2017). Hand gesture recognition based on deep learning for human-computer interaction. *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 4290–4294.

[23] Cippitelli, E., Gasparrini, S., & Spinsante, S. (2018). A human-robot interface for robotic hand control using Kinect-based gesture recognition. *Sensors*, 18(7), 1–14.

[24] Hussein, M. E., Elsayed, M. E., & El-Far, A. I. (2019). Real-time hand gesture recognition using artificial neural networks and Kinect V2. *Proceedings of the 2019 IEEE International Conference on Robotics and Automation (ICRA)*, 2275–2280.

[25] Wang, P., Li, W., Gao, Z., & Tang, C. (2016). Large-scale isolated hand gesture recognition using convolutional neural networks. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 75–82.

[26] Choi, Y. M., Lee, J., & Lee, K. (2013). Hand gesture recognition using deep learning. *Proceedings of the 2013 International Conference on Information and Electronics Engineering (ICIEE)*, 155–160.

[27] Abavisani, M., & Patel, V. M. (2018). Multimodal hand gesture recognition using deep learning. *Proceedings of the 2018 IEEE International Conference on Image Processing (ICIP)*, 401–405.

[28] Yin, X., & Liu, X. (2018). Hand gesture recognition and control for a robot arm based on vision using deep learning. *Proceedings of the IEEE International Conference on Image Processing*, 2348–2353.

[29] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceed of the IEEE Conference on Computer Vision and Pattern Recognition, 580–587.*

[30] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

[31] Qi, H., Brown, J., & Yao, S. (2017). Dynamic hand gesture recognition using CNNs for humanoid robot control. *Journal of Robotics and Automation*, 32(4), 452–460.

[32] Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). Highway networks. *arXiv preprint arXiv:1505.00387*.

[33] Gupta, A., & Chidambaram, K. (2020). Hand gesture recognition for robotic control using deep learning. *Proceedings of the International Conference on Robotics and Automation*, 3245–3250.

[34] Zhang, H., Yin, Z., & Li, C. (2019). A survey of vision-based hand gesture recognition. *IEEE Transactions on Multimedia*, 15(3), 249–263.

[35] Rastgoo, R., Kiani, K., & Escalera, S. (2018). Multi-modal deep hand gesture recognition. *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR)*, 3445–3450.

[36] Hjelm, R. D., & Lafleche, J. (2020). Hand gesture recognition and control for robotic systems using convolutional neural networks. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4567–4574.

[37] Sermanet, P., & LeCun, Y. (2019). Traffic sign recognition with multi-scale convolutional networks. *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 354–358.

[38] Zhou, X., Wei, T., & Wang, S. (2020). Real-time gesture recognition using deep learning techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(2), 490–502.

[39] Liu, W., Anguelov, D., Erhan, D., & Szegedy, C. (2016). SSD: Single shot multibox detector. *Proceedings of the European Conference on Computer Vision (ECCV)*, 21–37.

[40] Gu¨ler, R. A., Neverova, N., & Kokkinos, I. (2018). DensePose: Dense human pose estimation in the wild. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7297–7306.

[41] Yamashita, A., Suzuki, M., & Yamamoto, T. (2017). Application of convolutional neural networks in robotic hand gesture recognition. *Proceedings of the 2017 IEEE International Conference on Robotics and Automation*, 5675–5681.

[42] Chen, X., Yang, Y., Zhang, X., & Xu, S. (2019). Gesture recognition for real-time human-robot interaction using machine learning. *Journal of Robotics and Autonomous Systems*, 94, 93–105.

[43] Luo, J., Ye, G., & Shi, J. (2020). Vision-based hand gesture control system for robotic operations. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(3), 290–298.

[44] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

[45] Shukla, R., & Jain, A. (2018). Gesture recognition for controlling robots using deep learning techniques. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4568–4575.

[46] Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2261–2269.

[47] Zhang, X., Wang, W., & Luo, J. (2019). Visual-based hand gesture recognition for robotic control using transfer learning. *IEEE Transactions on Robotics*, 33(1), 23–34.

[48] Xu, X., Zhang, Q., & Hu, J. (2021). Gesture recognition-based robotic control using recurrent neural networks. *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 1225–1231.

[49] Paszke, A., Gross, S., & Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *Proceedings of the Advances in Neural Information Processing Systems*, 8024–8035.

[50] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the Advances in Neural Information Processing Systems*, 91–99.