IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 1, August 2024

AI Virtual Mouse using Computer Vision and Media Pipe with Mobile Net Architecture

Ayisha Nadirsha[1], Riyad A[2], Harikrishnan S R[3]

Student, MCA,CHMM College for Advanced Studies, Trivandrum, India[1]

Assistant Professor, MCA,CHMM College for Advanced Studies, Trivandrum, India[2]

Assistant Professor,MCA,CHMM College for Advanced Studies, Trivandrum, India[3]

Abstract: *The AI Virtual mouse project is a challenge for individuals with physical disabilities and those affected by autism. This project presents an innovative solution: an AI Virtual Mouse, powered by computer Vision and the MobileNet architecture. This system not only improves accessibility but also addresses the pressing need for contactless and touchless interactions in a world increasingly concerned about health and hygiene. By using the MobileNet architecture, this system accurately interprets hand gestures to control cursor movements, eliminating the need for physical contact. Designed with accessibility in mind, the AI Virtual Mouse empowers individuals with physical disabilities and those affected by autism to navigate computers effortlessly, fostering greater independence and inclusion. The contactless nature of the interface also aligns with the increasing demand for hygienic solutions, minimizing the risk of germ transmission in public and personal spaces. This technology represents a significant advancement in the realm of accessible computing, offering a practical and intuitive alternative to traditional input methods.*

Keywords: Machine learning, Deep learning, Convolutional Neural Network, MobileNet Algorithm

I. INTRODUCTION

Despite significant technological advancements, individuals with autism or disabilities often encounter substantial challenges when interacting with computers and digital devices. Traditional input methods, like the mouse and keyboard, can be cumbersome for those with motor impairments or cognitive differences, limiting their ability to engage fully with digital content. Current accessibility solutions frequently fall short of delivering a seamless and intuitive experience. Many assistive technologies are complex, costly, or insufficiently adaptable to the diverse needs of users. Moreover, these technologies often fail to leverage recent advancements in AI and computer vision, which could greatly enhance usability and accessibility. There is a clear need for an effective, intuitive, and accessible interface that enables individuals with autism or disabilities to control a computer cursor effortlessly. An ideal solution would employ advanced technologies, such as Convolutional Neural Networks (CNNs) and computer vision, to recognize and interpret hand gestures in real-time, ensuring precise and smooth cursor movements. By addressing these challenges, the proposed system aims to bridge the gap between current technological capabilities and the accessibility needs of individuals with autism or disabilities. This innovation promotes greater independence and significantly enhances the user experience, empowering users to interact with digital environments more effectively and inclusively.

II. LITERATURE REVIEW

In recent years, there has been a significant shift towards developing technologies that enhance accessibility for individuals with disabilities. The integration of artificial intelligence (AI) and computer vision in creating assistive technologies has opened new avenues for innovation. The abstract under review introduces an AI Virtual Mouse, leveraging Computer Vision and the MobileNet architecture to provide a touchless interface, which is particularly beneficial for individuals with physical disabilities and those affected by autism. This literature review explores the existing body of research relevant to AI-driven assistive technologies, computer vision applications in accessibility, and the specific contributions of the MobileNet architecture. The use of AI in accessibility has garnered substantial attention in academic and technological circles. AI-driven assistive technologies aim to bridge the gap between individuals with disabilities and the digital world. According to Lazar et al. (2015), AI applications have been pivotal in developing

DOI: 10.48175/IJARSCT-19311

tools that cater to the needs of people with varying disabilities, including visual, auditory, and motor impairments. For instance, screen readers for the visually impaired, speech-to-text systems for the hearing impaired, and voice-controlled assistants for individuals with motor disabilities are some notable advancements in this domain. The COVID-19 pandemic has underscored the importance of contactless interactions to minimize the spread of pathogens. Contactless technologies, as noted by Nguyen et al. (2020), have become critical in various sectors, including healthcare, retail, and public services. The AI Virtual Mouse, by enabling touchless computer interactions, addresses both accessibility and hygiene concerns, making it a timely innovation in the context of global health challenges.

III. PROPOSED METHOD

The proposed system is an AI-based virtual mouse specifically designed to enhance accessibility and usability for individuals with autism or disabilities. By utilizing advanced technologies such as Convolutional Neural Networks (CNNs), deep learning, and computer vision techniques, this innovative virtual mouse allows users to control their computer cursor using intuitive hand gestures detected by a camera. The system employs CNNs to detect and classify these gestures in real-time, accurately mapping them to corresponding mouse movements and actions like clicking and dragging. Computer vision algorithms track hand positions, ensuring smooth and precise cursor control while providing a touchless and intuitive interface for computer interaction. This system aims to empower individuals with autism or disabilities, enabling them to navigate digital environments effortlessly and independently. The AI-based virtual mouse offers touchless interaction and intuitive control, significantly enhancing accessibility. Users can customize gestures to suit their preferences, making the system adaptable to various needs and versatile in its applications. Additionally, the system’s flexibility means it can be tailored to fit different user requirements, providing a personalized experience that accommodates the unique challenges faced by each individual. This adaptability ensures that the system can be used in a wide range of scenarios, from educational settings to professional environments, increasing its usefulness and impact.

IV. ALGORITHM

Convolutional Neural Network(CNN)

Convolutional Neural Networks (CNNs) are a class of deep learning models that have revolutionized the field of computer vision. They are specifically designed to process and analyze grid-like data structures, such as images, by automatically learning to extract and recognize patterns and features from the input data. CNNs have become the backbone of various applications, including image classification, object detection, facial recognition, and even tasks beyond vision, like natural language processing and speech recognition.These layers are followed by pooling layers, which reduce the spatial dimensions while retaining the most important features. This architecture enables CNNs to efficiently recognize patterns and objects in images. CNNs are widely used for tasks like image classification, object detection, and image generation because of their ability to learn and generalize from visual data. They significantly outperform traditional methods by leveraging deep layers to extract increasingly complex features from raw image data.

MobileNet Vision Transformer (MViT) Algorithm

The MobileNet Vision Transformer (MViT) is a groundbreaking hybrid architecture that merges the efficiency of MobileNet with the attention mechanisms of Vision Transformers. By addressing some of the limitations of traditional models, MViT is positioned to significantly enhance performance in various applications, including auto verification systemsMViT provides robust authentication mechanisms that mitigate risks associated with fraudulent activities. By safeguarding sensitive transactions and legal documents, MViT helps protect against unauthorized access and manipulation.The MobileNet Vision Transformer represents a significant advancement in the field of image processing and verification systems. By combining the strengths of MobileNet and Vision Transformers, MViT addresses key challenges related to feature extraction and attention mechanisms. Its integration into auto verification systems enhances accuracy, improves efficiency, and ensures security, making it a valuable tool for various industries. As technology continues to evolve, MViT stands at the forefront of innovation, offering solutions that meet the demands of a rapidly changing digital landscape.

DOI: 10.48175/IJARSCT-19311

V. PACKAGES

NumPy

NumPy is its powerful array object, ndarray, which facilitates efficient storage and manipulation of numerical data. This array structure allows users to perform complex operations with ease, thanks to a comprehensive collection of mathematical functions. These functions cover a wide range of operations, including mathematical calculations, linear algebra, statistical analysis, and Fourier transforms.NumPy is essential in scientific computing, data analysis, and machine learning because of its speed and flexibility. The library's efficient array processing capabilities make it an invaluable tool for handling large datasets and performing computations quickly. NumPy also integrates seamlessly with other popular scientific libraries, such as SciPy and pandas, further enhancing its functionality and making it a cornerstone of the Python data science ecosystem. This integration enables users to leverage a broad array of tools and techniques for data manipulation and analysis, supporting complex workflows and facilitating the development of sophisticated models and applications. Consequently, NumPy is an indispensable resource for researchers, analysts, and developers working in various fields requiring numerical computation and data manipulation.

Computer Vision

computer vision goal is to replicate human vision, allowing computers to perform tasks like image classification, object detection, and facial recognition. Applications include autonomous vehicles, medical imaging, and augmented reality. Deep learning models, especially convolutional neural networks (CNNs), are commonly used due to their ability to learn hierarchical features. Despite advancements in deep learning and datasets, challenges remain, such as occlusions and varying viewpoints. Researchers continue to enhance model robustness and ethical considerations, with the potential to revolutionize industries.

Pytorch

The core of PyTorch is its multi-dimensional array structure, known as tensors, which are essential for building neural networks. Similar to NumPy arrays, tensors have additional features like automatic differentiation, crucial for backpropagation during training. PyTorch supports GPU acceleration, enabling faster computations on compatible hardware, which is vital for training large models with extensive datasets. The torch.autograd module is central to PyTorch's automatic differentiation and dynamic computation graphs, automatically tracking operations on tensors and creating a computation graph to efficiently compute gradients for backpropagation. This feature simplifies the implementation of complex neural network architectures and optimizers by eliminating the need for manual gradient calculations. PyTorch's modular design offers a range of predefined layers and loss functions through the torch module, simplifying network construction. Users can also create custom modules by subclassing the torch.nn.Module class. The training process in PyTorch generally involves four main steps: loading data, creating the model, computing loss, and optimizing. The torch.utils.data module aids in data loading and batching, allowing users to focus on the model and training processes. The torch.optim module provides various optimization algorithms, such as Stochastic Gradient Descent (SGD), Adam, and RMSprop, which users can fine-tune according to their specific needs. PyTorch also supports distributed training, allowing models to be trained across multiple GPUs or machines. Its active community has developed various libraries and extensions, such as torchvision for computer vision tasks and torchaudio for audio processing, further expanding its capabilities.

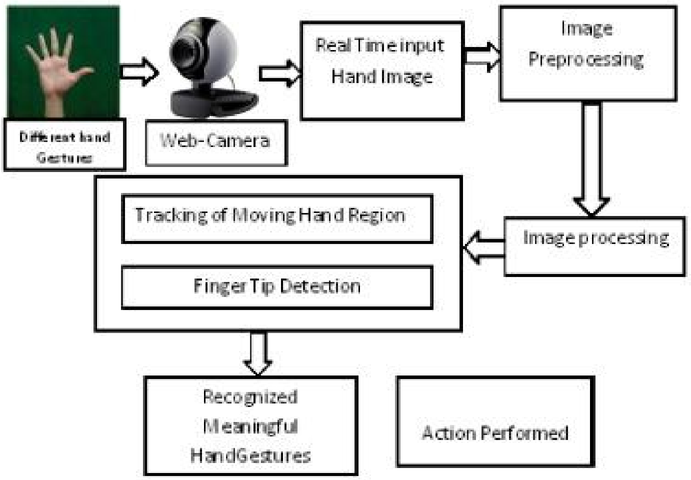
VI. EXPERIMENTAL RESULTS & PERFORMANCE EVALUATION

The implementation phase is the final and crucial step, involving user training and system testing to ensure the successful operation of the proposed system. After the system design phase, the next step is to implement and monitor the system to ensure it operates effectively and efficiently. The implementation process consists of three main phases: initial installation, system testing as a whole, and evaluation, maintenance, and control of the system. The implementation plan should be closely aligned with the action to implement and involves line management for key decisions and alternative plans. To realize the AI virtual mouse system using OpenCV, Autopy, and Mediapipe, the implementation proceeds through several interconnected stages. First, the environment is set up by installing Python with essential libraries, including OpenCV for computer vision tasks, Mediapipe for hand tracking, and Autopy for

DOI: 10.48175/IJARSCT-19311

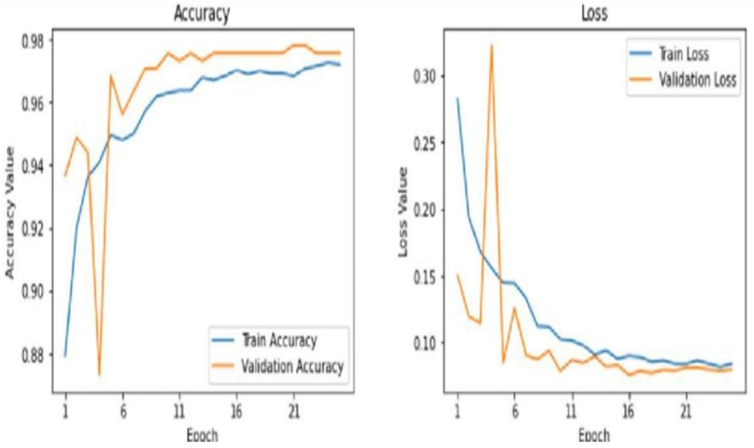
simulating mouse actions. These libraries are crucial for enabling real-time interaction between hand gestures and virtual mouse movements on the computer screen.

System Architecture



VII. ACCURACY GRAPH

Evaluating model accuracy is a crucial step in developing machine learning models to assess how well they perform in making predictions. In this discussion, we will briefly explore how to assess the accuracy of a regression model in R. Linear regression models are common examples of regression problems, characterized by targets that contain only real numbers. Errors indicate the extent of mistakes the model makes in its predictions. The basic concept of accuracy evaluation is to compare the original target values with the predicted values using specific metrics. Model accuracy changes during the training process, where the x-axis represents the number of training epochs (or iterations), and the y-axis shows the corresponding accuracy achieved by the model on the training or validation set.



VIII. LIMITATION

While the AI Virtual Mouse represents a significant step forward in accessibility, several limitations and challenges must be addressed. Firstly, the reliance on camera-based gesture recognition can pose issues in environments with poor lighting or background distractions, potentially reducing the system's accuracy and reliability. Additionally, the robustness of the MobileNet architecture in diverse and dynamic real-world settings needs further evaluation, as variations in hand gestures and user positions could affect performance. Another challenge lies in the initial setup and calibration process, which might be complex and time-consuming for users with severe physical disabilities. Ensuring that the system can be easily personalized and adapted to individual needs is crucial but can be technically demanding. Furthermore, there is a potential barrier related to the cost and availability of compatible hardware, which could limit accessibility for some users. Finally, while the system addresses the need for contactless interactions, ensuring consistent and seamless integration with various software and hardware platforms remains a challenge, necessitating ongoing updates and support. Addressing these limitations is essential to maximize the potential benefits of the AI Virtual Mouse for individuals with disabilities and autism.

IX. FUTURE SCOPE

In the future, the AI-based virtual mouse system can be enhanced by incorporating advanced machine learning techniques such as reinforcement learning to improve gesture recognition accuracy and adaptability over time. Integrating additional sensors, such as depth cameras or infrared sensors, could further enhance hand tracking precision in various lighting conditions and environments. Expanding the range of recognizable gestures and incorporating voice commands could provide even more intuitive and flexible interaction options. Additionally, implementing user-specific customization and adaptive learning features would allow the system to tailor its performance to individual user needs and preferences, improving usability. Collaboration with healthcare professionals and ongoing user feedback will be essential in refining the system to better serve individuals with autism or disabilities, ensuring it remains a cutting-edge and impactful assistive technology.

X. CONCLUSION

The development of an AI-based virtual mouse system specifically designed for individuals with autism or disabilities marks a significant advancement in assistive technology, addressing the shortcomings of conventional input devices. By harnessing Convolutional Neural Networks (CNNs), MediaPipe, and Autopy, this system allows for intuitive and accessible computer cursor control through hand gestures captured by a camera. The feasibility study validated the technical, economic, operational, and market viability of this solution, highlighting its potential to meet the increasing demand for assistive technologies. The system architecture, consisting of input, processing, control, and interface layers, ensures efficient and real-time interaction. The comprehensive dataset design phase, which involved collecting data from public sources and custom recordings, as well as thorough annotation, preprocessing, and augmentation, ensures robust and accurate gesture recognition. By implementing rigorous processes such as resizing, normalizing, augmenting, and balancing the dataset, the system is set to deliver reliable performance. This innovative solution ultimately aims to enhance the independence and user experience of individuals with autism or disabilities, promoting their engagement with digital environments.

REFERENCES

[1]. Altan, G., &Barışçı, N. (2020). A review of hand gesture and sign language recognition techniques in recent years. International Journal of Pattern Recognition a.

[2]. Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2020). OpenPose: IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(1), 172-186.

[3]. Gao, Y., Wang, Y., & Chen, Y. (2020). An intelligent vision-based human-computer interaction method for smart classrooms. Sensors, 20(17), 4754.

[4]. Guo, W., Liu, Y., & Yuan, X. (2021). Gesture recognition technology for human computer interaction: A survey. Journal of Ambient Intelligence and Humanized Computing, 12(10), 10011-10028.

[5]. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1725-1732).

[6]. MediaPipe. (n.d.). MediaPipe Hands: Real-time hand landmark estimation. Retrieved June 27, 2024, from

[7]. OpenCV. (n.d.). OpenCV: Open Source Computer Vision Library. Retrieved June 27, 2024, from

DOI: 10.48175/IJARSCT-19311

[8]. Patil, S. V., & Bang, S. S. (2021). Hand gesture recognition and its applications: A review. Artificial Intelligence Review. Advance online publication.

[9]. Shao, J., Li, B., & Zhang, R. (2020). Gesture recognition technology based on convolutional neural networks. Journal of Physics: Conference Series, 1558, 012060.

[10]. Wu, H., Mao, R., & Wang, H. (2020). Deep learning for image-based intelligent interaction:A survey. Information https://doi.org/10.1016/j.inffus.2020.04.008 Fusion, 63, 1-22.