Convolutional Neural Network for Hand Gesture Recognition using 8 different Gestures

Ms. B. Veena
Electronics and Communication Engineering
Institute of Aeronautical Engineering Dundigal, Hyderabad 500043

Vardhan Aluwala
Electronics and Communication Engineering
Institute of Aeronautical Engineering Dundigal, Hyderabad 500043

Vijay Kanth Marishetti
Electronics and Communication Engineering
Institute of Aeronautical Engineering Dundigal, Hyderabad 500043

Sravan Govardhanam
Electronics and Communication Engineering
Institute of Aeronautical Engineering Dundigal, Hyderabad 500043

Dhinakar Raj Bayikadi
Electronics and Communication Engineering
Institute of Aeronautical Engineering Dundigal, Hyderabad 500043

***Abstract*—This paper describes the implementation of hand gesture recognition which uses deep learning algorithm as CNN. Therefore, proposed technique is hassle free as control is not based on joysticks or switches. There are eight conditions considered for hand gesture as up, stop , right, No ,Left ,Down, clockwise and anticlockwise. There are many researchers worked on this area using different sensors, machine learning algorithms and deep learning algorithms. Limitations of the state of art techniques are studied in this paper and designed a new modified convolutional neural network (CNN) for gesture recognition. Dataset is created which generates 1000 Gray scale images for each type of gesture. Training modified CNN model gives prediction accuracy of 98.4024 percent while random forest machine learning classifier gives prediction accuracy of 69 ppercent. It is observed that proposed model gives better accuracy compared to state of art technique for hand gesture recognition. Obtained hand gesture class can be send to controller for further controlling in future like home automation robotic car movement etc.**

Keywords: Robot Car Movement, Gesture Recognition, Random Forest, Deep Learning, CNN, layer modification.)

I. Introduction

The use of hand gestures as a means of communication has been around for centuries, and it has become an es- sential part of human interaction. With the advancement of technology, hand gesture recognition has become a crucial aspect of human-computer interaction (HCI).[2] The ability to recognize and interpret hand gestures can enable comput- ers to understand human intentions and respond accordingly. This technology has numerous applications in various fields, including robotics, gaming, and healthcare.[4]

One of the primary applications of hand gesture recognition is in the field of robotics. Robots can be controlled using hand gestures, allowing humans to interact with them in a more natural and intuitive way.[8] This technology has the potential to revolutionize the way we interact with robots, making them more accessible and user-friendly. For instance, a robot can be programmed to perform tasks such as picking up objects or navigating through a space based on hand gestures.[1]

Another significant application of hand gesture recognition is in the field of gaming. Gesture-based gaming has become increasingly popular in recent years, with the introduction of gaming consoles such as the Nintendo Wii and the Xbox Kinect. These consoles use hand gesture recognition technol- ogy to allow players to control games using natural move- ments. This technology has opened up new possibilities for game developers, enabling them to create more immersive and interactive gaming experiences.[12]

Hand gesture recognition also has numerous applications in the field of healthcare. For instance, it can be used to control medical devices such as wheelchairs or prosthetic limbs.[14] This technology can also be used to develop systems that can recognize and interpret hand gestures made by patients with disabilities, enabling them to communicate more effectively with healthcare professionals.

In addition to these applications, hand gesture recognition has the potential to revolutionize the way we interact with computers. For instance, it can be used to develop systems that can recognize and interpret hand gestures made by users, en- abling them to control computers using natural movements.[2] This technology can also be used to develop systems that can recognize and interpret hand gestures made by people with disabilities, enabling them to interact with computers more effectively.

The development of hand gesture recognition technology has been made possible by advances in computer vision and machine learning. Computer vision algorithms can be used to recognize and interpret hand gestures, while machine learning algorithms can be used to develop systems that can learn from data and improve their performance over time.[6] These technologies have enabled researchers to develop systems that can recognize and interpret hand gestures with high accuracy, paving the way for a wide range of applications.

Despite the numerous applications of hand gesture recognition, there are still several challenges that need to be addressed. For instance, the accuracy of hand gesture recognition systems can be affected by factors such as lighting, background noise, and the complexity of the gestures being recognized[9]. Additionally, the development of hand gesture recognition systems requires large amounts of data, which can be time-consuming and expensive to collect. However, researchers are working to address these challenges, and the technology is rapidly advancing.

II. LITERATURE SURVEY

A. Representation learning: A review and new perspectives. IEEE Trans:

The efficacy of machine learning algorithms is often tied to how well data is represented, as different representations can either reveal or obscure underlying explanatory factors. While domain-specific knowledge can guide the design of these representations, generic priors also play a crucial role. This paper reviews recent advancements in unsupervised fea- ture learning and deep learning, exploring developments in probabilistic models, autoencoders, manifold learning, and deep networks. It highlights the need to address unresolved questions regarding the objectives for effective representation learning, the inference processes involved, and the connections between representation learning, density estimation, and manifold learning. The review also discusses three primary approaches—probabilistic models (including sparse coding and Boltzmann machines), reconstruction-based algorithms like autoencoders, and manifold-learning techniques. Understanding the interplay between these methods is a vibrant area of research that promises to enhance models by integrating the strengths of each paradigm.

B. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks:

The proliferation of image data online presents significant opportunities for advancing models and algorithms in image indexing, retrieval, and interaction. This paper introduces ”ImageNet,” a large-scale image database built on the WordNet ontology. ImageNet aims to cover a substantial portion of WordNet’s 80,000 synsets with 500-1000 high-quality, annotated images per synset, totaling tens of millions of images organized by semantic hierarchy. The database currently includes 12 subtrees with 5,247 synsets and 3.2 million images. Compared to existing datasets, ImageNet offers superior scale, diversity, and accuracy. The construction of this database involved data collection through Amazon Mechanical Turk. The paper also demonstrates ImageNet’s utility in object recognition, image classification, and automatic object clustering, highlighting its potential to significantly benefit the computer vision community.

C. Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition:

The proliferation of image data online presents significant opportunities for advancing models and algorithms in image indexing, retrieval, and interaction. This paper introduces ”ImageNet,” a large-scale image database built on the WordNet ontology. ImageNet aims to cover a substantial portion of WordNet’s 80,000 synsets with 500-1000 high-quality, anno- tated images per synset, totaling tens of millions of images organized by semantic hierarchy. The database currently in- cludes 12 subtrees with 5,247 synsets and 3.2 million images. Compared to existing datasets, ImageNet offers superior scale, diversity, and accuracy. The construction of this database in- volved data collection through Amazon Mechanical Turk. The paper also demonstrates ImageNet’s utility in object recog- nition, image classification, and automatic object clustering, highlighting its potential to significantly benefit the computer vision community.

D. Deep networks for early stage skin disease and skin cancer classification:

This study explores effective methods for robust skin dis- ease diagnosis, emphasizing two approaches: direct disease label classification and lesion characteristic identification. The authors advocate for focusing on lesion type tags as targets for automated diagnosis, arguing that this method provides a more practical and useful foundation for accurate disease diagnosis. Utilizing convolutional neural networks (CNNs), the study evaluates both disease-targeted and lesion-targeted classifiers. With a large dataset of 75,665 skin images, the research finds that lesion-targeted classification significantly outperforms disease-targeted approaches, achieving a mean av- erage precision (mAP) of 0.70 compared to 0.42, highlighting the effectiveness of focusing on lesion characteristics.

E. Batch normalization: Accelerating deep network training by reducing internal covariate shif:

Training deep neural networks is often hindered by internal covariate shift, where the distribution of each layer’s inputs changes as the parameters of preceding layers are updated. This issue necessitates lower learning rates and precise pa- rameter initialization, complicating model training. To address this, Batch Normalization normalizes layer inputs for each mini-batch, which stabilizes learning and allows for higher learning rates. This technique also acts as a regularizer, potentially reducing or eliminating the need for Dropout. Applied to a state-of-the-art image classification model, Batch Normalization achieved the same accuracy with 14 times fewer training steps and significantly improved performance on ImageNet classification, surpassing human-level accuracy.

III. RELATED WORK

Hand gesture recognition for controlling robotic systems has demonstrated various approaches and technologies to enhance accuracy and responsiveness. One significant area of research involves the use of Convolutional Neural Networks (CNNs) for gesture classification. Studies have shown that CNNs are highly effective in identifying and categorizing hand gestures due to their ability to learn complex features from image data.[4] For instance, research by Krizhevsky et al. (2012) demonstrated the power of deep CNNs in achieving state-of-the-art performance in image classification tasks, laying the foundation for their application in gesture recognition. Other works, such as those by Li et al. (2017), have utilized CNNs for safety helmet detection, illustrating the model's capability to handle various object detection tasks.[13] This approach has been adapted for hand gesture recognition, focusing on improving classification accuracy and real-time performance. In addition to CNNs, other machine learning techniques have

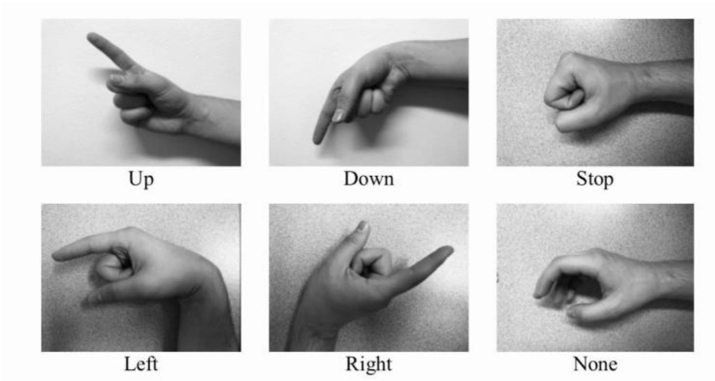

Fig. 1. Proposed Method Block Diagram

been employed for gesture recognition. For example, random forest classifiers have been used for hand gesture recogni- tion, as seen in studies like those by Vishnu et al. (2017). However, these methods often face limitations in accuracy and processing speed compared to CNN-based approaches. Recent advancements have focused on integrating background subtraction and pre-processing techniques to enhance the per- formance of gesture recognition systems. Techniques such as background subtraction, as used in the work by Doungmala and Klubsuwan (2016), help isolate hand gestures from static backgrounds, thereby improving recognition accuracy and reducing training time.

Furthermore, the integration of gesture recognition with robotic systems has seen significant progress. Research by Chen et al. (2015) explored augmented reality and gesture recognition to enhance social skills in individuals with autism, highlighting the potential for gesture-based control in various applications.[10] The combination of gesture recognition with robotic control systems has been exemplified in works like those by Peral et al. (2022), which focused on real-time hand gesture recognition for human-robot interaction. These studies underscore the ongoing efforts to refine gesture recognition technologies and their applications, paving the way for more sophisticated and user-friendly systems.

IV. PROPOSED METHOD

The proposed method for hand gesture control involves a systematic approach integrating deep learning and image processing techniques. The process begins with the creation of a specialized dataset consisting of 5,982 grayscale images categorized into six distinct gestures: 'Backward', 'Forward', 'Left', 'No-Motion', 'Right', and 'Stop'. Each gesture class contains 997 images. To build this dataset, a webcam is used to capture live images, followed by background subtraction using KNN to isolate the hand gestures. The images are then converted to grayscale, Gaussian Blur is removed, and hand segmentation is achieved through thresholding and contour detection. The final dataset is organized into folders based on the hand gestures, ensuring a robust collection for training. Pre-processing is a crucial step to ensure that the dataset is in a

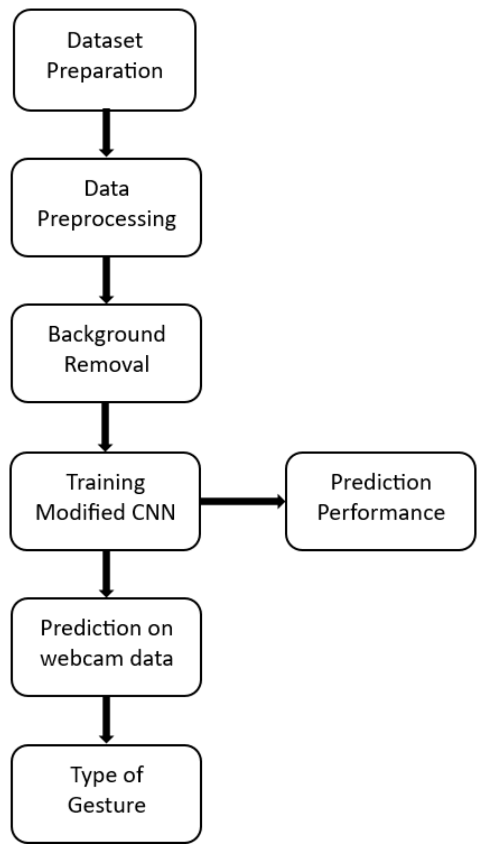

Fig. 2. Proposed Method Block Diagram

suitable format for training the Convolutional Neural Network (CNN). The images are resized to a consistent size and converted to grayscale to standardize the input. Background removal is achieved using the 'createBackgroundSubtractor- MOG2' method, which isolates moving objects from static backgrounds. This step simplifies the CNN's task by focusing solely on the hand gestures, thus enhancing the accuracy of gesture recognition.

Training the modified CNN with the pre-processed dataset is both efficient and effective. The CNN, which includes Con- volutional, Max-Pooling, Dense, and Flatten layers, achieves a high prediction accuracy of 99.6 percent with a training time of just 10 minutes. In contrast, a dataset without background removal would require approximately one hour for training. Once trained, the CNN model predicts hand gestures in real-time from webcam input. These predictions are used to control various functions, with results also communicated through a speaker for auditory feedback. This method highlights the efficiency of combining deep learning with image processing for accurate and responsive hand gesture recognition.

V. Conclusion

The results of the proposed modified Convolutional Neural Network (CNN) model demonstrate a significant improvement over traditional methods such as the random forest machine learning classifier for controlling a robot car. The modified CNN incorporates a unique layer structure that has been optimized to enhance accuracy, leading to an impressive prediction accuracy of 98.4024 percent. In contrast, the random forest classifier achieves only 69 percent accuracy when tested with real-time hand gestures captured through a webcam. This substantial increase in accuracy highlights the effectiveness of the modified CNN in interpreting hand gestures, which directs the robot car to move in five distinct directions, including a 'No Motion' gesture. The high accuracy of the CNN model underscores its potential for more precise and reliable control in robotic applications.

Fig. 3. Training accuracy and loss graph

Looking ahead, the system's capabilities can be further expanded by integrating it into a complete real-time application using a Raspberry Pi controller. This would enable the addition of GPS and other sensors to provide comprehensive location and environmental data, enhancing the robot car's functionality and adaptability. Additionally, the results obtained from this system could be shared and analyzed on a private cloud platform, facilitating further research and development. The integration of these elements promises to advance the state of the art in gesture-controlled robotics, making the system more versatile and applicable in various real-world scenarios. Overall, the proposed modifications and future enhancements signify a significant step towards more advanced and accurate robotic control systems.

References

[1] M. Peral *et al.*, "Efficient Hand Gesture Recognition for Human-Robot Interaction," *IEEE Robotics and Automation*, vol. 7, no. 4, 2022.
[2] F. Arce and J. M. G. Valdez, "Accelerometer-Based Hand Gesture Recognition Using Artificial Neural Networks," in *Soft Computing for Intelligent Control and Mobile Robotics, Studies in Computational Intelligence*, 2011.
[3] R. Wang and J. Popovic, "Real-time Hand-tracking with a Color Glove," *ACM Transactions on Graphics*, 2009.
[4] C.-C. Chang, I.-Y. Chen, and Y.-S. Huang, "Hand Pose Recognition Using Curvature Scale Space," in *IEEE International Conference on Pattern Recognition*, 2002.
[5] C. N. S., D. Prabha, and M. Z. Kurian, "Static Hand Gesture Recognition System for Device Control," *International Journal of Electrical, Electronics and Data Communication*, vol. 3, April 2015, ISSN: 2320-2084.
[6] A. Nale, A. Barkade, P. Ekhande, and S. Dayal, "Hand Gesture Recognition for Robotic Car Control," *International Journal of Advanced Research in Computer Science and Technology (IJARCST)*, vol. 3, no. 1, 2015.
[7] S. Jayaraman, S. Esakkirajan, and T. Veerakumar, *Digital Image Processing*, New Delhi: McGraw Hill Education, August 10, 2013.
[8] M. O. Qureshi and R. S. Syed, "The Impact of Robotics on Employment and Motivation of Employees in the Service Sector, with Special Reference to Health Care," *Safety and Health at Work*, vol. 5, pp. 198–202, 2014.
[9] Y. A. Fernando, Mathath, and M. A. Murshid, "Improving Productivity: A Review of Robotic Applications in Food Industry," *International Journal of Robotics Applications and Technology*, vol. 4, pp. 43–62, 2016.
[10] B. Fang, F. Sun, H. Liu, and C. Liu, "3D Human Gesture Capturing and Recognition by the IMMU-based Data Glove," *Neurocomputing*, vol. 277, pp. 198–207, 2018.
[11] Y. Shi, R. Taib, and S. Lichman, "Gesturecam: A Smart Camera for Gesture Recognition and Gesture-Controlled Web Navigation," in *Proceedings of the 9th International Conference on Control, Automation, Robotics and Vision*, Singapore, 5–8 Dec. 2006, pp. 1–6.
[12] T. S. Dahl, P. Gibbons, and O. Jones, "Identification and Production of Simple Tactile Gestures," in *Proceedings of IEEE International Conference on Human-Robot Interaction Workshop on Advances in Tactile Sensing and Touch-Based Human-Robot Interaction*, Boston, Massachusetts, Mar. 2012, doi: 10.13140/2.1.3165.344.
[13] N. K. Patel, S. B. Patel, and M. M. Ammar, "Accelerometer Based Gesture Controlled Wheelchair with GPS, GSM Navigation," *International Journal of Innovative Emerging Research in Engineering*, vol. 2, pp. 110– 113, 2015.
[14] C.-H. Chen, I.-J. Lee, and L.-Y. Lin, "Augmented Reality-Based Self- Facial Modeling to Promote the Emotional Expression and Social Skills of Adolescents with Autism Spectrum Disorders," *Research in Developmental Disabilities*, vol. 36, pp. 396–403, 2015.