International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Hand Gesture Recognition Using Convolutional Neural Networks

[1]Veluru Karthik Reddy, [2]Vanapalli Durga Prasanth, [3]R. Shiva Rama Krishna, [4]Naidu Sri lekha, [5]Jyothi N.M

Submitted: 29/01/2024 **Revised**: 07/03/2024 **Accepted**: 15/03/2024

Abstract: Hand gestures play a crucial role in communication and are essential in various scenarios where verbal communication is not possible. For instance, traffic policemen, newsreaders, airport staff, and gamers often rely on hand signals to communicate. Therefore, there is a growing need for robust hand pose recognition (HPR) methods that can identify hand gestures accurately. However, the current state-of-the-art HPR methods struggle with identifying hand gestures in the presence of cluttered backgrounds. To address this challenge, we propose a deep learning framework based on convolutional neural networks (CNNs) to identify hand postures regardless of hand size, location in the image, and background clutter. Our proposed CNN-based approach eliminates the need for feature extraction and learns to recognize hand poses without explicit foreground segmentation. This method effectively identifies hand poses, even in the presence of complex and varying backgrounds or poor lighting conditions. We have conducted several experiments, which demonstrate the superiority of our proposed method over state-of-the-art datasets. Our approach significantly improves the accuracy of hand pose recognition, making it more reliable and efficient for a wide range of applications. Therefore, our proposed method has significant potential for use in real-world scenarios, such as traffic management, sign language interpretation, and virtual reality gaming. Overall, our results suggest that deep neural networks can provide a robust and effective solution for hand gesture recognition tasks.

Keywords: *HPR,CNN,Segmentation,Background Clutter,Virtual Reality,Neural Network*

1. Introduction

Hand gesture recognition has become a key field of study as a result of its applications in a number of sectors, including human-computer interaction, sign language interpretation, gaming, and virtual reality. In particular, convolutional neural networks (CNNs) have excelled at picture classification tasks and have been expanded to recognise hand movements. Algorithms for hand position identification are, however, limited in their ability to recognise hands accurately due to problems including cluttered backgrounds, different-sized hands, and poor illumination. In this article, we present a CNN-based approach for hand position detection that is robust to these challenges and does not necessitate explicit feature extraction. We run comprehensive tests on benchmark datasets to demonstrate that our suggested solution performs better than cutting-edge methods.Our approach has potential applications in a variety of industries, including sports, healthcare, and more, in addition to sign language interpretation and human-computer interaction. Our proposed method for hand gesture detection using CNNs offers a number of advantages over traditional hand pose identification methods. First of all, it eliminates the need for time-consuming and expensive manual segmentation. Second, the ability of our approach to rapidly distinguish hand gestures is crucial for virtual reality and gaming applications. Our method is flexible and can account for differences in hand size and location, making it appropriate for a wider range of contexts. Our proposed strategy may alter how we interact with technology in general and make it more intuitive and natural.

[1]*Department of Computer Science and Information Technology, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522502, AP, India*
Email: 2000090063csit@gmail.com
[2]*Department of Computer Science and Information Technology, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522502, AP, India*
Email: 2000050008csit@gmail.com
[3]*Department of Computer Science and Information Technology Koneru Lakshmaiah Education Foundation, Vaddeswaram 522502, AP, India*
Email: 2000090062csit@gmail.com
[4]*Department of Computer Science and Information Technology, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522502, AP, India*
Email: 2000090025csit@gmail.com
[5]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522502, AP, India*
Email: jyothiarunkr@gmail.com

2. Related Work

Weijie Ke[1] developed a spiking convolutional neural network that reduces training and dataset processing time using EMG signal energy density maps. Jiabin Xu[2] proposed a wireless signal-based technique for dynamic hand gesture identification using the 802.11a preamble signal. Zhi-hua Chen[3] used background subtraction to recognize fingers and hand regions, while Heung-Il Suk[4] used a dynamic Bayesian network for continuous hand gesture recognition with a high recognition rate. Chenyang Zhang[5], Hong Cheng[6], Yu-Ting Li[7], Aashni Haria[8], Gerhard Rigoll[10], and Mahdi Abavisani[11] proposed various techniques for recognizing hand gestures. Zhang's[5] approach uses depth video and combines Edge Enhanced Depth Motion Map with Histogram of Gradient descriptor. Cheng's[6] Windowed Dynamic Time Warping technique introduces a parameterized searching window in

the cost matrix of the traditional DTW approach. Li's[7] hierarchical elastic graph matching method assigns different hierarchy levels to nodes in the graph. Haria's[8] system is marker less and cost-effective. Aparna Mohanty[9] suggested a convolutional neural network to accurately recognize hand positions without feature extraction. Rigoll's[10] Motion Fused Frames combines motion information with RGB images. Abavisani's[11] framework leverages knowledge from multiple modalities by training separate 3D-CNNs. MUNEER AL-HAMMADI[12] proposed a dynamic hand gesture recognition system using deep learning techniques that represent gestures using local and global body configuration data . Chunyu Xie [13] collected a large gesture database for mobile-based gesture recognition and proposed a Fisher criterion for F-BLSTM network to classify mobile hand gestures. George B. Mo [14] introduced Gesture Knitter, a tool for creating hand gestures for mixed reality applications. Priyanka Parvathy [15] and Haitham Hasan [16] proposed vision-based hand gesture recognition systems using machine learning and skin colour detection, respectively. Pavlo Molchanov [17] developed an effective method for dynamic hand gesture recognition with 3D CNNs. R.S. Jadon [18] proposed a vision-based gesture recognition system using a simple web camera and trained a neural network for hand detection, finger counting, and direction estimation. Pedro Neto's [19] method enables real-time and continuous hand gesture spotting for intuitive robot control, utilizing ANNs for gesture classification. Mohammad Mahmudul Alam [20] introduced a CNN-based approach for unified gesture recognition and fingertip position prediction in egocentric vision, simplifying the process into a single step.

3.Proposed Model

A . Data Flow of the Model :

The process of recognizing hand gestures involves capturing images or videos from a camera and processing them through an image processing pipeline. This pipeline applies various techniques, including thresholding, edge detection, and morphological operations, to segment the hand regions within the images. Object tracking algorithms such as Kalman filters or optical flow are used to track these segmented regions across frames. From the tracked regions, relevant features like hand shape, orientation, and motion

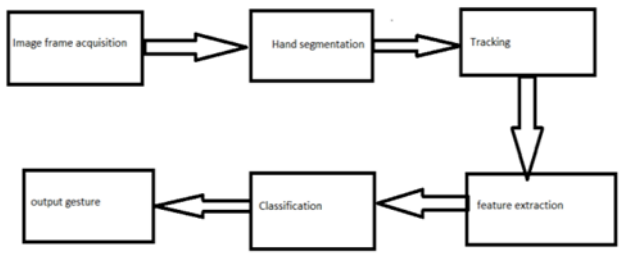

Fig. 1.Data Flow of the model

learning algorithms such as support vector machines or deep neural networks, which classify the hand gestures. The output of this classification process is a predicted gesture label, which can be used to control different applications or devices.

B .Model description :

To categorise gestures in the HaGRID dataset, the proposed CNN model applies four convolutional layers, four max pooling layers, two dense layers, and one flattening layer. The dataset is separated into training and testing sets depending on user-id, with 92% of the data utilised for training and 8% for testing. It contains 552,992 Full HD RGB photos classified into 18 gesture classes, including a "no gesture" class containing 123,589 images. Convolutional layers employ filters to examine the input images and find important details. In order to decrease the dimensionality of the data and increase the computational efficiency of the model, max-pooling layers down sample the output of the convolutional layers. Once the output of the max-pooling layers has been flattened, it is fed into the dense layers, which are fully connected layers. The output of the last dense layer is then sent into the soft max activation function, which gives each of the 18 gesture classes a probability value and enables the classification of the images. The model was trained using a suitable optimizer and loss function in addition to the architecture. By modifying how the model learns from the data, these hyperparameters aid in optimising the model's performance. The performance of the model can be further enhanced by modifying these hyperparameters as well as additional elements like activation functions and learning rate. Using the CNN model, we were able to achieve an overall accuracy of 99.96% on the Hagrid dataset.

C . Algorithm :

1. Gather a picture dataset of hand gestures.

2. Divide the dataset into training, validation, and test sets after normalising the pixel values and shrinking the photos to a standard size.

3. Choose the size, quantity, and kind of filters as well as the activation functions for each layer while designing the CNN's architecture. Other considerations include the amount of convolutional, pooling, and fully connected layers.

4. Assemble the CNN model by choosing an appropriate optimizer like Adam or RMS Prop, a loss function like categorical cross-entropy, and evaluation measures like accuracy.

5. To avoid overfitting, train the CNN model on the training dataset for the appropriate number of epochs while evaluating its performance on the validation dataset.

6. Using the evaluation measures, assess the model's

performance on the test dataset.

7. Test the performance of the fine-tuned CNN model by tweaking the hyperparameters, such as the learning rate, the number of filters, the sizes of the filters, and the number of layers.

8.Use a system or application for real-time hand gesture recognition to deploy the CNN model.

D . Model Architechture

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 150, 150, 32)	832
max_pooling2d_1 (MaxPooling2	(None, 75, 75, 32)	0
conv2d_2 (Conv2D)	(None, 75, 75, 64)	18496
max_pooling2d_2 (MaxPooling2	(None, 37, 37, 64)	0
conv2d_3 (Conv2D)	(None, 37, 37, 96)	55392
max_pooling2d_3 (MaxPooling2	(None, 18, 18, 96)	0
conv2d_4 (Conv2D)	(None, 18, 18, 96)	83040
max_pooling2d_4 (MaxPooling2	(None, 9, 9, 96)	0
flatten_1 (Flatten)	(None, 7776)	0
dense_1 (Dense)	(None, 512)	3981824
activation_1 (Activation)	(None, 512)	0
dense_2 (Dense)	(None, 10)	5130

Total params: 4,144,714
Trainable params: 4,144,714

Fig 2.Model Architecture

4. Experiments And Results

A Convolutional Neural Network (CNN) was trained using images of hand gestures from two standard datasets, with consideration given to both plain/uniform and complex backgrounds. The CNN architecture remained the same throughout all experiments, with weights trained using conventional backpropagation. Initially, the CNN model was trained on the ASL dataset without any data augmentation. Table provides details of the training procedure, including the training and testing sizes, batch size, and number of epochs. Fig. 1 shows the variation of accuracy with the number of epochs during CNN training. The efficacy of the CNN was demonstrated by achieving a test accuracy of 99.96% on the Hagrid dataset for images of hand gestures with complex backgrounds. This accuracy is higher than the state-of-the-art result of 99.92% reported on the ASL dataset.

4.1 Dataset and Execution.

ASL , HaGrid and other Dataset with 20 classes are the datasets used in this model for experimentation. ASL consists of 34627 images of different hand gestures (hand closing, hand opening, wrist flexion, wrist extension, index finger straightening, and thumb straightening.) with image dimension's 28x28. And HaGrid dataset consists of 552992 images with same dimensions and hand gesture categories of ASL. The Leap Gesture dataset consists of 24000 images depicting 20 distinct hand gestures. Specifically, there are 900 images per gesture category for training purposes, and 300 images per category for testing purposes. Execution of the model is done on the Google cloud platform called as goggle research colabs Which Provides 12GB NVIDIA Tesla K80 GPU , CPU: 1xsingle core hyper threaded Xeon Processors @2.3Ghz i.e.(1 core, 2 threads) and a disk space of 68.40GB.

Data	No.of classes	Training set	Testing set	Batch size	Epochs	Accuracy
ASL dataset	24	27,455	7,172	64	10	99.92
HaGrid	18	509,323	43,669	64	10	99.96
LeapGesture	20	18000	6000	64	10	100 (Overfitting model)

Fig 3. Tabular representation of Datasets

Performnce measures	Model-1	Model-2	Proposed Model
Precision	0.97	0.98	0.98
Recall	0.99	0.96	0.99
F1-score	0.98	0.98	0.99
Accuracy	99.92	100(overfitting model)	99.96

Fig 4. Performance metrics of models used

4.2 Graphical Representation of Results

Plots for accuracy and loss for Model-1

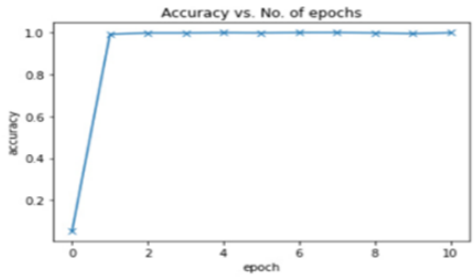

Fig 5. Accuracy graph of Model using Dataset

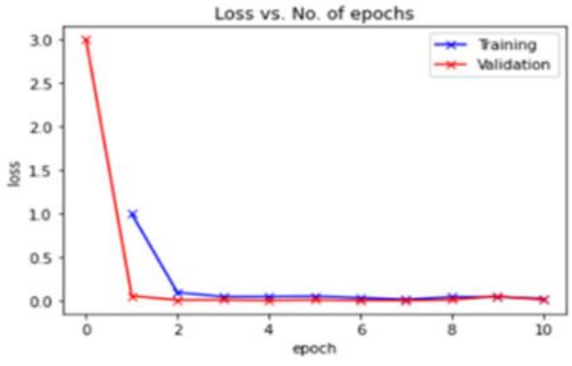

Fig 6. Loss graph of model using Dataset3

Plots for accuracy and loss for Model-2 using HaGrid dataset.

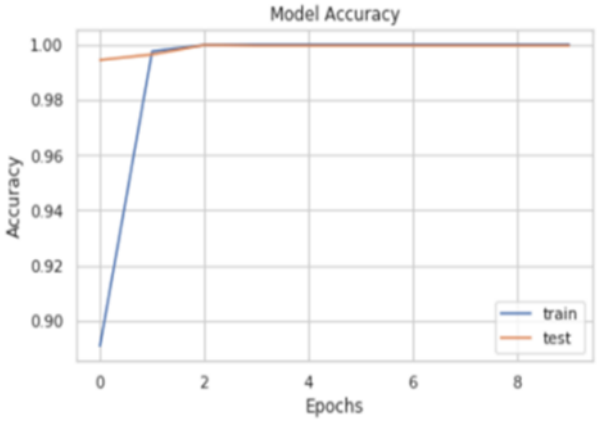

Fig 7. Accuracy graph of Model using HaGrid Dataset

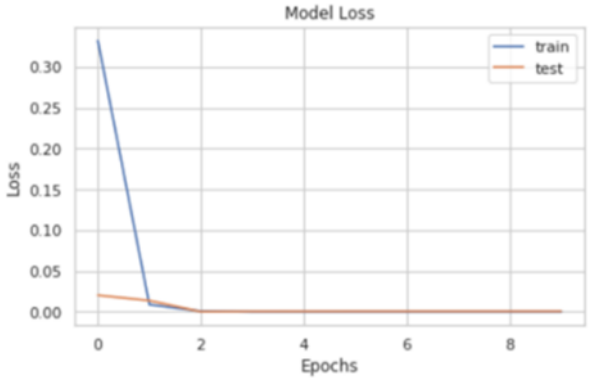

Fig 8. Loss graph of Model using HaGrid Dataset

Plots for accuracy and loss for Model-3 using ASL dataset.

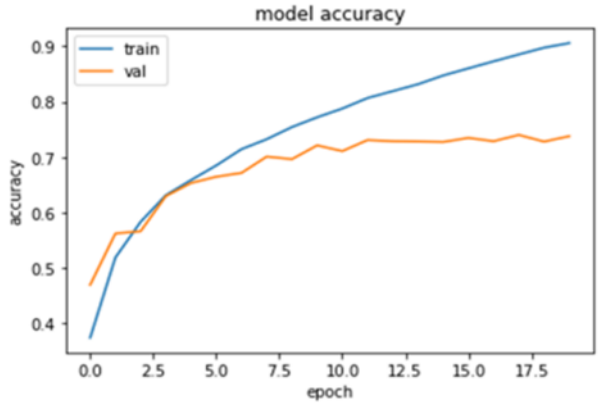

Fig 9. Accuracy graph of Model using ASL Dataset

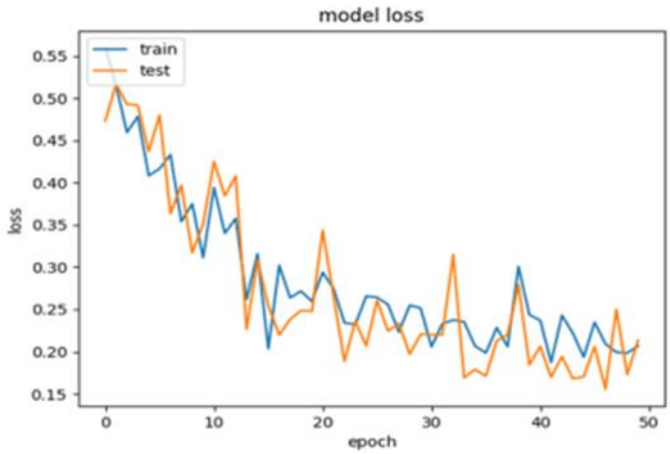

Fig 10. Accuracy graph of Model using HaGrid Dataset

5. Conclusion and Future Work

On difficult datasets, the proposed CNN model has proven to be extremely accurate at identifying hand gestures. It does not require manual segmentation or explicit feature extraction like other cutting-edge techniques. The suggested CNN model gets comparable results on the HaGrid dataset and outperforms existing approaches on datasets with complicated backgrounds despite having a simple design. This shows that the proposed approach may be used in fields like sign language interpretation and computer-human interaction. We intend to expand this work in the future to incorporate dynamic hand movements. In order to enable real-time processing on low-power devices, we strive to increase the model's computational efficiency using methods like pruning or compression. Model can be expanded to simultaneously recognise motions from several persons. Which has a use in applications like video conferencing or group interactions. To make the model more resilient to various variables, such as lighting conditions, hand orientations, and skin tone, use data augmentation or transfer learning.

References

[1] Ke, W., Xing, Y., Di Caterina, G., Petropoulakis, L., & Soraghan, J. (2020). Deep Convolutional Spiking Neural Network Based Hand Gesture Recognition. 2020 International Joint Conference on Neural Networks (IJCNN).

[2] Xu, J., & Jiang, T. (2017). Dynamic Hand Gesture Recognition Based on Parallel HMM Using Wireless Signals. Communications, Signal Processing, and Systems, 749–757.

[3] Chen, Z., Kim, J.-T., Liang, J., Zhang, J., & Yuan, Y.-B. (2014). Real-Time Hand Gesture Recognition Using Finger Segmentation. The Scientific World

Journal, 2014, 1–9.

[4] Suk, H.-I., Sin, B.-K., & Lee, S.-W. (2010). Hand gesture recognition based on dynamic Bayesian network framework. Pattern Recognition, 43(9), 3059–3072.

[5] Zhang, C., & Tian, Y. (2013). Edge Enhanced Depth Motion Map for Dynamic Hand Gesture Recognition. 2013 IEEE Conference on Computer Vision and Pattern Recognition

[6] Cheng, H., Luo, J., & Chen, X. (2014). A windowed dynamic time warping approach for 3D continuous hand gesture recognition. 2014 IEEE International Conference on Multimedia and Expo (ICME).

[7] Li, Y.-T., & Wachs, J. P. (2014). HEGM: A hierarchical elastic graph matching for hand gesture recognition. Pattern Recognition, 47(1), 80–88.

[8] Haria, A., Subramanian, A., Asokkumar, N., Poddar, S., & Nayak, J. S. (2017). Hand Gesture Recognition for Human Computer Interaction. Procedia Computer Science, 115, 367–374.

[9] Mohanty, A., Rambhatla, S. S., & Sahay, R. R. (2016). Deep Gesture: Static Hand Gesture Recognition Using CNN. Proceedings of International Conference on Computer Vision and Image Processing, 449–461

[10] Kopuklu, O., Kose, N., & Rigoll, G. (2018). Motion Fused Frames: Data Level Fusion Strategy for Hand Gesture Recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

[11] Abavisani, M., Joze, H. R. V., & Patel, V. M. (2019). Improving the Performance of Unimodal Dynamic Hand-Gesture Recognition With Multimodal Training. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[12] Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M. A., Alrayes, T. S., Mathkour, H., & Mekhtiche, M. A. (2020). Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation. IEEE Access, 8, 192527–192542.

[13] Li, C., Xie, C., Zhang, B., Chen, C., & Han, J. (2018). Deep Fisher discriminant learning for mobile hand gesture recognition. Pattern Recognition, 77, 276–288.

[14] Mo, G. B., Dudley, J. J., & Kristensson, P. O. (2021). Gesture Knitter: A Hand Gesture Design Tool for Head-Mounted Mixed Reality Applications. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.

[15] P. Parvathy, K. Subramaniam, G. K. D. Prasanna Venkatesan, P. Karthikaikumar, J. Varghese, and T. Jayasankar, "RETRACTED ARTICLE: Development of hand gesture recognition system using machine learning," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 6, pp. 6793–6800, Jul. 2020.

[16] Hasan, H., & Abdul-Kareem, S. (2012). RETRACTED ARTICLE: Static hand gesture recognition using neural networks. Artificial Intelligence Review, 41(2), 147–181.

[17] Molchanov, P., Gupta, S., Kim, K., & Kautz, J. (2015). Hand gesture recognition with 3D convolutional neural networks. 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

[18] Murthy, G. R. S., & Jadon, R. S. (2010). Hand gesture recognition using neural networks. 2010 IEEE 2nd International Advance Computing Conference (IACC). https://doi.org/10.1109/iadcc.2010.5423024

[19] Neto, P., Pereira, D., Pires, J. N., & Moreira, A. P. (2013). Real-time and continuous hand gesture spotting: An approach based on artificial neural networks. 2013 IEEE International Conference on Robotics and Automation.

[20] Alam, M. M., Islam, M. T., & Rahman, S. M. M. (2022). Unified learning approach for egocentric hand gesture recognition and fingertip detection. Pattern Recognition, 121, 108200.