Hand Gesture Recognition

Trusha R. Patel
Assistant Professor
Computer Eng. Dept.
CHARUSAT University
Gujarat, India
trushagajjar.ce@charusat.ac.in

Rikita Chokshi
Assistant Professor
Computer Eng. Dept.
CHARUSAT University
Gujarat, India
rikitachokshi.ce@charusat.ac.in

Yuvrajsinh Solanki
Computer Eng. Dept.
CHARUSAT University
Gujarat, India
21ce138@charusat.edu.in

Vishvdipsinh Umat
Computer Eng. Dept.
CHARUSAT University
Gujarat, India
21ce147@charusat.edu.in

[1]U & P. U. Patel Department of Computer Engineering, Chandubhai S Patel Institute of Technology (CSPIT), Faculty of Technology and Engineering (FTE), Charotar University of Science and Technology (CHARUSAT), Changa, India

Abstract:

This research paper present a hand gesture classification system utilize computer vision techniques and advance algorithm to find or comprehence the solution for accurately classifying hand gestures representing digits (0-5). In various applications like sign language translation and virtual keyboards, precise digit recognition from hand gestures is essential. Our approach involves a multi-stage process encompassing dataset collection, preprocessing, feature extraction, and classification using convolutional neural networks (CNNs).The collection of the dataset was manually done. Then we do the preprocess of the to reomve the noise and make images clear. After removing the noise from the dataset we perform the edge detection techniques to sharphen the edges and remove unwanted pixels from the images. After employing the data preprocessing we perform data augmentation techniques to increase the number of dataset with the help of it, we were able to generate a more diverse dataset with different types of features, which helps to learn features in a more diverse manner. Further, we use dataset to classify the gesture using convolutional neural networks (CNNs).Through the extraction of discriminative features and the utilization of CNNs, our system achieves high accuracy in digit classification. The results showcased the effectiveness and robustness of the proposed approach, achieving an impressive accuracy of 75% on the testing dataset.

1. Introduction

Hand gesture recognition (HGR) systems represent a promising avenue in human-computer interaction, offering natural and intuitive means of communication between humans and machines. Hand movements are one of the simplest forms of non-verbal communication among people which can express intentions or emotions without words spoken. Computational understanding of these signs opens many doors such as helping disabled persons or enabling control in dangerous places. By interpreting the spatial and temporal movements of the hand and fingers, HGR systems bridge the gap between computational operations and human non-verbal communication, revolutionizing the way we interact with technology [1].

The significance of HGR lies in its various applications, from sign language interpretation to virtual reality interfaces and remote device control [2]. Leveraging sophisticated algorithms and sensor technologies, these systems detect, track, and interpret the intricate movements of human hands, facilitating seamless interaction with digital interfaces. [3].

Recent advancements in artificial intelligence and machine learning, particularly deep learning models such as convolutional neural networks (CNNs), have significantly enhanced the accuracy and efficiency of HGR systems. CNNs have the ability learn hierarchical representations from raw images, enabling them to capture visual patterns and features that are difficult to manually define using those traditional methods.These models learn from vast datasets of hand gestures, improving their predictive capabilities over time and enabling robust recognition even in challenging environments. [4].

However, the development of HGR systems is challenging. Technical hurdles include variability in hand shape and appearance, occlusion, and robustness to environmental factors like lighting conditions [5]. Ergonomic considerations involve designing interfaces that accommodate diverse user preferences and physical abilities, ensuring accessibility for all users [6].

This paper aims to provide a comprehensive review of current HGR methodologies, including data acquisition techniques, feature extraction methods, and classification algorithms, while also proposing novel approaches to enhance system accuracy and efficiency.

2. Literature Review

This literature review gives a thorough overview of contemporary research efforts in the subject of hand gesture recognition, highlighting diverse techniques, methodology, and contributions made by different studies.

In this research paper [7] presents a low-cost hand gesture recognition system to develop an effective human-machine interface (HMI) for real-time scenarios1. It includes six stages: hand detection, gesture segmentation, feature extraction using pre-trained models, interactive HMI development, virtual mouse creation, and pointer smoothing with Kalman filter2. The inception-V1 model outperformed other CNN models and Vision Transformer in accuracy, precision, recall, and F-score for classifying hand gesture images3.

The paper [8] presents a real-time system for identifying human hand gestures from a live video feed, addressing challenges such as diverse lighting conditions, backgrounds, and gesture positions. A new dataset was created, featuring 4500 images of various hand gestures from individuals aged 10 to 50, under different conditions to enhance the model's robustness. The authors utilized deep learning architectures, particularly a modified VGG16 network, for gesture classification, achieving an impressive accuracy of 99.63%. The study's significant contributions include the development of an easy-to-implement real-time gesture classification system and the synthesis of a comprehensive dataset to train the model effectively.

The paper [9] presents a robust method for recognizing hand gestures using RGB-D data, aiming to enhance natural Human-Computer

Interaction (HCI) . For static hand gesture recognition, the method starts with contour extraction and uses algorithms like Distance Transform (DT) and K-curvature-convex defects Detection (K-CCD) for identifying palm centers and fingertips. It constructs a multimodal feature vector for gesture recognition. The paper proposes an improved dynamic time warping (IDTW) algorithm for recognizing dynamic hand gestures, combining Euclidean distance between hand joints and shoulder center with skeleton feature ratios to create a unified feature descriptor. It details an experiment using the UDLR-8 dataset for dynamic gesture recognition, highlighting the high average recognition rates of the proposed IDTW algorithm (96.5%) and DTW algorithm (91.4%).

This Paper [10] focuses on hand gesture recognition and classification using computer vision and deep learning techniques. It provides an in-depth exploration of methodologies employed to recognize and classify hand gestures, particularly within the context of computer vision. The dissertation likely discusses the utilization of deep learning algorithms such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs) to achieve robust recognition and classification performance. Mehatari's work contributes to the growing body of research aimed at enhancing gesture recognition systems, potentially offering practical applications in various fields such as human-computer interaction, sign language recognition, and healthcare.

3. Methodology

In this part, we will write about syntactic approach of a project.

3.1 Architecture Overview

Here Figure 1, shows the steps used by this model to classify the hand gesture.

Image Acquisition :- The retrieving of the imge is one of the difficult task. It is also one of the important task because without image we cannot proceed to next task.In this model the acquisition of the image are done manually. For each finger number their is around 100 or more image were captured. But still it will result in model overfitting so we increase the size of dataset by doing image augmentation.

Data Augmentation :- It enhances datasets by generating modified or synthetic data. Therfore, for to extend our dataset further we use data augmentation for our test and validation data.By augmenting the data we enhance the performance of the model and reduce overfitting.

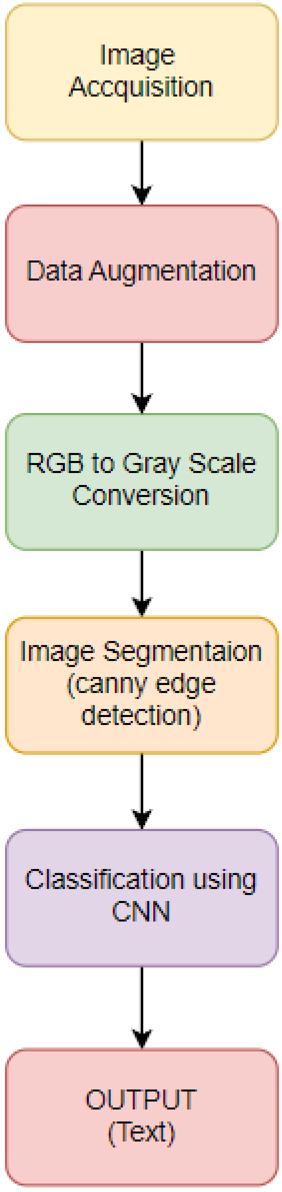

Figure 1. Steps for image Pre Processing

RBG Conversion :- In this the RGB color space is converted ino Gray Scale scale space.This transformation will help model to reduce complexity and noise reduction.

Image Segmentation :- It si used for partitioning an image into multiple segment.We use Canny Edge detection to perform edge based segmentatin.It is used to segement the image by grouping pixels that belong to same object.

CNN :- It is renowned for it effectiveness in handling complex visual data. It is also capable of learning hierarchical features from raw pixel data, capturing spatial connections between various portions of the hand and fingers. The use of data augmentation was critical in this work, as it improved CNN's capacity to generalize to new, previously unknown pictures. By exposing the model to a larger range of visual variances during training, the augmentation process created a more robust and flexible network capable of successfully identifying pictures other than those seen during training.

3.2 Data Pre-processing

Data preprocessing is one of the important task. It is used to reduce the complexities of the image and also it perform different task so that it is easy usable to extract feature.

Figure 2. Image

Color Space Transformation :- In this the image which where given are RGB color so it is complex do task and it required more memory so to reduce the usage of memory we convert it into Gray Scale.

Figure 3. Gray-scale Image

Edge Detection :- It is one of the fundamental technique which is used to identifies the edges of the image.In this work, the Canny edge detection technique was used on the grayscale version of the source photos.The subsequently generated edges found in the photos capture the key structural information of the hand, which would be helpful in our machine-learning model.

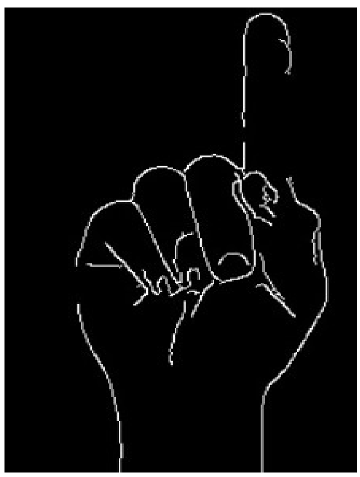

Figure 4. Edge Detected Image

Train-Validation-Test Split: The subsequent labeling of the dataset, discrete training, validation, and test sets are generated to train, confirm, and test the model. Stratified sampling can assure equitable representation of hands from each class in each division.Dataset was then split in 70- 30 ratio where 70% of the dataset was used for training and 30% of the dataset was used for testing the pre-trained model.

3.3 Dataset

Dataset used in this research work where manually gathered by our group using mobile phone. It comprises of 5 classes, starting from 1 to 5 digit using hand gesture. Each Images are well labeled and stored in different folder so to train a model become easy.Around 1000 images collected for hand gesture.

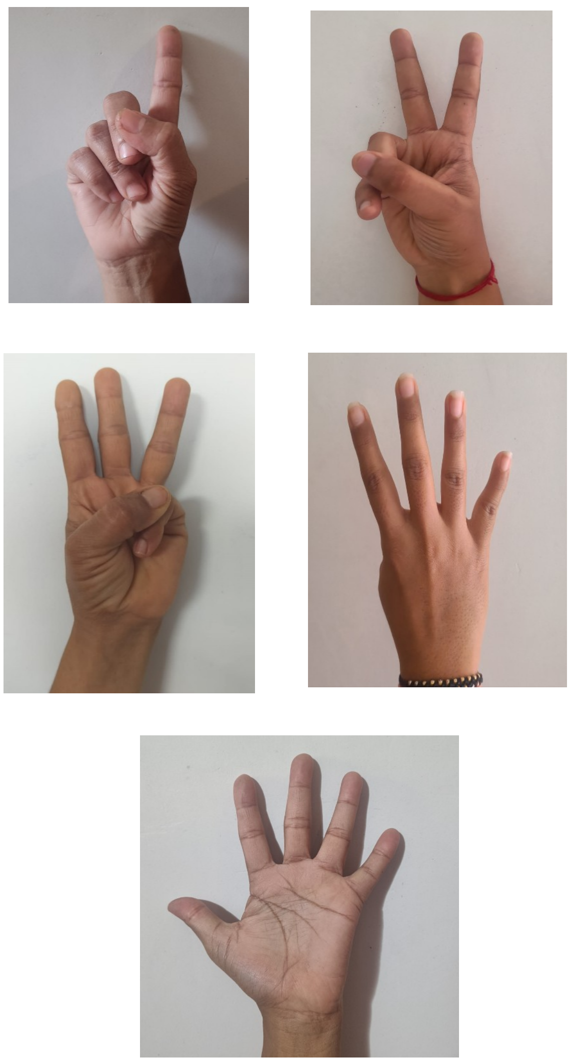

Figure 5. Dataset

4. Results

To assess the effectiveness of our hand gesture recognition system, we evaluated its performance on a curated dataset. This dataset comprised manually captured images featuring hand gestures from various individuals. And after performing operartion on dataset we plot graph for accuracy and loss and calculate classification metrics.

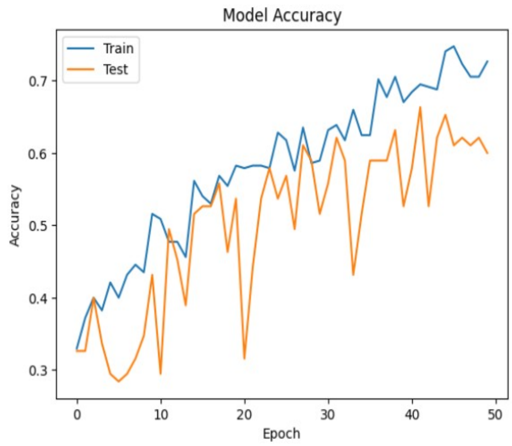

Figure 6. Model accuracy

we plotted a graph for the 50 epochs which is shown in figure 6. From the figure, it appears that the training accuracy increases steadily throughout the training process. The test accuracy also increases, but it seems to fluctuate more than the training accuracy. This could indicate that the model might be overfitting the training data.The graph we plotted is kniown as training accuracy curve.

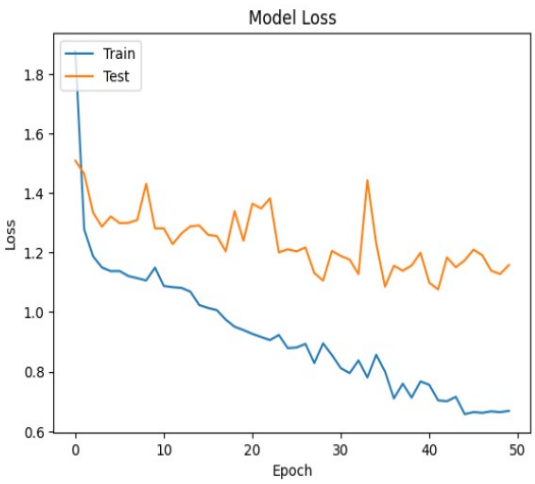

Figure 7. Model Loss

Furthermore, to observe the loss of the model, we also plotted another graph, as shown in Figure 7.We can clearly see that tthe train and test loss seem to be decreasing over epochs, which is a positive sign. This suggests the model is learning and improving its performance. However, gradually decreased as the epochs ran and the model learned the features more accurately. It appears that the test loss might be higher than the train loss throughout the epochs shown. This could be an indication of overfitting.

	precision	recall	f1-score	support
0	0.61	0.76	0.68	29
1	0.67	0.21	0.32	19
2	0.43	0.76	0.55	25
3	1.00	0.14	0.25	14
4	1.00	0.88	0.93	8
accuracy			0.57	95
macro avg	0.74	0.55	0.55	95
weighted avg	0.67	0.57	0.53	95

Table 1. Classification Metrices

```
Confusion Matrix:
[[22  2  5  0  0]
 [ 6  4  9  0  0]
 [ 6  0 19  0  0]
 [ 1  0 11  2  0]
 [ 1  0  0  0  7]]
```

Figure 8. Confusion matrix of testing data

4. Conclusion

In this research project we aimed to develop a better Hand Gesture Recognition Model. We are mainly focused on how to classify on to five digit using the hand gesture. The dataset used to trained the model are manually captured. We used image augmentation methods to make our dataset bigger and more diverse as it was manually made. This helped the model learn better and become more robust and efficient. We trained a deep learning model called CNN on the data with a 70-30 split on the dataset. The model achieved promising results, demonstrating the ability to classify various hand gesture with an accuracy of 75%.The limitation of the model is that it is provided with less number of datasets. Which result in overfitting of the model. So, Our future plan to improve our model is by collecting a larger or more diverse dataset and exploring different model.

5. References

1. Li, Y., & Chan, A. B. (2017). 3D Hand Shape and Pose Estimation from a Single RGB Image. In Proceedings of the IEEE International Conference on Computer Vision (ICCV).

2. Zhang, X., Wang, T., Xu, Y., & Li, H. (2019). A Review of Hand Gesture Recognition Techniques. In Proceedings of the International Conference on Robotics and Automation (ICRA).

3. Wu, Y., & Huang, T. (2019). A Survey on Deep Learning Based Hand Gesture Recognition. In Proceedings of the ACM Symposium on Applied Computing (SAC).

4. Song, S., Lan, S., & Xing, J. (2018). Dynamic Hand Gesture Recognition using

Temporal Deep Learning. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI).

5. Sharma, S., & Verma, V. (2016). Real-time Hand Gesture Recognition using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

6. Liu, Y., & Gong, J. (2017). Robust Hand Gesture Recognition using Depth and Skeletal Data. In Proceedings of the European Conference on Computer Vision (ECCV).

7. Sen, A., Mishra, T. K., & Dash, R. (2023). Deep Learning-Based Hand Gesture Recognition System and Design of a Human–Machine Interface. Neural Processing Letters, 55(9), 12569-12596.

8. Aggarwal, A., Bhutani, N., Kapur, R., Dhand, G., & Sheoran, K. (2023). Real-time hand gesture recognition using multiple deep learning architectures. Signal, Image and Video Processing, 17(8), 3963-3971.

9. Xu, J., Wang, H., Zhang, J., & Cai, L. (2022). Robust hand gesture recognition based on RGB-D Data for natural human–computer interaction. IEEE Access, 10, 54549-54562.

10. Mehatari, B. (2022). Hand Gesture Recognition and Classification using Computer Vision and Deep Learning Techniques (Doctoral dissertation, Dublin, National College of Ireland).