

Plagiarism Report

Originality Assessment

10%



Overall Similarity

Date: Oct 4, 2023

Matches: 342 / 3594 words

Sources: 14

Remarks: Low similarity detected, check with your supervisor if changes are required.

Verify Report:

ML BASED HAND GESTURE RECOGNITION

1st Nivedha V V

Department of Computer Science and Engineering, KCG College of Technology,

Chennai, India nivedhavijayaraghavan@gmail.com

2nd K S Suganya

Assistant professor

Department of Computer Science and Engineering, KCG College of Technology,

Chennai, India suganyaselvaraj06@gmail.com

Abstract—The projects focus on **Gesture recognition using** the mediapipe hands Lite framework where the custom dataset of 10 gesture Images is trained with an inbuilt mediapipe hands model which contains 2 CNN models - A palm detection model running on single Shot detection architecture and a hand landmark generator running on regression model architecture. The dataset is successfully trained and tested with **1 the proposed method** and an accuracy of 98 percent is obtained.

Index Terms—Gesture Recognition, Mediapipe, Hands Detection, CNN, Regression Model, Hand Landmark Detection, Machine Learning, Image Processing.

I. INTRODUCTION

In today's world, being part of something that constantly has an upgrade in life, something to always Update yourself on, and most importantly the impact technology has on our everyday life is just absolutely ecstatic. I have always been amused of how useful technology can be in Our day to day lives. Machine learning is one such Boone in the field of technology. In this project, we would be implementing gesture recognition- a highly interesting and important topic. **1 There have been** various research methods and algorithms proposed to recognise gestures using machine learning. We will be looking at one such method in this paper, **gesture recognition using** the framework mediapipe. We would be preparing a custom hand gesture dataset and then train the mediapipe hands model to give us recognised gestures. This project helps the hearing and speech impaired people communicate

effectively with others by showing the gestures in front of the webcam when the code is being run. This in turn enables the non-speech / hearing impaired person to communicate and understand with more clarity and precision. Although a lot of methods [1] have been proposed to enable the sign language recognition, there are not many in terms of emotions - gesture recognition.

II. RELATED WORK

In our pursuit of gaining comprehensive insights into Sign Language Recognition and the various techniques employed in this domain, we conducted a thorough review of several research papers. These papers collectively contribute to the overarching theme of our project:

a) KUMUD TRIPATHI, NEHA BARANWAL, AND G. C.

NANDI (2015) [1] proposed a continuous Indian Sign Lan-

guage (ISL) [3] gesture recognition system, addressing the challenge of recognizing sign language gestures within continuous

sequences. They employed a gradient-based key frame extraction method and achieved improved accuracy by using features obtained through Orientation Histogram (OH) and Principal Component Analysis (PCA).:

b) ARCHANA S. GHOTKAR AND GAJANAN K. KHARATE

(2015) [2] delved into **1 dynamic hand gesture**

recognition for Indian Sign Language and developed

algorithms for word recognition, including rule-based and Dynamic Time Warping-

based methods. Their innovative approach to sentence interpretation using inverted indexing tackled the challenges of continuous sentence recognition.:

c) M.M. GHARASUIE AND H. SEYEDARABI (2013) [3] introduced a real-

time system for recognizing dynamic hand

gestures using Hidden Markov Models (HMMs). They suc-

cessfully recognized key gestures related to English numbers in real-time, distinguishing between key and link gestures.:

d) KAIRONG WANG, BINGJIA XIAO, JINYAO XIA, AND DAN LI (2015) [4] presented

a **10 dynamic hand gesture** recognition algorithm that utilized an improved Codebook (CB) modeling method and spatial moments for feature extraction, achieving satisfactory recognition accuracy.:

e) H. FRANCKE, J. RUIZ-DEL-SOLAR, AND R. VER-SCHAE (2014) [5] proposed a

robust **2 real-time hand gesture detection and recognition** system that employed boosted classifiers, skin segmentation, and hand tracking. **3 Innovative training techniques**, such as **active learning and bootstrap**, significantly improved system performance.:

f) HARI PRABHAT GUPTA, HARESH S CHUDGAR, AND SIDDHARTHA (2016) [6]

addressed continuous **hand gesture recognition for** Human-Machine Interaction (HMI) using accelerometer and gyroscope sensors in smart devices. Their approach included gesture coding and spotting algorithms, facilitating the recognition of predefined gestures.:

g) NOOR TUBAIZ (2015) : [7] introduced a glove-based

Arabic sign language recognition system using sensor-based data gloves. This approach achieved high sentence recognition rates while eliminating the limitations of vision-based systems. h) M.K.

BHUYAN (2012) [8] proposed a gesture recog-

nition approach based on finite state machines (FSMs) and

video object planes. This method reduced gesture video se-

quences into representative key frames for efficient recogni-

tion.:

i) JOYEETA SINGHA AND KAREN DAS (2015) [9] pre- sented an automatic system for recognizing Indian Sign Lan-

guage alphabets in continuous video sequences, achieving a high recognition rate through a combination of preprocessing, feature extraction, and classification techniques.:

j) SUHARJITO, GUNAWAN, AND H. THIRACITTA

(2018) [10] explored **1 the use of** a modified Convolutional Neural Network (CNN) model **for**

Sign Language Recognition, incorporating transfer learning from Action Recognition.

While achieving high training accuracy, the paper

acknowledged challenges in validation accuracy.: These

referenced articles collectively contribute to the extensive

body of knowledge in Sign Language Recognition, forming the foundation for our project's research and development.

III. PROPOSED SYSTEM

A. TECHNIQUES FOR RECOGNITION AND TRAINING

1) MACHINE LEARNING: Machine learning is an art of computer science that comes under Artificial intelligence. As illustrated in the figure 1, This is **1 used in the** development of analytical models that use different forms of input data

in creating meaningful and understandable insights. Machine

learning models constantly learn from the data to find the optimal point of accuracy in its

outputs.

At its core, machine learning involves training a model on a dataset, allowing it to learn patterns and relationships in the data. The model can then use this learning to make predictions or decisions on new data that it has not seen before. There are several different types of machine learning, including supervised learning, unsupervised learning, and reinforcement learning.

Fig. 1. Machine Learning.

2) CONVOLUTION NEURAL NETWORK: The convolutional neural network (CNN) is a basic ML model which uses in the images, the training input datasets and then assigns weights to each and every parameter based on the situation considered and thus, based upon the assigned scores it differentiates one image from the other. These weights denote the importance of an object in an image.

This Image recognition feature performed by the CNN can be broken down into fewer steps. To begin with, the Convolutional layer captures all the low-level features such as the edge detection, sharpness, brightness, color exhibited by the image and other observable features. Then, when there are further more layers added on top of this, the architecture can read the high-level features as well, which helps the model understand every aspect of the picture considered along with its parameters of consideration add formula for CNN. There

Fig. 2. Convolution Layer

Fig. 3. Pooling Layer

are multiple layers in the CNN: Convolutional layer, Pooling Layer, Fully Connected layer.

CONVOLUTIONAL LAYER:

The convolutional layer as illustrated in Figure 2, is a part of CNN in which the majority of the task lies in recording and analyzing the characteristics in a given input dataset. It will have an output dataset, which will be high if among the two matrices considered the one having a high value is in the same place. This calculation is because the layer functions by considering a dot product of the filter with the random size of the input image. As a result, dot product of the output can inform us if the pixel pattern in the underlying image corresponds to the pixel pattern described by our filter or just a dot product of the weights.

POOLING LAYER:

The Pooling Layer, as illustrated in Figure 3, Similar to the way dictionary mapping works, image processing also works by having multiple feature maps where the entire image will be replicated and have a mapping structure of the original input data image. Whenever there is an operation of the convolutional layer on these datasets, these feature maps will associate the input data with the filter. They are majorly used in performing

two features in image recognition: Max pooling and Average pooling.

FULLY CONNECTED LAYERS:

These layers are usually not very significant but also at the same time is used in improving the overall class scores. They are contained at the end of the network layer and stays hidden.

These layers kick in as soon as the Pooling layer data maps the input data with the output and the final image recognition model is ready to go. If the layer finds some anomalies in the mapping, it will optimize the scores of various parameters and either re-run the algorithm or adapt the model to the new

Fig. 4. Fully Connected Layers

Fig. 5. Image Processing Structure

scores as illustrated in the Figure 4.

B. IMAGE PROCESSING

Image processing is a crucial technique used to analyze and extract valuable information from images, playing a pivotal role in various industries undergoing automation during the fourth industrial revolution. Many companies are turning to

machine learning for its precision and cost-cutting potential. This technology finds applications in tasks like image information extraction, object classification, categorization, visualization, and pattern recognition. Today's advanced cameras and image-capturing devices rival human perception. Smartphones employ facial recognition for real-time 3D object overlays, while cars use 360-degree cameras for constant environmental monitoring and autonomous driving. Digital image processing is the preferred method due to its speed, allowing computers to process raw image data swiftly, identify features, characteristics, masks, and points within images. In contrast, analog methods are slower and less efficient. The focus here is on digital image processing, particularly using Convolutional Neural Networks (CNNs) due to their effectiveness. CNNs address a common challenge in image processing, where input data must match the model's resolution standards to prevent excessive hidden layers and parameter complexity. CNNs employ three key layers: Convolution, Pooling, and fully connected layers. The specific project scope revolves around gesture recognition, utilizing CNNs to achieve the desired results. As human gestures are the input, this topic is explored further. The image processing workflow involves these layers, ensuring efficient and accurate recognition of **6 gestures in a** wide range of applications.

An illustration of the Image Processing Structure is shown in the Figure 5.

C. GESTURE RECOGNITION

Gesture recognition involves identifying human gestures and triggering actions accordingly, from symbol depiction to movement or even dancing. This process captures and analyzes complex gestures, interprets them, and translates them into actionable commands. Instead of traditional input methods like typing, gestures can be used, with motion sensors in cameras translating the actions into machine language for ML models. The success of gesture recognition relies on the chosen pipeline or architecture. The project employs Google's efficient and hassle-free Mediapipe framework, utilizing hand tracking neural network pipelines for accurate gesture recognition.

1) STEPS INVOLVED IN RECOGNITION: The process of gesture recognition involves several key steps to ensure the development of an accurate [4 machine learning model](#) for image recognition.

Creating a Dataset: The first crucial step is to prepare and curate the right datasets. These datasets should have images with resolutions compatible with the machine learning model's requirements. Two essential types of datasets are used:

1) Training Data: This labeled dataset is used to train the [6 machine learning model](#). The model uses this data to learn and adjust its parameters iteratively. Typically, this dataset accounts for up to 80 percent of the total data.

2) Testing Data: This dataset is used to evaluate the trained model's performance on unseen data. Adjustments and fine-tuning are done based on the model's performance with this dataset.

Choosing a Model: The choice of [4 machine learning model](#) depends on the specific problem, type of dataset, and desired outputs. Various machine learning algorithms are available, such as K-means for unlabeled data, regression for prediction, and categorical classification. In real-world scenarios with large datasets and complex problems, there is no one-size-fits-all model. Parameters, variable types, and data nature determine the suitable algorithm.

Supervised Learning: This paradigm involves the model learning from input-output pairs through

historical data. It encompasses regression for continuous output and classification for discrete datasets. Common algorithms include Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), linear regression, and decision trees. Supervised learning excels when there is ample input data, as the model refines its parameters to match historical outputs.

In the context of this project, CNN models have proven highly effective for image processing and classification, specifically for gesture recognition. These models learn to recognize gestures by comparing input images to labeled historical data, making them a suitable choice for this application.

D. ARCHITECTURES

Single Shot Detector Model (SSD): The SSD model is a powerful and efficient solution ⁹ for object detection across multiple categories. It comprises two main components: the backbone model, which extracts features from the input data using a pre-trained image classification network, and the SSD head, consisting of stacked convolutional layers that generate bounding boxes around objects. The SSD model detects different object classes and aspect ratios at each grid in its feature map. It assigns scores to these detections and applies non-maximal suppression to obtain the final detections. ³ The training of SSD involves optimizing a loss function that incorporates localization and classification components, making it a robust choice for object detection tasks.

Regression Model: Regression models, including neural network regression, are statistical models that predict relationships between dependent and independent variables. They use lines or curves to estimate future values based on input data. Neural network regression introduces non-linearity by mapping node outputs to multiple values, making it suitable for complex prediction tasks. Each neuron in the model is

connected to others through weighted connections, allowing them to collectively determine outputs efficiently. This approach is useful for applications like gesture recognition, where predicting precise outcomes based on input data is essential.

E. TRAIN THE ALGORITHM TO CONSTRUCT THE MODEL

Training the machine **6 learning algorithm is** a crucial step in the model development process. It involves splitting the dataset into training and testing portions, and then feeding the training data into the chosen algorithm. The model iteratively adjusts its parameters until it achieves high accuracy. **14 In supervised learning,** the **model learns from** labeled input-output pairs, minimizing differences between predicted and documented results. The process aims for minimal human intervention, focusing on selecting the right model parameters and refining the input dataset to optimize results. **4 The model's performance** is evaluated by comparing its outputs to the expected results, allowing for continuous improvement.

F. TESTING AND IMPROVISATIONS

Once the model has been trained successfully, the next important step is **5 to use the** input dataset and test if the model produces the required outputs. It is also important to check if the model **4 that we have** developed using the test and train can handle the adversarial data, as what seems normal to a human eye might cause the biggest misclassification to **a machine learning model**. Once we check **6 the testing data** with the outputs, we can then check for the output accuracy and **model performance using** various factors.

For testing our model, we use the following formulae and calculate the accuracy. One can say **4 that a model can** perform well, by checking the accuracy value. we would be checking **6 the**

accuracy of our model for the custom dataset based on the formula given below:

There are a few parameters to be taken into consideration while calculating the accuracy:

Fig. 6. Hand Lank Marks

Precision - Precision gives the percentage of correct positive results out of the total positive results

predicted by the model
$$\text{precision} = \frac{\text{truepositives}}{\text{truepositives} + \text{falsepositives}}$$

Recall - it Recall gives the percentage of correct positive results **7 out of all positive** results

$$\text{Recall} = \frac{\text{truepositives}}{\text{truepositives} + \text{falsepositives}}$$

G. LIBRARY FUNCTION

Mediapipe

2 The ability to perceive the coordinates and movement of hands is a very important aspect in improving the success rate and efficiency of data modeling. In our case, it is highly crucial **3 for sign language** understanding and **hand gesture recognition**. Mediapipe is a framework by Google that enables open-source cross-platform for various **machine learning methods** and models in Image processing. It has detection methodologies built for various action tracking. however, **9 we will be** using the framework pipeline in hand gesture recognition from Mediapipe.

Hands – Mediapipe

MediaPipe Hands utilizes an ML pipeline consisting of two models working together. we are using the hand tracking neural network pipeline - lite on our custom dataset to predict 2D and 3D hand landmarks on our custom images.

The proposed model contains two inbuilt CNN models- **2 A**

palm detection model which uses the single shot detection model architecture and **A hand landmark** generator **which uses a** regression model architecture.

Methodology of mediapipe-hands

First, when the palm detection model detects an image, it creates a box around the image. then, the palm detector removes the background and returns a cropped image. 11 The palm detection model processes the image in the initial frame or if the hand is absent from subsequent frames in an effort to simplify the model and boost its real-time performance. 2 The hand landmark model then generates 3D hand key points from the cropped image. The landmark model works through 21 key point coordinates of the hand by using the regression algorithm. The 21 coordinate points can be seen below in Figure 6. The landmark consists of the coordinates (x, y) and the relative depth (z), where z represents the camera's distance from the wrist and larger values indicate closer proximity to the camera. [17] This in turn predicts the final coordinates of the image. The final returned coordinates from 9 the model are then matched with the stored coordinates of our custom images and the model returns the gesture name.

A. Output

IV. EXPERIMENTAL RESULTS

The Recognitions using the technique of interest - "Medi- apipe Hands" were documented and are given below in Figures

7 and 8

Fig. 7. Recognition of gestures

Fig. 8. Recognition of gestures

5 We can see from the above images that the model has given us accurate recognition outputs making this a successful attempt. The accuracy 4 of the model will be discussed in the next section to document the efficiency of the model.

B. Accuracy

The model was trained and tested for 5 accuracy and the results are as follows:

From Figure 9, we can see the training and testing accuracy to be precise and unwavering. Further, to calculate the overall

Fig. 9. Accuracy Graph.

Fig. 10. Accuracy Table.

Accuracy 7 of the model, The result of the calculation is as follows:

From the table depicted in 10, we can see that the model is 98 percent accurate and has been found to be very efficient.

V. CONCLUSION

As an image processing technique, 11 Gesture recognition using mediapipe has been successfully implemented providing us with promising accuracy. The coordinate-based recognition system has proved 5 to be an accurate way to distinguish and recognise images as each angle and position is saved as a new coordinate. This in turn has provided us with maximum number of correct recognitions. The importance of the topic was to enable good usability and simple understanding of the gestures 9 and it has shown significant results

and has paved way for massive enhancements in the future for the speech/hearing impaired. With advancing technologies and improving environments, this proposed project does have scope for additional implementations and future enhancements. The methodology can further be carried out in an app exclusively for the speech / hearing impaired empowering them to flourish in their fields of interest.

REFERENCES

- [1] Mahesh, B. Machine learning algorithms-a review. [13 International Journal of Science and Research \(IJSR\)](#) 9., pp.381-386, (2020).
- [2] Vihar Kurama-Blog ML-based Image Processing, (2021)
- [3] Geethu G Nath and Arun C S, "Real Time Sign Language Interpreter," 2017 [1 International Conference on Electrical, Instrumentation, and Communication Engineering \(ICEICE2017\)](#).
- [4] Kumud Tripathi, Neha Baranwal and G. C. Nandi, "Continuous Indian Sign Language Gesture Recognition and Sentence Formation", Eleventh International Multi- Conference on Information Processing- 2015 (IMCIP-2015), Procedia Computer Science 54 (2015) 523 – 531. [5] Manasa Srinivasa H S and Suresha H S, "Implementation of Real Time Hand Gesture Recognition," International Journal of Innovative Research in Computer and Communication Engineering, Vol. [3, Issue 5, May 2015](#).
- [6] [Joyeeta Singha and Karen Das](#), "Automatic [Indian Sign Language Recognition for](#) Continuous Video Sequence," ADBU Journal of Engineering Technology 2015 Volume 2 Issue 1.
- [7] [Archana S. Ghotkar and Gajanan K. Kharate](#), "Dynamic [Hand Gesture Recognition and Novel Sentence Interpretation Algorithm for Indian Sign Language Using Microsoft Kinect Sensor](#)," [Journal of Pattern Recognition Research](#) 1 (2015) 24-38.

[8] M.K. Bhuyan, "FSM-based recognition of dynamic hand gestures via gesture summarization using key video object planes," World Academy of Science, Engineering and Technology Vol: 6 2012-08-23. 42

[9] M.M.Gharasuie and H.Seyedarabi, "Real-time [10 Dynamic Hand Gesture Recognition using Hidden Markov Models](#)," 2013 8th Iranian Conference on [Machine Vision and Image Processing \(MVIP\)](#).

[10] [8](#) Kairong Wang, Bingjia Xiao, Jinyao Xia, and Dan Li, "A Dynamic Hand Gesture Recognition Algorithm Using Codebook Model and Spatial Moments," 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics.

[11] Francke H., Ruiz-del-Solar J. and Verschae R., "Real-Time Hand Gesture [3 Detection and Recognition Using Boosted Classifiers and Active Learning](#)," Advances in [Image and Video Technology](#). PSIVT 2007.

Lecture Notes in Computer Science, vol 4872. Springer, Berlin, Heidelberg.

[12] Jagreet Kaur, Blog [4 Machine Learning Model Testing Training and Tool](#), February 2022

[13] Unext-blog 10 Popular Regression Algorithms [In Machine Learning](#) of 2022,2022

[14] Harris, Moh, and Ali Suryaperdana Agoes. In [12 2nd International Seminar of Science and Applied Technology \(ISSAT 2021\)](#), Atlantis Press. Applying Hand Gesture Recognition for User Guide Application Using MediaPipe.,pp. 101-108,2021

Sources

1	https://www.ijraset.com/research-paper/sign-language-recognition-system-for-hearing-and-speech-impaired-people INTERNET 2%
2	https://chuoling.github.io/mediapipe/solutions/hands.html#:~:text=MediaPipe Hands utilizes an ML pipeline consisting of,palm detector and returns high-fidelity 3D hand keypoints. INTERNET 1%
3	https://www.semanticscholar.org/paper/Real-Time-Hand-Gesture-Detection-and-Recognition-Francke-Ruiz-del-Solar/93302e6f5ff83ebd43ebb383281835dc05ae9f86 INTERNET 1%
4	https://medium.com/@datasciencewizards/a-guide-to-data-splitting-in-machine-learning-49a959c95fa1#:~:text=In data science or machine learning, data splitting,a model can get trained, tested and evaluated. INTERNET 1%
5	https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/ INTERNET 1%
6	https://www.applause.com/blog/training-data-validation-data-vs-test-data INTERNET 1%
7	https://machinelearningmastery.com/fbeta-measure-for-machine-learning/ INTERNET <1%
8	https://www.ijcdr.org/papers/IJCDR2201036.pdf INTERNET <1%
9	iq.opengenus.org/single-shot-detector/ INTERNET <1%
10	https://www.ijcaonline.org/proceedings/iccc2016/number1/26156-cc57 INTERNET <1%
11	https://developers.google.com/mediapipe/solutions/vision/gesture_recognizer INTERNET <1%
12	https://www.atlantispress.com/proceedings/issat-21/articles INTERNET <1%
13	https://www.ijsr.net/archive/v9i1/v9i1.php INTERNET <1%
14	https://builtin.com/machine-learning/supervised-learning INTERNET <1%

EXCLUDE CUSTOM MATCHES OFF

EXCLUDE QUOTES OFF

EXCLUDE BIBLIOGRAPHY OFF