iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.39103

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Real-Time Sign Language Detection and Recognition

Sarthak Sharma[1], Preet Kaur Nagi[2], Rahul Ahuja[3], Poorti Rajani[4], Senior Asst. Prof. Kavita Namdev[5]
[1, 2, 3, 4, 5]*Acropolis Institute of Technology and Research*

Abstract: Sign language is one of the oldest and most natural form of language for communication, but since most people do not know sign language and interpreters are very difficult to come by we have come up with a real time method using neural networks for fingerspelling based American sign language. In our method, the hand is first passed through a filter and after the filter is applied the hand is passed through a classifier which predicts the class of the hand gestures.

Keywords:

- *Feature Extraction and Representation:* The representation of an image as a 3D matrix having dimension as ofheight and width of the image and the value of each pixel as depth (1 in case of Grayscale and 3 in case of RGB). Further, these pixel values are used for extracting useful features using CNN.
- *Artificial Neural Networks:* Artificial Neural Network is a connections of neurons, replicating the structure of human brain. Each connection of neuron transfers information to another neuron. Inputs are fed into first layer of neurons which processes it and transfers to another layer of neurons called as hidden layers. After processing of information through multiple layers of hidden layers, information is passed to final output layer.

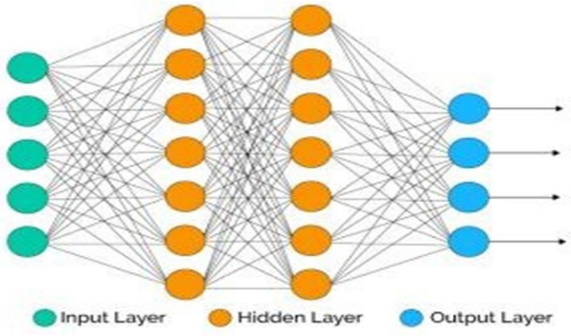


Figure 5.1: Artificial neural networks

There are capable of learning and they have to be trained. There are different learning strategies:

- Unsupervised Learning
- Supervised Learning
- Reinforcement Learning.

- *Convolution Neural Network:* Unlike regular Neural Networks, in the layers of CNN, the neurons are arranged in 3 dimensions: width, height, depth. The neurons in a layer will only be connected to a small region of the layer (window size) before it, instead of all of the neurons in a fully-connected manner. Moreover, the final output layer would have dimensions (number of classes), because by the end of the CNN architecture we will reduce the full image into a single vector of class scores.

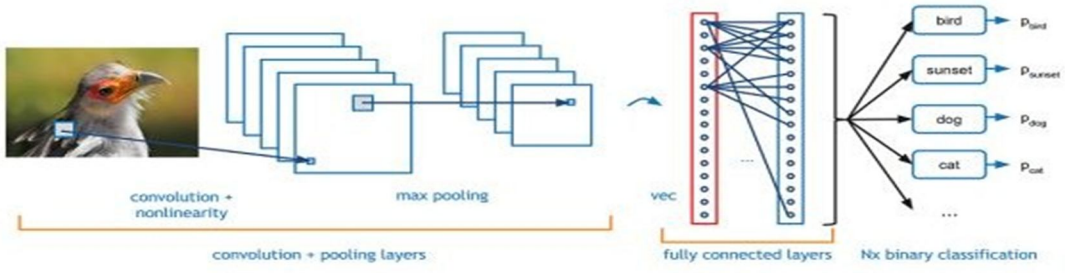


Figure 5.2: Convolution neural networks

- *Convolution Layer:* In convolution layer we take a small window size [typically of length 5*5] that extends to the depth of the input matrix. The layer consist of learnable filters of window size. During every iteration we slid the window by stride size [typically 1], and compute the dot product of filter entries and input values at a given position. As we continue this process well create a 2-Dimensional activation matrix that gives the response of that matrix at every spatial position. That is, the network will learn filters that activate when they see some type of visual feature such asan edge of some orientation or a blotch of some color.
- *Pooling Layer:* We use pooling layer to decrease the size of activation matrix and ultimately reduce the learnable parameters. There are two type of pooling :
  - *Max Pooling:* In max pooling we take a window size [for examplewindow of size 2*2], and only take the maximum of 4 values. Well lidthis window and continue this process, so well finally get a activation matrix half of its original Size.
  - *Average Pooling:* In average pooling we take average of all Values in a window.

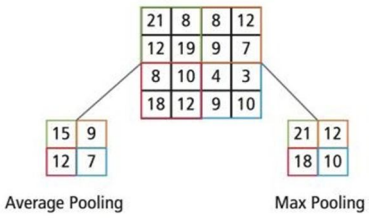


Figure 5.3: Types of pooling

- *Fully Connected Layer:* In convolution layer neurons are connected only to a local region, while in a fully connected region, well connect the all the inputs to neurons.

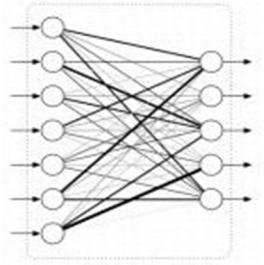


Figure 5.4: Fully Connected Layer

- *Final Output Layer:* After getting values from fully connected layer, well connect them to final layer of neurons [having count equal to total number of classes], that will predict the probability of each image to be in different classes.
- *TensorFlow:* Tensorflow is an open source software library for numerical computation. First we define the nodes of the computation graph, then inside a session, the actual computation takes place. TensorFlow is widely used in Machine Learning.
- *Keras:* Keras is a high-level neural networks library written in python that works as a wrapper to TensorFlow. It is used in cases where we want to quickly build and test the neural network with minimal lines of code. It contains implementations of commonly used neural network elements like layers, objective, activation functions, optimizers, and tools to make working with images and text data easier.
- *OpenCV:* OpenCV(Open Source Computer Vision) is an open source library of programming functions used for real-time computer-vision. It is mainly used for image processing, video capture and analysis for features like face and object recognition. It is written in C++ which is its primary interface, however bindings are available for Python, Java, MATLAB/OCTAVE.

I. INTRODUCTION

American sign language is a predominant sign language Since the only disability D&M people have is communication related and they cannot use spoken languages hence the only way for them to communicate is through sign language. Communication is the process of exchange of thoughts and messages in various ways such as speech, signals, behavior and visuals.Deaf and dumb (D&M) people make use of their hands to express different gestures to express their ideas with other people. Gestures are the nonverbally exchanged messages and these gestures are understood with vision. This nonverbal communication of deaf and dumb people is called sign language.

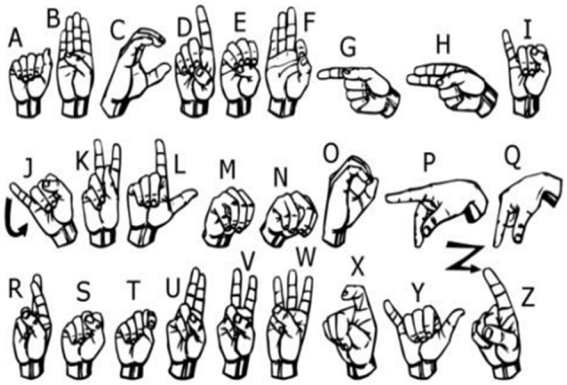
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Sign language is a visual language and consists of 3 major components

Fingerspelling	Word level sign vocabulary	Non-manual features
Used to spell words letter by letter .	Used for the majority of communication.	Facial expressions and tongue, mouth and body position.

.

In our project we basically focus on producing a model which can recognize Fingerspelling based hand gestures in order to form a complete word by combining each gesture. The gestures we aim to train are as given in the image below.



II. PROBLEM FORMULATION

Deaf is a disability that impair their hearing and make them unable to hear ,while mute is a disability that impair their speaking and make them unable to speak. Both are only disabled at their hearing and/or speaking, therefore can still do much other things.

The only thing that separate them and the normal people is communication. If there is a way for normal people and deaf-mute people to communicate, the deaf-mute people can easily live like a normal person. And the only way for them to communicate is through sign language.

While sign language is very important to deaf-mute people, to communicate both with normal people and with themselves, is still getting little attention from the normal people. We as the normal people, tend to ignore the importance of sign language, unless there are loved ones who are deaf-mute. One of the solution to communicate with the deaf-mute people is by using the services of sign language interpreter. But the usage of sign language interpreter can be costly. Cheap solution is required so that the deaf-mute and normal people can communicate normally.

The problems of developing sign language recognition ranges from the image acquisition to the classification process. Researchers are still finding the best method for the image acquisition. Gathering images using camera gives the difficulties of image pre-processing. Meanwhile, using active sensor device can be costly. Wide choice of recognition method makes researchers unable to focus on one best method. Choosing one method to be focused on, tends to make other method that may be better suit for Sign Language Recognition, not being tested. Trying out other methods makes researchers barely develops one method to its fullest potentials.

Hence,in our Project we will be creating a real-time automatic sign language gesture recognition system, using different tools (Deep Learning, OpenCV, and Tensorflow), which will be helpful for deaf and disabled people We are going to create our own dataset which uses raw images to match our requirements. While other systems used MNIST dataset for implementation of this project that affects the accuracy of the model.

III. LITERATURE REVIEW

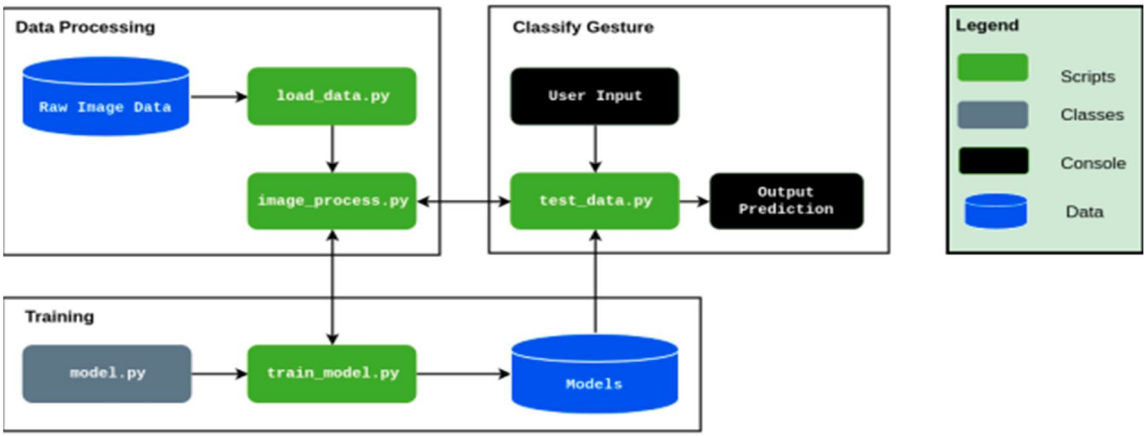
We have gone through other similar works that are implemented in the domain of sign language recognition. The summaries and drawbacks of each of the project works are mentioned below:

A Survey of Hand Gesture Recognition Methods in Sign Language Recognition: Issues were faced relating to the accuracy of the model. The accuracy achieved was roughly estimated to be around 81.64%, which didn't match the required criterion so as to abide by the standard characteristics of detecting and pre-processing image dataset.

Communication between Deaf-Dumb People and Normal People: They build a recognition system for Flemish sign language using convolutional neural networks and achieve an error rate of 2.5%. It is a non-real-time application that assures fluctuating accuracy due to use of background subtraction algorithms.

A recognition model is built using hidden Markov model classifier and a vocabulary of 30 words and they achieve an error rate of 10.90%.

IV. METHODOLOGY



Let's understand the flow of our project: Data Processing, Training, and Classify Gesture.

1) *Data Processing:* The load data.py script contains functions to load the Raw Image Data and save the image data as numpy arrays into storage. The process data.py script will load the image data from data.py and preprocess the image by resizing the image, and applying filters and ZCA whitening to enhance features. During training the processed image data was split into training, validation, and testing data and written to storage. Training also involves a load dataset.py script that loads the relevant data split into a Dataset class. For use of the trained model in classifying gestures, an individual image is loaded and processed from the system.
2) *Training:* The training loop for the model is contained in train model.py. The model is trained with hyperparameters obtained from a config that lists the learning rate, batch size, image filtering, and number of epochs. The configuration used to train the model is saved along with the model architecture for future evaluation and tweaking for improved results. Within the training loop, the training and validation datasets are loaded as Dataloaders and the model is trained using Adam Optimizer with Cross Entropy Loss. The model is evaluated every epoch on the validation set and the model with best validation accuracy is saved to storage for further evaluation and use. Upon finishing training, the training and validation error and loss is saved to the disk, along with a plot of error and loss over training.
3) *Classify Gesture:* After a model has been trained, it can be used to classify a new ASL gesture that is available as a file on the system. The user inputs the file path of the gesture image and the test data.py script will pass the file path to process data.py to load and preprocess the file the same way as the model has been trained.

V. RESULT DISCUSSIONS

Using Inception v3, we are successfully able to use convolutional neural networks for correctly recognizing images of static sign language gestures. The results obtained are showing a consistent high accuracy rate. This goes on to show that given a proper dataset, and correctly cropped images, Inception v3 is an apt model for static sign language gesture recognition.

Choosing our own dataset-model combination was tricky as there were models which tended to overfit to a particular dataset.

VI. CONCLUSION

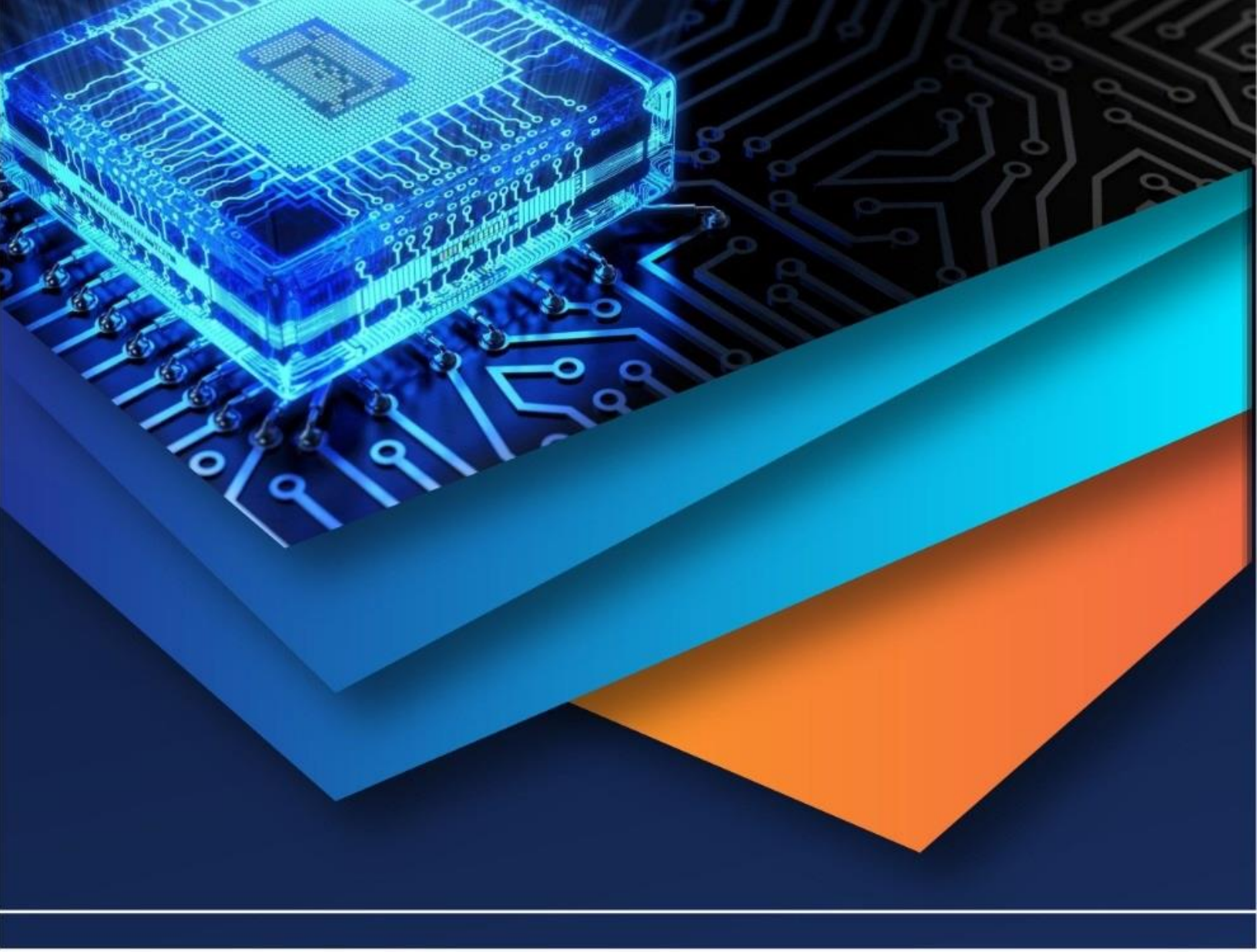
In this Project, we have gone through an automatic sign language gesture recognition system in real-time, using different tools. Although our proposed work expected to recognize the sign language and convert it into the text which can be useful for deaf and dumb people there's still a lot of scope for possible future work. In this project, a survey on sign language recognition is presented and various techniques have been studied and analysed for the same. In the recognition process, segmentation plays a crucial part in which skin region is separated from the background which usually affects the recognition accuracy. Besides segmentation, classification also depends on the feature extraction techniques which performs dimensionality reduction and reduces the computation cost. Study of various classification techniques concludes that deep neural networks (CNN, Inception model, LSTM) performs better than traditional classifiers such as KNN and SVM.

VII. ACKNOWLEDGMENT

We would like to thank the open-source community on the Internet, in particular the tutors at TensorFlow, whose help was indispensable to our learning process.

REFERENCES

[1] https://data-flair.training/blogs/sign-language-recognition-python-ml-opencv/
[2] https://www.analyticsvidhya.com/blog/2021/06/sign-language-recognition-for-computer-vision-enthusiasts/
[3] T. Bohra, S. Sompura, K. Parekh and P. Raut, "Real-Time Two Way Communication System for Speech and Hearing Impaired Using Computer Vision and Deep Learning," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 734-739, doi: 10.1109/ICSSIT46314.2019.8987908.
[4] Singha, Joyeeta & Das, Karen. (2013), "Recognition of Indian Sign Language in Live Video," International Journal of Computer Applications. 70. 10.5120/12174-7306.
[5] H. Muthu Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862125.
[6] J. Zhou and J. Hoang, "Real Time Robust Human Detection and Tracking System," in IEEE Computer Society Conference, 2005.
[7] C. P. Papageorgiou, M. Oren and T. Poggio, "A General Framework for Object Detection," in Center for Biological and Computational Learning Artificial Intelligence Laboratory, Cambridge, 2005.
[8] R. D. Charette and F. Nashashibi, "Real TimeVisual Traffic Lights Recognition Based on Spot Light Detection andAdaptive Traffic Lights Template," in IEEE, 2009.
[9] S. K. Nayar, S. A. Nene and M. Hiroshi, "Real-Time 100 Object Recognition System," in IEEE International Conference, 1996.
[10] A. Adam, E. Rivlin, S. Ilan and D. Reinitz, "Robust Real-Time Unusual Event Detection Using Multiple Fixed-Location Monitors," in IEEE, 2008.
[11] C. Bahlmann, Y. Zhu, R. Vishwanathan, M. Pelkoffer and T. Koehler, "A System for Traffic Sign Detection, Tracking, and Recognition Using Color, Shape, and Motion Information," in IEEE, 2005.
[12] M. Betke, E. Haritaoglu and L. S. Davis, "Real-time multiple vehicle detection and tracking from a moving vehicle," Machine Vision and Applications, p. 12, 2000.
[13] Q. Chen, N. D. Georganas and E. M. Petriu, "Real-time Vision-based Hand Gesture Recognition Using Haar-like Features," in IEEE Conference, 2007.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY